

L Padma Suresh
Subhransu Sekhar Dash
Bijaya Ketan Panigrahi *Editors*

Artificial Intelligence and Evolutionary Algorithms in Engineering Systems

Proceedings of ICAEES 2014, Volume 1

Advances in Intelligent Systems and Computing

Volume 324

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagrass, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at <http://www.springer.com/series/11156>

L Padma Suresh · Subhransu Sekhar Dash
Bijaya Ketan Panigrahi
Editors

Artificial Intelligence and Evolutionary Algorithms in Engineering Systems

Proceedings of ICAEES 2014, Volume 1

Editors

L Padma Suresh
Electrical and Electronics Engineering
Noorul Islam Centre for Higher Education
Kumaracoil, Tamil Nadu
India

Bijaya Ketan Panigrahi
Electrical Engineering
IIT Delhi
New Delhi, Delhi
India

Subhransu Sekhar Dash
Electrical and Electronics Engineering
SRM Engineering College, SRM University
Kattankulathur, Tamil Nadu
India

ISSN 2194-5357

ISSN 2194-5365 (electronic)

ISBN 978-81-322-2125-8

ISBN 978-81-322-2126-5 (eBook)

DOI 10.1007/978-81-322-2126-5

Library of Congress Control Number: 2014950644

Springer New Delhi Heidelberg New York Dordrecht London

© Springer India 2015

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 *International Conference on Artificial Intelligence and Evolutionary Algorithms in Engineering Systems (ICAEES)* held on 22 and 23 April, 2014 at Noorul Islam Centre for Higher Education, Noorul Islam University, Kumaracoil. ICAEES 2014 received 526 paper submissions from various countries across the globe. After a rigorous peer-review process, 235 full length articles were accepted for oral presentation at the conference. This corresponds to an acceptance rate of 45 % and is intended for maintaining the high standards of the conference proceedings. The papers included in this LNAICS volume cover a wide range of topics on Genetic Algorithms, Evolutionary Programming, and Evolution Strategies such as AIS, DE, PSO, ACO, BFA, HS, SFLA, Artificial Bees and Fireflies Algorithm, Parallel Computation, Membrane, Grid, Cloud, DNA, Mobile computing, Computer Networks and Security, Data Structures and Algorithms, Data Compression, Data Encryption, Data Mining, Digital Signal Processing, Digital Image Processing, Watermarking, Security and Cryptography, AI methods in Telemedicine and eHealth, Document Classification and Information Retrieval, Optimization Techniques, and their applications for solving problems in these areas.

In the conference, separate sessions were arranged for delivering the keynote address by eminent members from various academic institutions and industries. Eight keynote lectures were given in two different venues as parallel sessions on 22 and 23 April, 2014. In the first session, Dr. Shivashankar B. Nair, Professor and Head, Department of Computer Science and Engineering, Indian Institute of Technology, Guwahati gave a talk on “Emulating Bio-Inspired Mechanisms Using Mobile Agents” in Venue 1 and Dr. D. Deva Raj, Professor, Department of Electrical and Electronics Engineering, Kalasalingam University gave his lecture on “Multi Objective Evolutionary Algorithms” and covered various evolutionary algorithms and its applications to Power Systems in Venue 2. In the second session, Dr. Rusli Haji Abdullah, Software Engineering and Information System Department Faculty of Computer Science and Information Technology, Universiti Putra Malaysia gave his talk on “Artificial Intelligent in Knowledge Engineering: A Case of Knowledge Management System with Agent Technology Environment”

at Venue 1 and Dr. S.S. Dash, Professor, SRM University, Chennai gave his lecture on “Intelligence Computing Toward Renewable Energy Applications” and clarified the queries raised by the participants at Venue 2. He has explained the importance of soft computing in solar plant, wind mills, etc. Similarly on the second day in the first session, Dr. Ramachandran Kaimal, Professor and Chairperson, Department of Computer Science and Engineering, Amirta School of Engineering delivered his keynote address on “High Dimensional Data Analysis” at Venue 1 and the lecture by Prof. B.K. Panigrahi of IIT Delhi, India, on “Recent Advances in Bio Inspired Computing” gave an overall idea about the need for bio-inspired computing applications to the society at Venue 2. Dr. P. Karuppanan, Associate Professor, from National Institute of Technology, Allahabad gave a talk on “Artificial Intelligence Applications in Embedded Systems” at Venue 1 and Dr. Swagatam Das, from Indian Statistical Institute, Kolkata gave his lecture on “Computational Swarm Intelligence in Engineering Systems” and discussed the applications of Intelligent Systems in real-life problems at Venue 2. All these lectures generated great interest among the participants of ICAEES 2014 in paying more attention to these important topics in their research work.

We take this opportunity to thank the authors of all the submitted papers for their hard work, adherence to the deadlines, and suitably incorporating the changes suggested by the reviewers. The quality of a refereed volume depends mainly on the expertise and dedication of the reviewers. We are indebted to the Program Committee members for their guidance and coordination in organizing the review process.

We would also like to thank our sponsors for providing all the support and financial assistance. We are indebted to the Chairman, Vice-Chancellor, Advisors, Pro-Vice-Chancellor, Registrar, Faculty members, and Administrative Personnel of Noorul Islam Centre for Higher Education, Noorul Islam University, Kumaracoil for supporting our cause and encouraging us to organize the conference on a grand scale. We would also like to thank all the participants for their interest and enthusiastic involvement. Finally, we would like to thank all the volunteers whose tireless efforts in meeting the deadlines and arranging every detail meticulously made sure that the conference could run smoothly. We hope the readers of these proceedings find the papers useful, inspiring, and enjoyable.

April 2014

L Padma Suresh
Subhransu Sekhar Dash
Bijaya Ketan Panigrahi

Contents

Improvement in Hungarian Algorithm for Assignment Problem	1
Kartik Shah, Praveenkumar Reddy and S. Vairamuthu	
Vertex Cover Problem—Revised Approximation Algorithm.	9
Kartik Shah, Praveenkumar Reddy and R. Selvakumar	
A Simple Control Strategy Technique for a Single-phase Bridgeless Active Rectifier with High Power Factor and Voltage Stabilization Using Partial Digital Implementation	17
Rahul Ganpat Mapari and D.G. Wakde	
Ensemble Neural Network Algorithm for Detecting Cardiac Arrhythmia	27
S. Aruna and L.V. Nandakishore	
An Efficient Invasive Weed Optimization Algorithm for Distribution Feeder Reconfiguration and Loss Minimization	37
K. Sathish Kumar, K. Rajalakhsmi, S. Prabhakar Karthikeyan and R. Rajaram	
Implementation of Generative Crossover Operator in Genetic Algorithm to Solve Traveling Salesman Problem.	47
Devasenathipathi N. Mudaliar and Nilesh K. Modi	
A 4-bit 9 KS/s Distortionless Successive Approximation ADC in 180-nm CMOS Technology	55
P. Dipu, B. Saidulu, K. Aravind, Johny S. Raj and K. Sivasankaran	
EEG-based Automatic Detection of Drowsy State	65
Jinu Jai, Geevarghese Titus and S. Purushothaman	

WIDS Real-Time Intrusion Detection System Using Entrophical Approach	73
Kamalanaban Ethala, R. Sheshadri and S. Sibi Chakkaravarthy	
Optical Character Recognition for Alphanumerical Character Verification in Video Frames	81
Sheshank Shetty, Arun S. Devadiga, S. Sibi Chakkaravarthy, K.A. Varun Kumar, Ethala Kamalanaban and P. Visu	
Modified AODV Routing Protocol for Multi-hop Cognitive Radio Ad Hoc Networks	89
Melvin Mathew, G. Shine Let and G. Josemin Bala	
Theoretical Framework of the Algorithm to Thwart MAC Spoofing DoS Attack in Wireless Local Area Infrastructure Network	99
M. Durairaj and A. Persia	
PCA-Based Feature Selection for MRI Image Retrieval System Using Texture Features	109
N. Kumaran and R. Bhavani	
A Survey of Location Prediction Using Trajectory Mining	119
B.A. Sabarish, R. Karthi and T. Gireeshkumar	
Survey on Router Policies Providing Fairness and Service Differentiation Favoring Real-Time Data Transfers in Internet	129
Jyothish K. John and R.V. Siva Balan	
Exploration of the Effect of Demographic and Clinical Confounding Variables on Results of Voxel-Based Morphometric Analysis in Schizophrenia	139
Anupa A. Vijayakumari, Priyadarshini Thirunavukkarasu, Ammu Lukose, Vikram Arunachalam, Jitender Saini, Sanjeev Jain, Bindu M. Kutty and John P. John	
A Novel Method for Secure Image Steganography	151
S. Anjana and P.P. Amritha	
Techniques for Enhancing the Performance of TCP in Wireless Networks	159
MD. Sirajuddin, Ch. Rupa and A. Prasad	

Analysis of Classification Models Using Image Statistics and Data Miner for Grade Prediction of Astrocytoma 169
 M. Monica Subashini, Sarat Kumar Sahoo, S. Prabhakar Karthikeyan and I. Jacob Raglend

Object Detection in Cluttered Environment Using 3D Map 181
 Deepesh Jain, Renuka Ramachandran, Anuhya Vunnam and P. Vignesh

Smart Energy Meter with Instant Billing and Payment 187
 Dhananjayan Ravi, J. Shibu and E. Shanthi

GA-Based Compiler Parameter Set Tuning 197
 N.A.B. Sankar Chebolu, Rajeev Wankar and Raghavendra Rao Chillarige

An Intelligent Intrusion Detection System Using Average Manhattan Distance-based Decision Tree 205
 R. Selvi, S. Saravan Kumar and A. Suresh

Intelligent Medical Diagnosis System Using Weighted Genetic and New Weighted Fuzzy C-Means Clustering Algorithm 213
 P.S. Jeetha Lakshmi, S. Saravan Kumar and A. Suresh

Security Analysis in Cloud Environment 221
 M.S. Akshay, Ashina Kakkar, K. Jayasree, P. Prudhvi and Prathibha Shridhara Metgal

A Bi-level Clustering Analysis for Studying About the Sources of Vehicular Pollution in Chennai 229
 Gunaselvi Manohar, S. Prasanna Devi and K. Suryaprakasa Rao

Investigation of Fault Detection Techniques for an Industrial Pneumatic Actuator Using Neural Network: DAMADICS Case Study 237
 V. Elakkiya, K. Ram Kumar, V. Gomathi and S. Rakesh Kumar

Design of Portable Security System Using Face Recognition with Back-Propagation Algorithm and MATLAB 247
 M. Hareesh Babu, M. Bala Naga Bhushanamu, B. Benarji and M. Purnachandra Rao

Model-Based Control for Moisture in Paper Making Process 257
 C. Karthik, K. Suresh, K. Valarmathi and R. Jacob Rajesh

Elimination of Harmonics in Seven-Level Cascaded Multilevel Inverter Using Particle Swarm Optimization Technique	265
W. Razia Sultana, Sarat Kumar Sahoo, S. Prabhakar Karthikeyan, I. Jacob Raglend, Pasam Harsha Vardhan Reddy and Gangireddy Taraka Rajasekhar Reddy	
Optimization and Quality-of-Service Protocols in VANETs: A Review	275
K.R. Jothi and A. Ebenezer Jeyakumar	
Object Detection Using Robust Image Features	285
Khande Bharath Kumar and D. Venkataraman	
Service-Adaptive Fuzzy Multi Criteria Based Intelligent Vertical Handover Decision Algorithm for Heterogeneous Wireless Networks.	297
V. Anantha Narayanan, A. Rajeswari and V. Sureshkumar	
Automatic Traffic Classification Using Machine Learning Algorithm for Policy-Based Routing in UMTS–WLAN Interworking	305
V. Anantha Narayanan, V. Sureshkumar and A. Rajeswari	
Sequential Decision Making Using Q Learning Algorithm for Diabetic Patients	313
Pramod Patil, Parag Kulkarni and Rachana Shirsath	
Design of Quantum Cost and Delay-Optimized Reversible Wallace Tree Multiplier Using Compressors	323
A.N. Nagamani and Vinod Kumar Agrawal	
A Novel Authentication Framework for Hadoop	333
P.K. Rahul and T. GireeshKumar	
Fuzzy ART-Based User Behavior Trust in Cloud Computing	341
M. Jaiganesh, M. Aarthi and A. Vincent Antony Kumar	
Semi-supervised Learning Algorithm for Online Electricity Data Streams	349
Pramod Patil, Yogita Fatangare and Parag Kulkarni	

An Efficient Continuous Speech Recognition System for Dravidian Languages Using Support Vector Machine 359
 J. Sangeetha and S. Jothilakshmi

Video Surveillance Based Tracking System 369
 R. Venkatesan, P. Dinesh Anton Raja and A. Balaji Ganesh

Stroke Detection in Brain Using CT Images 379
 S. Neethu and D. Venkataraman

Preemptive Appliances Scheduling in Smart Home Using Genetic Algorithm 387
 A. Ranjini and B.S.E. Zoraida

Credible Secure Data Aggregation in Wireless Sensor Networks 395
 M.P. Anuradha and Gopinath Ganapathy

Zumkeller Labeling Algorithms for Complete Bipartite Graphs and Wheel Graphs 405
 B.J. Balamurugan, K. Thirusangu and D.G. Thomas

An Unidentified Location-Based Efficient Routing Protocol in VANET. 415
 P. Dharani, S. Sibi Chakkaravarthy, M. Ganesan, Ethala Kamalanaban, P. Visu, Pravin R. Patil and C. Mahesh

A Secure and Efficient Binding Update Scheme with Decentralized Design for Next Generation IP Mobility. 423
 Senthil Kumar Mathi, M.L. Valarmathi and Srilakshmy

Stego with Color Cue. 433
 N.R. Raajan, G. Balasubraminayan, B. Barath Krishna, S. Ramya, M. Malligaraj and K. Karthikeyan

Secure Seed-Based Sturdy OTP via Convenient Carry-on Device. 447
 Ashok Kumar Mohan and T. Gireesh Kumar

A (t, n) Secure Sum Multiparty Computation Protocol Using Multivariate Polynomial Secret Sharing Scheme 457
 K. Praveen and Nithin Sasi

Small-World Particle Swarm Optimizer for Real-World Optimization Problems 465
 Megha Vora and T.T. Mirmalinee

A Comparative Study of Feature Ranking Methods in Recognition of Handwritten Numerals 473
 Abhinaba Roy, Nibaran Das, Amit Saha, Ram Sarkar, Subhadip Basu, Mahantapas Kundu and Mita Nasipuri

Performance Evaluation of PET Image Reconstruction Using Radial Basis Function Networks 481
 T. Arunprasath, M. Pallikonda Rajasekaran, S. Kannan and Shaeba Mariam George

Clustering for Knowledgeable Web Mining. 491
 B.S. Charulatha, Paul Rodrigues, T. Chitralekha and Arun Rajaraman

Effective Path Discovery Among Clusters for Secure Transmission of Data in MANET. 499
 P. Madhavan, P. Malathi and R. Abinaya

Quality-of-Service Analysis of AOMDV and AOMDV-MIMC Routing Protocols for Mobile Ad hoc Networks 511
 P. Periyasamy and E. Karthikeyan

Particle Swarm Optimization-Based SONAR Image Enhancement for Underwater Target Detection 523
 P.M. Rajeshwari, G. Kavitha, C.M. Sujatha and Dhilsha Rajapan

Intelligent Modeling and Optimization of ECM Process Parameters 533
 T.M. Chenthil Jegan, D. Ravindran and M. Dev Anand

An Effective Automation Testing Framework for OATS Tool 543
 Gobi Ramasamy and Sathishkumar Ramalingam

Multimodal Biometric Authentication System Based on Score-Level Fusion of Palmprint and Finger Vein 551
 C. Murukesh, K. Thanushkodi, Padmanabhan Preethi and Feroze Naina Mohamed

Synergistic Clinical Trials with CAD Systems for the Early Detection of Lung Cancer 561
 G. Vijaya and A. Suhasini

A Unified Framework for Network Bandwidth and Link Latency Detector Based on Cloud Computing 569
 S. Suguna and A. Suhasini

Development of Concatenative Syllable-Based Text to Speech Synthesis System for Tamil 585
 B. Sudhakar and R. Bensraj

Design of Low-Power Blink Detector for Minimally Invasive Implantable Stimulator (SoC) Using 180 nm Technology 593
 J. Joselyn Priyadarshini and S. Ravindrakumar

Energy- and Trust-Based AODV for Quality-of-Service Affirmation in MANETs 601
 Sridhar Subramaniam and Baskaran Ramachandran

Classification of Remote Sensing Image Based on Different Similarity Measures 609
 Kartik Shah, Shantanu Santoki, Himanshu Ghetia and D. Aju

A Start to Fail Frequency Technique for Detecting Hardware Trojan 621
 Sharmila Durai, Prasanna Kumar and Srinivasan Ramasamy

A Novel Approach Privacy Security Protocol Based SUPM Method in Near Field Communication Technology 633
 S. Kannadhasan, M. Isaivani and G. Karthikeyan

Gabor Transform for the Time–Frequency Localization of Impulse Faults in a Transformer 645
 N. Vanamadevi, S. Santhi and M. Arivamudhan

A Modified Priority-Based Multischeduler (PBMS) for Optical Network 657
 A. Adaikalam, S. Manikandan and V. Rajamani

Comparative Analysis of Digital Watermarking in Discrete Wavelet Transform and Mojette Transform 667
 Chandini Rajeev and K.P. Girish

A Three Factor Authentication System for Smartcard Using Biometric, Visual Cryptography and OTP 673
 Akhitha S. Kumar and K.P. Girish

Identifying Sound of RPW In Situ from External Sources	681
Betty Martin, P.E. Shankaranarayanan, Vimala Juliet and A. Gopal	
VNS-Based Heuristic for Identical Parallel Machine Scheduling Problem	693
S. Bathrinath, S. Saravana Sankar, S.G. Ponnambalam and I. Jerin Leno	
Green Algorithm for Virtualized Cloud Systems to Optimize the Energy Consumption	701
P. Prakash, G. Kousalya, Shriram K. Vasudevan and K.S. Sangeetha	
Lecture Notes in Computer Science: S Transform for the Analysis of Impulse Faults in Transformer	709
N. Vanamadevi, S. Santhi and R. Saranya	
Defensive Mechanism to Guard Against Packet Droppers in Mobile Ad Hoc Network	721
S. Madhurikkha and R. Sabitha	
Real-Time Intrusion Prediction Using Hidden Markov Model with Genetic Algorithm	731
T. Divya and Kandasamy Muniasamy	
Detection of Power Quality Disturbances Based on Adaptive Neural Net and Shannon Entropy Method	737
D. Kavitha, P. Renuga and M. Seetha Lakshmi	
Texture Feature Extraction Using MGRLBP Method for Medical Image Classification.	747
Suganya Ramamoorthy, R. Kirubakaran and Rajaram Siva Subramanian	
A Novel Lightweight Protocol for Address Assignment in Ad Hoc Networks Based on Filters.	755
M. Anusuya Shyamala and R. Velayutham	
Structural Refinement: An Effective OCL-Based Testing Approach . . .	765
A. Jalila and D. Jeya Mala	
Dynamic Architecture and Performance Analysis of Secure and Efficient Key Management Scheme in Multicast Network	775
N.M. Saravanakumar, R. Keerthana and G.M. Mythili	

Mining Undemanding and Intricate Patterns with Periodicity in Time Series Databases 785
 S. Sridevi, P. Saranya and S. Rajaram

Group-Based Access Technique for Effective Resource Utilization and Access Control Mechanism in Cloud 793
 Lavanya Selvaraj and Saravana Kumar

Design of Fuzzy Logic-Based pH Controller for High-Pressure-Rated Modified CSTR System 803
 Jithin Kannangot, Ponnusamy Lakshmi and Keppayan Thirupathi

Level Control of Quadruple Tank Process with Finite-Time Convergence Using Integral Terminal Sliding Mode Controller 813
 Sekaran Sankaranarayanan, Lakshmi Ponnusamy and Sangapillai Sutha

An Enhanced Security Framework for a Cloud Application 825
 B. Balamurugan and P. Venkata Krishna

Enhanced Role-Based Access Control for Cloud Security 837
 B. Balamurugan and P. Venkata Krishna

Model Predictive Controllers for Nonminimum-phase Quadruple-tank Process 853
 Keerthi Chacko, Lakshmi Ponnusamy and Sangapillai Sutha

About the Editors

Dr. L Padma Suresh obtained his doctorate from M S University and Dr. M.G.R University, respectively. He is presently working as a Professor and Head in Department of Electrical and Electronics Engineering, Noorul Islam University, Kumaracoil, India. Dr. Suresh is well known for his contributions to the field in both research and education contributing over 50 research articles in journals and conferences. He is the editorial member of International Journal of Advanced Electrical and Computer Engineering and also served as reviewer in various reputed journals. He has been a life member of the Indian Society for Technical Education. He also served in many committees as Convener, Chair, and Advisory member for various external agencies. His research is currently focused on Artificial Intelligence, Power Electronics, Evolutionary Algorithms, Image Processing, and Control Systems.

Dr. Subhransu Sekhar Dash is presently working as a Professor in the Department of Electrical and Electronics Engineering, SRM Engineering College, SRM University, Chennai, India. He received his Ph.D. degree from College of Engineering, Guindy, Anna University. He has more than 17 years of research and teaching experience. His research areas are Power Electronics and Drives, Modeling of FACTS Controller, Power Quality, Power System Stability, and Smart Grid. He is a Visiting Professor at Francois Rabelais University, POLYTECH, France. He is the chief editor of International Journal of Advanced Electrical and Computer Engineering.

Dr. Bijaya Ketan Panigrahi is an Associate Professor in the Electrical and Electronics Engineering Department in Indian Institute of Technology Delhi, India. He received his Ph.D. degree from Sambalpur University. He is serving as a chief editor to the International Journal of Power and Energy Conversion. His interests include Power Quality, FACTS Devices, Power System Protection, and AI Application to Power System.

Improvement in Hungarian Algorithm for Assignment Problem

Kartik Shah, Praveenkumar Reddy and S. Vairamuthu

Abstract Hungarian method for assignment problem is generally used in parallel environment for the assignment of job to a processor. If the number of processors and number of jobs are same, then we can assign each processor 1 job with less cost using Hungarian method. If the number of jobs is larger compared to number of processors, then this method does not work (another approach is using dummy processors, but it is not implementable). In this paper, we proposed an alternate approach same as Hungarian method for assignment of more jobs to lesser processors.

Keywords Assignment problem · Hungarian method · Improvement

1 Introduction

The assignment problem can be applied in real-time examples where the assignment of jobs to processors is needed. Consider the n tasks with benefits a_{ij} such that i task associates with j processor. We want to assign task to processor on one-to-one basis, which minimizes the total cost. The method for solving such problem is called as Kuhn's Hungarian method [1]. There are a lot of implementation and discussion about the Hungarian algorithm, and some of them are sequential shortest path methods [2–12]. It can be applied to the assignment of objects to the person. In each iteration, an unassigned person is assigned with the object. The basic method of Hungarian is serial in nature. It can be parallelized to achieve faster computation [13].

K. Shah (✉) · P. Reddy · S. Vairamuthu
School of Computing Science and Engineering, VIT University, Vellore 632014, India
e-mail: kartikshah@mail.com

P. Reddy
e-mail: praveenkumar.reddym2012@vit.ac.in

S. Vairamuthu
e-mail: svairamuthu@vit.ac.in

1.1 Assignment Problem

Suppose we have ‘ n ’ resources and we have to assign ‘ n ’ tasks to those resources on one-to-one basis. Also, we know the cost of assigning a resource to a processor. We have interest in finding an optimal assignment which minimizes the total cost.

Theorem *If a number is added or subtracted from all the entry of any 1 row or 1 column of a cost matrix, then an optimal assignment for the resulting cost matrix is also an optimal assignment for the original cost matrix.*

Based upon the above discussion, we will consider the classical Hungarian method for problem solving and also we will introduce another approach which will be variant of the original one. It is discussed in the below sections.

2 Existing Work

Hungarian method can be used to solve the assignment problems. General algorithm for assignment problem is as follows:

Steps:

1. Obtain the cost matrix from the past history.
2. Find the row minimum of each row and subtract it from the other elements of the corresponding row.
3. Find the column minimum of the reduced matrix and subtract it from the other elements of the corresponding column.
4. Write the 0's and draw the minimum number of lines either horizontal or vertical to cover all the 0's. If the number of lines is equal to the order of cost matrix, then go to Step 6. Else go to Step 5.
5. Find the minimum cost from the uncrossed cross and edit the junction points, whereas subtract it from the uncrossed cost. Go to Step 4.
6. Write 0's obtained and allocate the task to the processors evenly.

Algorithm 1 There are 3 possible cases where the assignment problem can be used. These 3 cases are as follows:

Case 1 The number of processors and number of tasks are same.

If the number of processors and number of tasks are same, then each processor is allocated with 1 task based upon the Algorithm 1. Consider the example for the case.

Example 1 A programmer wishes to assign 3 tasks to 3 processors in such a way that each task is assigned to more than 1 task. The cost of assigning task to processor is given by the following matrix:

$$\begin{array}{ccc} \left[\begin{array}{ccc} 8 & 7 & 6 \\ 5 & 7 & 8 \\ 6 & 8 & 7 \end{array} \right] & \begin{array}{l} P1 \\ P2 \\ P3 \end{array} \\ T1 & T2 & T3 \end{array}$$

Step 1: After Step 1, the cost matrix will look like the one shown above.

Step 2: After second step of evaluation, it will look like

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 3 \\ 0 & 2 & 1 \end{bmatrix}$$

Step 3: Find the column minimum and subtract it from the other elements of the corresponding column. The reduced matrix is given below:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 1 & 1 \end{bmatrix}$$

Step 4: Write the 0's and draw the minimum number of lines to cover all the 0's.

$$\begin{bmatrix} \overline{2} & 0 & 0 \\ 0 & \textcircled{1} & 3 \\ 0 & 1 & 1 \end{bmatrix}$$

Step 5: Find the minimum cost from the uncrossed costs, sum it with the junction point, and subtract it from the uncrossed costs. The reduced matrix is as follows:

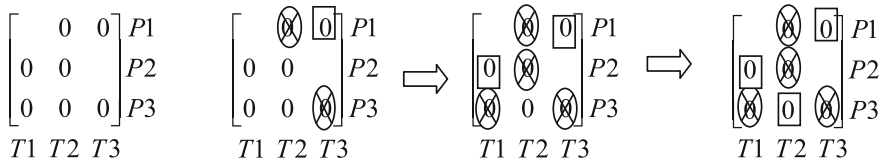
$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Step 6: Write all zeros and draw the minimum number of lines.

$$\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array}$$

Number of lines = 3 = Order of cost matrix

Final Process:



Final Allocation 1

Processor	Task
P1	T3
P2	T1
P3	T2

Total cost: 6 + 5 + 8 = 19 units

Final Allocation 2 (another possibility)

Processor	Task
P1	T2
P2	T1
P3	T3

Total cost: 7 + 5 + 7 = 19 units

Here, we have 2 possibilities of allocation. It can be more than 2 also, and we can have the same final cost.

Case II The number of processors is greater than the number of tasks.

In Case I, we can see that if the number of processors is equal to the number of tasks, then we can have 1 processor with the single task. But if the number of processors and number of tasks are not the same, then we need to go for some modification in algorithm 1. If the number of processors is greater than the number of tasks, then we have to add a dummy task (i.e., do nothing). In such case, 1 processor will remain idle. Consider the following example.

Example 2 A programmer wishes to process 3 tasks to 4 processors in such a way that each task must be allocated to some processor. The cost matrix is as follows:

$$\begin{matrix}
 \begin{bmatrix} 8 & 6 & 7 \\ 3 & 8 & 4 \\ 8 & 9 & 8 \\ 9 & 6 & 4 \end{bmatrix} & \begin{matrix} P1 \\ P2 \\ P3 \\ P4 \end{matrix} \\
 \begin{matrix} T1 & T2 & T3 \end{matrix} &
 \end{matrix}$$

It can be rewritten as follows:

$$\begin{array}{cccc}
 \left[\begin{array}{cccc}
 8 & 6 & 7 & 0 \\
 3 & 8 & 4 & 0 \\
 8 & 9 & 8 & 0 \\
 9 & 6 & 4 & 0
 \end{array} \right] & P1 & & \\
 & P2 & & \\
 & P3 & & \\
 & P4 & & \\
 T1 & T2 & T3 & dT4
 \end{array}$$

where dT4 represents the dummy task. Now, we can apply Algorithm 1 as shown in Example 1.

Case III The number of processors is less than the number of tasks.

Like in case II, the number of processors is greater than the number of tasks. So, we added a new dummy task. Similarly, in this case, we will add a dummy processor which can be assigned to a task. Consider the following example.

Example 3 A programmer wishes to process 4 tasks to 3 processors in such a way that each task must be allocated to some processor. The cost matrix is as follows:

$$\begin{array}{cccc}
 \left[\begin{array}{cccc}
 8 & 6 & 7 & 9 \\
 3 & 8 & 4 & 5 \\
 8 & 9 & 8 & 4
 \end{array} \right] & P1 & & \\
 & P2 & & \\
 & P3 & & \\
 T1 & T2 & T3 & T4
 \end{array}$$

It can be rewritten as:

$$\begin{array}{cccc}
 \left[\begin{array}{cccc}
 8 & 6 & 7 & 9 \\
 3 & 8 & 4 & 5 \\
 8 & 9 & 8 & 4 \\
 0 & 0 & 0 & 0
 \end{array} \right] & P1 & & \\
 & P2 & & \\
 & P3 & & \\
 & dP4 & & \\
 T1 & T2 & T3 & T4
 \end{array}$$

where dP4 represents the dummy processor. Now, we can apply Algorithm 1 as shown in Example 1.

3 Methodology

Section 2 shows the existing approach which is available to work with the different cases. Now, consider Case III. Here, we are finally assigning a task to a processor which is just a dummy processor. It is not a realistic approach. We cannot assign a task to a processor which actually does not exist. We can improve the performance of the above-mentioned algorithm by giving 1 processor to more than 1 task with least final cost. So, if there are 4 tasks and 3 processors, then we can have each processor 1 task and out of 3 and 1 processor will have 2 tasks. This can be done in such a way that the final cost will be reduced. Consider the Algorithm 1 and compare it with modified Algorithm 2 shown below:

1. Obtain the cost matrix from the past history.
2. Find the row minimum of each row and subtract it from the other elements of the corresponding row.
3. Find the column minimum of the reduced matrix and subtract it from the other element of the corresponding column.
4. Write 0's obtained and assign the tasks to processors till each processor is assigned with 1 task
5. Now, for the remaining tasks, consider the processor which is having less work from cost matrix and assign the remaining tasks to those processors evenly.

Algorithm 2 This algorithm will work for the condition when the number of processors will be very less compared to the number of tasks. It can be more understood by the following example:

Example 4 Consider the Example 3 in given Case III.

$$\begin{array}{cccc} \left[\begin{array}{cccc} 8 & 6 & 7 & 9 \\ 3 & 8 & 4 & 5 \\ 8 & 9 & 8 & 4 \end{array} \right] & P1 & & \\ & P2 & & \\ & P3 & & \\ T1 & T2 & T3 & T4 \end{array}$$

Applying the Algorithm 2,

After Step 1, it will look like the one shown above.

After Step 2, it will look like,

$$\left[\begin{array}{cccc} 2 & 0 & 1 & 3 \\ 0 & 5 & 1 & 2 \\ 4 & 5 & 4 & 0 \end{array} \right]$$

After Step 3, it will look like,

$$\left[\begin{array}{cccc} 2 & 0 & 0 & 3 \\ 0 & 5 & 0 & 2 \\ 4 & 5 & 4 & 0 \end{array} \right]$$

After Step 4, it will look like,

$$\left[\begin{array}{cccc} - & 0 & 0 & - \\ 0 & - & 0 & - \\ - & - & - & 0 \end{array} \right]$$

Finally, assignment is as shown below:

$$\begin{bmatrix} - & 0 & 0 & - \\ 0 & - & 0 & - \\ - & - & - & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} - & \textcircled{0} & 0 & - \\ \textcircled{0} & - & 0 & - \\ - & - & - & \textcircled{0} \end{bmatrix} \Rightarrow \begin{bmatrix} & 0 & 0 & \\ 0 & & \textcircled{0} & \\ & & & 0 \end{bmatrix}$$

Final allocation is shown in the following table:

Task	Processor
T1	P2
T2	P1
T3	P2
T4	P3

Final cost can be given as follows: $3 + 6 + 4 + 4 = 17$ units. Here, final cost for processor 2 will be 7 units.

4 Conclusions and Future Work

From the above Algorithm 2 and Example 4, we can conclude that we can assign more number of tasks to less number of processors using the Hungarian algorithm with some modification mentioned in Algorithm 2. Also, we can have very less amount of cost. We can utilize processors more efficiently. It can work for all types of assignment problems where we can apply Hungarian algorithm. The complexity of the algorithm will be same as the complexity of Hungarian algorithm. The benefit of using modified Hungarian algorithm is that in case of less number of processors with more number of jobs are present. In reality, the situation is like the same as more number of jobs are assigned to less number of processors. Due to that reason, the algorithm mentioned above will be more useful.

4.1 Comparison

As per existing methodology, if we go for computing cost with dummy processor, we can get the minimum cost using Hungarian method. But the problem with the algorithm is that the approach is not practical (i.e., ideally dummy processors are not possible). The more realistic approach is that we can use modified algorithm for job distribution. The modified algorithm is more practical compared to existing approach by means of not introducing any dummy processors. Here, we can assign jobs based on the load given to each processor. Depending upon load, the final cost will be calculated and the cost will be minimum.

4.2 Future Work

Job assignment problem is problem of assigning jobs to processors appropriately, so that the jobs can be assigned properly and cost will be minimum. We can extend this method for large number of jobs with very less number of processors. As the number of jobs increases compared to number of processors, the calculation becomes trickier and requires more calculations. Also, the steps of assignment increase as the number of jobs increases. The work can be extended by finding such techniques which have less processors and more number of jobs.

Acknowledgments The authors would like to thank the School of Computer Science and Engineering, VIT University, for giving them the opportunity to carry out this project and also for providing them with the requisite resources and infrastructure for carrying out the research.

References

1. W. Kuhn, The hungarian method for assignment problem. *Nav. Res. Logistics Q.* **2**, 83–87 (1955)
2. D.P. Bertsekas, *Linear network optimization algorithms and codes* (MIT Press, Cambridge, 1991)
3. G. Carpaneto, S. Martello, P. Toth, Algorithms and codes for assignment problem. *Ann. Oper. Res.* **13**, 193–223 (1988)
4. D.A. Castanon, B. Smith, A. Wilson, *Performance of parallel assignment algorithms on different multiprocessor architectures*. Alphatech report TP-1245, Burlington, MA, (1989)
5. U. Derigs, The shortest augmenting path method for solving assignment problem-motivation and computational experience. *Ann. Oper. Res.* **4**, 57–102 (1985)
6. M. Engquist, A successive shortest path algorithm for the assignment problem. *INFRO.* **20**, 370–384 (1982)
7. F. Glover, R. Glover, D. Klingman, *Threshold assignment algorithm*. Centre for Business Decision analysis report CBDA 107, Graduate School of Business, University of Texas at Austin (1982)
8. J.R.M. Hall, An algorithm for distinct representatives. *Am. Math. Monthly* **51**, 716–717 (1956)
9. R. Jonkar, A. Volgenant, A shortest augmenting path algorithm for the dense and sparse linear assignment problems. *Computing* **3**, 92–106 (1987)
10. E. Lawler, *Combinational Optimization: Networks and Matroids* (Holt, Rinehart & Winston, New York, 1976), p. 206
11. L.F. McGinnis, Implementation and testing of a primal dual algorithm for the assignment problem. *Oper. Res. Int. J.* **31**, 277–291 (1983)
12. C.H. Papadimitriou, K. Steiglitz, *Combinational Optimization: Algorithm and complexity* (Prentice-Hall, Englewood Cliffs, 1982)
13. E. Balas, D. Miller, J. Pekny, P. Toth, A parallel shortest path algorithm for assignment problem. *J. ACM* **38**(4), 985–1004 (1991)

Vertex Cover Problem—Revised Approximation Algorithm

Kartik Shah, Praveenkumar Reddy and R. Selvakumar

Abstract This paper is aimed to present the solution to vertex cover problem by means of an approximation solution. As it is NP complete problem, we can have an approximate time algorithm to solve the vertex cover problem. We will modify the algorithm to have an algorithm which can be solved in polynomial time and which will give near to optimum solution. It is a simple algorithm which will be based on articulation point. Articulation point can be found using the Depth First Search algorithm.

Keywords Articulation point · Vertex covering problem · Optimization · Approximation algorithm

1 Introduction

A problem (P) is called NP complete if P is in class NP and every problem P' is polynomially reducible to P . Vertex cover problem is that problem in which we take an undirected graph ($G = (V, E)$) with ' V ' vertices and ' E ' edges and will have a set which contains the vertices which can cover all the edges of the graph. According to Garey and Johnson, vertex cover problem is one of the six basic NP complete problems [1].

It is observed that sometimes, there exist some problems for which we do not have an optimal solutions. However, we can have an approximate solution which

K. Shah (✉) · P. Reddy · R. Selvakumar
School of Computing Science and Engineering, VIT University, Vellore 632014, India
e-mail: kartikshah@mail.com

P. Reddy
e-mail: praveenkumar.reddym2012@vit.ac.in

R. Selvakumar
e-mail: rselvakumar@vit.ac.in

will be closer to the optimal solution. The algorithm which provides closer solution to the optimal solution is known as approximation algorithms.

Let $G = (V, E)$ be a graph where V is number of vertices and E be number of edges. The set of vertices $V' \subseteq V$ is said to be cover if for each edge $(u, v) \in E$, either $u \in V'$ or $v \in V'$ or both. The number of vertices in V' is known as the size of the cover V' . The problem of finding minimum size is known as vertex cover problem.

2 Related Work

For vertex cover problems, we have many real-life problems in which vertex cover problem solution can be applied. Like, to find population growths taken into polynomial time and for that we can take bi-parted graph and also take articulation points [2]. Also, vertex cover problem for network base routing delays in tolerance network [3], for network traffic measurement [4]. Polynomial space parameterized vertex cover can be solved in $O(1.2738^k + kn)$ time [5]. F. Delbot and C. Laforest had proposed an algorithm in which vertexes are scanned from left to right. Condition is as follows: “ u is added to vertex cover if and only if it has at least one neighbor not in the cover” [6]. This algorithm is called as LIST LEFT which is given as follows [7]:

Labeled graph $L(G) = (L(V), E)$

1. $C \leftarrow \phi$
2. For each vertex, $u \in L(V)$ do
3. If u has at least one right neighbor, then
4. $C \leftarrow C \cup \{u\}$
5. Return C

Sorted LL is another technique of finding vertex cover. It takes decision as “if there is at least one $v \in N(u)$ with lower degree, select u ; otherwise, if u has only neighbor with higher degree, u is not selected.” Algorithm can be represented as follows [7]:

Labeled graph $L(G) = (L(V), E)$

1. $C \leftarrow \phi$
2. For each vertex, $u \in L(V)$ do
3. If u has at least one right neighbor with a lower degree or a right neighbor with the same degree, then
4. $C \leftarrow C \cup \{u\}$
5. Return C

One more approach is to use ANTI SORTED LL. In this algorithm, degree of vertex’s neighbor is taken into consider. Below is the algorithm shown [7]:

Labeled graph $L(G) = (L(V), E)$

1. $C \leftarrow \phi$
2. For each vertex, $u \in L(V)$ do
3. If u has at least one right neighbor with a larger degree or a left neighbor with the same degree, then
4. $C \leftarrow C \cup \{u\}$
5. Return C

These all algorithms can be used for finding vertex cover set of a graph.

It is possible to approximate the weight of the vertices which results in local approximation and called as local-ratio theorem and is so called local-ratio theorem for weighted vertex cover problem [8]. Further, it states that it is possible to put together the Nemhauser–Trotter algorithm (local optimization algorithm) and local-ratio theorem to get new approximation techniques that improve performance [8]. Kernelization is the process of applying polynomial time preprocessing to the instance of graph $G(V, E)$ and obtaining other instance $G'(V', E')$ where $V' \leq V$ and G' will have vertices of vertex cover V' [9]. It is observed that we can find up to $3/2$ vertices of minimum vertex covers using collection of graph transformations. The algorithm guarantees an approximation ratio of $3/2$, for finding large number of randomly created graphs. The reductions are extremely fast. The problem has best-case and worst-case approximation ratio as $2-O(1)$ [10]. The random graphs that we generally use are simple random graph model proposed by y Erdos and Renyi [11]. Chromatic number is defined as minimum number of colors used for vertex to cover all with distance 1. Chromatic number can be used to find the vertex over of the graph. Kuhn and Mastrolilli [12] investigated weighted vertex cover problem for graphs when a locally bounded coloring is given.

3 Approach

There are many approaches to solve vertex cover problem.

3.1 Classical Approach

The approximation algorithm which is available to solve the vertex cover problem is a polynomial time algorithm, and it can be solved in $O(V + E)$ time complexity where V is number of vertices and E is number of edges. The approximation vertex cover is a polynomial time 2 approximation algorithm.

That is, let V be a vertex cover and V^* be the optimal cover to that problem. Then, we can prove that

$$|V| \leq 2|V^*|$$

So, in this case, if number of vertices increases, the solution to vertex cover may also diverse far from optimal solution.

Algorithm 1.1 Approximation Algorithm [13]

APPROX VERTEX COVER(G)

1. $C \leftarrow \phi$
 2. $E' = E[G]$
 3. While $E' \neq \phi$
 4. Let (u, v) be an arbitrary edge of E'
 5. $C \leftarrow C \cup \{u, v\}$
 6. Remove from E' every edge incident on either u or v
 7. Return C
-

3.2 Our Approach

As we noted above that approximate vertex cover problem solution is less than or equal to 2 times the optimal solution. We proposed the algorithm which will take near to optimal solution.

In our approach, we will find the articulation point of that particular graph and then add those vertices in the vertex cover. From remaining edges, we will take the common vertex which can cover 2 or more edges. After that if some more edges are not covered, then we will take 1 of the vertex from the edge. In order to find the articulation point, we can go for Depth First Search (DFS) algorithm, which takes $O(V + E)$ time to find the articulation points. There are many more algorithms available for finding articulation points. DFS is one of the simplest algorithms.

Proposed algorithm to find the vertex cover of graph is shown below.

Algorithm 1.2 Proposed Vertex Cover Algorithm

-
1. $C \leftarrow \phi$
 2. $C' = \text{ArticPointDFS}(\text{vertex});$
 3. $C = C \cup C'$
 4. Repeat Step 2 and Step 3 until all the Articulation Point is found;
 5. $E' = E[G];$
 6. $E' = E' - \{\text{set of edges covered by } C'\}$
 7. While $E' \neq \phi$
- Let (a, b) be an arbitrary edge, check $(a, b) \cap \{\text{take each edge of } E'\} \neq \phi$ Then,
 add any of vertices ' a ' or ' b ' in C .
 Remove edge (a, b) from E'
 Else add the common vertex.
 Remove edges containing that vertices from E'
-
8. Return C .
-

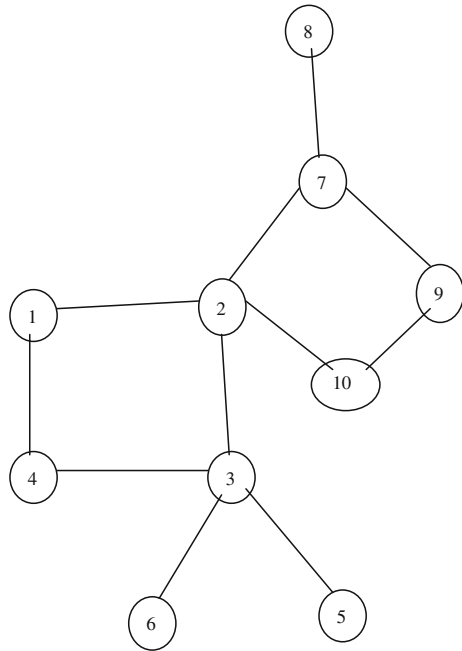
4 Analysis

Now comparing both the algorithms, let us take 1 case of finding vertex cover. Consider the following example. It contains 10 vertices and 13 edges. All vertices are connected with edges as shown in Fig. 1. Now, we will apply the vertex cover algorithms to the following example with both the approaches (i.e., existing approach and proposed approach). After that result will be compared for example, and output from each algorithm will be compared. The algorithm is computed step by step as per the algorithms discussed above.

Using existing algorithm:

1. $C \leftarrow \phi$
2. $E' = \{(1, 2), (2, 3), (3, 4), (1, 4), (2, 7), (2, 9), (2, 10), (3, 5), (3, 6), (7, 8), (7, 9), (7, 10), (9, 10)\}$
3. Let us take an arbitrary edge (9, 10)
4. $C = \{9, 10\}$
5. $E' = \{(1, 2), (2, 3), (3, 4), (1, 4), (2, 7), (2, 9), (2, 10), (3, 5), (3, 6), (7, 8)\}$
6. Let us take another arbitrary edge (1, 4)
7. $C = \{1, 4, 9, 10\}$
8. $E' = \{(2, 3), (2, 7), (2, 9), (2, 10), (3, 5), (3, 6), (7, 8)\}$

Fig. 1 Vertices are connected with each other, and a complex graph is formed



9. Let us take another arbitrary edge (3, 5)
10. $C = \{1, 3, 4, 5, 9, 10\}$
11. $E' = \{(2, 7), (7, 8)\}$
12. Let us take another arbitrary edge (7, 8)
13. $C = \{1, 3, 4, 5, 7, 8, 9, 10\}$

Finally, vertex cover contains 8 elements.

Using proposed algorithm:

1. $C \leftarrow \phi$
2. $E' = \{(1, 2), (2, 3), (3, 4), (1, 4), (2, 7), (2, 9), (2, 10), (3, 5), (3, 6), (7, 8), (7, 9), (7, 10), (9, 10)\}$
3. Find articulation points, $C' = \{2, 3, 7\}$
4. $C = \{2, 3, 7\}$
5. $E' = E' - \{\text{set of edges covered by } C'\}$
6. $E' = \{(1, 4), (9, 10)\}$
7. Take (1, 4) as an arbitrary edge.
8. $C = \{1, 2, 3, 7\}$
9. $E' = \{(9, 10)\}$
10. Take (9, 10) as an arbitrary edge.
11. $E' = \phi$
12. $C = \{1, 2, 3, 7, 9\}$

Finally, vertex cover contains 5 elements.

As we can see the result from existing approach and from proposed approach, it is clear that the existing approach can give us the true result but that will not be minimal. Some vertices that are not needed will also be included in the vertex cover list, while the proposed approach will have less number of vertices in the vertex cover list and which covers all the edges. Using this approach, first we will collect all the articulation point details and will include that vertices into vertex cover list. After that we will go with normal approach for remaining vertices, and the result will be as shown in above example. The implementation of the above approach is discussed in following section.

5 Implementation

Implementation of proposed algorithm can be done in 2 steps:

- Step 1: Find the articulation points of the graph given using articulation point algorithm (Using DFS algorithm). Add articulation points in the set of vertex cover. For example, given above, after applying algorithm, we can get following output (Fig. 2).

Fig. 2 Step 1 of implementation

```

Enter Number of vertices and Edges Respectively
10 13
Enter edges(e.g vertex1 vertex2)
1 4
1 2
3 4
2 3
3 5
3 6
2 9
2 10
2 7
7 9
7 10
7 8
7 10
Articulation Points are:2 3 7

```

Step 2: Remove the edges which are adjacent to all the vertices listed in Step 1. Apply approx vertex cover algorithm on remaining graph (Disconnected Graph). For example, given above, we have 4 choices $\{(1, 9) \text{ or } (1, 10) \text{ or } (4, 9) \text{ or } (4, 10)\}$. We can choose any combination.

Step 3: Take union of vertices found in Steps 1 and 2.

6 Conclusion

From above example, we can see that using existing algorithm, we are getting 8 elements in vertex cover set. While using proposed algorithm, we are getting 5 elements. Also, it is same as minimal vertex cover.

If V is a vertex cover set derived from Algorithm 1.1 and V^* is an optimal vertex cover set, then

$$|V| \leq 2|V^*|$$

If C is a vertex cover set derived from Algorithm 1.3 and C^* is an optimal vertex cover set, then

$$|C| \approx |C^*|$$

Proposed algorithm takes $O(2(V + E))$ time complexity. $O(V + E)$ for DFS and $O(V + E)$ for finding vertex cover. So, total $O(V + E)$ time to compute vertex cover is same as the available algorithm, but our algorithm provides much nearer solution to optimal solution.

7 Future Work

In this proposed algorithm, we are using DFS algorithm to find the articulation points of the graph. This algorithm takes $O(V + E)$ time to compute the articulation points. We can go for some other techniques to find the articulation points which can take less time compared to DFS. Also, we can find some other techniques which always provide the exact solution to optimal solution. Also, some more techniques can be used to provide nearer solution. There is also possible to find algorithm which runs faster than this algorithm.

Acknowledgment We would like to thank the School of Computer Science and Engineering, VIT University, for giving us such an opportunity to carry out this research work and also for providing us the requisite resources and infrastructure for carrying out the research.

References

1. M. Garry, D. Johnson, *Computers and Intractability: A User Guide to the Theory of NP Completeness* (San Francisco, 1979)
2. P.S. Oliveto, X. Yao, J. He, Analysis of Population-based Evolutionary Algorithms for the Vertex Cover Problem (IEEE, 2008), pp. 1563–1570
3. L. Ding, B. Gu, X. Hong, B. Dixon, Articulation node based routing in delay tolerant networks, in IEEE International Conference (2009), pp. 1–6
4. Y. Zeng, D. Wang, W. Liu, A. Xiong, An approximation algorithm for weak vertex cover problem in IP network traffic measurement, IEEE International Conference (2009), pp. 182–186
5. J. Chen, I.A. Kanj, G. Xia, Improved parameterized upper bounds for vertex cover. 31st International Conference on Mathematical Foundations of Computer Science (2006)
6. F. Delbot, C. Laforest, A better list heuristic for vertex cover. *Inf. Process. Lett.* **107**, 125–127 (2008)
7. E. Angel, R. Campigotto, C. Laforest, Algorithm for the vertex cover problem on large graphs. IBISC Research report (2010)
8. R. Bar-Yehuda, S. Even, A local-ratio theorem for approximating the weighted vertex cover problem. *Ann. Discrete Math.* **25**, 27–46 (1985)
9. G.L. Nemhauser, L.E. Trotter, Vertex packing: structural properties and algorithms. *Math. Program.* **8**, 232–248 (1975)
10. E. Asgeirsson, C. Stein, *Vertex Cover Approximations on Random Graphs* (Springer, Berlin, 2007), pp. 285–296
11. P. Erdos, A. Renyi, On random graphs. *Publ. Math. Debrecen* **6**, 290–297 (1959)
12. F. Kuhn, M. Mastrolilli, Vertex cover in graphs with locally few colors (2011), pp 498–509
13. D. Hochbaum, Approximation algorithms for the set covering and vertex cover problems. *SIAM J. Comput.* **11**(3), 555–556 (1982)

A Simple Control Strategy Technique for a Single-phase Bridgeless Active Rectifier with High Power Factor and Voltage Stabilization Using Partial Digital Implementation

Rahul Ganpat Mapari and D.G. Wakde

Abstract A partial digital implementation approach to improve the power factor of single-phase rectifiers and to regulate the output voltage against the change in line voltage and load is presented in this paper. A two-leg configuration, which has single IGBT in each leg, is adopted to reduce the number of switching devices compared with conventional AC–DC converter. This converter topology is evaluated on the basis of performance, and its salient features such as simplicity, low cost, and high performance are discussed to analyze its applicability. The proposed control strategy using continuous switching pulse width modulation (CSPWM) is bridgeless and transformer-less. A control technique and operational procedure are also developed, both theoretically and experimentally. The experimental results clearly verify the theoretical analysis from the prototype connected to grid unity.

Keywords Active rectifier (AC–DC) · Boost rectifiers · Continuous switching pulse width modulation (CSPWM) · Power factor · Single phase · Voltage regulation

1 Introduction

In this paper, a proposed approach to improve the power factor of single-phase rectifiers and to regulate the output voltage against the change in grid voltage and load is presented. This converter topology is evaluated on the basis of performance, and its salient features such as simplicity, low cost, and high performance are

R.G. Mapari (✉)

Electronics Department-Director, Sant Gadgebaba Amravati University, Amravati, India
e-mail: rahul_mapari272153@yahoo.com

D.G. Wakde

P.R. Patil College of Engineering, Amravati, Maharashtra, India
e-mail: dr_dgwakde@rediffmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_3

discussed to analyze its applicability. The proposed control strategy is bridgeless, transformer-less, and output current sensor-less and consists of only two bi-directional IGBTs and two diodes. The voltage regulation is achieved by a simple voltage divider to communicate to a controller to control the duty cycles of PWM. A control technique and operational procedure are also developed, both theoretically and experimentally. The experimental results clearly verify the theoretical analysis from the prototype connected to grid unity.

The single-switch rectifier has one of the simplest circuit structures. Typical voltage and current waveforms for the circuit, using hysteresis current control, are represented in [5–7]. Hysteresis band is made large in the figure for illustrative purposes. The two-switch rectifier [8, 9] performs the same switching action as the single-switch rectifier but has the advantage of higher efficiency.

Conventionally, AC–DC converters, which are also called rectifiers, are developed using diodes and thyristors to provide controlled and uncontrolled DC power with unidirectional and bi-directional power flow [10]. They have the demerits of poor power quality in terms of injected current harmonics, caused voltage distortion, and poor power factor at input AC mains and slow varying rippled DC output at load end, low efficiency, and large size of AC and DC filters [11]. The simplest line-commutated converters use diodes to transform the electrical energy from AC to DC. The use of thyristors allows for the control of energy flow. The main disadvantage of these naturally commutated converters is the generation of harmonics and reactive power.

Vienna rectifier is a three-switch, unity power factor boost rectifier. This rectifier operates by having the input stage creating a DC voltage across the two switches connected to the primary transformer [12–14]. The Vienna rectifier, even though it operates with only three switches, endures higher stresses than that of six-switch converter. This type of converter, however, has issues with start-up over current, as well as lack of current limiting during overload conditions.

Generally, the control structure of a three-phase six-switch PWM boost converter consists of an inner current control loop and outer voltage control loop [15]. The current controller senses the input current and compares it with a sinusoidal current reference. To obtain the current reference, the phase information of the utility voltage or current is required. This information is obtained by employing a phase lock loop (PLL), which creates transients if the frequency ratio changes [16]. To simplify the control structure, one-cycle-control (OCC)-based AC-to-DC converter has been proposed [17–19]. However, the scheme based on OCC exhibits instability in operation when magnitude of the load current falls below a certain level or when the converter is operating in the inverting mode of operation. To avoid it a modified OCC, bi-directional high-power factor AC-to-DC converter is proposed in [20]. This scheme uses saw-tooth wave to generate PWM pulses which incorporate low-frequency harmonics [21].

OCC presents some drawbacks intrinsic with its physical realization: the controller and its parameters cannot be modified without hardware re-design; moreover, they are influenced by temperature drifts, typical of analog systems. Another disadvantage is the need of both voltage and current measurements [22]. To overcome

these limitations, the OCC technique is implemented digitally using field programmable gate array (FPGA) [23]. This system uses PLL to find phase information of utility voltage and current. Another drawback of this system is that controller takes integer numbers only. The split operation is limited only to dividing number by a power of two.

This paper presents a simple control strategy which removes most of the drawbacks present in the classical methods for a single-phase active rectifier consists of only two bi-directional switches and two diodes. The grid voltage is directly delivered to the switches without using bridge and transformer.

The use of bridge and transformer created losses which reduces the efficiency of total system. Also in most of the paper, output voltage regulation is achieved using output current sensor; in this prototype without using current sensor, results are achieved. The prototype model is validating for 0.1, 0.3, 0.5, 0.7, and 1 kW experimentally. The voltage regulation is achieved for 250 VDC output voltage.

2 Circuit Configuration

The circuit shown in Fig. 1 consists of a prototype AC–DC converter with two bi-directional switches and two diodes. The bi-directional switches are connected at the lower side of the legs.

It is assumed that the anode-to-cathode voltage of each diode of the rectifier as well as the voltage drop through bi-directional switches is of negligible quantity. Figure 2 shows the waveforms of the considered power circuit. The bi-directional switches, $T1$ and $T2$, are turned on at an appropriate interval, conducting a partial line current. As a result, the input voltages, V_{Ao} , V_{Bo} —the voltages between each input terminal (A , B) and the zero point of the main source—become the staircase waveforms, and the input current waveforms become similar to the sinusoidal input voltages; in this case, it is clear that the power factor is improved.

Fig. 1 Circuit of prototype AC–DC converter with two bi-directional switches

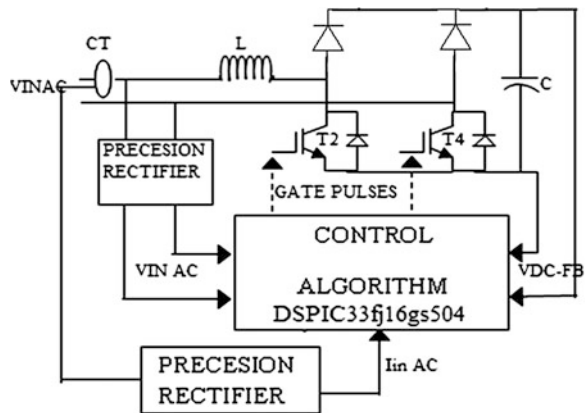
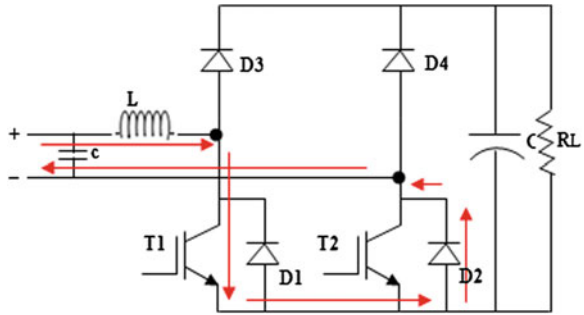


Fig. 2 Configuration of PFC technique when $T1$ is ON



The characteristics of the rectifier can be controlled by adjusting the duty cycle D , where D is a constant value between 0 and 1 which determines the pulse width for $Q1$ and $Q2$. The phase angle between the supply voltage and the fundamental component of the rectifier input voltage is also controlled by the duty ratio. Therefore, we can improve the characteristics of the considered rectifier by selecting a suitable D and the phase angle.

3 Operation Sequence

Based on line frequency, period of each half cycle of line voltage is 10 ms. To achieve more accurate results and high switching frequency, each half cycle is divided into 200 parts. Each part is of 50- μ s period. Four different configurations describing the sequence of operation to achieve high power factor and regulating the output voltage are as follows.

3.1 Case 1: Positive Half Cycle, $T1$ is ON

When $T1$ switch is turned on, the configuration is shown in Fig. 2. In this case, during the ON period of 200 parts, current IL is flowing through the inductor to charge it. The voltage across the inductor (Ldi/dt) is less than the DC output voltage.

3.2 Case 2: Positive Half Cycle, $T1$ is OFF

The configuration is shown in Fig. 3. During the OFF period of 200 parts, $T1$ is turned OFF. Maximum peak voltage across the inductor is greater than the DC output voltage. This voltage is moved toward the load through diode $D3$. The inductor acts as a boost inductor.

Fig. 3 Configuration of PFC technique when $T1$ is OFF

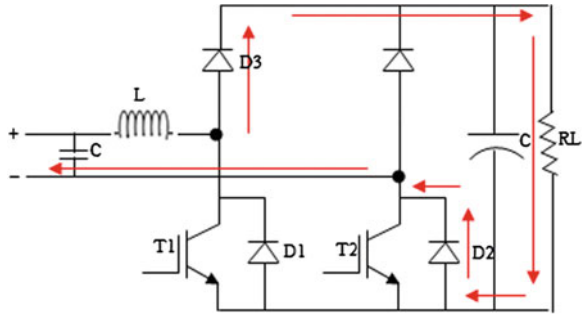
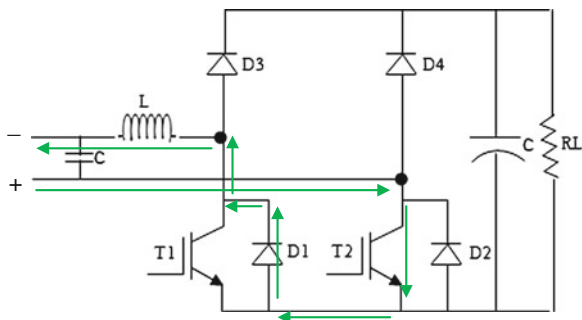


Fig. 4 Configuration of PFC technique when $T2$ is ON



3.3 Case 3: Negative Half Cycle, $T2$ is ON

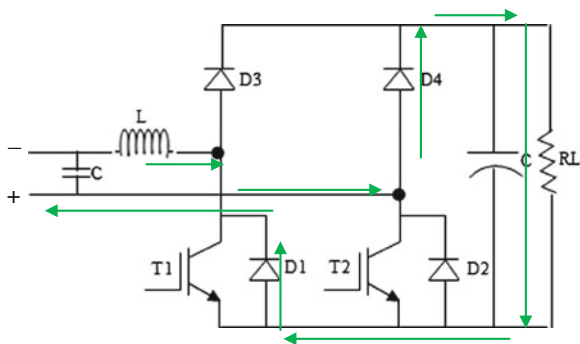
When $T2$ switch is turned on and $T1$ is turned OFF, the configuration is shown in Fig. 4. In this case, during the ON period of 200 parts, current I is flowing from positive terminal through switch $T2$ toward the inductor to charge it. The voltage across the inductor (Ldi/dt) is less than the DC output voltage.

3.4 Case 4: Negative Half Cycle, $T2$ is OFF

The configuration is shown in Fig. 5. During the OFF period of 200 parts, $T2$ is turned OFF. Maximum peak voltage across the inductor is greater than the DC output voltage. This voltage is move toward the load through diode $D4$. The inductor acts as a boost inductor.

The output DC voltage is controlled using a continuous switching digital technique. By taking the feedback of actual DC output voltage, the controller takes the action to regulate it. Here, we perform a simple program to adjust the duty ratio D according to the actual DC voltage. The DSPIC33fj16gs504 controller is used to generate the switching signals. A 10-bit ADC is used; hence, total count is 1,024.

Fig. 5 Configuration of PFC technique when T_2 is OFF



For the safe margin, we consider it is 70 % as 700. The regulated voltage we consider is 200 V. The threshold of 10 V is provided. The feedback voltage VDCFB can be mathematically represented as,

$$VDCFB = \frac{\text{ADC result count}}{\text{ADC full scale count}} \times 500 \quad (1)$$

The duty ratio D is directly proportional to the VDCFB. The ON time of the duty cycle is scaled by VDCFB factor.

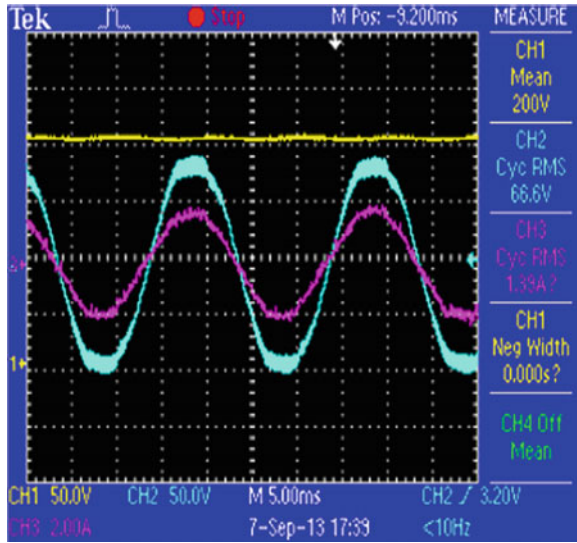
$$T_{ON} = K \times D \quad (2)$$

According to the ZCD status (positive or negative half cycle), controller decides switching of IGBTs. By default, duty cycle is 10 %. As per VDCFB, the duty cycle increases step by step in each 50 μ s to achieve desired VDC.

4 Experimental Results

The following are details of the experimental laboratory prototype: The rated output power (P_{out}) = 1 kW, V_{rms} = 90 V, L = 2 mH, C = 440 μ F, and utility grid frequency f = 50 Hz.

Fig. 6 The input voltage V_s and current waveforms I_s



4.1 Input Power Factor and Current Waveforms

The prototype is implemented at different levels of loads 0.2, 0.4, 0.7, and 1.0 kW. In Fig. 6, the input current is and the input voltage V_s is depicted. Notice that the experimental waveforms are similar to the ones obtained with the simulation.

Figure 7 shows that the line current is becoming a more sinusoidal, and it follows the line voltage at any load. Figure 8 describes the power factor remains approximately same for all loads.

4.2 Regulating the Output Voltage at Load Variance

The proposed control method keeps the output voltage stable even though the load is suddenly changed to different levels during operation. This is evaluated by the settling time in Fig. 9, where the settling time at the load variance from 100 to 300 W is approximately 55 ms. The experimental results illustrate that the proposed approach has been accurately analyzed with respect to efficiency and stability.

Fig. 7 The input voltage V_s and current waveforms I_s with different loads, **a** 400 W, **b** 700 W, **c** 1,000 W

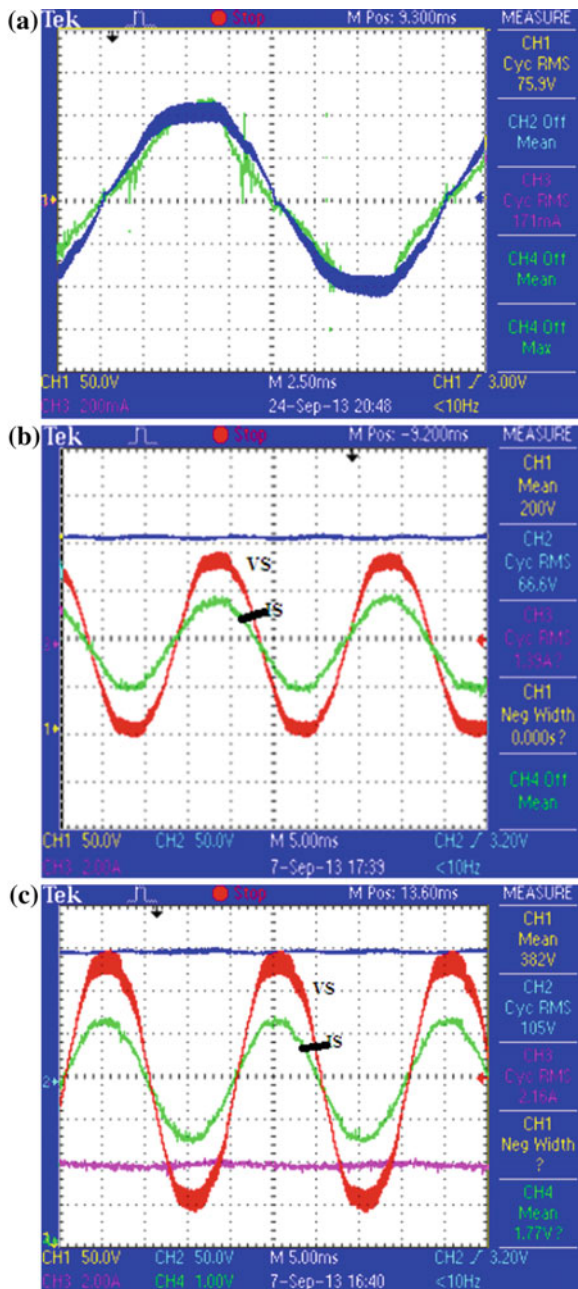


Fig. 8 Theoretical results of power factor versus load

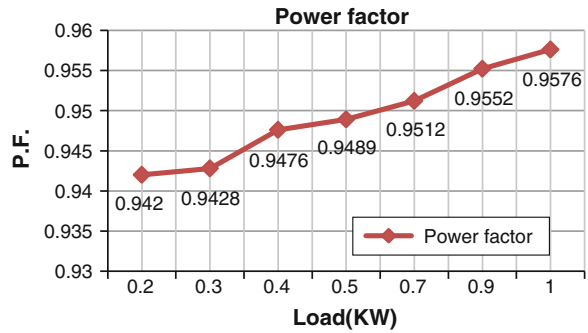
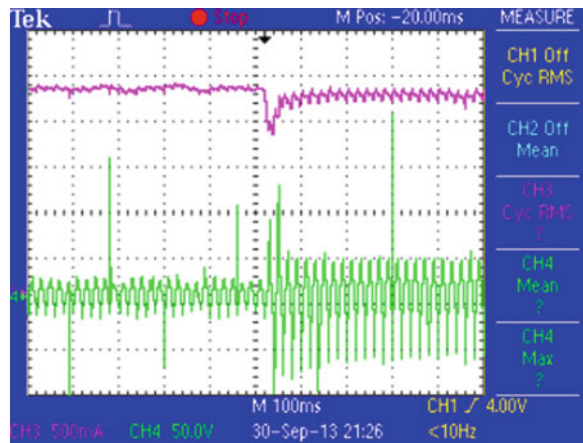


Fig. 9 Step load changes from 100 to 300 W, upper part is DC voltage and lower part is source current



5 Conclusion

In this paper, a proposed approach to improve the power factor of single-phase rectifiers and to regulate the output voltage is presented. The proposed control strategy is bridgeless, transformer-less, and output current sensor-less, and consists of only two bi-directional IGBTs and two diodes. As a result, the proposed control method is able to improve the power factor of the single-phase rectifier connected with grid unity, increase the quality of the harmonics of the input current, and stabilize the output voltage.

References

1. Z. Yang, P.C. Sen, Recent developments in high power factor switch mode converters, in *IEEE Proceedings CECE'98* (1998), pp. 477–480
2. R.W. Erickson, S. Cuk, R.D. Middlebrook, Large-signal modelling and analysis of switching regulators, in *Proceedings IEEE PESC'82* (1982), pp. 240–250
3. R. Erickson, M. Madigan, S. Singer, Design of a simple high-power-factor rectifier based on the flyback converter, in *Proceedings IEEE APEC'90* (1990), pp. 792–801
4. D. Simonetti, J. Sebastian, J. Uceda, A small-signal model for sepic, Cuk, and flyback converters as power factor pre-regulators in discontinuous conduction mode, in *Proceedings IEEE PESC'93*, pp. 735–741 (1993)
5. I.F. Schlect, B.A. Miwa, Active power factor codon for switching power supplies. *IEEE Trans. P.E.* **2**(4), 273–281 (1987)
6. M. Kazerani, P.D. Ziogas, G. Joos, A novel active c-t wave shaping technique for solid-state input power factor conditioners. *IEEE Trans. Ind. Electron.* **38**(1), 72–78 (1991)
7. A.R. Rasad, P.D. Ziogas, S. Manias, A novel passive wave shaping method for single-phase diode rectifiers. *IEEE Trans. Ind. Electron.* **37**(6), 521–530 (1990)
8. R. Itoh, K. Ishizaka, Single-phase sinusoidal convertor using MOSFETS. *IEE PIOC* **136**(5), 237–242 (1989)
9. A.W. Green, J.T. Boys, Cumnt forced single-phase reversible rectifier. *IEE PIOC* **136**(5), 205–212 (1989)
10. W.M. Grady, M.J. Samotyj, A.H. Noyola, Survey of active power line conditioning methodologies. *IEEE Trans. Power Delivery* **5**, 1536–1542 (1990)
11. B. Singh, B.N. Singh, A. Chandra, A. Pandey, A review of single-phase improved power quality AC–DC converters. *IEEE Trans. Ind. Elect.* **50**(5), 962–981 (2003)
12. J.W. Kolar, U. Drofenik, F.C. Zach, VIENNA rectifier II-a novel single-stage high-frequency isolated three-phase PWM rectifier system. *IEEE Trans. Ind. Electron* **46**(4), 674–691 (1999)
13. A.D. Pathak, R.E. Locher, H.S. Mazumdar, 3-phase power factor correction using vienna rectifier approach and modular construction for improved overall performance, efficiency and reliability. *Power electronics conference in long bench* (2003)
14. J. Kolar, F. Zach, A novel three phase utility interface minimizing line current harmonics of high power telecommunication rectifier modules. *IEEE Trans. Ind. Electron.* **44**(4), 456–467 (1997)
15. D.C. Lee, D.S. Lim, AC voltage and current sensorless control of three-phase PWM rectifiers. *IEEE Trans. Power Electron.* **17**(6), 883–890 (2002)
16. H.W. Van Der Broeck, H.-C. Skudelny, G.V. Stanke, Analysis and realization of a pulsewidth modulator based on voltage space vector. *IEEE Trans. Ind. Appl.* **24**(1) (1988)
17. M.K. Smedley, S. Cuk, One cycle control of power converters. *IEEE Trans. Power Electron,* **10**(6), 625–633 (1995)
18. Y. Chen, K.M. Smedley, One-cycle-controlled three-phase grid connected inverters and their parallel operation. *IEEE Trans. Ind. Appl.* **44**(2), 663–671 (2008)
19. C. Yang, K.M. Smedley, Parallel operation of one-cycle controlled three-phase PFC rectifiers. *IEEE Trans. Ind. Electron.* **54**(6), 3217–3224 (2007)
20. D. Ghodke, K. Chattarjee, Modified one cycle controlled bidirectional high-power-factor AC-to-DC converter. *IEEE Trans. Ind. Electron.* **55**(6), 2459–2472 (2008)
21. D. Ghodke, E.S. Shreeraj, K. Chattarjee, B.G. Farnandis, One-cycle controlled bi-directional Ac–Dc converter with constant power factor. *IEEE* (2008)
22. D.V. Ghodke, E.S. Shreeraj, K. Chatterjee, B.G. Fernandes, One cycle controlled bi-directional Ac to Dc converter with constant power factor, in *Proceedings IEEE Power Electronics Specialist Conference* (2008)
23. M. Barbati, C. Calusi, C. Cecati, One-cycle controlled active rectifier for full digital implementation, in *Proceedings IEEE* (2010)

Ensemble Neural Network Algorithm for Detecting Cardiac Arrhythmia

S. Aruna and L.V. Nandakishore

Abstract Cardiac arrhythmias are electrical malfunctions in rhythmic beating of the heart. Sometimes, they cause life-threatening conditions. Hence, they need to be diagnosed quickly and accurately to save life and prevent further complications and effective management of the disease. In this paper, we propose an ensemble neural network algorithm to detect arrhythmia. Bagging approach with multilayer perceptron and radial basis neural networks is used to classify the standard 12-lead Electrocardiogram (ECG) recordings in the cardiac arrhythmia database available in UCI Machine Learning Repository. The classification performance of the diagnostic model was analyzed using the following performance metrics, namely precision, recall, F-measure, accuracy, mean absolute error, root mean square error, and area under the receiver-operating curve. The classifier accuracy obtained for the ensemble neural network (ENN) model is 93.9 and 94.9 % for ENN-RBFN and ENN-MLP, respectively.

Keywords Bagging · Cardiac arrhythmia · Correlated feature selection · Multi-layer perceptron · Radial basis function neural networks

1 Introduction

Cardiac arrhythmias are abnormal rhythmic activity of heart. Arrhythmias are categorized into two types, those starting in the upper two chambers (atria or auricles) and those starting in the lower two chambers (ventricles). Arrhythmias

S. Aruna (✉)

Department of Computer Science, A.M. Jain College, Meenambakkam,
600114 Chennai, India
e-mail: arunalellapalli@yahoo.com

L.V. Nandakishore

Department of Mathematics, M.G.R Educational and Research Institute University,
Maduravoyal, 600095 Chennai, India
e-mail: lvnandaishore@gmail.com

© Springer India 2015

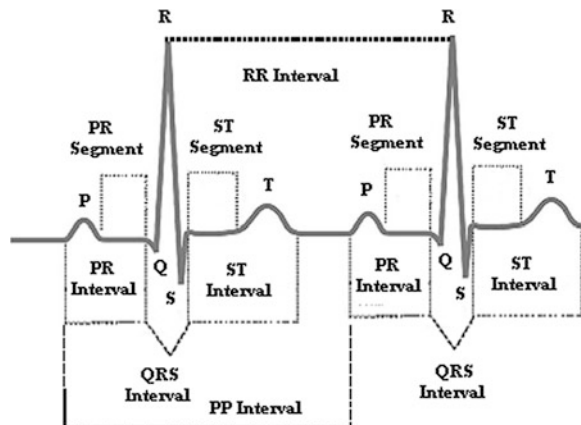
L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_4

starting at ventricles are more serious than those starting at auricles. The normal resting heart rate is about 60–100 beats per minute (bpm). According to the speed of the heart rate, arrhythmias can be further categorized into bradycardia, tachycardia, premature contraction, and fibrillation. Arrhythmias represent a serious threat to the patient recovering from acute myocardial infarction, especially ventricular arrhythmias such as ventricular tachycardia (VT) and ventricular fibrillation (VF) [1]. In the USA, more than 850,000 people are hospitalized every year for arrhythmia [2]. Ventricular tachycardia kills an estimated 120,000 people in the UK each year [3]. The detection of arrhythmia is an important task in clinical reasons which can initiate life-saving operations [4]. Some patients do not have any symptoms of arrhythmia. They are diagnosed during routine examination. Treatment varies depending on the type of arrhythmia. Several methods for automated arrhythmia detection have been developed in the past few decades to simplify the monitoring task [5].

Electrocardiogram (ECG) records the electronic activities of the heart and has been widely adopted for diagnosing cardiac arrhythmia [6]. The state of cardiac health is generally reflected in the shape of the ECG waveform and heart rate and contains important pointers to the nature of the disease attacking the heart [7]. Figure 1 shows the normal ECG signal. The normal ECG signal is described by P -QRS- T waves. P wave records the atrial depolarization. QRS complex records the ventricular depolarization. T wave records the repolarization of the ventricles. PP interval records the atrial rate. RR interval records the ventricular rate. PR interval measures the AV node function. ST interval is measured from the point at which QRS complex finishes with the end of T wave. PR segment connects P and QRS complex. ST segment represents the period of depolarization of the ventricles.

In this paper, we propose an ensemble neural network (ENN) algorithm based on bagging to diagnose cardiac arrhythmia from 12-lead ECG recordings. The rest of the paper is organized as follows. Section 2 gives details about the dataset used for the experiment and ensemble neural network algorithm. Section 3 gives the results obtained. Concluding remarks are given in Sect. 4 to address further research.

Fig. 1 Normal ECG signal



2 Materials and Methods

2.1 Dataset Description

The cardiac arrhythmia database used in this study was obtained from the UCI Machine Learning Repository [8]. The dataset has 452 instances of 16 classes. Class 01 is normal, classes 02–15 represent different types of cardiac arrhythmias, and unclassified data come under class 16. Each instance has 279 attributes, of which first four attributes, namely age, sex, height, and weight, give the general description of the patient. Remaining attributes are extracted from the standard 12-lead ECG recordings. There are 206 linear-valued attributes, and the rest is nominal-valued attributes.

2.2 Ensemble Neural Network Algorithm

The ENN algorithm is based on the bagging approach of the ensemble classification method. Bagging is a statistical resample and combine technique [9] based on bootstrapping and aggregating techniques. Bootstrap resampling technique generates multiple versions of the predicting model. Then, the aggregating technique combines those together [10]. Bagging reduces the variance for the classifier. Artificial neural networks (ANN), namely multilayer perceptron (MLP) and radial basis function neural networks (RBFN), were used as base classifiers for the diagnostic model. ANN is mathematical models inspired by biological neural networks where nodes represent neurons and arcs represent axons, dendrites, and synapses. MLP is a feedforward neural network with three layers, the input layer, one or more hidden layers, and output layer. The training and testing vectors presented to the input layer are processed by hidden and output layers. The computational capabilities of MLP are presented by Lippman [11]. RBF networks have a static Gaussian function as the nonlinearity for the hidden layer processing elements and the Gaussian function, responds only to a small region of the input space where the Gaussian is centered [12]. The key to a successful implementation of these networks is to find suitable countries for the Gaussian functions [13]. Figure 2 shows the ENN diagnostic model.

The sequence of steps in the ENN algorithm is as follows:

Step 1: Calculation of mean and mode for all the attributes in the training set.

Step 2: Replacement of missing values of the attributes using the values obtained in Step 1.

Step 3: Calculation of correlation coefficients for all the attributes.

Step 4: Removal of attributes with low-class correlation coefficients.

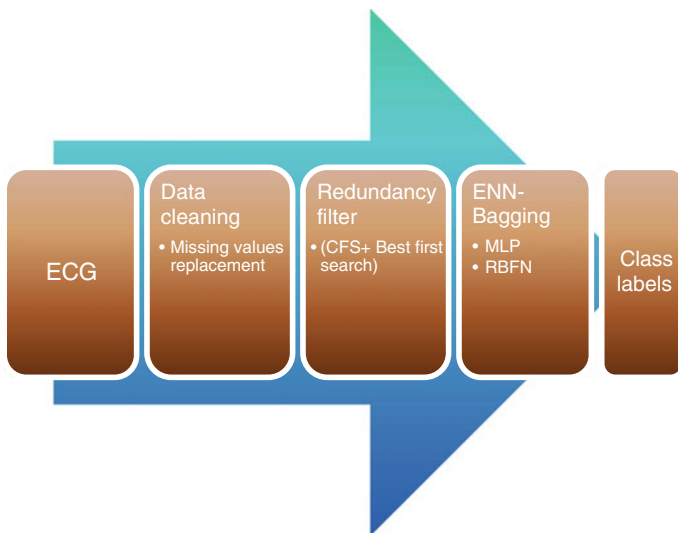


Fig. 2 ENN diagnostic model

Step 5: Selection of best feature subsets of the remaining attributes using best first search criterion.

Step 6: Building predictor model from the feature subset obtained in Step 5 using bootstrap resampling with aggregation for MLP and RBFN classifiers.

Step 7: Prediction of class labels using the model built in Step 6.

The ENN diagnostic model has two phases, data preprocessing phase and classification phase. In the data preprocessing phase, two filters are used. The arrhythmia database has noisy, redundant, and about 0.33 % of missing values which may cause error in classification. Hence, attribute-based filter is first applied to replace the missing values with the modes and means of the training data. Then, to remove the redundant attributes from the dataset correlation feature selection (CFS), redundancy filter is applied. Correlation coefficients for all the attributes were computed using Eq. 1 where CL is the correlation between the summed feature subsets and the class variable, N is the total number of subset attributes, A_C is the average of the correlations between the class variable and the subset of attributes, and A_I is the average intercorrelation between a subset of attributes.

$$CL = \frac{N\overline{A_C}}{\sqrt{N + N(N - 1)\overline{A_I}}} \quad (1)$$

Feature subsets having high class correlation and low intercorrelation are selected using best first search criterion after removing features with low CL values. Finally, the resultant feature subset is classified using the bagging-based ENN approach with MLP and RBFN as base classifiers.

3 Results

WEKA [14], Java-based data mining tool, is used for conducting the experiments with tenfold cross-validation. Tenfold cross-validation has been proven to be statistically good enough in evaluating the performance of the classifier [15].

3.1 Performance Metrics

The performance criterion for the diagnostic model is analyzed by computing precision, recall, F -measure, accuracy, mean absolute error (MAE), root mean square error (RMSE), and area under the ROC (AUC) from the confusion matrix. The precision is the computational measure of predictive accuracy of a particular class. Recall is the measure of positive samples predicted as positive. Accuracy is the ratio of the predictions that are correct. F -Measure. MAE is the statistical measure of how far is the estimated value from the actual value. RMSE measures the difference between the actual values and the values predicted by the model. Precision, recall, F -measure, accuracy, MAE, and RMSE are calculated using Eqs. 2–7.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (3)$$

$$F - \text{Measure} = 2 * \frac{\text{precision} - \text{recall}}{\text{precision} + \text{recall}}. \quad (4)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \quad (5)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |P_i - C_i|. \quad (6)$$

$$\text{RMSE} = \sqrt{\langle (M_i - C_i)^2 \rangle}. \quad (7)$$

TP is the true positives, FP is the false positives, TN is the true negatives, FN is the false negatives, N is the number of instances, P_i is the prediction value of the instance i , C_i is the class value of the instance i , and M is the measured value of the instance i . AUC analyzes the variance independent of the decision's sensitivity. The AUC is obtained by the nonparametric method based on the Wilcoxon's trapezoidal rule to approximate the area [4].

3.2 Diagnosis of Cardiac Arrhythmia

Cardiac arrhythmia database is classified by both linear and ensemble classifiers for neural networks. From 450 attributes after replacing the missing values, CFS selected 14 attributes among 279 attributes. The total number of subsets examined is 3,289. Merit of the best subset found is 0.592, and then, bagging-based ENN-MLP and ENN-RBFN are used for classification of the dataset with 14 attributes. Bag size percent used is 100. Five hundred epochs are used for ENN-MLP. For ENN-RBFN, the ridge value for regression set is $1.0E-8$, minimum standard deviation is 0.1, and the number of k -means cluster is 2. Table 1 shows the comparison results for arrhythmia classification by MLP, RBFN, ENN-MLP, and ENN-RBFN classifiers. Figure 3 shows the ROC for ENN-MLP and ENN-RBFN classifiers. A receiver-operating characteristic (ROC) curve is constructed by plotting false-positive rate versus the true-positive rate for varying cutoff values.

ROC analysis originated in electrical engineering in the early 1950s where the technique was developed to assess the performance of signal detection devices (receivers) and later spread into other fields, finding useful applications in both psychology and medical diagnosis [16]. From the results, apparently ENN-MLP achieved better classification performance than all the other classifiers. The results infer that ensemble approach improved the classification performance of both MLP and RBFN classifiers.

Table 1 Cardiac arrhythmia classification results

Performance metrics	Linear classifiers		Ensemble classifiers	
	RBFN	MLP	ENN-RBFN	ENN-MLP
Precision	0.89	0.71	0.94	0.95
Recall	0.88	0.62	0.94	0.95
F -measure	0.88	0.66	0.94	0.95
Accuracy (%)	88.14	62.03	93.9	94.9
MAE	0.15	0.45	0.11	0.06
RMSE	0.30	0.47	0.23	0.22
AUC	0.82	0.49	0.92	0.94

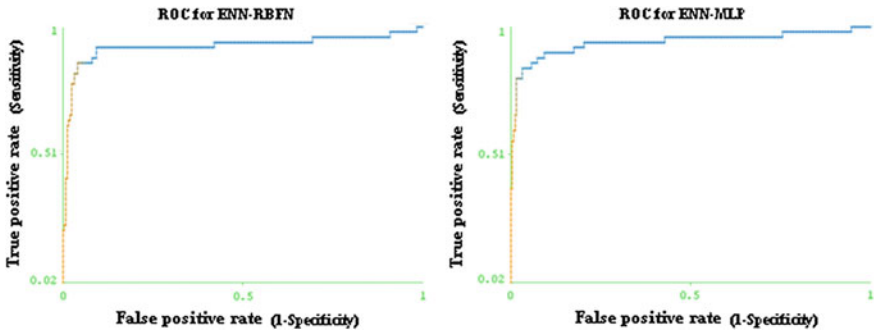


Fig. 3 ROC for ENN-RBFN and ENN-MLP for ten iterations

3.3 Related Work

In [17], the authors used neural network with weighted fuzzy membership function to classify cardiac arrhythmia with an accuracy of 81.32 %. In [18], Elsayad used learning vector quantization neural networks and obtained the classification accuracy of 76.92 %. Zuo et al. [4] used kernel difference KNN to detect arrhythmia. They achieved an accuracy of 70.66 %. In [20], authors used weighted fuzzy artificial immune recognition system for medical diagnosis. They obtained an accuracy of 80.71 % for ECG arrhythmia database. Uyar et al. [22] classified arrhythmia using a serial fusion of support vector machines and logistic regression with an accuracy of 76.1 %. In [23], authors used a novel pruning approach using expert knowledge for data running for diagnosing the arrhythmia with an accuracy of 68.47 %. Ozcan [24] used fuzzy support vector machines for ECG arrhythmia classification. For adaptive neuro fuzzy inference system (ANFIS) and fuzzy

Table 2 Comparison of classification accuracies of our method and other methods in the literature

Model	Accuracy (%)	
NEWFM [17]	81.32	
HLVQ [18]	76.92	
KDFW-KNN [19]	70.66	
Fuzzy weighted AIRS [20]	80.71	
SVM with Gaussian kernel [22]	76.10	
A novel pruning approach [23]	68.47	
ANFIS	79.43	
FSVM-DTCM [24]	83.33	
MNN model	82.22	
GFFNN model	82.35	
MLP model [25]	86.67	
Our method	ENN-RBFN	93.90
	ENN-MLP	94.91

support vector machines distance to class mean method they got accuracy of 79.43 and 83.33 %, respectively. In [25], authors proposed an effective ANN-based approach for cardiac arrhythmia classification. They got classification accuracy of 82.22, 82.35, and 86.67 % for modular neural network model, generalized feed-forward neural network model, and multilayer perceptron model, respectively (Table 2).

4 Conclusion

Cardiac arrhythmias are the irregular rhythm of the heart show a serious medical problem sometimes threatening the life. Detection of arrhythmia is an important task in effective management of the disease. Automated system for arrhythmia detection helps the physicians in simplifying the task. In this paper, we propose an ensemble neural network algorithm based on the bagging approach for detecting the presence or absence of arrhythmia. Multilayer perceptron and radial basis function neural networks are used as base classifiers for the proposed diagnostic model. The performance of the model was evaluated using the following metrics, namely precision, recall, F -measure, accuracy, mean absolute error, root mean square error, and area under ROC. The experiments were conducted using the WEKA data mining tool with tenfold cross-validation. Cardiac arrhythmia database from the UCI Machine Learning Repository is used for the study. Ensemble methods achieved a classification accuracy of 94.9 and 93.9 % for ENN-MLP and ENN-RBFN, respectively. In the present work, the ensemble classifier is used for detecting the presence or absence of arrhythmia. The future work will be concentrated in classifying different types of arrhythmias. Different types of ensemble approaches with neural networks will be compared with the present model for a broader experimental evaluation and further enhancement of the algorithm.

References

1. K.H. Ryu, A data mining approach and framework of intelligent diagnosis system for coronary artery disease prediction, IEICE technical report (2008), pp. 33–34
2. H.G. Lee, K.H. Ryu, A data mining approach for coronary heart disease prediction using HRV features and carotid arterial wall thickness. *Bio. Eng. Inform.* 200–206 (2008)
3. R.D. Raut, S.V. Dudul, Arrhythmias classification with MLP neural network and statistical analysis. First IEEE International Conference on Emerging Trends in Engineering and Technology (2008), pp. 553–558
4. W.M. Zuo, W.G. Lu, K.Q. Wang, H. Zhang, in *Computers in Cardiology* Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier (2008), pp. 253–256
5. D.M. Dubin, Rapid Interpretation of EKG's. (2001)
6. L. Breiman, Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996)

7. J. Carney, P. Cunningham, The neural BAG algorithm: optimizing generalization performance in Bagged Neural Networks, in *Proceedings of the 7th European Symposium on Artificial Neural Networks* (1999), pp. 35–40
8. R. Lippman, An introduction to computing with neural nets. *IEEE Trans. ASSP Mag.* **4**, 4–22 (1987)
9. M.D. Buhmann, *Radial Basis Functions: Theory and Implementations Cambridge Monographs on Applied and Computational Mathematics* (Cambridge University Press, Cambridge, 2003)
10. S.V. Chakravarthy, J. Ghosh, Scale based clustering using radial basis function networks, in *Proceeding of IEEE International Conference on Neural Networks* (1994), pp. 897–902
11. P. Baldi, S. Brunak, Y. Chauvin, Assessing the accuracy of prediction algorithms for classification: and overview. *Bioinformatics* **5**, 412–424 (2000)
12. J.A. Hanley, B. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) Curve. *Radiology* **143**, 29–36 (1982)
13. T. Fawcett, *ROC Graphs: Note and Practical Considerations for Data Mining Researchers*, HP Labs Technical Report, (2003)
14. S.H. Lee, J.K. Uhm, J.S. Lim, Extracting input features and fuzzy rules for detecting ecg arrhythmia based on NEWFM. *International Conference on Intelligent and Advanced Systems*
15. A.M. Elsayad, Classification of ECG arrhythmia using learning vector quantization neural networks, in *Proceedings of Computer Engineering Systems* (2009), pp. 139–144
16. K. Polat, S. Şahan, S. Güneş, A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. *Expert Syst. Appl.* **31**, 264–269 (2006)
17. A. Uyar, F. Gurgun, Arrhythmia classification using serial fusion of support vector machines and logistic regression, in *Proceedings of Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2007)* (2007), pp. 560–565
18. A.M. Mahmood, M. Rao Kuppa, A novel pruning approach using expert knowledge for data specific pruning. *Eng. Comput.* **28**, 21–30 (2012)
19. N.O. Ozlem, F. Gurgun, Fuzzy support vector machines for ECG arrhythmia detection, in *Proceedings of Pattern Recognition (ICPR 2010)* (2010), pp. 2973–2976
20. S.M. Jadhav, S.L. Nalbalwar, A.A. Ghatol, Artificial neural network models based cardiac arrhythmia disease diagnosis from ECG signal data. *Int. J. Comput. Appl.* **44**, 8–13 (2012)

An Efficient Invasive Weed Optimization Algorithm for Distribution Feeder Reconfiguration and Loss Minimization

K. Sathish Kumar, K. Rajalakshmi, S. Prabhakar Karthikeyan
and R. Rajaram

Abstract The distribution network carries electricity from the transmission system and delivers it to consumers. Distribution losses account for major part of power system losses. The low-voltage operation in the distribution system is a major reason for higher technical losses due to inherent properties of the network. In this paper, a method based on invasive weed optimization algorithm (IWOA) is proposed for distribution network reconfiguration with the objective of real power loss minimization. The feeder reconfiguration problem is formulated as a nonlinear optimization problem, and IWOA is used to find the optimal solution. The proposed method is implemented on standard 16-bus test system [1]. Test results show that the proposed feeder reconfiguration method can effectively ensure the loss minimization [2].

Keywords Invasive weed optimization algorithm (IWOA) · Feeder reconfiguration

1 Introduction

The part of power system in which electric power is distributed among various consumers for their local use is known as distribution system. Electricity distribution is the final stage in the delivery of electricity to end users. The distribution network includes medium-voltage power lines, substations and pole-mounted

K. Sathish Kumar (✉) · S. Prabhakar Karthikeyan · R. Rajaram
School of Electrical Engineering, VIT University, Vellore, India
e-mail: kansathh21@yahoo.co.in

S. Prabhakar Karthikeyan
e-mail: spk25in@yahoo.co.in

K. Rajalakshmi
Department of Electronics Engineering, Alpha College of Engineering, Anna University,
Chennai, India
e-mail: rajii_sure@yahoo.co.in

transformers, low-voltage distribution wiring, and sometimes meters. Distribution network is designed with mesh structure but operated in a radial configuration. Reconfiguration also relieves the overloading of the network components [3]. Switches in distribution network are categorized into sectionalizing switches (normally closed) and tie switches (normally open). Reconfiguration of feeder entails altering the topological structure of distribution feeders by changing the open/close status of the switches under both normal and abnormal operating conditions [4].

In this paper, invasive weed optimization algorithm (IWOA) [5] is suggested to reconfigure the distribution system with the objective of reducing the real power losses. IWOA is a recently developed evolutionary algorithm inspired from growth of plant weeds in agricultural land. The algorithms include mathematical programming and artificial intelligent methods, such as refined genetic algorithms (RGA) [6], ant colony search algorithm (ASCA) and GA [7], heuristic approach (HA) [8], adaptive genetic algorithms (AGAs) [9], bacterial foraging optimization algorithm (BFOA) [10], honeybee mating optimization (HBMO) [11], and fuzzy adaptive particle swarm optimization (FAPSO) [12], are proposed to reconfigure the distribution feeders with the objective of minimizing real power losses while avoiding transformer and feeder overloads and inadequate voltages.

A radial distribution load flow method [13] is used to compute voltage profile and power flows under steady-state operating conditions for the chosen test systems. This procedure is followed by symmetrical fault analysis which yields fault voltages along with angles at each bus in the corresponding bus systems. A post-fault load flow solution is carried before the inception of IWO algorithm. After implementation of IWOA, the losses are found to be reduced in both the systems.

2 Problem Formulation

The main aim of current optimization problem is real power loss minimization which can be expressed in mathematical form [14] as follows:

$$\Delta P = \operatorname{Re}\left\{2 \sum_{i \in D} I_i (E_m - E_n)^*\right\} + R_{\text{loop}} \left| \sum_{i \in D} I_i \right|^2 \quad (2.1)$$

subject to the following constraints

1. No feeder section is left out of service.
2. Radial structure of system should always be retained.

$$\phi(i) = 0 \quad (2.2)$$

3. Bus voltage magnitude should be within the limits.

$$E_{\min} \leq E \leq E_{\max} \quad (2.3)$$

where D set of buses that are disconnected from feeder-II and connected to feeder-I M tie bus of feeder-I through which loads from feeder-II will be connected n tie bus of feeder-II that will be connected to bus m through a tie switch I_i complex bus current at bus i . R_{loop} series resistance connecting the two substation buses of feeder-I and feeder-II via closure of the specified tie switch.

E_m complex voltage at node m and E_n complex voltage at node n .

Equation (2.1) represents the objective function to be minimized using the optimization technique. In Eq. (2.1), ΔP refers to net power loss, whereas E_m and E_n correspond to voltages of nodes at higher and lower potential. Equation (2.2) gives the voltage constraints of each node, and Eq. (2.3) preserves the radial characteristics in 16-bus test systems.

3 Invasive Weed Optimization Algorithm

The IWO algorithm is a common phenomenon in agriculture that is colonization of invasive weeds [15]. According to the common definition, a weed is a plant growing where it is not wanted. Any tree, vine, shrub, or herb may qualify as a weed, depending on the situation; generally, however, the term is reserved for those plants whose vigorous, invasive habits of growth pose a serious threat to desirable, cultivated plants. This is a simple and effective algorithm in converging to optimal solution by employing basic properties, such as seeding, growth, and competition, in a weed colony.

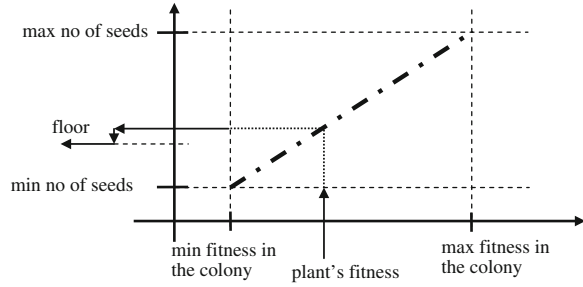
3.1 Steps Involved in IWO Algorithm

To simulate colonizing behavior of weeds, the basic properties of the process to be considered are as follows: (1) initializing a population, (2) reproduction, (3) spatial dispersal, and (4) competitive exclusion. The process is explained in detail as follows:

3.1.1 Initialize a Population

A population of initial solutions is spread over the d -dimensional problem space in random positions.

Fig. 1 Seed production procedure in a colony of weeds



3.1.2 Reproduction

A member of the population of plants is allowed to produce seeds on its own with colony's lowest and highest fitness: The number of seeds produced by each plant increases linearly from minimum possible seed production to its maximum. In other words, a plant will produce seeds based on its fitness, the colony's lowest and highest fitness with increase in linearity. Figure 1 illustrates this procedure.

3.1.3 Spatial Dispersal

Randomness and adaptation in the algorithm are given in this part. The generated seeds are randomly distributed in d -dimensional search space by normally distributed random numbers with mean equal to zero, by varying the variance. This means that seeds will be randomly distributed such that they abode near to the parent plant. However, standard deviation (SD), σ , of the random function will be minimized from previously defined initial value, σ_{initial} , to a final value σ_{final} , in every step (generation). In simulations, a nonlinear alteration shows satisfactory performance, which is given in Eq. (3.1) where iter_{max} is the maximum number of iterations, σ_{iter} is the standard deviation in the present time step, and n is the nonlinear modulation index.

$$\sigma_{\text{iter}} = \frac{(\text{iter}_{\text{max}} - \text{iter})^n}{(\text{iter}_{\text{max}})^n} (\sigma_{\text{iter}} - \sigma_{\text{final}}) + \sigma_{\text{final}} \quad (3.1)$$

3.1.4 Competitive Exclusion

If a plant leaves no offspring, then it would go extinct; otherwise, they would take over the world. Thus, there is a need of some competition between plants for limiting maximum number of plants in a colony. After passing some iterations, the number of plants in a colony will reach its maximum by fast reproduction; however, it is expected that the fitter plants will reproduce more-than-undesirable plants. By reaching the maximum number of plants in the colony, P_{max} a mechanism for eliminating the plants of poor fitness in the generation activates. The elimination

mechanism works as follows: When the maximum number of weeds in a colony is reached, each weed is allowed to produce seeds according to the mechanism mentioned in Sect. 3.1.2. The produced seeds are then allowed to spread over the area according to Sect. 3.1.3. When all seeds are found in that search area, they are ranked together with their parents (as a colony of weeds). Next, weeds with lower fitness are eliminated to reach the maximum allowable population in a colony. In this manner, plants and offspring are ranked together and the one having better fitness survives, which then allowed to replicate. This mechanism gives a chance for the plants with lower fitness to reproduce, and if their offspring has a good fitness in the colony can survive. The population control mechanism is also applied to their offspring at the end of a given run, realizing competitive exclusion.

4 Implementation of Invasive Weed Optimization Algorithm to Reduce the Power Loss

The steps in IWO algorithm are implemented for loss minimization as follows:

Step 1: Initializing the parameters used in optimization process:

d is the dimensional search space represents the number of switching configurations for 16-bus test systems. n is the nonlinear modulation index and σ_{initial} is the initial standard deviation. σ_{final} is the final standard deviation that is specified as [0.0000000000000001].

$$\sigma_{\text{final}} \ll 0.01\sigma_{\text{initial}} \quad (4.1)$$

S_{max} is the maximum allowable seed population. S_{min} is the minimum allowable seed population. iter_{max} is the maximum number of iterations. y is the initial size of seed matrix, i.e., number of busses and $y =$ matrix size [16] for 16-bus system.

Step 2: Obtain post-fault voltage values of each switching configuration from the load flow program.

Step 3: Initialization of seed matrix containing voltage values from load flow solution with the tolerance of 5 %.

Step 4: Check for voltage constraints, if $V > V_{\text{max}}$, and make $V = V_{\text{max}}$.

Step 5: After the adjustment of voltage magnitude, load flow is carried out to determine bus currents I_i at each bus in the test system.

Step 6: Calculation of initial fitness function (ΔP) from Eq. (2.1), which determines net power loss.

Step 7: Rank the ΔP values of the fitness matrix in descending order.

Step 8: Reproduction loop calculates the number of seeds for each fitness value depending upon the rank.

$$S_{\Delta P(i)} = S_{\max} * \frac{(\Delta P_{\text{best}} - \Delta P(i))}{(\Delta P_{\text{best}} - \Delta P_{\text{worst}})} \quad (4.2)$$

Step 9: Calculation of σ_{iter} from Eq. (3.1) for spatial dispersion of seeds produced in the previous step.

$$\sigma_{\text{iter}} = \frac{(\text{iter}_{\max} - \text{iter})^n}{(\text{iter}_{\max})^n} (\sigma_{\text{initial}} - \sigma_{\text{final}}) + \sigma_{\text{final}}$$

Step 10: Spatial dispersal loop from $i = 1, 2, \dots$ and where ind represents the final seed of corresponding ΔP value in the fitness matrix.

$$\text{Computer}(i) = \text{rand} * \sigma_{\text{iter}}, \quad (4.3)$$

where rand represents any value between 0 and 1 and also $\text{se} = 1, 2, \dots, \text{ind}$.

$$\text{compute seed}(\text{se}) = \text{seed}(\text{se}) + r(i) \quad (4.4)$$

Step 11: Competitive exclusion for every value of Δf in the fitness function and check whether $S > S_{\max}$ if true $S = S_{\max}$.

Step 12: Colonization with the new reproduced seeds, colony of weeds along with their parents, is formed as follows.

$$V_{\text{new}} = V + \text{seed} \quad (4.5)$$

Step 13: Perform Step 4 again.

Step 14: Perform Step 5 again.

Step 15: Calculate the final fitness value (ΔP) from Eq. (2.1). Also, search for global best.

Step 16: Check for maximum number of iterations, and if exceeded, print global best and stop; otherwise, go back to Step 3.

5 Test System

A standard 16-bus test system is chosen for the application of proposed method. The system diagram is shown in Fig. 2. The system data for the chosen 16-bus test system are tabulated in Table 1.

This system operates with tie switches that are open in normal conditions [16–18]. The loads are energized by all the feeders, and changing the on/off status of tie switches will cause transfer of loads from one feeder to the other [19–21].

Fig. 2 16-bus distribution network

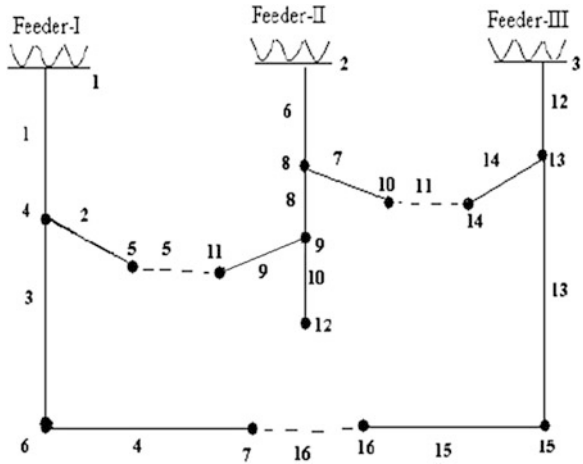


Table 1 System data for 16-bus distribution network

From	To	Resistance (Ω)	Reactance (Ω)	Load (MW)	Load (Mvar)
1	4	0.075	0.1	2.0	1.6
4	5	0.08	0.11	3.0	1.5
4	6	0.09	0.18	2.0	0.8
6	7	0.04	0.04	1.5	1.2
2	8	0.11	0.11	4.0	2.7
8	9	0.08	0.11	5.0	3.0
8	10	0.11	0.11	1.0	0.9
9	11	0.11	0.11	0.6	0.1
9	12	0.08	0.11	4.5	2.0
3	13	0.11	0.11	1.0	0.9
13	14	0.09	0.12	1.0	0.7
13	15	0.08	0.11	1.0	0.9
15	16	0.04	0.04	2.1	1.0
5	11	0.04	0.04	-	-
10	14	0.04	0.04	-	-
7	16	0.09	0.12	-	-

6 Results and Discussion

The proposed method is implemented on 16-bus test systems. The results in terms of power loss with various switching positions are tabulated for 16-bus test system in Table 2. In the 16-bus system, reconfiguration has been performed to reduce the power loss without disturbing the radial structure of the system.

The results thus obtained are compared with various existing methods. For effective understanding, the comparison is tabulated for 16-bus system in Table 3.

Figure 3 depicts the convergence graph of power loss in 16-bus system. The computational time for each iteration is determined as 2.33 s in case of 16-bus test system.

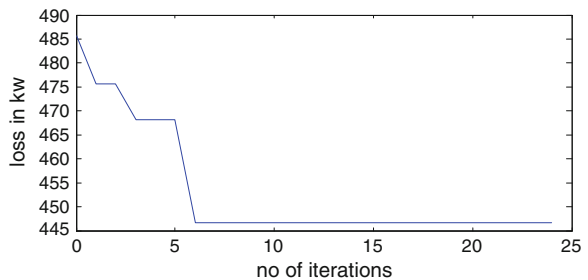
Table 2 Power loss results in 16-bus test system using IWOA with various switch statuses

Switch status		Power loss (kW)
Closed	Open	
1, 2, 3, 4, 6, 8, 10, 11, 14, 12, 13, 15	5, 7, 16	475.66
1, 2, 3, 6, 7, 8, 10, 12, 13, 14, 15, 16	5, 4, 11	445.14
1, 2, 3, 6, 8, 10, 11, 12, 13, 14, 15, 16	5, 7, 4	468.11

Table 3 Comparison with other methods using 16-bus network

Loss in base configuration—511.41 kW		
Methods	Switches open	Power loss (kW)
Proposed method	5, 4, 11	445.14
ACSA [20]	6, 9, 11	466.11
RGA [6]	6, 9, 11	466.11

Fig. 3 Convergence characteristics of power loss in 16-bus test system



7 Conclusion

In this paper, IWOA is proposed to reconfigure the distribution system with the objective of loss reduction. The objective function is formulated with voltage and security constraints. The two standard test systems have been chosen for implementation of IWOA and the obtained considerable minimization in power loss. The results obtained are validated with existing methods. These test results conclude that IWOA approach can be effectively used to reduce the power loss in a distribution system by reconfiguration.

Acknowledgments The authors would like to thank the management of VIT University for their encouragement and the support given during this work.

References

1. C. Lakshminarayana, M.R. Mohan, An improved technique for service restoration in distribution systems using non-dominated sorting genetic algorithm. *Int. J. Electr. Power Energy. Syst* **31**(3), 162–170 (2011)
2. E.R. Sanseverino, Minimum losses reconfiguration of MV distribution networks through local control of tie-switches. *IEEE Trans. Power Deliv.* **18**(3), 762–771 (2003)
3. Q. Zhou, D. Shirmohammadi, W.-H.E. Liu, Distribution feeder reconfiguration for service restoration and Load balancing. *IEEE Trans. Power Syst.* **12**(2), 724–729 (1997)
4. D. Das, A fuzzy multiobjective approach for network reconfiguration of distribution systems. *IEEE Trans. Power Deliv.* **21**(1), 202–209 (2006)
5. A.R. Mehrabian, C. Lucas, A novel numerical optimization algorithm inspired from weed colonization. *Int. J. Ecol. Inform.* **1**, 355–366 (2006)
6. J.Z. Zhu, Optimal reconfiguration of electric distribution network using refined genetic algorithm. *Electr. Power Syst. Res.* **62**, 37–42 (2002)
7. Y.K. Wu, C.Y. Lee, L.C. Liu, S.H. Tsai, Study of reconfiguration for the distribution system with distributed generators. *IEEE. Trans. Power Deliv.* **25**(3) (2010)
8. J.A. Martin, A.J. Gil, A new heuristic approach for distribution systems loss reduction. *Electr. Power Syst. Res.* **78**(11), 1953–1958 (2008)
9. A. Swarnkar, N. Gupta, K.R. Niazi, Minimal loss configuration for large scale radial distribution systems using adaptive genetic algorithms, in 16th National Power Systems Conference. (2010)
10. K. Sathish Kumar, T. Jayabarathi, Power system reconfiguration and loss minimization for an distribution systems using bacterial foraging optimization algorithm. *Int. J. Electr. Power Energy Syst.* (2011)
11. T. Niknam, An efficient multi-objective HBMO algorithm for distribution feeder reconfiguration. *Int. J. Expert Syst. Appl.* **38**, 2878–2887 (2011)
12. T. Niknam, E. Azadfarsani, M. Jabbari, A new hybrid evolutionary algorithm based on new fuzzy adaptive PSO and NM algorithms for distribution feeder reconfiguration. *Int. J. Energy Convers. Manage.* **54**, 7–16 (2012)
13. T. Thakur, J. Dhiman, A new approach to load flow solutions for radial distribution system. *IEEE PES Transmission and Distribution Conference and Exposition Latin America* (2006)
14. S. Civanlar, J.J. Grainger, H. Yin, S.S.H. Lee, Distribution feeder reconfiguration for loss reduction. *IEEE Trans. Power Deliv.* **3**(3), 1217–1223 (1988)

15. D. Shirmohammadi, W.H. Hong, Reconfiguration of electric distribution networks for resistive line loss reduction. *IEEE Trans. Power Deliv.* **4**(1), 1492–1498 (1989)
16. F.V. Gomes, S. Carneiro, J.L.R. Pereira, M.P. Vinagre, P.A.N. Garcia, A new heuristic reconfiguration algorithm for large distribution systems. *IEEE Trans. Power Syst.* **20**(3), 1373–1378 (2005)
17. S.K. Goswami, S.K. Basu, A new algorithm for the reconfiguration of distribution feeders for loss minimization. *IEEE Trans. Power Deliv.* **7**(3), 1484–1491 (1992)
18. T.E. McDermott, I. Drezga, R.P. Broadwater, A heuristic nonlinear constructive method for distribution system reconfiguration. *IEEE Trans. Power Syst.* **14**(2), 478–483 (1999)
19. F.V. Gomes, S. Carneiro Jr., J.L.R. Pereira, M.P. Vinagre, P.A.N. Garcia, L.R. De Araujo. A new distribution system reconfiguration approach using optimum power flow and sensitivity analysis for loss reduction. *IEEE Trans. Power Syst* **21**(4) (2006)
20. C.T. Su, C.F. Chang, J.P. Chiou, Distribution network reconfiguration for loss reduction by ant colony search algorithm. *Int. J. Electr. Power Syst. Res.* **75**, 190–199 (2005)
21. D.P. Kothari, R. Ranjan, K.C. Singal, *Renewable energy sources and technology* (Prentice Hall India 2011)

Implementation of Generative Crossover Operator in Genetic Algorithm to Solve Traveling Salesman Problem

Devasenathipathi N. Mudaliar and Nilesh K. Modi

Abstract The research work aims to solve symmetric traveling salesman problem more efficiently. In this research paper, a different crossover operator is proposed, which produces 18 valid offsprings from two parents. The performance of proposed crossover operator is compared with three other existing crossover operators by maintaining the selection technique, mutation technique, and fitness function identical. This crossover operator is tested with data from TSP dataset. The intercity distance table of cities in which distance is measured with L1 norm formed the input to the coded C program that implemented the proposed crossover operator. The same dataset was used to compare the performance of this crossover operator with other three crossover operators. The comparative study indicates that proposed crossover operator performs well compared to other crossover operators in solving traveling salesman problem.

Keywords Symmetric traveling salesman problem • Multiple offspring producing crossover operator • Performance of crossover operator • Intercity distance table • Fitness function

1 Introduction

Traveling salesman problem is an NP hard, combinatorial optimization problem, where a salesman must visit all the cities in his territory exactly once by covering least total distance. The distance between any two given cities (to and fro) is same,

D.N. Mudaliar (✉)
MCA Department, SVIT, Vasad, India
e-mail: devas_mca@yahoo.co.uk

D.N. Mudaliar
R & D Centre, Bharathiar University, Coimbatore, India

N.K. Modi
MCA Department, SVICS, Kadi, India
e-mail: drnileshmodi@yahoo.com

and so the traveling direction is not a hindrance. There exist many problems such as student group formation problem, genome sequencing, and vehicle routing that have traveling salesman problem structure [1–3]. Efficiently solving traveling salesman problem would solve many other problems related to it as no polynomial time algorithm can be formulated for this. Considering a bruteforce approach to solve this problem is infeasible.

Exact algorithms, tour construction, and tour improvement are some of the approaches to solve traveling salesman problem. Linear programming and branch and bound form the types of exact algorithms, while insertion heuristics, closest neighbor heuristics, and greedy heuristics are types of tour construction approach. Finally, tour improvement approach consists of genetic algorithms, ant colonization algorithms, tabu search, etc. However, the above-mentioned approaches and their types have their own set of opportunities and challenges.

Many researchers have applied genetic algorithm to solve the traveling salesman problem or problems having traveling salesman problem structure. Three operators, viz. selection, crossover, and mutation, are used in solving a problem by genetic algorithm. As the first step, some defined number of random feasible solutions (called population) is generated by the genetic algorithm. The selection operator in the genetic algorithm then selects the most fittable solutions (of some defined proportion from the randomly generated population using some fitness function) through various selection techniques such as tournament selection and roulette wheel selection. The filtered (or selected) solutions are now paired to produce new breed of chromosomes. A pair of parent solutions cross over to produce another set of offspring solutions. This process called crossover has various techniques to achieve the task. In case of traveling salesman problem, famous crossover techniques called partially mapped crossover, order crossover, cycle crossover, etc., exist. To bring variation in the offspring population (so that the solutions do not get trapped in the local minima), mutation operator is executed, which randomly changes a gene or two of offspring solutions. The above process from selection to mutation is repeated till a definite number of times or for the time indicative improvement occurs in the population set [4].

In this research work, the authors have tried to present a different crossover operator in genetic algorithm that is able to solve traveling salesman problem. The significance of this crossover operator is that it takes two valid parent solutions and produces 18 valid offspring solutions. Even though published research work exists for crossover work in genetic algorithm that produces more than two offspring solutions, they did not focus to solve the traveling salesman problem. In addition to this, it has to be brought to notice that traditional crossover operators cannot be applied to solve traveling salesman problem as they may end up with invalid solutions. The next section of the paper represents the work done by different researchers in connection with solving traveling salesman problem or its variants through different approaches. The third section details on the actual implementation of the research work. The fourth section elucidates on the results obtained and comparison of the results. The last section concludes the research paper with achievable future directions.

2 Proposed Methodology

Most of the two-point crossover operator in genetic algorithm takes in two parent chromosomes as input and produces two valid offspring chromosomes as output. Additionally, the two offspring chromosomes contain the features of both the parents. The authors propose a two-point crossover approach that produces 18 valid offspring chromosomes and all the offspring chromosomes contain the features of both the parents.

2.1 Example of M-Crossover Operator

We try to solve the Traveling Salesman Problem with 9 cities using two-point crossover operator.

First cut point—after third gene (even though a different value less than second cut point can be set).

Second cut point—after sixth gene

Parent 1—1 2 3 4 5 6 7 8 9

Parent 2—9 1 2 8 7 4 5 6 3

Using the given cut points, the parent chromosomes can be cut into three parts, viz.

Parent 1—[1 2 3] [4 5 6] [7 8 9]

Parent 2—[9 1 2] [8 7 4] [5 6 3]

2.2 Creating the First Offspring Chromosome

Inserting the first part of Parent 2 before the first part of Parent 1, we get

[9 1 2] [1 2 3] [4 5 6] [7 8 9]

Striking the matching chromosomes of Parent 2 part from Parent 1 parts, we get

[9 1 2] [~~1 2 3~~] [1 2 3] [4 5 6] [7 8 9]

Deleting the striked genes and grouping the rest of the genes according to the cut points, we get

[9 1 2] [3 4 5] [6 7 8]

By simply removing the partition, we get the first valid offspring—9 1 2 3 4 5 6 7 8.

2.3 Creating the Second Offspring Chromosome

Inserting the first part of Parent 2 after the first part of Parent 1 and before the second part of Parent 1, we get

[1 2 3] [9 1 2] [4 5 6] [7 8 9]

Striking the matching chromosomes of Parent 2 part from Parent 1 parts, we get

[~~1~~ ~~2~~ 3] [9 1 2] [4 5 6] [7 8 ~~9~~]

Deleting the striked genes and grouping the rest of the genes according to the cut points, we get

[3 9 1] [2 4 5] [6 7 8]

By simply removing the partition, we get the second valid offspring—3 9 1 2 4 5 6 7 8.

2.4 Creating Other Offspring Chromosomes

In the above manner, we create the remaining 7 more chromosomes with the help of our proposed crossover. The change that is to be followed in getting the remaining chromosomes is identifying which part of Parent 2 chromosome is to be put before which part of Parent 1 chromosome. This is illustrated with the help of Fig. 1.

Figure 1 represents that the first part of Parent 2 be inserted before the first part of Parent 1 and we get the first chromosomes after following the above steps. Following the above steps, the remaining 7 offspring chromosomes are as follows: 345 691 278, 874 123 569, 123 874 569, 123 568 749, 563 124 789, 125 634 789, and 124 563 789.

To obtain the next set of 9 offspring chromosomes, just interchange the contents of Parent 1 and Parent 2 and perform the above-mentioned steps. The following offspring chromosomes will be obtained by performing this step:

123 987 456, 912 387 456, 987 412 356, 456 912 873, 912 456 873, 912 874 563, 789 124 536, 127 894 563, and 124 789 563.

Once all 18 valid offspring chromosomes are obtained, two best chromosomes with respect to fitness value are selected and sent for further stage and the remaining chromosomes are simply ignored.

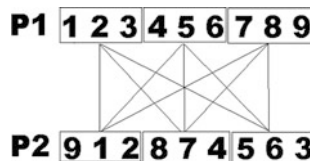


Fig. 1 Part of Parent 2 (P2) to be kept before part of Parent 1 (P1)

3 Actual Experiment

The proposed crossover operator was tested with test data from TSPLIB dataset. Four different C programs were developed to implement and compare the proposed crossover operator in genetic algorithm. All contents of the C programs were same except the crossover operator implementation part. We coded the following functions in C language to accomplish the above tasks.

1. Random initialization of population
2. Selection of chromosomes
3. Cloning of chromosomes
4. Crossover of chromosomes
5. Mutation of chromosomes.

The following steps describe C program coded to implement the experiment.

1. Initialize a random population of 100 valid chromosomes and perform steps 2–8 for 50 iterations.
2. Obtain the fitness value of each chromosome of population by the fitness function.
3. Select the best 50 % of chromosomes with respect to fitness value.
4. Clone the selected chromosomes by merely creating a copy of those chromosomes and add them to the population.
5. Randomly create pairs of chromosomes and send them for crossover.
6. The crossover results in two new valid offspring chromosomes for every pair of chromosomes sent for crossover.
7. Send 2 % of the entire newly obtained offspring chromosome for mutation (in our case, displacement mutation).
8. Obtain the fitness value of each chromosome and return to step 2.

The developed C programs were tested for test dataset (fri26_d.txt) which provides the intercity distance table for 26 cities. The TSP test data in intercity distance table of the dataset obtained from TSPLIB formed the input to the C programs [5]. The different crossover techniques were proposed: crossover (discussed in methodology section), order crossover, partially mapped crossover, and cycle crossover. The output and results of the experimental work are discussed in the next section.

4 Results and Discussion

As stated in the previous section, we have tried to evaluate the efficiency of our proposed crossover operator by comparing the results of the experiment with the existing crossover operators. Roulette wheel selection technique was used to select the chromosomes. Crossover rate was set to 0.9, and mutation rate was set to 0.02. The initial population was set to 100 valid chromosomes.

Table 1 Comparison of performance of proposed crossover with other crossover operators for 26-city traveling salesman problem

Number of iterations (generations)	Best fitness value obtained by proposed crossover C program	Best fitness value obtained by partially mapped crossover C program	Best fitness value obtained by order crossover C program	Best fitness value obtained by cycle crossover C program
0–10	1,261	1,857	1,724	1,802
11–20	1,144	1,674	1,665	1,721
21–30	1,054	1,606	1,602	1,871
31–40	937	1,616	1,567	1,849
41–50	1,051	1,541	1,549	1,790

Table 1 represents the best fitness value obtained up to a given number of iterations (generations) for 26-city problem (fri26_d.txt). It could be noted that output given by the C program of our proposed crossover approach gives best results quickly. The shortest distance for the 26-city problem (for this test data) obtained till now by researchers is 937, and within 50 iterations (generations), our proposed crossover approach is better close to the optimal solution compared to the other crossover techniques.

The obtained optimal path by m-crossover operator for 26-city problem (FRI26) is as follows: 8-7-5-6-4-2-3-14-15-12-13-11-10-16-9-19-20-18-17-21-26-22-24-23-25-1. The length of this path is 937 km according to the values given in the dataset.

5 Conclusion and Future Work

In this research paper, the authors have tried to propose a new crossover operator of genetic algorithm to solve symmetric traveling salesman problem. Alternately, the performance of the proposed work is compared with other existing crossover operators that aid to solve symmetric traveling salesman problem. C programs were coded to implement and compare the performance of proposed crossover operator and other three crossover operators. A test data from TSPLIB (fri26_d.txt 26-city problem) were considered for the same. The results of the experiment positively proved our proposed crossover approach to solve symmetric traveling salesman problem.

As part of the future work, we plan to implement the proposed crossover operator for more number of cities for symmetric traveling salesman problem. In addition to this, the same crossover approach can be used to solve asymmetric traveling salesman problem as well.

References

1. R. Agarwala, D.L. Applegate, D. Maglott, G.D. Schuler, A.A. Schäffer, A fast and scalable radiation hybrid map construction and integration strategy. *Genome Res.* **10**, 350–364 (2000)
2. D.N. Mudaliar, N.K. Modi. Evolutionary algorithm approach to pupils' pedantic accomplishment, in *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)*, Advances in intelligent systems and computing, vol. 199 (Springer, Berlin, 2013), pp. 415–423
3. R. Matai, S.P. Singh, M.L. Mittal, in *Traveling salesman problem: an overview of applications, formulations, and solution approaches*, Traveling Salesman Problem, Theory and Applications (InTech, Croatia, 2010), pp. 1–24
4. N. Bansal, A. Blum, S. Chawla, A. Meyerson, Approximation Algorithms for Deadline-TSP and Vehicle Routing with Time-Windows, in *Proceedings of ACM STOC* (2004), pp. 166–174
5. M. Ünal, Ak. Ayça, V. Topuz, H. Erdal, Genetic algorithm optimization of PID controllers using ant colony and genetic algorithm, *Studies in computational intelligence*. vol. 449 (Springer Berlin, 2013), pp. 19–29

A 4-bit 9 KS/s Distortionless Successive Approximation ADC in 180-nm CMOS Technology

P. Dipu, B. Saidulu, K. Aravind, Johny S. Raj and K. Sivasankaran

Abstract In recent years, analog-to-digital converters are the crucial part of many applications. In this paper, we proposed a 1.8 V capacitor-array-based successive approximation ADC. This SAR ADC uses bootstrapped switch to decrease distortion, and comparison is done using a pre-amplifier preceding a latched comparator. A 4-bit SAR ADC with high resolution was designed in 180-nm CMOS process. This paper aims at describing the design of a discrete-component, successive approximation register analog-to-digital converter (SAR ADC). The performance evaluation was done using Cadence ADE tool.

1 Introduction

In the present-day scenario, most of the applications in the electronic world exist in the digital domain. Since naturally occurring waveforms are truly analogous in nature, a need for converting these analog waveforms into its digital counterparts is of utmost importance. Thus, mixed-signal circuit theory plays an important role in the analysis and implementation of digital signal processing systems. Among mixed-signal digital systems, analog-to-digital conversion (ADCs) circuits are considered to be the inevitable part of system design. Through this paper, we are focusing on the implementation of successive approximation register ADC, which is a general type of ADC used for low-power and medium-resolution applications. Compared to other analog-to-digital conversion types, SAR ADC gives a variety of advantages. As compared to Flash ADC, which requires a huge number of comparators, SAR ADC requires only one comparator. Since only one comparator is used in its design, the demand for low power consumption has been achieved successfully.

P. Dipu (✉) · B. Saidulu · K. Aravind · J.S. Raj · K. Sivasankaran
School of Electronics Engineering (SENSE), VIT University, Vellore, Tamilnadu, India
e-mail: dipugovind@gmail.com

B. Saidulu
e-mail: bellamkonda.saidulu@gmail.com

SAR ADC provides an added advantage that it has very high accuracy and moderate resolution. It can be achieved by using a capacitor array in the place of a resistive array, used during data conversion. Considering the above-mentioned factors, SAR ADC can be regarded as an ideal component for some portable or battery-powered instruments. For example, SAR ADCs are the crucial part of many medical applications.

They are commonly used in implantable pacemakers for controlling the pattern and speed of heartbeats. The sole aim of this paper is to design and implement discrete component in SAR ADC in Cadence Virtuoso ADE Tool. Section 1 describes about the overall architecture of the SAR ADC and its components which includes switches, comparator, SAR logic, and the DAC. Section 2 consists of simulation results and its analysis, and the final section consists of conclusion and future aspects of the proposed design.

2 Architecture

The Fig. 1 shows the general block diagram of SAR architecture. The analog input is given to the sample and hold circuit switch which is a bootstrap switch, which samples the input at a rate of 9 KS/s. The held value is given as input to the comparator whose other input is DAC voltage. The output of the comparator is properly switched by a flip-flop whose clock period is double that of comparator clock period. These values are fed to the SAR logic whose values are controlled by proper set and clear. The outputs of the SAR logic are properly given to the DAC circuitry, capacitive array. In a single hold period, the DAC produces for different voltages depending upon the SAR output, which is compared with the hold value.

The different blocks of SAR ADC architecture are explained below.

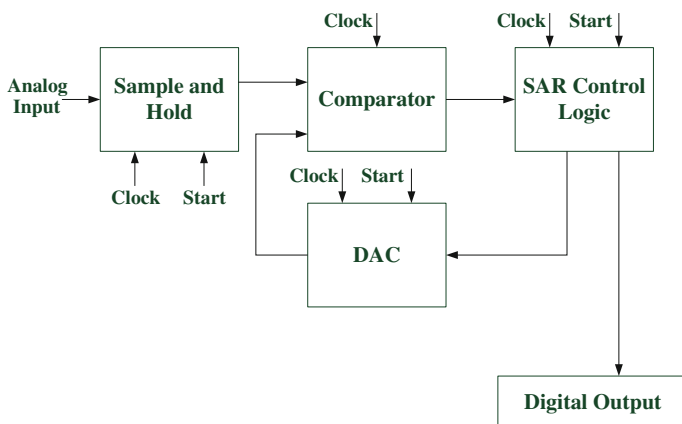
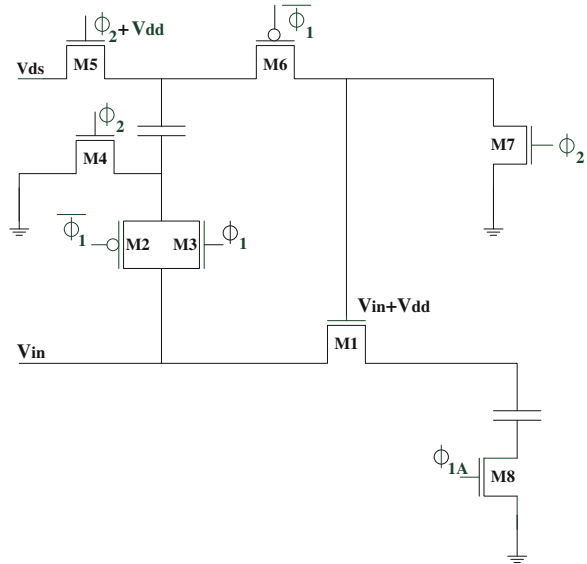


Fig. 1 SAR ADC architecture

Fig. 2 Bootstrap switch



2.1 Sample and Hold

The sample and hold circuit can be implemented using NMOS switch. But there are certain drawbacks in NMOS since the resistance (R_{on}) depends on the input voltages. Due to this reason, there is a possibility of getting harmonics in the sampled value (Fig. 2).

Bootstrap switch is one of the best options to avoid harmonics, where the resistance is independent of the input voltage (Fig. 3).

During the hold phase, Φ_1 goes low and Φ_2 goes high. The transistors M2, M3, and M6 are OFF. Transistors M4 and M5 are ON, and this charges the capacitor to V_{dd} . Here, the transistor M1 is OFF, and hence, the capacitor holds the signal. The Nakogome Charge pump generates the $\Phi_2 + V_{dd}$. These hold period values are compared with the DAC outputs in respective clock periods.

2.2 Comparator

There are various comparator designs proposed in order to compare the analog input with the DAC output. Here, we tried to implement a dynamic two-stage latched comparator. The comparator mainly consists of a voltage amplifier and a latch. When clock goes to zero, the transistor M1 gets OFF, at the same time M7 and M8 gets ON. The voltage at the respective drain terminal A and B of M7 and M8 is connected to the gate terminal of M10 and M11 and goes to the OFF state, so the output is obtained as zero when clock is zero. When clock goes high, the

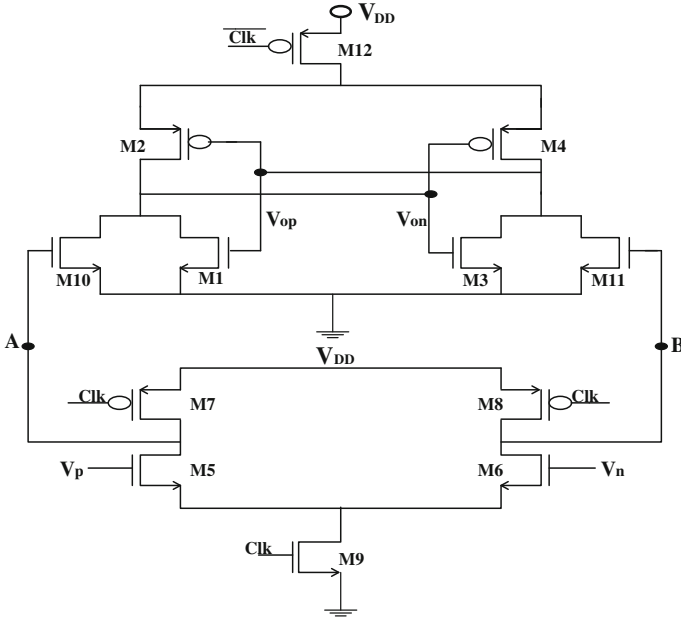


Fig. 3 Comparator

transistor M9 gets ON and M7 and M8 get OFF. Depending upon the input given at the Vp and Vn of M5 and M6, the voltage at A and B is fed back to gate terminal of M10 and M11. Depending upon the strength of the voltage at A and B, either M10 or M11 goes ON first and gets the respective output at Vop and Von (Fig. 4).

2.3 Successive Approximation Register Control Logic

The successive approximation register logic mainly consists of set and clear controlled D-FF. The D-FF is implemented as per the TSPCR architecture. The first section of SAR logic consists of a shift register whose output is connected to the D-FFs in control sections. The set and clear are given in such a way that the D-FF outputs are set to 1,000 initially. Depending upon the clock given the outputs are further shifted such as 0100, 0010, 0001 and at the same clock period, these outputs are given to the control section in order to set the D-FFs. Each input of the D-FF in the control section is connected to the comparator output terminal whose output is obtained by comparing the input hold values with the DAC outputs.

The control section logics are implemented in such a way that the output of each flip-flop is fed back to input of the preceding D-FF clock terminal and the outputs are obtained at the respective clock periods.

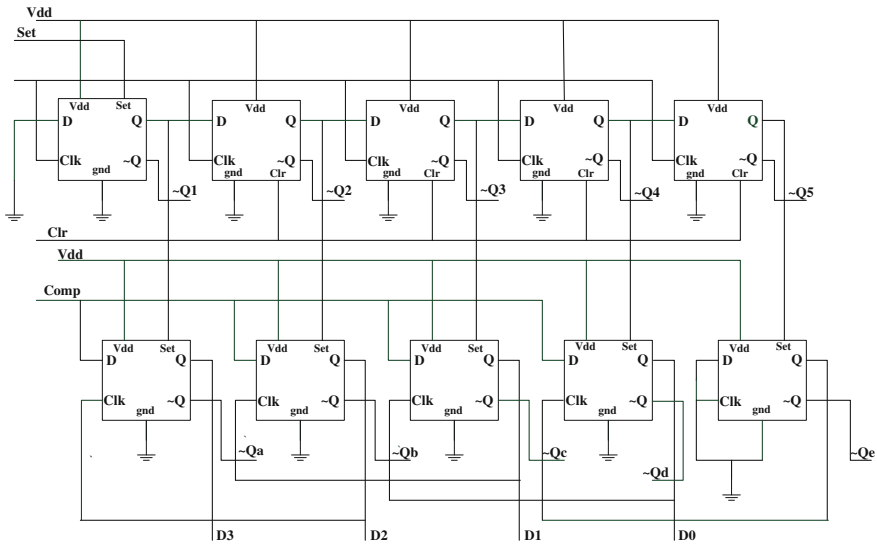


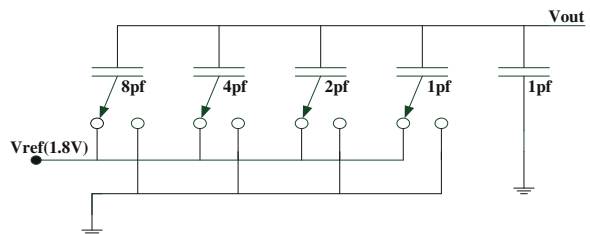
Fig. 4 SAR logic

2.4 Digital-to-analog Converter

DAC is one of the main components in SAR ADC. The SAR ADC speed mainly depends on setting time of DAC. Here, we have designed a DAC using capacitive array to get a better performance. Capacitive array is much faster than the resistive array. The capacitors are arranged in such a way that the value of the adjacent capacitor is half of the preceding one. The capacitors are driven by the SAR logic output. Depending upon the SAR logic output, the capacitors are either connected to V_{DD} or V_{SS} , and we will get different ratios of voltages. The sizing of capacitor array is done in such a way that kT/c noise is less than that of quantization noise of ADC. The capacitive array for a 4-bit ADC is shown below (Fig. 5).

The output of the DAC is given as input to the comparator whose value is compared with the input. The output voltage V_{out} can be represented as $V_{out} = KV_{ref}D$, where K is a scaling factor, V_{ref} is the reference voltage and D can be expressed as $D = D = \frac{b_1}{2^1} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \dots + \frac{b_N}{2^N}$, where N is the total number of bits.

Fig. 5 Capacitor array



3 Simulation Results

3.1 Sample and Hold Waveforms

The inputs are sampled at 9 KS/s, and the hold values are clocked to the comparator within the respective clocks. The output for the sampled and hold waveform is shown below (Fig. 6).

3.2 Comparator Waveforms

The clock period of the comparator is chosen as 15 us, and the respective V_{op} and V_{on} are shown below. It is observed that the V_{op} and V_{on} are always switched to zero whenever the clock value is zero (Fig. 7).

3.3 SAR Logic Waveform

The SAR logic clock period is taken as 30us, and by giving proper set and clear at 150 us, the cycles are repeated as shown below (Fig. 8).

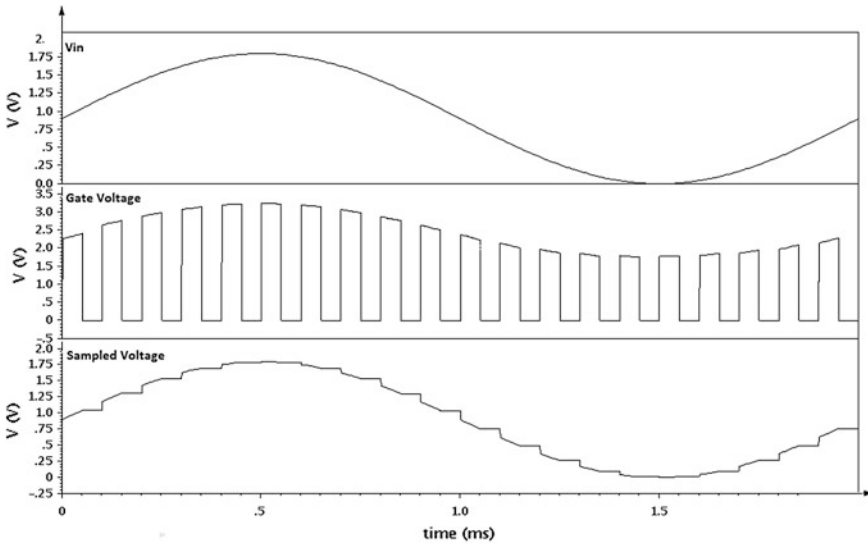


Fig. 6 Sample and hold waveform

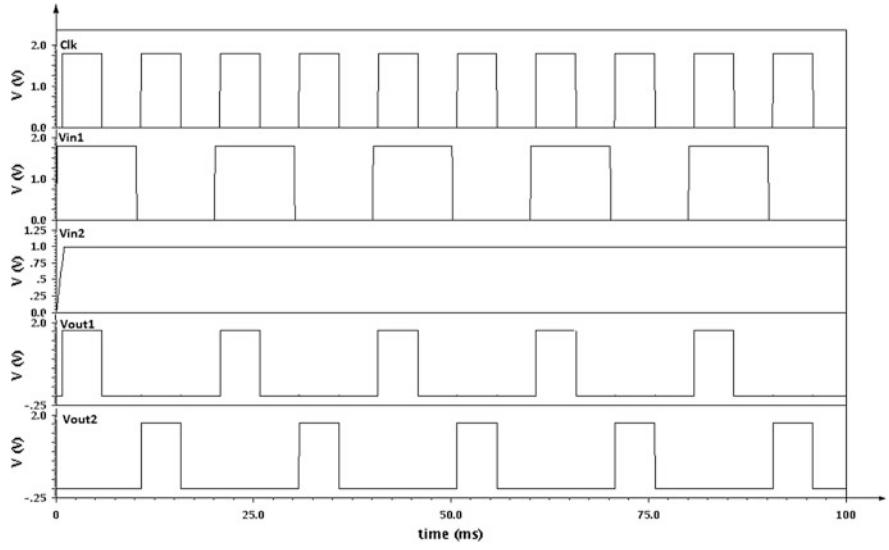


Fig. 7 Comparator waveform

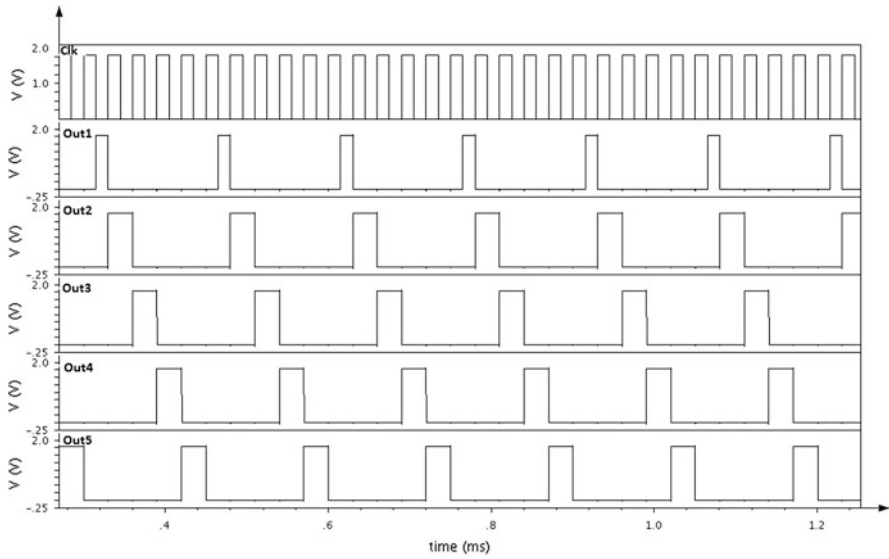


Fig. 8 SAR logic waveform

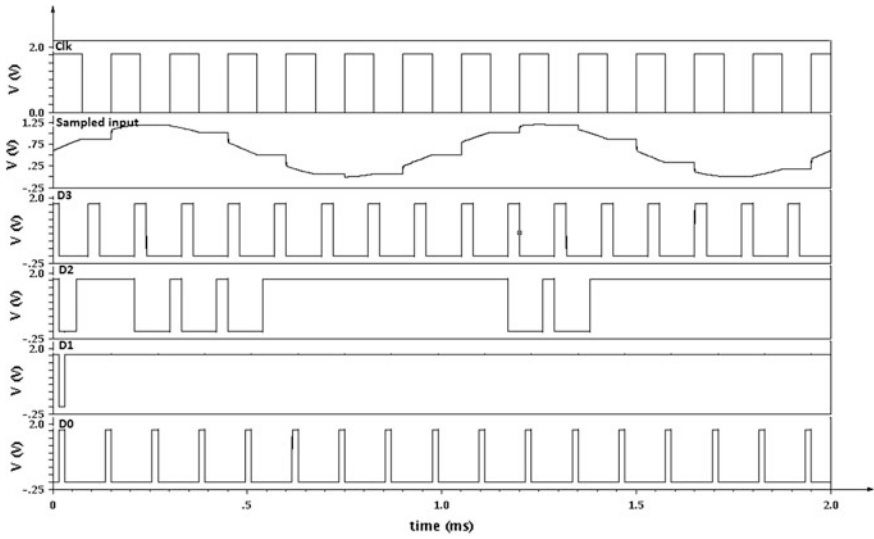


Fig. 9 SAR ADC output waveforms

3.4 SAR ADC Output

The digital outputs corresponding to respective input analog signals are shown in the waveforms (Fig. 9).

A full conversion of a single analog value requires 120 us clock period. It is observed that in this clock period, four different DAC outputs are compared with the hold values.

3.5 Power Dissipation Comparison

See Fig. 10.

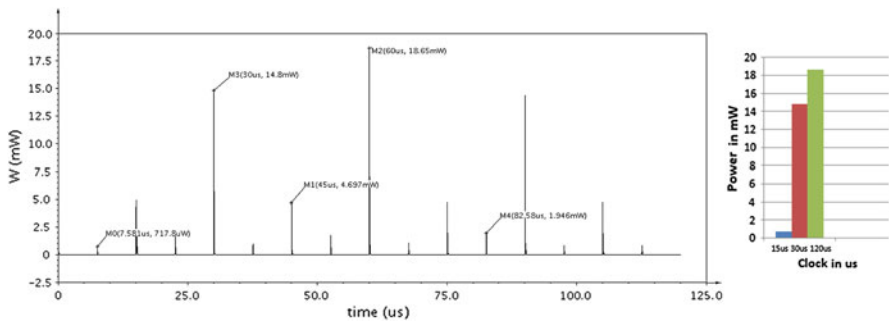


Fig. 10 Power dissipation with different clocks in ADC

4 Conclusion

A 4-bit 9 KS/s SAR ADC was designed in 180-nm CMOS technology. It is observed that the bootstrap switch that we used provides distortionless output. The DAC which we have used the capacitive array without any switches which help us to reduce the overall power dissipation. The bootstrap switch used for S&H circuit gives the distortionless, accurate and faster response. Power analysis was done, and the power dissipation for different clocks was observed.

References

1. Y. Chen, S. Tsukamoto, T. Kuroda, A 9b 100 MS/s 1.46 mW SAR ADC in 65 nm CMOS. IEEE Asian Solid-State Circuits Conference (2009)
2. R. Hedayati, A Study of Successive Approximation Registers and Implementation of an Ultra-Low Power 10-bit SAR ADC in 65 nm CMOS Technology, Master's thesis performed in Electronic Devices (2011)
3. T. Hong†, J.M. Chileshe, J. Lu, B. MO, C. Lai, Design and Implementation of SAR ADC, J. Comput. **6**(12) (2011)
4. C. Jun, R. Feng, X. Mei-hua, in *IC Design of 2 Ms/s 10-bit SAR ADC with Low Power*, Microelectronic Research and Development Center, Shanghai University, IEEE (2007)
5. D.A. Johns, K. Matrin, *Analog Integrated Circuit Design*. Wiley, New York (2008)
6. A.C. Kailuke, V. G. Nasre, M. Shojaei-Baghini, D. K. Rajendra, Design of low power integrated SAR ADC in 0.18 μm CMOS process, in *Proceedings of SPIT-IEEE*
7. J. Park, H.J. Park, J.W. Kim, S. Seo, P. Chug, A 1 mW 10-bit 500 KSPS SAR A/D Converter. **ISCAS 2000**—IEEE International Symposium on Circuits and Systems (2000)
8. J. Zhang, Design a 10-bit SAR A/D Converter Based on CMOS Technology. University of Electronic Science and Technology of Xi'an, Chengdu (2008)

EEG-based Automatic Detection of Drowsy State

Jinu Jai, Geevarghese Titus and S. Purushothaman

Abstract Electrical signal generated by the brain represents not only the brain function but also the status of the whole body. This paper focuses on finding the relation between EEG signal and human drowsiness, for which we require efficient algorithms. In the drowsiness state, a decrease of vigilance is generally observed. Identification was done by giving the preprocessed signal to a trained ANN to identify correctly the sleep condition of the person under observation. Different back-propagation algorithms are used for the study and the best one chosen by using the MSE estimation. Then using this system, classification is done and the drowsy signal sample is identified from given input samples.

Keywords EEG signal · Central nervous system · ERS · IIR digital filters · Feed-forward neural network

1 Introduction

Electroencephalogram analysis (EEG) is a very useful technique to investigate the activity of the central nervous system (CNS). It gives information related to the brain activity based on electrical recordings taken on the scalp of the subjects. It is produced by the bombardment of neurons within the brain. The measurement is taken for a short duration of 20–40 min by placing multiple electrodes over the scalp. The researchers have found that the brain function and status of the whole

J. Jai (✉)
AJCE, Kanjirapally, India
e-mail: jinujai@gmail.com

G. Titus · S. Purushothaman
VIT University, Vellore, India
e-mail: geevarghesetitus@amaljyothi.ac.in

S. Purushothaman
e-mail: purushothaman@vit.ac.in

body is indicated by these signals. Because of the clinical applications, sleep EEG is a very important research branch of medicine [1].

Commonly the EEG waves range from 0.5–500 Hz. The following five frequency bands are clinically relevant: (i) delta, (ii) theta, (iii) alpha, (iv) beta, and (v) gamma. Delta waves frequency is up to 3 Hz. Theta waves frequency ranges from 4 to 7 Hz. It is slowest wave having highest amplitude. It is the dominant rhythm in adults in deep sleep. It emerges with closing of the eyes and with relaxation. Alpha waves have a frequency range from 7 to 12 Hz. It is most commonly seen in adults. Alpha activity occurs rhythmically on the occipital regions. Alpha wave appears with closing eyes and disappears normally with opening eyes. It is regarded as a normal waveform. Beta wave has frequency ranges from 14 to 30 Hz, having small amplitude. It is the dominant rhythm during alert or anxious states and is usually observed from the frontal and central portions on the brain. It is generally a normal rhythm and is observed in all age groups. Gamma wave is fastest wave of brain and having a frequency ranges from 30 to 45 Hz and is normally called as fast beta wave. It also has a very low amplitude, and its presence is rarely felt. These waves are occurred in front central part of the brain. It suggests the event-related synchronization (ERS) of the brain [2].

Since the EEG signals are very weak, they can easily be contaminated by other sources. The unwanted signals in an EEG are known as artifacts. Artifacts can be divided into two categories: physiologic and non-physiologic. The physiologic artifacts include the signals originated from heart, eyes, muscle, and tongue. Sweating can also alter the impedance at the electrode scalp interface and produce an artifact. Non-physiologic artifacts include 50-Hz interference from electric equipment. Non-physiologic artifacts are caused by body or electrode movements [1].

Drowsiness appears into the EEG spectrum, predominantly in the central regions of the brain as an increase and decrease of activity in the frequency band observed. EEG is so efficient in detecting drowsiness that it is often used as a reference indicator. The reference is built by expert doctors who visually observe the proportion of signal activity on a short-time window as in. Usually the signals are taken from a large number of channels. The computation will be easier and faster when using a few number of channels. But by using large number of channels, it is easier to know how EEG energy is shifted from one frequency band to another and how frequently the fluctuations will repeat.

The raw EEG signal is obtained from the sample dataset from EEGLab. All recordings were measured using standard electrode placement scheme also called as International 10–20 system. Data set contains the 32 single channel recordings. The data were sampled at a rate of 128 samples per second [3].

Section 2 deals with EEG extraction and artifact removal techniques. Section 3 covers drowsy-state identification using neural network, Sect. 4 deals with the results and findings of the work, and final section provides the conclusion.

2 EEG Extraction and Artifact Removal Techniques

EEG is generally described in terms of its frequency band. The amplitude of the EEG shows a great deal of variability depending on external stimulation as well as internal mental states. Delta, theta, alpha, beta, and gamma are the names of the different EEG frequency bands which relate to various brain states. All these rhythms change with the physical and mental activity. So in order to study the various changes, the extraction of these rhythms is important. Frequency analysis or filtering is an established and classical way to deal with a single channel signal. When we are interested in a specific frequency range, time-invariant band-pass filtering or the Fourier transform (FT) will be extracted from the target frequency component. EEG can be filtered by IIR digital filters to extract different frequency band. Different types of filters are as follows:

1. Butterworth filter, normally referred to as a maximally flat magnitude filter. The frequency response of the Butterworth filter is maximally flat, i.e., has no ripples in the passband, and rolls-off toward zero in the stopband is used.
2. Chebyshev type 1 and type 2 filters having a steeper roll-off and more passband ripple (type I) or stopband ripple (type II) than Butterworth filters. These filters have the property that they minimize the error between the idealized and the actual filter characteristic over the range of the filter, but with ripples in the passband.

A bank of four band-pass filters is used, and each one is used for extracting one of the rhythms. First band-pass filter is used for filtering theta rhythm in the range of 4–7 Hz with a sampling frequency of 25 Hz. Second band-pass filter is used for filtering alpha rhythm in the range of 7–12 Hz with a sampling frequency of 30 Hz. Third band-pass filter is used for filtering delta rhythm in the range of 0.5–3 Hz with a sampling frequency of 10 Hz and the fourth band-pass filter is used for filtering beta rhythm in the range of 14–30 Hz with a sampling frequency of 75 Hz. For the same filter specification like pass-band ripple, stop band ripple and sampling frequency Chebyshev type 1 and type 2 filters takes less order compared to Butterworth filter.

The FFT (fast Fourier transform) is a mathematical process which is used for the EEG analysis to investigate the composition of an EEG signal. Since the FFT transforms a signal from the time domain into the frequency domain, frequency distributions of the EEG can be observed. Figure 1, represents the frequency-domain representation of EEG rhythms using Chebyshev type 1 filter.

For selecting the filter which gives better performance, root-mean-square error (RMSE) of each filter output is calculated. Based on the result, it can be seen that the filter which gives minimum RMSE is Chebyshev type 1 and that it gives a comparatively better performance compared to other two filters.

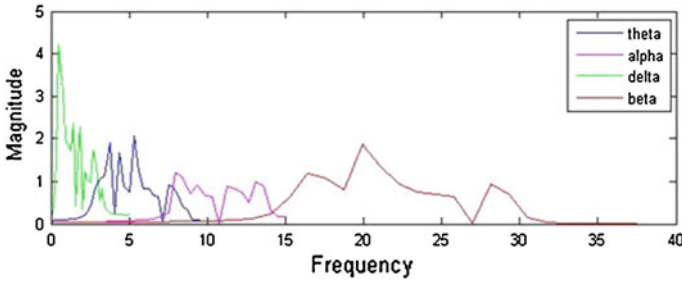


Fig. 1 Rhythms extracted using Chebyshev type 1 filter using frequency-domain representation

Table 1 RMSE comparison

Rhythms	RMSE (Chebyshev type 1)	RMSE (Chebyshev type 2)	RMSE (Butterworth)
Alpha	8.9416	9.1009	9.1524
Beta	5.424	5.6503	5.6779
Theta	9.4846	10.9139	10.3354
Delta	16.4659	18.1593	17.25

Once the signals have been extracted, the next stage would be to apply signal processing techniques to reduce or remove the artifacts present in an EEG. The removal of biological artifacts is very hard without loss of some of the EEG data or information. To remove the eye blink, ECG, EMG, and other biological artifacts, advanced signal processing techniques based on independent component analysis is used. In order to reduce the power line artifacts and other band limited biological artifacts, various digital filtering techniques can be applied (Table 1).

Here, FastICA based on kurtosis is used for the removal of artifact. Before applying the ICA algorithm, some preprocessing is done to make the ICA estimation simpler and better conditioned. The preprocessing steps are centering, whitening, and measure of non-Gaussianity [4, 5].

3 Drowsy Identification Using ANN

In order to identify drowsy state, artificial neural network is used. The neural network architecture consists of a feed-forward neural network with five hundred neurons in the input layer and one output layer.

3.1 Neural Network Training Algorithm

In a feed-forward neural network, neurons are connected only in forward direction. Each layer of the network is connected to the next layer, without any feedback. Hence, the network is trained using the back-propagation training algorithm, and the obtained results are compared with the desired results. The difference between the actual results and the desired results gives the error. In a back-propagation technique, the weights and input threshold of the neural network are changed in such a way that the error can be reduced.

In the neural network model, there were several training algorithms with different computation and storage requirements. However, a specific algorithm cannot be used in all applications. Here, the system training has been implemented using scaled conjugate gradient, conjugate gradient, resilient back-propagation algorithms [3].

3.2 Preprocessing Steps

The recorded signal obtained from the human brain contains many artifacts. Before any further processing, the artifacts must be removed. One of the most commonly used methods is independent component analysis (ICA). The key to estimating the ICA model is non-Gaussianity, without which the estimation is not possible. Non-Gaussianity is the correction that modifies the expected Gaussian function for the measurement of a physical quantity. To measure the non-Gaussianity, different methods such as kurtosis and negative entropy are used. Here, the artifact removal technique involves FastICA based on kurtosis. Kurtosis is a measure of the peakedness of the probability density function, of real-valued random variable. This portion has been dealt in Sect. 2.

3.3 ANN Creation

Artificial neural network is created by using different combinations of input and target sample set, and there will be one output. The artifact-free data set is represented in a matrix form, where each row represents the number of channels used for the observations and each column represents the obtained signal data. Then, each column of data (contains approximately 30,000 samples) is selected and is divided into a set of 100, 500, 1,000 samples and each set is represented as a column of another matrix. Then by visual inspection, sample sets are created based on the presence and absence of considerable peak in the entire data set. Corresponding to the input sample size, different target sets are selected containing 300, 60, and 30 samples, respectively.

3.4 ANN Training

The ANN network is trained to give an output of one when there is a peak in the given input signal and an output of zero otherwise. Different back-propagation algorithms such as scaled conjugate gradient, conjugate gradient, and resilient algorithms have been used.

3.5 ANN Testing

The trained ANN network is tested for different input samples and from the output obtained the drowsy data set is identified. The input sample is trained using the different back-propagation algorithms and the algorithm which gives better result is found out by considering the mean square error (MSE) estimation. MSE is the default function for feed-forward network, which is the average squared error between the network output and the target output.

3.6 ANN Classification

After creation, training, and testing of network, the next step is to identify the drowsy signals. For this, ANN classification is used. Different combinations of input and target sample set are considered and the best classification rate is obtained for the combination of 500 input samples and 60 targets.

4 Results and Findings

The artifact-free signal is given to the training algorithms and the mean square error rate of each algorithm is calculated. The results are shown in the Table 2. This table shows different set of inputs and the MSE for different back-propagation algorithm. From this table, we can see that scaled conjugate gradient (SCG) algorithm gives better performance compared to other algorithms. As a result, SCG back propagation is used as the training algorithm for proposed system.

In order to see how the network responds to different samples of inputs and targets, classification rate is calculated for 100, 500, and 1,000 input sample set. This is shown in the Table 3. It is seen that the combination of 500 inputs and 60 targets gives better classification rate. The next step is to classify signals according to drowsy and not drowsy signals. Table 4 shows classification rate for different set of inputs (combinations of 500 input and 60 targets), and best classification rate of 90 % has been obtained.

Table 2 Mean-squared error comparison

Inputs	RP	SCG	CGB
Set 1	0.5055	0.4147	0.5055
Set 2	0.5482	0.4963	0.5482
Set 3	0.5240	0.4546	0.5240
Set 4	0.5154	0.4343	0.5154
Set 5	0.5183	0.4544	0.5183
Set 6	0.5264	0.5022	0.5264

Table 3 Comparison of classification rate for different input samples

Inputs	RP %	SCG %	CGB %
Set 1 (100 Samples)	53.3	75	50.3
Set 2 (500 Samples)	66.7	90	70
Set 3 (1000 Samples)	53.3	66.7	66.7

Table 4 Classification rate for 500 input samples

Inputs	RP %	SCG %	CGB %
Set 1	66.7	90	70
Set 2	56	70	56.7
Set 3	60	80	71.7
Set 4	63.3	73.3	73.3
Set 5	60	73.3	68.3
Set 6	50	65.3	55

5 Conclusion

This work describes the identification of drowsy signal using neural network. As a preprocessing step, the artifacts in the EEG signals are removed using FastICA. Then, the artifact-free EEG signal is trained using different neural network back-propagation algorithms and the performance is compared by considering mean square error. From the comparison, SCG algorithm has minimum MSE, and hence, it is used as the training algorithm for proposed system. The trained network can now be used to identify a drowsy pattern for the received input. A maximum classification rate of 90 % has been obtained using this network.

Mean square estimation has been used as a general estimation technique for all the back-propagation algorithms and has been employed in our network. Different estimation techniques such as mean absolute error (MAE) and mean relative error (MRE) can be studied to check whether a better classification of signals can be done.

References

1. S. Sanei, J. Chambers, *EEG Signal Processing* (Wiley, NJ, 2007)
2. M.S. Amin, M.R. Azim, F.M. Hasan, *Spectral Analysis of Human Sleep EEG Signal*, in 2nd International Conference on Signal Processing System ICPS (2010)
3. A. Vuckovic, V. Radivojevic, A.C.N. Chen, Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. *Med. Eng. Phys.* **24**(5), 349–360 (2002)
4. A. Hyvarinen, E. Oja, Independent component analysis: algorithm and applications. *Neural Netw.* **13**(4–5), 411–430 (2000)
5. Ella Bingham, A. Hyvarinen, A fast fixed point algorithm for Independent component analysis of complex valued signals. *Int. J. Neural Syst.* **10**(1), 1–8 (2000)
6. T. Tanaka, Y. Saito, Rhythmic component extraction for multichannel EEG data analysis. *ICASSP 2008*, 425–428 (2008)
7. S. Monto, S. Palva, J. Voipio, J.M. Palva, Very slow EEG fluctuation predict the dynamics of stimulus detection and oscillation of amplitude in humans. *J. Neurosci.* **28**(33), 8272–8286 (2008)

WIDS Real-Time Intrusion Detection System Using Entrophical Approach

Kamalanaban Ethala, R. Sheshadri and S. Sibi Chakkaravarthy

Abstract Nowadays, threats, worms, virus, and malwares in the Internet and security breaches such as intrusion and penetration testing in the network are quite common and lead to the loss of huge amount data. In recent decades, various researchers revealed their perceptions on security and security-related issues. In this paper, we propose a robust intrusion detection system based on Entrophical approach. Here, our system monitors the normal behavior of the network by means of probabilistic system with monitoring active ARP protocol in all PCAP files captured by packet analyzer and detects the intrusion by means of deviation in the PCAP. Entrophical approach deals with profiling strategy; here, data logs of users are classified as profiles such as base, daemon, and user. Various IDS are compared with the Entrophical model-based IDS. Experimental results compared with snort, security onion, and our methodology show that Entrophical model is a level head through many phases, and the comparison outstrips with reliable performance. Real-time results have also been enhanced. This is the first claim for designing an IDS model to combat the real-time attacks such as aircrack-ng, airmon-ng, and airodump-ng from the operating system “BACKTRACK.”

Keywords BACKTRACK · WLAN · Snort IDS · Entrophical approach · Kali Linux

K. Ethala (✉) · S. Sibi Chakkaravarthy
Department of Computer Science and Engineering, Vel Tech University, Chennai, India
e-mail: kamalanaban2009@gmail.com

S. Sibi Chakkaravarthy
e-mail: sb.sibi@gmail.com

R. Sheshadri
Department of Computer Science and Engineering, Sri Venkateswara University, Tirupathi,
Andhra Pradesh, India

1 Introduction

WLAN is one of most advanced wireless networking technology for this sophisticated world. The advancement in WLAN and in wireless technology is increasing day by day. The biggest issues in the wireless as per concern are security and its flaws. Various researchers were proposing a new algorithm at every periodicity. A unique and first-hand solution ensured in order to combat this issue based on SIDS.

1.1 Basics About Snort IDS

Snort IDS is one of the efficient open source intrusion detection system and intrusion prevention system. Snort is a threaded signature-based IDS which filters packet based on protocols defined in it. Each protocol checks with the packet header along with the socket address of TCP/IP. Payload is also examined for all the packets and validates whether the packet belongs to the appropriate source and destination. Several classified patters are compared and associated in order to predict the better results. Pattern count of each string will be in terms of K (1,000).

1.2 Previous Work

In previous methodology such as CCT which deals with Intrusion detection using 6° separation and multipath navigation. Signature-based intrusion detection system is subjected to an objective model followed by modeling technique which predicts the behavior of the system notifying that the system is compromised to any attacks or not. Modeling techniques used in anomaly-based IDS require active data of users. SIDS mines all necessary features from the input user data (PCAP—that is from captured packet file) and compresses them into actual behavior status. After these actions, IDS can actively monitor whether the system is compromised to any attacks or not.

2 Entrophical Approach: Working Methodology

In Entrophical approach, data logs are classified into profiles, and each holds on particular data and privileges about the particular user in the active networks. Profiling deals with the authorized privileges for the authorized users. This type of profile scheming can be used to deploy Entrophical model-based IDS in any type of

networks as well as it can be used by any complex system without having any prior knowledge about the model. Consider the following with input node value and weight rates.

$$Q = f(P_1, P_2, P_3, \dots, P_n)$$

Given for various training inputs and the system function of each weighted rates will be approximate for all the functions including normal behavior of the system and abnormal behavior of the system.

$$Q = p_0 + \sum_{i=1}^m p_i q_i + \sum_{j=1}^m p_{ij} q_i q_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m p_{ijk} q_i q_j q_k \dots$$

where i, j , and k are various co-efficient of complex systems. Corresponding model for the complex system can be deployed from the single-time polynomial where P_i can be used as weight ratio inputs for feature extraction during profiling. Run-time polynomial scheme is denoted as follows

Formal Expression Definition:

Hidden states $Q = \{q_i\}$, $i = 1, \dots, N$.

Transition probabilities $A = \{a_{ij} = P(q_j \text{ at } t + 1 | q_i \text{ at } t)\}$, where $P(a|b)$ is the conditional probability of a given b , $t = 1, \dots, T$ is time, and q_i in Q . Informally, A is the probability that the next state is q_j given that the current state is q_i .

Observations (symbols) $O = \{o_k\}$, $k = 1, \dots, M$.

Emission probabilities $B = \{b_{ik} = b_i(o_k) = P(o_k | q_i)\}$, where o_k in O . Informally, B is the probability that the output is o_k given that the current state is q_i .

Initial state probabilities $\Pi = \{p_i = P(q_i \text{ at } t = 1)\}$.

Initialization—Interface level at mon0

$$\alpha_1(i) = p_i b_i(o(1)), i = 1, \dots, N$$

Recursion: Capturing and analyzing packets

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^N \alpha_t(j) a_{ji} \right] b_i(o(t+1)) \quad (1)$$

here $i = 1, \dots, N$, $t = 1, \dots, T - 1$

Pattern: prediction of attack

$$P(o(1)o(2)\dots o(T)) = \sum_{j=1}^N \alpha_T(j)$$

Obviously, both forward and backward algorithms must give the same results for total probabilities $P(O) = P(o(1)o(2)\dots o(T))$.

2.1 Profiling Environment

Organizing the profiles for each privileged users is a bit toughest job. Since we use random ordering technique to daemon the profiles. The host and logs are updated in the profiles and maintained thoroughly for all centralized domains. Profiling is done for two main categories namely.

2.2 Base Category

In base category, the system processes and individual user log are captured generally. In this profile, the system calls and its frequency are ranked for individually. Large value occurs for rare occurring and smaller value for reputations.

2.3 Daemon Category

Daemon profile registers huge structures of system calls, UID processes and system processes in a daemon process running on a system. A daemon process waits for a call from external processes [11]. Daemon profile treats idle state as delimiter of profiles. A profile starts recording when a client connected to a daemon. Again, the base profile is used to make the daemon profile [11]. Daemon process can be recorded in a small amount of data. This method also simplifies the comparison of profiles [11].

2.4 User Category

User category is used at interface level, and privileges to the users are given by the administrators. Various user privileges are set by the DBA's.

2.5 Algorithm

```

Begin
Function y. = EntrophicalIDS P, X, Q, D.
Input = Data logs, Network (N) ;
Output = Y, Classifier
For i = 1; i ++
get (Data log, classifier) ;
End for

```

```

Hash (N) ;

Begin Classifier:

    Apply K means p(pcap) ;
    Find Centroid (p()) ;
    Pos_Cent (p()) ;

    Begin Profile:
        Privilege (USER, BASE, DAEMON) ;

Select (classifier, profile) ;
Hash (select ()) ;
Result (behavior.csv) ;

    End Profile ;

End Classifier ;

End

```

3 Comparison: Results and Performance

See Fig. 1 and Tables 1 and 2.

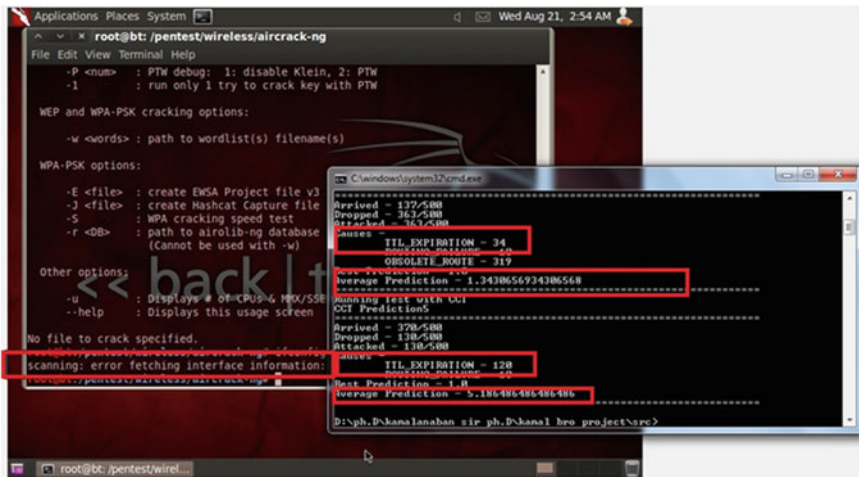


Fig. 1 Results of Entrophical approach—demonstrated using BACKTRACK5 R3 in AMD A6 3240 GHZ CPU. *Left* Shows the performance of the Entrophical model. Final output is denoted *right* which shows that our proposed methodology restricts the illegal scanning of ports in WLAN (IEEE 802.11—Wi-Fi)

Table 1 Comparison of result prediction for various IDS

S. no	IDS	NMAP	Spoofing	Armitage
1	Snort	Yes	Yes	No
2	Entrophical model	Yes	NA	Yes
3	Security onion	No	Yes	No

Table 2 Comparison of result prediction for various operating systems

S. no	IDS	Back track	Kali Linux	Oph crack	Katana
1	Snort	No	No	Yes	Yes
2	Entrophical model	Yes	Yes	Yes	Yes
3	Security onion	No	No	Yes	No

4 Conclusion

In this paper, we propose an ensured unique algorithm for detecting intrusions in WLAN environments. Various stages of optimization are used in order to extract features and to create the work model to predict the intrusions. Here, certain degree has been allowed in order to optimize the certain parameter in the classifiers. The work model and methodology used in this paper is tested with the limited number of attacks such as DDoS, Armitage framework, and Nmap applications. Here, we have tested our proposed model in various working environment other than Windows. We intend to test our methodology in combatting operating systems such as Back Track and Kali Linux with positive achievements. Figure 1 and Tables 1 and 2 show that our proposed approach is better in most of the cases for real-time attacks using remote tools and frameworks.

References

1. K. Ethala, R. Sheshadri, Combatting cyber terrorism-assessment of log for malicious signatures. *Am. J. Appl. Sci.* 1660–1666 (2013)
2. R. Di Pietro, L.V. Mancini, *Intrusion Detection Systems*. Series: Advances in Information Security **38** (2008)
3. I.A.B. Bazara, H. Anthony Chan, in *Handbook of Information and Communication Security*. Intrusion detection systems (Springer, Berlin 2010)
4. W. Kanoom, N. Cuppens-Boulahia, F. Cuppens, F. Autrel, Advanced reaction using risk assessment in Intrusion detection system. *Crit. Inf. Infrastruct. Secur LNCS* **5141**, 58–70 (2008)
5. A.A. Ghorbani, W. Lu, M. Tavallae, Network intrusion detection and prevention. Series: Advances in Information Security (eBook, 2010)
6. N. Tuck, T. Sherwood, B. Calder, G. Varghese, in *Deterministic memory-efficient string matching algorithms for intrusion detection*. In Proceedings of IEEE INFOCOM (2004), pp. 2628–2639

7. K.A. García, R. Monroy, L.A. Trejo, C. Mex-Perera, E. Aguirre, Analyzing log files for postmortem intrusion detection. *IEEE Trans. Syst. Man Cybern.* (2012)
8. C. Cowan, P. Wagle, C. Pu, S. Beattie, J. Walpole, in *Buffer overflows: attacks and defenses for the vulnerability of the decade*. Proceedings of DARPA Information Survivability Conference Exposition (1999), pp. 154–163
9. Y. Wang, W. Fu, D.P. Agrawal, Intrusion detection in gaussian distributed wireless sensor networks. *IEEE* (2009)
10. B. Liu, P. Brass, O. Dousse, P. Nain, D. Towsley, in *Mobility improves coverage of sensor networks*. Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) (2005)
11. S. Janakiraman, S. Rajasundaran, P. Narayanasamy, The model—dynamic and flexible intrusion detection protocol for high error rate wireless sensor networks based on data flo. *IEEE*
12. A. Modirkhazeni, N. Ithnin, O. Ibrahim, in *Secure multipath routing protocols in wireless sensor networks: a security survey analysis*. *IEEE International Conference* (2010)
13. Z. Mingqiang, H. Hui, W. Qian, in *A graph-based clustering algorithm for anomaly intrusion detection*, ICCSE 2012. *IEEE Conference* (2012)
14. Y. Guan, A.A. Ghorbani, N. Belacel, in *Y-means: a clustering method for intrusion detection*. Canadian Conference on Electrical and Computer Engineering (2003), p. 14
15. U. Prathap, P. Deepa Shenoy, K.R. Venugopal, in *Wireless sensor networks applications and routing protocols: survey and research challenges*. *IEEE Symposium* (2012)
16. G.G. Xie, J. Gibson, in *A network layer protocol for UANs to address propagation delay induced performance limitations*. Proceedings of the MTS/IEEE Conference and Exhibition (OCEANS 2001). (2011) 20872094
17. M. Patil, R.C. Biradar, in *A survey on routing protocols in wireless sensor networks*, ICON 2012. *IEEE Conference* (2012)
18. Q. Ren, Q. Liang, in *A contention-based energy-efficient MAC protocol for wireless sensor networks*. Proceedings of 2006 IEEE Wireless Communications and Networking Conference (WCNC 06). (2006), pp. 1154–1159
19. W. Zhu, Q. Wang, in *Improving intrusion detection through merging heterogeneous IP data*. Proceeding of the IEEE International Conference on Information and Automation Shenyang (2012)
20. B. Zhang, Research on intrusion detection based on heuristic genetic neural network. *Adv. Intell. Soft Comput.* **149**, 567–573 (2012)

Optical Character Recognition for Alphanumerical Character Verification in Video Frames

Sheshank Shetty, Arun S. Devadiga, S. Sibi Chakkaravarthy,
K.A. Varun Kumar, Ethala Kamalanaban and P. Visu

Abstract In real world, optical character recognition (OCR) is one of the key terms challenging image processing stream. Various applications are emerging based on OCR; one fine good example of these terms was recognizing vehicle's number plate in tolls. Since various researchers are under strong discussion in this area, here we proposed a new algorithm for recognizing the characters from the motion pictures. Our proposed model is subjected to two major segregations: One is for character mapping and another one is for character verification. Experimental results have been enclosed with an accuracy level of 97.08 %.

Keywords Optical text extraction · Optical character recognition · Canny filter · RGB image · Gabor filters

S. Shetty (✉) · A.S. Devadiga
Department of Computer Science and Engineering, NMAM Institute of Technology,
Nitte, Karnataka, India
e-mail: sheshankshetty06@gmail.com

A.S. Devadiga
e-mail: arundevadiga1@gmail.com

S.S. Chakkaravarthy · K.A. Varun Kumar · E. Kamalanaban · P. Visu
Department of Computer Science and Engineering, Vel Tech University, Chennai, India
e-mail: sb.sibi@gmail.com

K.A. Varun Kumar
e-mail: varun.kumar300@gmail.com

E. Kamalanaban
e-mail: ethalakamalanaban2009@gmail.com

P. Visu
e-mail: pandu.visu@gmail.com

1 Introduction

In the last years, the problem of text detection, recognition, and extraction from a video frame that received a significant attention and lot of work has been proposed. Since this is an era of digitalization, the videos act as an efficient source of information. The one important information hidden inside the video is text, e.g., the title of the movie and scene text hidden in the random video (i.e., shop name and number plate of the vehicle). This text embedded inside the videos may be in small quantity, but these always carry important information of the multimedia content. Hence, retrieving a text from the video makes it more useful and highly desirable.

The proposed method for recognizing and extracting a text from a video frame is optical text extraction (OTE)–optical character recognition (OCR)-based method. This OTE–OCR-based method has three main modules: division of videos into an individual frame, text detection and text recognition and extraction. The major objective of this proposed method OTE–OCR-based text recognition and extraction from the video frames is to efficiently extract a text from video frames. Henceforth, there will be an extension of the application of the OCR systems into the wider area.

Many problems have been faced during the text extraction from the video frames. These problems were due to low resolution of the video, due to complex backgrounds, and also due to the unknown colors. Lot of methods were proposed to overcome these problems by incorporating different architectures and different methods to detect a text out of videos, and these methods have been discussed in the related work section.

The document is organized as follows: In Sect. 2, we discuss various related works on text detection from videos. Section 3 describes the basics of OCR. Section 4 reviews the methodologies involved in the proposed method. In Sect. 5, we present the experimental results of this project compared to the previous work. Finally, we conclude by giving a brief overview of the proposed method and also give a glimpse on future works.

2 Related Work

In this section, we review related works on text retrieval from videos. Lienhart and Stuber [1] briefly discuss automatic text recognition in digital videos. This work comes under the category of bottom-up method where images are segmented to form regions, and later, character regions are grouped to form words. In this work, they deal with text segmentation method, where the aim of segmentation step focuses on dividing frames into regions that contain a text and regions that do not contain a text. Regions that do not contain a text are discarded since these regions are not useful in the recognition process. The regions with text are called as character candidate region, which are passed to recognition process. Motion analysis method is used to detect the character candidate region in consecutive frames. Since this is a

bottom-up method, its complexity relies on the image content and also on the segmentation algorithm.

The text localization is one major task performed in this proposed system. In recent years, it has got a significant attention [2, 3]. Chen et al. [2] in their work review about feature extraction using Gabor filters. The text embedded in a video can be a superimposed text and scene text [1]. The motion pictures captured can be from some random scene. Hence, detecting a text from a scene image or video is also an important area to be considered. Ohya et al. [4] in their work briefly discuss about the text detection from the scene image. Scene image is one of the complex images because the images captured occur in a three-dimensional space and might be distorted due to the illumination of light, image might be tilted or slanted, and some part of the image might be partially visible. Hence, these are the major problems faced in the scene images. Since in scene image, the text exists in different orientations, Ohya et al. focused mainly on monochrome images and also on still images rather than on motion pictures or videos. Our main focus will be on detecting a text embedded in a video.

The existing works [1–4] have one or more limitations while retrieving a text from images or videos. There exists sensitivity to different fonts, font sizes, and colors and also restrict to the type of text retrieved (i.e., titles or subtitles only), and not able to handle videos, rather restricting to still images. In our proposed method, we not only restrict to detecting the textual information in the motion picture. But, we present an efficient computational method of extracting the text and passing the segmented characters to the OCR system, and the final outcome will be an editable text in a Word format.

3 Working Methodology

The OCR-based text recognition and extraction is a novel computation scheme which involves three major tasks. The proposed system is divided into three phases: division of videos into an individual frame, text detection phase, and text recognition and extraction phase. As a first phase of our proposed system, we take video as an input. The input video is divided into individual frames, each individual frames are passed through the rest of the two phases, and the individual frame represents the RGB image. The conversion of RGB image to grayscale image is done.

The second phase of the proposed system is text detection; here, we perform two main operations that are text localization and text verification. Text localization is performed on individual frames. Here, feature extraction from an individual frame is done. Henceforth, edge map of the individual frames must be created. There are many methods explained previously for creating the edge map [6]. The edge detector used is a Canny filter [5]. The edge detection (i.e., horizontal and vertical edge detections) is done to the grayscale image using the Sobel and Canny masks. Using edges as the prominent feature of our system gives the opportunity to detect

characters with different fonts and colors since every character presents strong edges, despite its font or color, in order to be readable. Canny edge detector is applied to grayscale images. Canny uses Sobel masks in order to find the edge magnitude of the image, in gray scale, and then uses non-maxima suppression and hysteresis threshold. The two operations performed by the Canny edge detector manage to remove non-maxima pixels, without effecting the connectivity of the contour. The resulted edge map will be a binary image with background 0 (black) and connectivity of the contour in 1 (white). Later, dilation on the resulted image is done to find the textlike region. Dilation by a cross-shaped element is performed to connect the character contours of every text line. The dilation is process of increasing the size of pixel and preserving the connectivity of the contour 3×13 rectangular structuring element, and octagon structuring element is applied horizontally and vertically, respectively, on the edge map. Different cross-shaped structuring elements (i.e., disk, line) were tried. For more effectiveness of detection, rectangular and octagon were used. The common edge between the vertical and the horizontal edges is extracted, and it is dilated again to get the accurate text like regions.

Morphological binary open image operation is performed to remove the small objects and this operation removes the binary image all connected components that have fewer than 600 pixels. The groove filling mechanism is applied to fill the gap between the non-connected pixels.

3.1 Algorithm—Alphanumeric Character Extraction

```

Begin
Function OCR
    Input:Video file(.avi)
    Output:Image/frames
    Step:Pre-processing
Convert RGB -> grey;
Sobel(horizontal,videoframes);
Canny(vertical,videoframes);
Plot(octagon);
Plot(rectangle);
Dilate(videoframes);
    for i=1:m
        for j=1:n
            FindText(min(n),max(m),min(m),max(n));
        End
    End
End
Joincharparts(FindText);
End

```

3.2 Algorithm—Alphanumeric Character Verification

```

Begin
Function Text Or Not
    Input: Four coordinates of the dilated regions
    (X,X1,Y,Y1) and the edge map image (H)
    Output: Text Or Not
        for i= X:X1
            for j=Y:Y1
                hcount=hcount+1
            end
        end
        tcount=(X1-X) * (Y1-Y)
        Ratio=hcount/tcount
        If (Ratio >=0.065)
            Result= TRUE ( Its Text)
        end
end
end

```

4 Experiment and Results

The ten sample images were passed to evaluate the performance of the OCR-based system. According to the evaluation, the detection percentage was 99.35 %, and the recognition and extraction percentage was 92.45 %. By using the proposed system, the detection of the text from the image was exceptionally good since there was small miss rate. There was a minor decrease in the recognition and extraction percentage. This was because the template file used for the OCR system was basically trained for Times New Roman and Arial Black fonts. Hence, scene images had a lower recognition rate compared to the normal monochrome images due to the unknown fonts and the colors. This can be overcome by training the template file for more fonts (Fig. 1).



Fig. 1 Results of our proposed methodology. **a** Scene image, **b** representing the extracted region, and **c** segmented text extracted using our proposed methodology

Table 1 Result analysis

Number of video frames	12
Detection percentage	97.08 %
Extracted and recognized percentage	91.60 %

By using the OCR-based text recognition and extraction method, we evaluated the system by passing 12 video frames, the detection percentage was 97.08 %, and the extraction and recognition percentage was 91.60 %. Since large fonts and scene objects (alphanumeric) were used, the processing speed was of average speed (Table 1).

The coordinates of the dilated regions are passed to text verification module. This is performed to check whether the textlike regions extracted are text or not. To verify whether these regions are text or not, the *hcount* of the textlike region of the image is calculated where *hcount* is the total number of white pixels in the detected image. Next, the *tcoun*t of the textlike region of the image is calculated where *tcoun*t is the total number of the pixels in the detected image. The ratio of *hcount* to *tcoun*t is performed, and if the ratio is greater than or equal to a threshold value 0.065, then the coordinates of the dilated regions passed are assumed to be a text, or else the regions extracted are discarded, since it is not useful.

Table 2 Performance analysis of test data

Sample images (with scene text and superim- posed text)	Number of characters in an image	Total number of characters detected, recognized, and extracted			
		Text detection phase	Detection percentage (%)	Text recog- nition and extraction phase	Recognition and extrac- tion percent- age (%)
10.avi	14	14	100	12	85.71
5.avi	93	90	96.77	84	90.32
mg1.avi	41	41	100	40	97.56
s2.avi	110	110	100	101	91.81
2.avi	11	11	100	11	100
sample.avi	62	61	98.38	61	98.38
blank.jpg	10	10	100	10	100
txt.jpg	18	18	100	16	88.88
s3.jpg	61	60	98.36	54	88.52
6.avi	24	24	100	20	83.33

The overall detection percentage of OCR-based system for 10 sample images is 99.35 %

The overall recognition and extraction percentage of OCR-based system for 10 sample images is 92.45 %

$$\text{Text ratio} = \frac{(\text{hcount})}{(\text{tcount})} \geq 0.065$$

hcount the count of white pixels in an extracted dilated region

tcount the total number of pixels in an extracted dilated region.

5 Conclusion

OCR is one of the emerging areas where most of the researchers are interested in finding the optical version of a character from a digital image to the wrapped version of editable text. Our proposed algorithm is used to perform OCR efficiently with an accuracy of 97.08 %. We have tested nearly about 10 data sets with various video frames/rates/fps. Table 2 denotes clearly the performance analysis of the video frames for various data sets. We have tested our proposed model in images also and achieved the accuracy level better than the accuracy level of video frames. The overall detection rate is very high when compared to recognition rate. Table 2 clearly states the performance metrics of our proposed model.

References

1. R. Lienhart, F. Stuber, Automatic text recognition in digital videos, in *Proceedings SPIE, Image and Video Processing IV*. (1995) pp. 2666–2675
2. X. Chen, J. Yang, J. Zhang, A. Waibel, Automatic detection and recognition of signs from natural scenes. *IEEE Trans. Image Process.* **13**, 87–99 (2004)
3. V. Wu, R. Manmatha, E.M. Riseman, Text finder: an automatic system to detect and recognize text in images. *IEEE Trans. Pattern Anal. Mach. Intell.* (1999)
4. J. Ohya, A. Shio, A. Akamatsu, Recognition of characters in scene images. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 214–220 (1994)
5. J. Canny, A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 679–698 (1986)
6. R. Gonzalez, R. Woods, *Digital Image Processing* (Addison Wesley, Boston, 1992), pp. 414–428

Modified AODV Routing Protocol for Multi-hop Cognitive Radio Ad Hoc Networks

Melvin Mathew, G. Shine Let and G. Josemin Bala

Abstract Cognitive radio network (CRN) is a next generation network which has been studied and receiving significant research due to the underutilization and congestion in radio spectrum. Cognitive radio technology possesses self-adaptivity, and routing is one of the key issues for designing a self-adaptive networks. It has been received greater attention due to the tremendous growth in spectrum utilization there by reusing the spectrum efficiently and sharing the licensed band in the absence of primary users (PUs). In this paper, we have proposed a modified AODV routing protocol which can be used for multi-hop cognitive radio ad hoc networks that can efficiently utilizes the spectrum in an intelligent manner. The proposed routing method selects the best path to transmit the packets considering the PUs activity and switches to new route if there is any interference of PUs present. In this method, the routing table periodically updates its information and can be used to find the optimal route.

Keywords Cognitive radio network · Secondary users · Multi-hop · Routing · Radio spectrum

1 Introduction

With rapid growth of wireless network applications, nowadays there is a great demand for spectrum radio resources. The fixed radio spectrum is allocated by government agencies for various applications. The unlicensed band or ISM band is widely utilized and congested. As per the report by Federal Communications

M. Mathew (✉) · G.S. Let · G.J. Bala
Department of Electronics and Communication Engineering, Karunya University,
Coimbatore, India
e-mail: tkmmelvin@gmail.com

G.S. Let
e-mail: shinelet@gmail.com

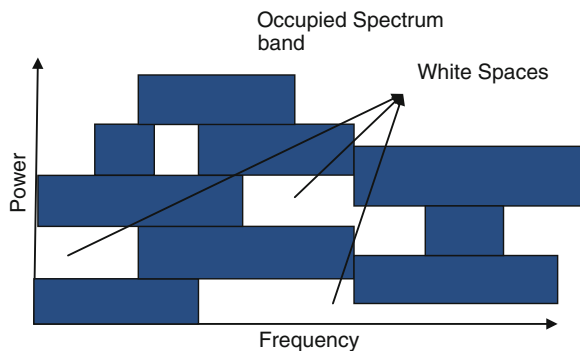
Commission (FCC), the usage of licensed spectrum varies between 15 and 85 % as time and location vary [1]. To overcome this scenario, cognitive radio is introduced as the next generation wireless technology in [2]. Cognitive radio is an intelligent radio which can change its transmitter and receiver parameters depending on the spectrum environment. A node integrated with cognitive radio technology can sense different spectrum range (Hz) and transmit the data based on the available bandwidth.

The main aim of cognitive radio is to select the best available licensed radio spectrum without giving interference to the licensed users. The temporally vacant licensed radio spectrum is called white space or spectrum hole or spectrum opportunities [2, 3]. Figure 1 shows the spectrum opportunity concept.

Cognitive radio network (CRN) is the collection of two or more devices equipped with wireless communication, networking and cognitive radio capability. In CRN, the unlicensed users are called cognitive users or secondary users (SUs), and the licensed users are called primary users (PUs). For SUs, no fixed spectrum will be allocated [4], and they will be using licensed spectrum when PUs are not using it. Guaranteed quality of service (QoS) is provided to PUs, and best effort QoS is given to SUs in CRN. Based on the architecture of cognitive network, it is classified as infrastructural CRN and infrastructureless CRN [5]. In infrastructural CRN, cognitive users communicate each other in centralized manner through fixed infrastructure or base station. Infrastructural CRN is also called as centralized CRN. In infrastructureless CRN, SUs communicate each other in ad hoc manner. The infrastructureless CRN is also called as cognitive radio ad hoc network.

For efficient communication between various cognitive users, various routing protocols are used. Section 2 gives an overview of the existing routing protocols in CRNs. Section 3 presents the proposed modified AODV routing protocol for cognitive radio ad hoc network. Next, Sect. 4 presents the performance evaluation, and finally, conclusion and remarks are presented in Sect. 5.

Fig. 1 Spectrum opportunity concept (*source* Elsevier computer networks 50, p. 2130)



2 Related Work

Routing in CRN is difficult task because of its dynamic variation of the available channels, data rates, and bandwidth. The routing protocols in CRN can be classified as proactive, reactive, hybrid, and adaptive per hop, and the routing model can be centralized or distributed [6]. Depending upon the protocol operation, the routing protocols for CRN can be classified as spectrum-aware-based, multi path-based, local coordination-based, reactive source-based, and tree-based [7]. In [8–11], route discovery is incorporated with spectrum sensing. Sampath et al. [8] have proposed high-throughput packet transmission using spectrum-aware routing (SPEAR) protocol. SPEAR is an on demand routing protocol, and it makes routing decisions based on the collaboration of physical and MAC layers. Ma et al. [9] modify AODV protocol, and route decision is done by intermediate SU node such that it increases the available time for data transmission between SU source and destination in order to reduce switching delay. This routing protocol introduces extra overhead of broadcasting RREQ messages, and deafness introduces extra delay to RREQ messages.

The authors in [10] have proposed a routing protocol called SAMER which considered spectrum availability and quality. SAMER protocol enhances route robustness and improves throughput performance of the network. This protocol establishes route based on periodically collected information, and best path is selected depending on minimum hop count and spectrum availability. Cheng et al. [11] have also proposed an on demand routing protocol for CRN called spectrum-aware on demand routing protocol (SORP), where the best channel is selected based on minimum channel switching delay and back-off delay. In local coordination-based routing, nodes choose the flow direction based on neighborhood interaction. The authors in [12] have presented a routing protocol, which minimizes channel switching delay for SUs and help to minimize channel contention among SUs. This protocol improves the end-to-end performance of the network.

The [13, 14] considers multipath routing in CRN. Multipath routing discovers multiple routes between source and destination. This minimizes inter-path contention and interference and enhances route reliability. The authors in [15, 16] proposed a tree-based routing protocol. In this type of routing protocols, a tree structured network is enabled by configuring a root. The tree-based algorithm works along with channel selection mechanism. The tree-based scheme suffers from scalability problems because of overhead incurred in establishing and maintaining tree structure. Gymkhana, a distributed routing protocol proposed in [17], collects key parameters for routing between source and destination, and mathematical framework is modeled and evaluated to find the connectivity of different paths by considering PUs activity. The authors in [18] considered AODV as a pure on demand route acquisition system, as nodes do not maintain routing information or participate in exchange of routing tables. In ad hoc network, AODV routing protocol minimizes the number of required broadcasts for creating the routes. So, the modification of AODV routing protocol for CRN is considered in this paper.

3 Modified AODV Routing Protocol

In this paper, we are using a modified form of AODV routing protocol. In AODV routing protocol, the network remains in a idle state if there is no other communication. If it needs a connection to other node, which is the destination node, it broadcasts a message to its neighboring nodes to update the routing table. After updating the routing table, each node contains the neighboring node information, and a route is established to the destination node. AODV routing protocol is a combination of DSDV and DSR routing protocol. In this proposed work, that is modified form of AODV, there consists of detecting any obstacle in its path. In cognitive radio, the PU has the full access to the spectrum. Even in the absence of the PU, the SU can access the spectrum and transmits the information. Here, we are mainly focusing on routing in the cognitive radio between the SU. In this, route between the SU can be established in the absence of PUs. If there is a need for the connection between two SUs in the network, at first, it broadcasts the message to its neighboring nodes and updates the routing table. The route is established based upon the path where there are no PUs active. After the route is established between the SUs, it can transmit information to the destination nodes.

In the proposed method, it also checks whether there are any PUs active during the routing. If any PUs detected along the routing path, the route to the destination is changed since the PUs have full control to the spectrum. The PUs and the licensed users in the spectrum cannot be interfered under any circumstances. When a PU is detected along the routing, the nodes can broadcast the messages to the neighboring nodes and update the routing table. The new routing table contains the information to the destination where there are no PUs present. The new route can transfer packets to the destination. The main idea behind is that it continuously checks the presence of PUs in the routing path, and if any PUs activated, the route can be diverted to another one in which there is no PUs interference.

Algorithm

```

1  begin
2  if a packet exists to send from source to destination then
3    broadcasts the messages to all other nodes in network.
4    routing table updated which contains information of all neighbouring nodes.
5    route has been established by checking the absence of the primary users in the path.
6  if there is any primary user interference then
7    broadcasts the messages to all nodes to update the routing table
8    new route has been established without interfering the primary users
9  end endend

```

4 Simulation Results

In this, simulations are carried out using ns2 software. Table 1 shows the simulation parameters set in the simulation software. Figures 2 and 3 show the simulation results running at different times. The simulations process has been implemented

Table 1 Simulation parameters

Scenario dimension	1,000 × 1,000
Traffic application	CBR
Routing protocol	AODV
Antenna type	Omni antenna
Wireless MAC interface	IEEE 802.11
Propagation model	Two ray ground
Number of nodes	50
Number of primary users	25
Number of secondary users	25
Number of channels per radio	2

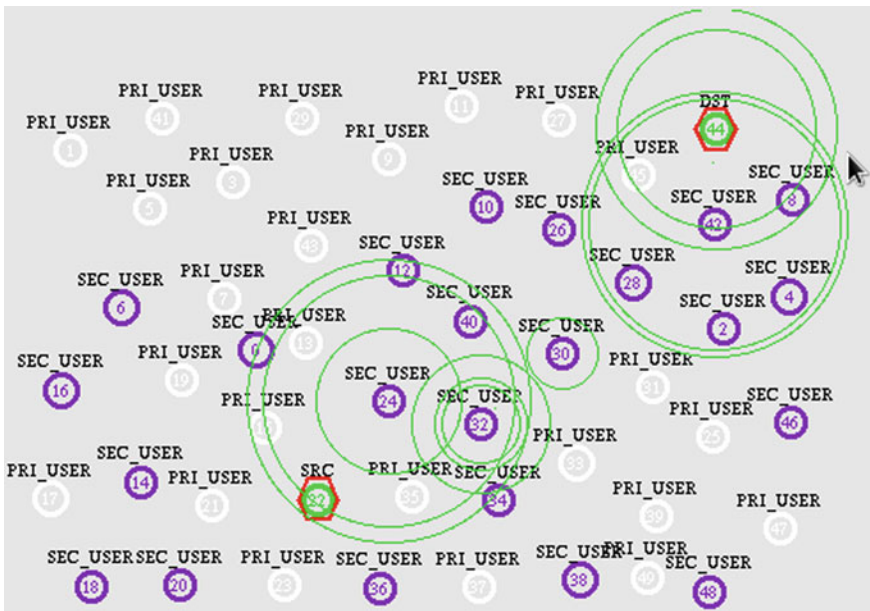


Fig. 2 Routing path established between source and destination

using ns2 software with tcl scripting language. We have considered 25 PUs and 25 SUs in the network to carry out the simulation. The protocol we have used in this scenario is modified AODV.

After starting, the simulation nodes will send the broadcast message to its neighboring nodes. The broadcast messages are useful to update the routing table. The routing table contains the information about its neighboring nodes and the distances from it. After routing table has been updated, the route has been established between the source and destination. Figure 2 shows the route which has been established between source and destination. The nodes are labeled as PRI_USER and SEC_USER in the simulation. The PUs are represented by white rounded

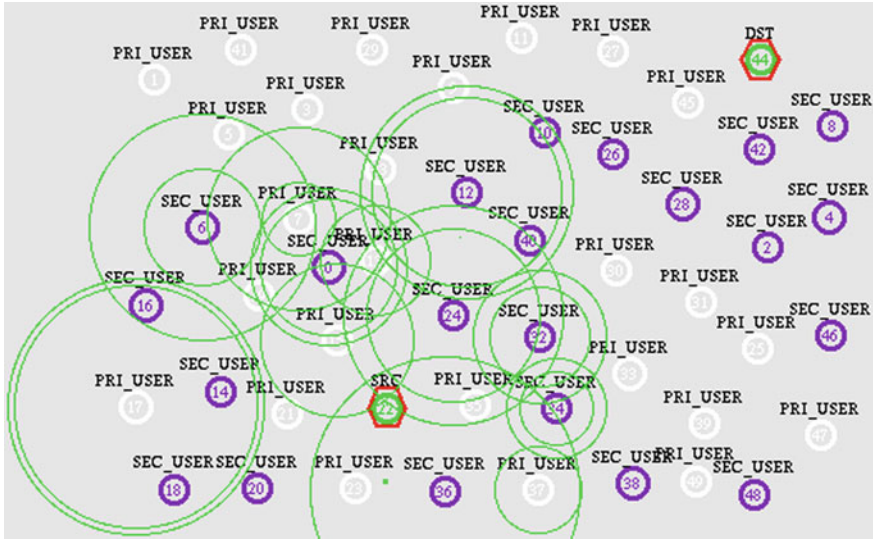


Fig. 3 Primary node 30 takes up its position, and secondary user 22 changes the routing path

nodes which are represented by odd numbers, and the SUs are represented in blue color, and they are in even numbers. In this scenario, route has been established between SU 22 which is the source and the SU 44 which is the destination one. The packets have been delivered from source to destination and its path which is shown in the Fig. 2. After some time, the PU is activated and takes the position of SU 30, and the routing cannot take place across the mentioned path which will cause interference to PUs. This is shown in the Fig. 3. The source node updates the routing table and checks the new path to transmit the packets from source to destination. After the new path has been constructed, routing can occur in that path.

After completing the simulations, the graphs have been plotted. The throughput and packet delivery ratio (PDR) graphs have been plotted. Figure 4 shows the throughput graph in terms of packet size. We have changed the packet size to 256 bytes and 512 bytes. By varying the packet size, the throughput also changes. It can be seen that throughput can be achieved higher when the packet size increases. The throughput can be achieved around 170 kbps in the case of packet size 512 bytes, whereas throughput can be achieved 87 kbps in case of packet size 256 bytes.

Figure 5 shows that throughput plotted against single-hop and multi-hop communication. Single-hop communication can be achieved higher throughput, whereas in case of multi-hop, throughput reduced to a reasonable level. The reason is that throughput can be achieved higher if there is no interference or there must be a direct path. If there are large number of paths between the throughput reduced to a low level.

In Fig. 6, it shows the PDR versus time. PDR is the ratio of packets received in terms of number of packets send. For small packet size of 256 bytes, PDR remains constant after some time, and when packet size increases, PDR value tends to increase.

Fig. 4 Throughput analysis for varying packet size

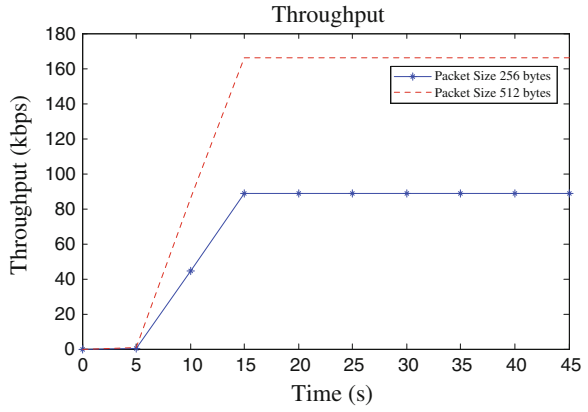


Fig. 5 Throughput analysis for different number of hops

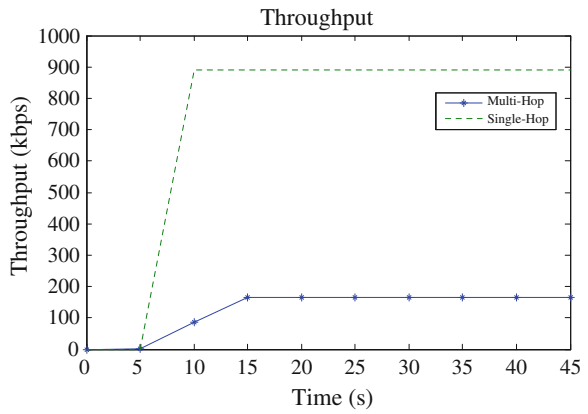
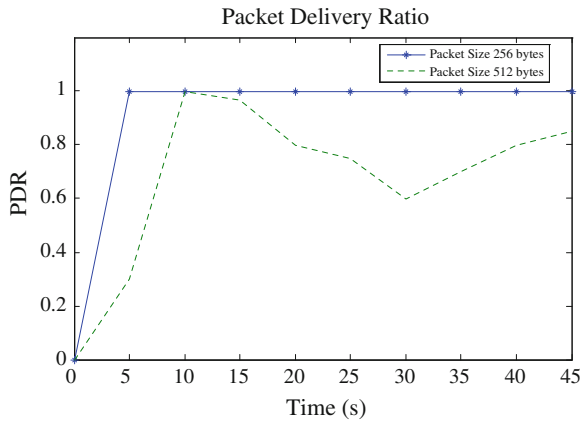


Fig. 6 Packet delivery ratio versus time



5 Conclusion

In this paper, modified AODV routing protocol for cognitive radio ad hoc networks is implemented. The main concept in this protocol is that to establish an optimal route between SUs by taking into account the PUs activity. The protocol provides better adaptivity since the path can be switched if there is any interfering PUs present. Simulation results show that the throughput can be increased if the packet size increases. Also, throughput performance is compared between single-hop and multi-hop. The PDR is also compared with the packet size. The PDR will be constant after 5 s for packet size of 256 bytes, and it varies in case of 512 bytes. The proposed algorithm focuses on PUs' activity and provides better path for SUs to forward packets. It reduces the routing overhead by dynamically changing the route based upon interfering nodes, and overall, it provides better performance.

References

1. Federal Communications Commission, Spectrum policy task force. Technical report (2002)
2. I.F. Akyildiz, W.-Y. Lee, M.C. Vuran, S. Mohanty, NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Comput. Netw.* **50**, 2127–2159 (2006)
3. S. Haykin, Cognitive radio: brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **23**, 201–220 (2005)
4. E. Buracchini, The software radio concept. *IEEE Commun. Mag.* **38**, 138–143 (2000)
5. B. Wang, K.J.R. Liu, Advances in cognitive radio networks: a survey. *IEEE J. Sel. Top. Signal Process.* **5**, 5–23 (2011)
6. H.A.A. Al-Rawi, K.L.A. Yau, Routing in distributed cognitive radio networks: a survey. *Wireless Pers. Commun.* **69**, 1983–2020 (2013)
7. A. Ali, M. Iqbal, A. Biag, X. Wang, Routing techniques in cognitive radio networks: a survey. *Int. J. Wireless Mobile Netw.* **3**(3), 69–110 (2011)
8. A. Sampath, L. Yang, L. Cao, H. Zheng, B.Y. Zhao, High throughput spectrum-aware routing for cognitive radio networks, in *Proceedings of the Third International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)* (2008)
9. H. Ma, L. Zheng, X. Ma, Y. Luo, Spectrum aware routing for multi-hop cognitive radio networks with a single transceiver, in *Proceedings of 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)* (2008)
10. I. Pefkianakis, S.H.Y. Wong, S. Lu, SAMER: spectrum aware mesh routing in cognitive radio networks, in *Proceedings of 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)* (2008)
11. G. Cheng, W. Liu, Y. Li, W. Cheng, Spectrum aware on demand routing in cognitive radio networks, in *Proceedings of 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)* (2007)
12. Z. Yang, G. Cheng, W. Liu, W. Yuan, W. Cheng, Local coordination based routing and spectrum assignment in multi-hop cognitive radio networks. *Mobile Netw. Appl.* **13**, 67–81 (2008)
13. X. Wang, T.T. Kwon, Y. Choi, A multipath routing and spectrum access (MRSA) framework for cognitive radio systems in multi-radio mesh networks, in *Proceedings of the 2009 ACM Workshop on Cognitive Radio Networks* (2009)

14. L. Lin, A.P. Wang, X.W. Zhou, X.N. Miao, Noncooperative differential game based efficiency-aware traffic assignment for multipath routing in CRAHN. *Wireless Pers. Commun.* **62**(2), 443–454 (2012)
15. B. Zhang, Y. Takizawa, A. Hasagawa, A. Yamauchi, S. Obana, Tree-based routing protocol for cognitive wireless access networks, in *Proceedings of IEEE Wireless Communications and Networking Conference* (2007)
16. D. Chen-li, Z. Guo-an, G. Jin-yuan, B. Zhi-hua, A route tree-based channel assignment algorithm in cognitive wireless mesh networks, in *Proceedings of International Conference on Wireless Communications and Signal Processing (WCSP)* (2009)
17. A. Abbagnale, F. Cuomo, Gymkhana: a connectivity-based routing scheme for cognitive radio ad hoc networks, in *Proceedings of INFOCOM IEEE Conference on Computer Communications Workshops* (2010)
18. C.E. Perkins, E.M. Royer, Ad-hoc on-demand distance vector routing, in *Proceedings of 2nd IEEE Workshop on Mobile Computing Systems and Applications* (1999)

Theoretical Framework of the Algorithm to Thwart MAC Spoofing DoS Attack in Wireless Local Area Infrastructure Network

M. Durairaj and A. Persia

Abstract A major threat in wireless local area infrastructure network is denial-of-service (DoS) attacks. It makes the resources unavailable for its anticipated user which can be accomplished through spoofing legitimate client/AP's medium access control (MAC) address. Less protection in MAC address led to get easy spoofing. Since the management frame is unencrypted, adversary sends the management frame to the victim using spoofed MAC address. This prerequisite goaded to offer an effective prevention mechanism for DoS attack. Even though several preventing mechanisms are available, no one provides complete solution in preventing MAC layer DoS attack. This paper proposes a theoretical framework of threshold value (ThreV) algorithm, which is based on setting up ThreV for the management frame, to effectively address this issue.

Keywords Denial of service · Medium access control · Spoofed MAC address · Threshold value algorithm · Wired equivalent protocol

1 Introduction

Security issues in wireless network increases as popularity increases. In wireless local area network (WLAN) infrastructure architecture, communication takes place with the help of access point (AP). In general, infrastructure network does not have firewall to defend the entire network. Physical protection of wired medium such as firewalls and shields cannot be applied to wireless networks. This makes the intruders to easily enter into the network and do malicious harm to the network.

M. Durairaj · A. Persia (✉)
School of Computer Science, Engineering and Applications, Bharathidasan University,
Tiruchirappalli, India
e-mail: persia_paradise@yahoo.co.in

M. Durairaj
e-mail: durairajum@gmail.com

There are number of attacks possible in infrastructure network [1]. Many people are not aware of the denial-of-service (DoS) attack [2] on their own network. Sending continuous stream of forgery frames by an attacker can easily slow down the network, which prevents the availability for authenticated clients. Several protocols were developed to protect wireless network. Everyone has security deficiencies. Wired equivalent protocol (WEP) is a basic part of IEEE 802.11 standard for the protection of wireless network which uses RC4 algorithm. There are several safety deficiencies like two messages encrypted by the same key stream. To overcome these deficiencies, Wi-Fi Protected Access (WPA) and 802.1x were developed. The 802.1x is a security protocol based on the frame structure of 802.11. It attempts to provide strong authentication, access control, and WEP key management for Wireless LANs. Unfortunately, 802.1x misses its goals in access control DoS attacks [3]. Currently, as literatures say there are no effective IEEE-approved ways to solve the security hole. This paper proposes a theoretical framework of ThreV algorithm to address the issues of preventing attacks in an infrastructure network.

In this paper, Sect. 2 presents the background review and related work, which are to understand the paper. Section 3 explains the architecture of the proposed technique and experimentations carried out. Section 4 presents the theoretical framework of ThreV algorithm which is to prevent attacks in an infrastructure network. Section 5 concludes the paper.

2 Related Work

To detect DoS attacks in its early stages before it reaches the victim using stateful and stateless signature, John Haggerty et al. propose Distributed DoS Detection Mechanism (DiDDeM). It provides a natural way of tracking the attack sources without requiring the use of any trace-back techniques. The DiDDeM offers a distributed and scalable approach to attack responses [4].

Samra et al. propose an algorithm to enhance the performance of the correlation of two wireless intrusion detection techniques (WIDTs) such as Received Signal Strength Detection Technique (RSSDT) and Round Trip Time Detection Technique (RTTDT) for detecting medium access control (MAC) spoofing DoS attacks. The experiments were demonstrated with the absence of false negatives and low number of false positives [5].

Ding describes an efficient solution to avoid DoS attacks in WLAN using Central Manager (CM). CM acts as a back-end server which maintains three tables and timer to detect DoS attacks. Apart from preventing DoS attack, this mechanism can be used to improve the performance of WLAN. This proposed solution is evaluated by five different DoS attacks such as large number of association requests (LASO), EAP failure, EAP start, EAPOL logoff, and MAC disassociation [6].

Sheng et al. propose Gaussian Mixture Modeling (GMM) for Received Signal Strength (RSS) profiling to detect spoofing attacks using multiple air monitors (AMs) which sniffs wireless traffic passively without the cooperation of APs and clients.

In this method, accurate detection of MAC spoof is obtained using GMM mechanism [7].

Saelim et al. provide a MAC spoofing detection algorithm in IEEE802.11 networks. To differentiate an attacker station from a genuine station, the proposed algorithm utilizes physical layer convergence protocol (PLCP) header of IEEE 802.11 frames. Experimental results provide cent percentage of MAC spoofing DoS detection when two monitoring stations are located at an appropriate location [8].

3 Proposed Solution to Prevent DoS Attack

Many security techniques were introduced to prevent DoS attacks; still, effective solution for DoS attack is needed. In this paper, multiple techniques are introduced to address the drawback of existing solutions.

Hybridization of multiple detection techniques is proposed to develop as an effective tool for preventing DoS attack in an infrastructure network. The detection technique called as computerized monitoring system (CMS) [9] is an integration of three algorithms which are ThreV, alternative numbering mechanism (ANM), and traffic pattern filtering and letter envelop protocol (TPatLetEn). The architectural frame work of the proposed model is illustrated in Fig. 1. Experimentation on the DoS attacks consists of two stages, i.e., *evaluation of attack* and the *preventive mechanism*. Effectiveness of the attacks is evaluated by measuring delay time, packet loss and throughput. Packet loss occurs when one or more packets of data travelling across a computer network fail to reach their destination. Packet delay time refers to the time taken for a packet to transmit across a network from source to destination. Throughput is the average rate of successful message delivery over a communication channel. It is usually measured in bits per seconds and sometimes in data packets per second. As a preventing mechanism, integration of threshold value (ThreV), ANM, and Traffic Pattern Filtering with (TPatLetEn) is used to detect and block DoS attack.

3.1 Computerized Monitoring System (CMS)

CMS integrates three detection techniques such as ThreV, ANM, and TPatLetEn. It maintains an intruder table (InT) contains MAC address of intruders, and basic identity check (BIC) table contains MAC addresses of WLAN users. When client sends a request to AP, it first checks in InT whether it has particular MAC address or not. If address is presented, the request is considered as spoofing attack. If not, the request goes to BIC for basic check. If the MAC address is not presented in the table, CMS blocks the user from further communication, whereas if the user's identity was found in BIC, the ThreV mechanism takes over this. After this process, the request goes to ANM and TPatLetEn. When above three conditions are satisfied,

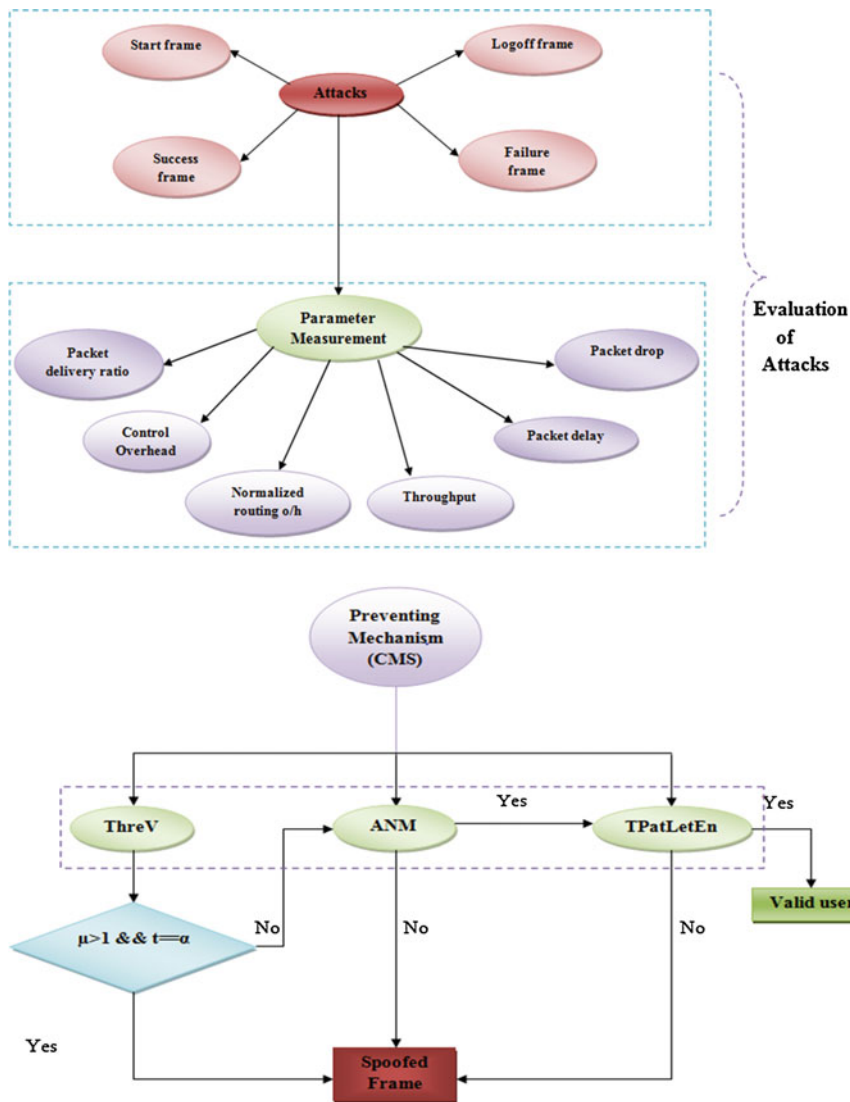


Fig. 1 Architecture of proposed model

CMS allows the AP/Client to start their communication. If any one of the prescribed solutions fails to satisfy, it can be considered as a spoofing attack. The operations of CMS are as illustrated in Fig. 2.

Basic Identity Check (BIC). BIC table contains MAC address of WLAN users who are all in the network. Once the request is processed, it checks in BIC whether the client/AP's MAC address presents or not. If not, it blocks the sender and stores

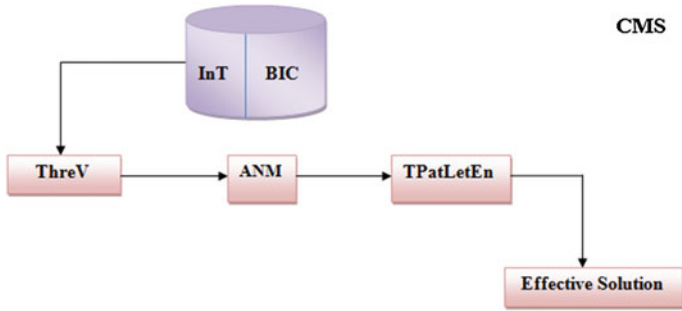


Fig. 2 Operation of CMS

its original MAC address in InT. If the MAC is presented, it redirected to ThreV algorithm to detect the MAC spoofing attack for effective detection of DoS.

Intruder Table (InT). Hackers are identified by the CMS, and it blocks the intruders, then find out the intruders MAC address, and stores its InT which stores the MAC address of the intruder.

Threshold Value (ThreV). By receiving login request from client, AP should respond by sending response message. Here, the threshold is assigned as 4 ms. AP should response to client at 4 ms. If AP receives more than one request within the certain ThreV, this request is considered as a spoofed frame (SF). In the case of AP, if client receives response message from AP before the threshold stage, it will be taken as a SF. There are some cases where AP cannot respond to client within threshold, and it may take more than 4 ms as a result of traffic overload. In this case, it cannot be assumed as an attacker. So the ANM detection technique will be used.

Alternative Numbering Mechanism (ANM). ANM is used instead of sequence number field. Odd number should be given for this such as 1, 3, 5, 7, 9, ..., n . Sequence number can be easily tracked by hacker. If intruder randomly guesses the sequence number, attacks can be launched in a simple manner. Maintaining ANM, hackers cannot assume the exact sequence number which presents in the numbering field. It is difficult for an intruder to assume the apt ANM to make MAC spoofing attack. If it finds out the exact ANM of client/AP, it should be redirected to TPatLetEn.

Traffic Patter Filtering with Letter Envelop Protocol (TPatLetEn). The client randomly generates two prime numbers $p1$ and $q1$ and then computes $N1 = p1 * q1$. In the same way, AP generates $p2, q2$ and computes $N2$ [10].

- During the authentication, the client sends an “envelop” containing $N1$ to the AP. The AP stores it and sends $N2$ to the client. The $N2$ sent by AP is common for all clients.
- When the client wants to disconnect, it sends the de-authentication frame to the AP, along with the $p1$. The AP will compute $p1/N1$ and finds whether it corresponds with the $N1$ which was already stored.

- If it is correct, the client will be disconnected. Otherwise, the frame will be rejected assuming that it is from the hacker.

Similar procedure is followed for AP when it wants to disconnect from the client. When the attack is vigorous, this protocol is not enough to save the client from the attackers. In the case of vigorous attacks, we propose the combination of traffic pattern filtering (TPF) with LEP where AP uses TPF along with LEP. The TPF works as follows:

- If a request is received more than five times at a particular time from a client, it infers that the request is from the hacker and ignores it.
- Since the hacker continuously sends request, AP is unable to process the request from the legitimate clients.

4 Preventing DoS Attack Using ThreV

ThreV is a mechanism which thwarts DoS attack in an infrastructure network. This section describes that how the attacks are took place and the way ThreV prevents the intruders.

4.1 *ThreV in EAP Start Frame Attack*

Client sends a start frame request to AP. After AP received the request, CMS checks the client's identity using InT. If the client's MAC address is presented in InT, it reject and blocks the intruder. If not, BIC takes over this packet which stores all the WLAN user's MAC addresses. If the client's MAC is presented, then ThreV process the request. Here, ThreV is set to 4 ms, i.e., $\text{ThreV}(\alpha) = 4 \text{ ms}$. At the 4th ms, the transmitted packet (μ) will be calculated. After calculating μ , the requisite conditions will be checked.

If the μ is greater than one within the threshold time (α), then it considered as a SF. If μ is equal to 1 but it exceeds α , this will be redirected to second prevention algorithm which is called ANM. This is not a simple process considering SF, because heavy traffic may be a reason for delayed packet. If μ is equal to one at threshold time α , this will be redirected to ANM which provides additional security to defend against DoS attacks. After subjected to ANM and TPatLetEn (ANM and TPatLetEn is not covered in this paper), CMS will decide whether it responds back to the user with yes or no. If it replies with yes, the communication between AP and client gets starts. The user will be blocked when any of the mechanisms are not satisfy the conditions and spoofing the intruders' original MAC addresses and store it in InT.

4.2 ThreV in EAPOL Logoff Frame Attack

During the communication period, when AP receives logoff request from client, it sends logoff packet to client by asking whether client want to logoff or not. If a client replies to AP with α and continue logoff message, ANM and TPatLetEn take control over the packet. If μ is greater than one within α period, this is considered as SF, and CMS spoofs the attackers' original MAC address and stores it in InT.

4.3 ThreV in EAP Success Frame Attack

After a login request is evaluated by three algorithms and find that the request is from legitimate, then it proceed with yes or no message to AP for further communication. If success frame μ is greater than two within its α period, this is considered as spoofing attack then the attackers MAC address is spoofed and stored it in InT. If μ is equal to one, but it exceeds α , this will automatically redirected to ANM and then TPatLetEn. This delay may happen due to traffic overhead which is a reason for not blocking them. If μ is equal to one at α , this will also be redirected to next prevention algorithm which enhances the security by hybridizing multiple techniques.

4.4 ThreV in EAP Failure Frame Attack

In some cases, the AP responds to client with failure message which represents congestion overhead in an infrastructure network. By sending failure frame, If μ is greater than two within its α period, this will be considered as spoofing attack and then the attackers' MAC address is spoofed and stored it in InT. If μ is equal to two but exceeds α , this will be automatically redirected to ANM and then TPatLetEn. This delay may happen due to traffic overhead which is a reason for not blocking them. If μ is equal to two at α , the next prevention algorithm which enhances the security by using hybridizing multiple techniques will be exploited. Sample ThreV algorithm to detect Start frame attack target as client is as follows.

Algorithm: ThreV

1. initialize $\alpha = 4 \text{ ms}$, $\mu = 1$, $t = 0$
2. if $\mu == 1$ && $t < \alpha$ then
3. Reject the packet, spoof and store it in Intruder Table
4. if $\mu == 1$ && $t > \alpha$ then
5. redirect it to ANM
6. if $\mu > 1$ && $t == \alpha$ then
7. Reject the packet, spoof and store it in intruder Table

8. *if $\mu == 1$ && $t == a$ then*
9. *redirect it to ANM*

The above algorithm describes that by receiving login request from client, AP should respond back by sending response message. Here, the threshold is assigned as 4 ms. AP should response to client at 4 ms. If AP receives more than one request within the certain ThreV, this request is considered as a SF. In the case of AP, if client receives response message from AP before the threshold, it will also be taken as a SF. There are some cases where AP cannot respond to client within threshold. It may take more than 4 ms because of traffic overload. In this case, it cannot be assumed as an attacker. So the ANM detection techniques will be used.

5 Conclusion

WLAN offers increased wireless access to the client with the help of AP. Since the popularity of wireless increases, the security issues also increases as well. DoS is a great threat in wireless infrastructure network which is an immense challenge to defend against it. Theoretical framework of ThreV algorithm is proposed in this paper to detect MAC spoofing DoS attack. This framework would provide a better solution for preventing the intruder's attack on WLAN. As a future work, ThreV algorithm is to be deployed in NS2 [11] and the throughput, drop rate, control overhead, end-to-end delay, and jitter will be measured, and optimization of the parameters will be proved with the aid of results. This paper suggests CMS is an effective solution for detecting and preventing MAC spoof attack in wireless local area infrastructure network.

References

1. A. Celik, P. Ding, in *Improving the Security of Wireless LANs by Managing 802.1x Disassociation*. Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC04) (2004), 53–58
2. J. Haggerty, Q. Shi, M. Merabti, Early detection and prevention of denial-of- service attacks: a novel mechanism with propagated traced-back attack blocking. *IEEE J. Sel. Areas Commun.* **23**(10), 1994–2002 (2005)
3. A.A. samra, R. Abed, Enhancement of passive MAC spoofing detection techniques. *Int. J. Adv. Comput. Sci. Appl.* **1**(5) (2010)
4. P. Ding, A solution to avoid denial of service attacks for wireless LANs. *Int. J. Netw. Secur.* **4** (1), 35–44 (2007)
5. Y. Sheng, K. Tan, G. Chen, D. Kotz, A. Cambell, in *Detecting 802.11 MAC Layer Spoofing Using Received Signal Strength*. The 27th Conference on Computer Communications IEEE (2008)
6. S. Sivagowry, A. Persia, B. Vani, L. Arockiam, in *A Solution to Prevent Resource Flooding Attacks in 802.11 WLAN*. Lecture Notes in Computer Science Communication in Computer and Information Science (CCIS 269) (2012), pp. 607–616

7. The ns Manual, the VINT Project (2009). <http://www.scribd.com/doc/52291274/Network-Simulator-2-Manual>
8. T. Saelim, P. Chumchu, C. Sriklauy, A new MAC address spoofing detection algorithm using PLCP header. IEEE ICOIN. 48–53 (2011)
9. A. Persia, S. Sivagowry, L. Arockiam, B. Vani, Inhibition of denial of service attack in WLAN using the integrated central manager. Int. J. Comput. Appl. **29**(8), 28–33 (2011)
10. A. Gupta, M. Garg, DoS Attacks on IEEE 802.11 Wireless Networks and its Proposed Solutions. Soc. Sci. Res. Netw. (2010)
11. A. Persia, S. Sivagowry, M. Durairaj, Study of thwarting DoS attacks by detecting MAC spoof in WLAN infrastructure network. IEEE Xplore 264–268 (2012)

PCA-Based Feature Selection for MRI Image Retrieval System Using Texture Features

N. Kumaran and R. Bhavani

Abstract Due to the vast number of medical technologies and equipments, the medical images are growing at a rapid rate. This directs to retrieve efficient medical images based on visual contents. This paper proposed the magnetic resonance imaging (MRI) scan image retrieval system using co-occurrence matrix-based texture features. Here, the principal component analysis (PCA) is applied for optimized feature selection to overcome the difficulties of feature vector creation with Haralick's texture features. Then, K-means clustering and Euclidean distance measure are used to retrieve best MRI scan images for the query image in medical diagnosis. The experimental results demonstrate the efficiency of this system in clusters accuracy and best MRI scan image retrieval against using all the fourteen familiar Haralick's texture features.

Keywords Co-occurrence matrix · Euclidean distance · K-means clustering · PCA · Texture features

1 Introduction

Content-based image retrieval (CBIR) is a technique in which different visual contents have been measured to search and retrieve images from the mass amount of image databases based on the input image. The content-based medical image retrieval (CBMIR) systems are medical domain-specific search engine for medical image databases, which indexing and retrieving medical images according to their visual contents such as texture, shape, and other information [1–3].

N. Kumaran (✉) · R. Bhavani
Department of Computer Science and Engineering, Annamalai University,
Annamalainagar, Chidambaram 608002, Tamil Nadu, India
e-mail: kumaran81@gmail.com

R. Bhavani
e-mail: shahana_1992@yahoo.co.in

The significance of new technologies such as X-ray radiography, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and picture archiving and communication systems (PACS) has resulted in an explosive growth in the number of medical images stored in the database. Medical images classifying, indexing, and retrieval in manual methods are very expensive and time consuming. This will lead various systems for storage, organization, indexing, and retrieval of the medical images.

The most important objective of the CBMIR system is to retrieve the images from the huge volume of medical databases with high accuracy by performing feature extraction, classification, and similarity measure process. So the retrieved images are used for various medical diagnostic purposes.

Generally, the medical image database contains a lot of texture-based information capable for retrieval purpose. This paper proposed the MRI scan image retrieval system in two parts of the human body such as the spine and brain using co-occurrence matrix [4]-based texture features with principal component analysis (PCA) [5] feature selection transformation, K-means clustering [6], and Euclidean distance measure [7]. The accuracy, precision, and recall rate of this system are high compared with using all fourteen Haralick's texture features [8].

The next section of the paper describes related works of the system. The brief discussion on the proposed work and Haralick's texture features is given in Sects. 3 and 4. Sections 5 and 6 explain about feature selection and K-means clustering. Section 7 deals the image retrieval. Section 8 shows experiments and results. In Sect. 9, the conclusion of the work with future prospects is given.

2 Related Works

Medical images have become a key investigation tool for medical diagnosis. With the growth of medical databases, new applications committed to statistical analysis of medical data have appeared. There are many existing systems that provide different methods and algorithms for CBMIR. The most important intention of all these systems is to prove the improvement of results so as to give support to the doctors and radiologists in diagnosis of treatments.

In [9], they described a medical image retrieval system using low-level features and high-level semantic features with 90 % of precision and recall rate. Here, medical images were segmented into several sub-images using fuzzy C-mean clustering algorithm and extracting 3 gray-level features using color moments. Then, the sub-images were changed to binary image, and seven shape features and four texture features were extracted using co-occurrence matrix. Then, the genetic algorithm was used to select optimal features, and the text information in the medical image was chosen for the semantic content of the report of radiologists.

Selvarani and Annadurai [10] illustrated a system by combining low-level content features and high-level semantic features for medical image retrieval. The semantic information was extracted from the DICOM header which was used to

perform the initial search. This pre-filtering of the images reduced the number of images to be searched. Then, texture features and shape features were found by Gabor filter and the fixed block resolution format. Image retrieval is performed by Euclidean distance measure. This system reduced the time taken to search the entire medical image database. Also, the average precision rate of 80 % is achieved.

Horsthemke et al. [11] explained two different texture feature-based CBMIR systems. The first system can be used to provide context-sensitive tools for computer-aided diagnosis with pixel-level co-occurrence matrices. The second system can be used directly as a computer-aided diagnosis system for case-based and evidence-based medicine with pixel-level and global-level co-occurrence matrices.

Zhang and Zhu [12] proposed a method using co-occurrence matrix to extract texture feature and edge histogram to extract shape feature of medical images. Then, Euclidean distance was used for medical image retrieval. Results of experimentation showed that the system had a recall rate about 90 % and applied to medical image retrieval with promising effect. In [13], the authors presented an evaluation of the diagnosis of dementia using texture analysis of brain MRI with Gabor wavelets and further classified by the back propagation network. Here, three different types of texture features were extracted: The first had the gray-level co-occurrence matrix (GLCM) features, the second had the Haralick's features, and the third had Gabor wavelet-based Haralick's features. From the comparison of the average efficiency, the statistical features extracted from Gabor wavelets provided better efficiency of 97 % than the other two methods.

Prasad and Krishna [14] evaluated the performance of two statistical methods of texture features proposed by Haralick's and Tamura for retrieving similar cases for CT scan brain images. To speed up the search process, selected features were extracted and indexed using hash structure. The Euclidean distance measure was used for similarity measurement. Both the methods were compared based on precision and recall. Tamura features were found to provide better retrieval results for CT scan brain images. In our previous work [15], the performance measures for spine MRI scan image retrieval proved that texture features (black-white symmetry, geometric symmetry, degree of direction, orientation features and central symmetry) based on the texture spectrum were somewhat good compared to Haralick's texture features (contrast, angular second moment, coarseness, entropy) and the combination of both features of image retrieval was the best.

3 Proposed Work

The block diagram of the proposed CBMIR system is shown in Fig. 1 in which two parts of the human body MRI scan images such as the spine and brain are used to construct the training data set.

In our work, we find GLCM to each MRI scan image in the training image database. Then, fourteen Haralick's textures feature vector values are extracted as

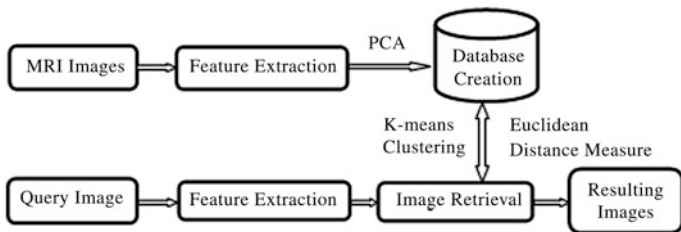


Fig. 1 The design of the proposed retrieval system

feature components, and PCA feature selection transformation is used to create an optimized database.

After that, using k-means clustering, the training images are clustered by means of the selected texture feature-based database. When a testing MRI scan image is submitted, the same texture feature extraction and feature vector value construction process have been applied to obtain the feature vector values for the testing image. Next, for similarity comparison between the query MRI scan image and the clustered MRI scan images, a Euclidean distance function is used. The closest Euclidean distance values for the query image are ranked, and best MRI scan images are retrieved for medical diagnosis.

4 Haralick’s Texture Features

Since we are interested in the statistical approach, we make use of the most suitable Haralick’s features. The major advantage of using the texture attributes is obviously their simplicity. The most common features used in practice are the measures derived from GLCM. These features have been widely used in the analysis, classification, and interpretation of medical images. The following fourteen Haralick’s texture features are extracted for the training MRI scan images.

1. Angular second moment: $\sum_i \sum_j p(i,j)^2$
2. Contrast: $\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}, |i - j| = n$
3. Correlation: $\frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y \mu}{\sigma_x \sigma_y}$
4. Sum of squares: variance: $\sum_i \sum_j (1 - \mu)^2 p(i,j)$
5. Inverse difference moment: $\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$

6. Sum average: $\sum_{i=2}^{2N_g} ip_{x+y}(i)$
7. Sum variance: $\sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$
8. Sum entropy: $-\sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} = f_8$
9. Entropy: $-\sum_i \sum_j p(i,j) \log(p(i,j))$
10. Difference variance: $\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$
11. Difference entropy: $-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$
12. Information measure of correlation 1: $\frac{HXY-HXY1}{\max\{HX, HY\}}$
13. Information measure of correlation 2: $(1 - \exp[-2(HXY2 - HXY)])^{1/2}$ where,

$$HXY = -\sum_i \sum_j p(i,j) \log(p(i,j)),$$

$$HXY1 = -\sum_i \sum_j p(i,j) \log\{p_x(i)p_y(j)\},$$

$$HXY2 = -\sum_i \sum_j p_x(i)p_y(j) \log\{p_x(i)p_y(j)\}$$
14. Maximum correlation coefficient: $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$.

Figure 2 shows the sample Haralick's feature extraction output screen of our proposed work.

5 Feature Selection

We are using PCA as a feature selection algorithm. PCA is useful when we have obtained features on large number of attributes and believe that there is some redundancy in those features. In our case, redundancy means that some of the features are correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, it should be possible to reduce the observed attributes into a smaller number of principal components (artificial attributes) that will account for most of the variance in the observed attributes.

After extracting fourteen texture features, the database is normalized using the z-transform and rescales the feature values. Then, using PCA transformation, we have selected five best features (i.e.) ASM, entropy, inverse difference moment,

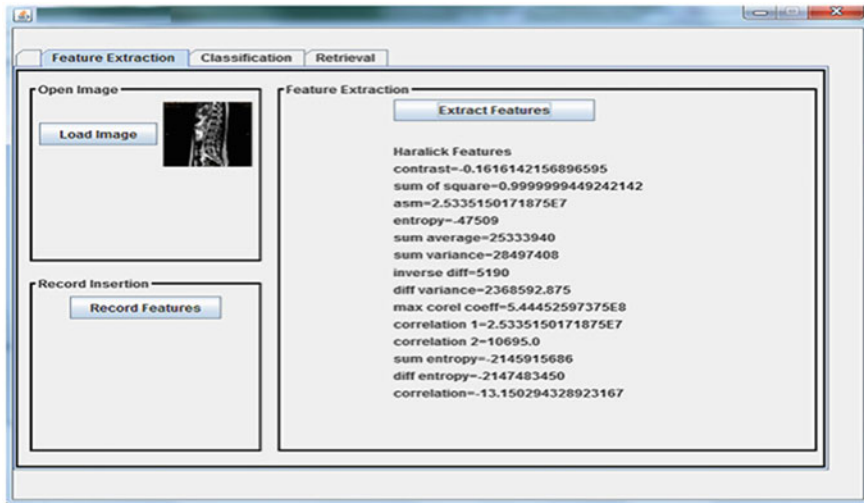


Fig. 2 Output screen for Haralick's texture feature extraction

inertia, and correlation for MRI scan image retrieval using variance as 0.95. So, instead of using fourteen Haralick's texture features, we are using only five texture features for best MRI image retrieval.

6 K-means Clustering

K-means clustering is one of the simplest unsupervised learning algorithm that solves the clustering problem. The procedure follows a simple and easy way to classify a given data set through 'K' number of clusters. In this work, given 1,250 normal and abnormal MRI scan images as training images in a 5-dimensional metric space, determine a partition of the images into maximum 20 clusters and 100 iterations, such that the images in a cluster are more similar cases to each other than two images in different clusters.

We initialize 20 clusters by arbitrarily selecting one image to represent each cluster. Each of the remaining images is assigned to a cluster, and the clustering criterion is used to calculate the cluster mean. These means are used as the new cluster points, and each image is reassigned to the cluster that it is most similar to. This continues until there is no longer change when the clusters are recalculated.

7 Image Retrieval

The Euclidean distance is calculated between the query image and the clustered images. If x_i and y_i are 2D feature vectors of the clustered training images and query image, respectively, then the distance measure is defined as,

$$d_{E(X,Y)} = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$$

The calculated distances are sorted in increasing order and display the first N images as the best similar MRI scan images for medical treatment. The sample output screen for MRI brain image retrieval is shown in Fig. 3.

8 Experiments and Results

This method is implemented on a computer system using Java as the programming language and MS Access as the backend. In this work, we used around 1,250 MRI scan images as a training set and 100 MRI scan images as testing set in BMP format with the size of 256×256 as a database. Two parts of the human body MRI scan images such as 900 spine and 450 brain images are used. The nature of clustering of the system is to cluster the normal and abnormal human body MRI scan images for spine and brain using 1,250 training data set. The effectiveness of the K-means clustering algorithm can be measured by accuracy, sensitivity, and specificity.

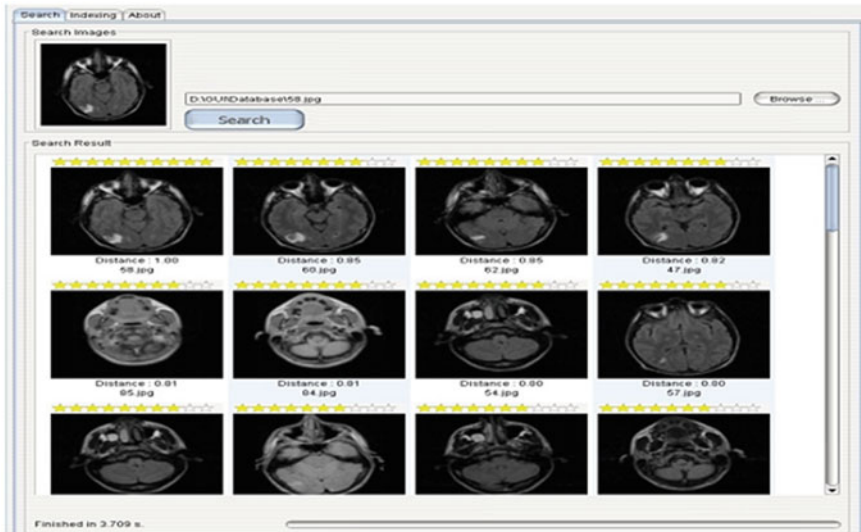


Fig. 3 Best retrieved MRI brain images

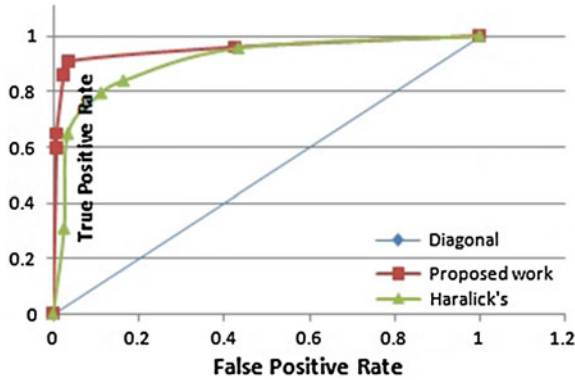


Fig. 4 Empirical ROC curves

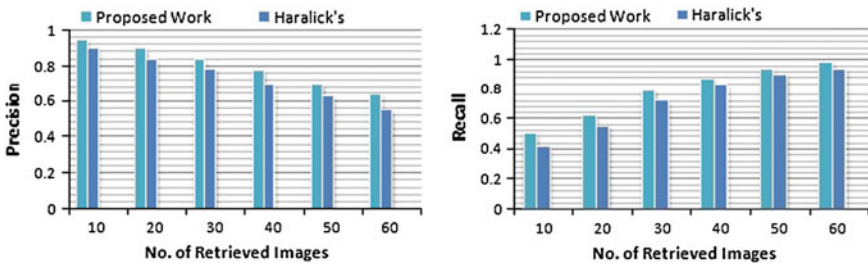


Fig. 5 Performance measure graphs based on precision and recall

The K-means clustering algorithm gave a test accuracy of 85.2 % while using fourteen Haralick’s texture features. The proposed PCA-based feature selection transformation with five Haralick’s features gave the test accuracy of 95.6 %. Figure 4 shows the empirical receiver operating characteristic curves in support of various cutoff points with a false-positive rate on the X-axis and true-positive rate on the Y-axis.

The effectiveness of the proposed method can be measured by precision and recall, which are often referred together since they measure the different aspects of the system performance. The results are given in the following Fig. 5.

9 Conclusion

In this paper, we have proposed an efficient MRI scan image retrieval system using PCA-based optimized Haralick’s texture features. The experimental results demonstrate that the proposed method has the best accuracy, precision, and recall rate than usual Haralick’s texture feature-based MRI scan image retrieval methods. We have planned to extend our work with all types of human body scan images.

References

1. H. Muller, N. Michoux, D. Bandon, A. Geissbuhler, A review of content-based image retrieval systems in medical applications: clinical benefits and future directions. *Int. J. Med. Inform.* **73**, 1–23 (2004)
2. X.S. Zhou, S. Zillner, M. Moeller et al., Semantics and CBIR: a medical imaging perspective, in *ACM Conference on Content-based Image and Video Retrieval* (2008) pp. 571–580
3. C.B. Akgül, D.L. Rubin, S. Napel, C.F. Beaulieu et al., Content-based image retrieval in radiology: current status and future directions. *J. Dig. Imag.* **24**(2), 208–222 (2011)
4. B. Ramamurthy, K.R. Chandran, Content based medical image retrieval with texture content using gray level co-occurrence matrix and K-means clustering algorithms. *J. Comput. Sci.* **8** (7), 1070–1076 (2012)
5. U. Sinha, H. Kangarloo, Principal component analysis for content-based image retrieval. *Radiographics* **22**, 1271–1289 (2002)
6. G.N. Lee, H. Fujita, K-means clustering for classifying unlabelled MRI data, in *IEEE Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications* (2007) pp. 92–98
7. S. Ayyachamy, V.S. Manivannan, Distance measures for medical image retrieval. *Int. J. Imag. Syst. Technol.* **23**(1), 9–21 (2013)
8. R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.* **3**(6), 610–621 (1973)
9. H. Shao, W.C. Cui, H. Zhao, Medical image retrieval based on visual contents and text information, in *IEEE International Conference on Systems* (2004) pp. 1098–1103
10. A.G. Selvarani, S. Annadurai, Medical image retrieval by combining low level features and Dicom features, in *IEEE International Conference on Computational Intelligent and Multimedia Applications* (2007), pp. 587–589
11. W. Horsthemke, D. Raicu, J. Furst, Task-oriented medical image retrieval, in *MICCAI Work Shop Proceedings* (2007) pp. 31–44
12. P. Zhang, H. Zhu, Medical image retrieval based on co-occurrence matrix and edge histogram, in *IEEE conference on multimedia technology* (2011) pp. 5434–5437
13. T.R. Sivapriya, V. Saravanan, P. Ranjit Jeba Thangaiyah, Texture analysis of brain MRI and classification with BPN for the diagnosis of dementia. *Eng. Inf. Technol. Commun. Comput. Inf. Sci.* **20**(4), 553–563 (2011)
14. B.G. Prasad, A.N. Krishna, Statistical texture feature-based retrieval and performance evaluation of CT brain images, in *IEEE International Conference on Electronics Computer Technology* (2011) pp. 1–4
15. N. Kumaran, R. Bhavani, Spine MRI image retrieval using texture features. *Int. J. Comput. Appl.* **46**(24), 1–7 (2012)

A Survey of Location Prediction Using Trajectory Mining

B.A. Sabarish, R. Karthi and T. Gireeshkumar

Abstract This paper is a research and analysis on the prediction of location of moving objects that gained popularity over the years. Trajectory specifies the path of the movement of any object. There is an increase in the number of applications using the location-based services (LBS), which needs to know the location of moving objects where trajectory mining plays a vital role. Trajectory mining techniques use the geographical location, semantics, and properties of the moving object to predict the location and behavior of the object. This paper analyses the various strategies in the process of making prediction of future location and constructing the trajectory pattern. The analyses of various mechanisms are done based on various factors including accuracy and ability to predict the distant future. Location prediction problem can be with known reference points and unknown reference points, and semantic-based prediction gives an accurate result whereas the probability-based prediction for unknown reference points.

Keywords Location-based services • HMM • Personal communication system • GMPMINE and cluster ensemble algorithm • Trajectory mining algorithms

B.A. Sabarish (✉)

Department of Information Technology, Amrita Vishwa Vidyapeetham,
Coimbatore, India
e-mail: sabarishpm@gmail.com

R. Karthi

Department of Computer Science Engineering, Asian College of Engineering
and Technology, Coimbatore, India
e-mail: karthiamrita@gmail.com

T. Gireeshkumar

Department of Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: gireeshkumart@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms
in Engineering Systems*, Advances in Intelligent Systems and Computing 324,
DOI 10.1007/978-81-322-2126-5_14

1 Introduction

Moving objects include vehicles, human, and mobile devices, which traverse across the geographical region (complete trajectory). Trajectory is the path of moving object traverse along and can be represented with the reference points or position in the path. It is essential to know the approximate location of the objects to be known to provide the location-based services (LBS). Many methods predict the location using the linear functions, but practically it follows a nonlinear dynamic function. Prediction can be classified as personal-based prediction and group-based prediction. Personal-based prediction involves the process of collecting information about the individuals independent of each other. This process produces unique trajectory pattern [1] for every individual, and this increases the number of trajectory pattern and complexity. In the other case, group-based mining process identifies the common trajectories and clusters the objects together and creates a general trajectory pattern. It provides a better result because the human tends to move in crowd than as individual. The issue arises in group-based mining is that there may be a chance of leaving out some interesting pattern when it is generalized into groups.

Figure 1 represents the general architecture of the trajectory mining process. The process starts from the collection of trajectory data and applying data mining techniques to gain the information and analyze the association between them. The gained knowledge can be used to predict the location behavior of the objects with the help of various prediction techniques including the statistical and probabilistic approaches (HMM), semantic-based prediction of various locations.

1.1 Personal-Based Prediction

In personal communication system (PCS), movement of mobile users is recorded in the database and service provider maintains it. The resources in communication can be dynamically allocated if the movement of the objects is predicted, which can improve the effective utilization of resources. Future prediction of mobile users can be easily done by efficient processing of their location-dependent queries regarding hotel and health care application. In the case of analyzing the individual movement pattern, mining can use the property of sequential data mining process. Individual movement will follow a pre-defined sequence of association rules, which can differ for working day or holiday or functions. For example, the movement of a human on

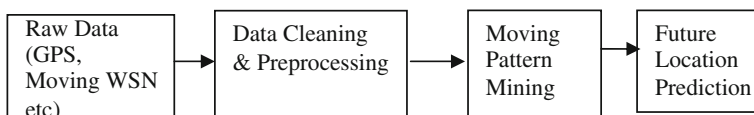


Fig. 1 General architecture of moving objects trajectory mining

a working day will be like Home → Bus Stop → Office → Bus stop → Home, etc. From this individual, mining information also the groups can be formed using the spatiotemporal attributes available in the historical data. The prediction accuracy again depends on the historical data collected and processed before the prediction process [2].

1.2 Group-Based Prediction

“Group-based data mining” is a data mining method involving the search for the association between individuals. Groups are formed based on the common behavior of the moving objects; it can be geographically closer or stay together for meaningful duration of time. Group mining is tedious when compared to individual pattern mining because the group behavior tends to have loss of information due to generalization. Normally, the process of finding frequency-pattern prediction involves GMPMINE and Cluster Ensemble algorithm for prediction of group movement patterns. In which, GMPMINE algorithm extracts the local association rules in the groups formed and CE algorithm puts the local association rules together to improve the association rule mining process. In Group-based prediction, the accuracy is calculated using punctual score and path score. Punctual score is the measure of relation of the moving object with respect to the region. Path score is an aggregate of the punctual score of all nodes along the path.

This paper is organized as follows. Section 2 describes motivation for this problem; Sect. 3 introduces terminology and needed preliminaries used in this paper. Section 4 gives some introduction about the trajectory mining techniques and analysis of various methodologies. Finally, Sect. 5 presents our conclusions.

2 Motivation

Predicting the future locations motivates the research issues in trajectory because of the dynamic behavior of the objects. Prediction of popular locations where most people visit leads to loss of information and leads to imbalanced data problem. And these prediction algorithms mainly predict only when the previous location (trajectory prefix) is known, which leads to loss of recall predictions. The performance of the prediction algorithm is affected by an increase in the number of moving objects, trade-off between the speed of prediction and accuracy, and ability to predict the distant future. The quality can be increased by means of partitioning and following client server architecture. The trajectory can be partitioned and clustered then prediction can be done at the cluster level. Method of client server architecture is proposed in which each individual object measures its own movement and server indexes the location at various level and uses queries using the filtering mechanism. Existing space partitioning approaches leads to two major issues answer-loss

problem and granularity problem. The moving objects tend to move in groups, which leads to data-loss problem and representation semantics of various reference points. The success of the LBS is based on area location information in the particular time stamps accurately. It will be a challenge to predict the behavior of the system, if the object movement is fast and dynamic.

3 Preliminaries

3.1 Movement Pattern Mining

Spatial and temporal relations and similarities can be identified from the trajectory dataset which is normally represented as a sequence of reference points. The trajectories point can be compared against the prototypes, which are usually generated by the available information and previous history. Classification of usual and unusual behavior is done based on the similarity-level measure with the remaining behaviors. Peng et al. introduced a prediction model based on the incremental model of predicting and identifying the mobile user, which can be used for resource allocation problem in limited resource-based networks.

3.1.1 Periodic Pattern Mining

Periodic pattern means identifying the repetition or replica of the pattern happening after some duration of time. For example, the visit of tourists, migration of birds and animals during a particular season for the year and used to find the behavior of the moving objects. It can be used to identify the peak hours of traffic during weekdays and weekends. From this periodic identification the prediction of the movement, a prediction can be done on the expectation of crowd and population details, etc. The main issue is the selection of appropriate, optimum period for predicting the behavior, which may range from an hour to a year.

3.2 Trajectory Clustering

Clustering of mobile objects has gained importance because of its dynamic behavior of movement. Trajectory clustering is the process of grouping similar paths. While comparing two trajectories, the middle some reference points (sub-trajectory) can be similar but not the whole path. This sub-trajectory has been identified as the impact sub-trajectories since it is used by many trajectories. These sub-trajectories can be used for specific area analysis. Trajectory clustering algorithm (TRACCLUS) is proposed by Lee et al. [3] which is partition and group framework.

TRACCLUS is divided into two phases: partitioning and grouping. In partitioning, each trajectory is divided into regions using the line segments. In grouping, the line segments are represented using minimum description length. Then, similar line segments should be grouped together using DBSCAN algorithm. TRACCLUS measures the accuracy in terms of the preciseness and conciseness properties. Trajectory can be classified based on the features of the regions identified [3, 4].

Since the trajectory mining process is dynamic, the size of data will be huge so it will be nice option to choose the sample and approximation/summarization techniques to solve the problem of scalability. Makris et al. proposed a machine learning-based approach to identify the usual behavior of the user. The concept of route is used, which in turn has identified by the pair of source and destination with some control points in the middle of the trajectory. Similarity can be measured by comparing the trajectory with the route. If a new trajectory is found, then it cannot be directly classified as usual behavior it will monitored whether that trajectory is used most often based on that a new route is created or an unusual route is identified [5].

4 Trajectory Mining Algorithms

Trajectory mining is the process of mining the path traversed by various objects and applies the data mining principles to identify the frequent matching pattern and predicts the path the nodes may travel [6].

4.1 Prediction Using Hidden Markov Models

Predicting the future locations with HMM proposed by Wesley Mathew et al., implemented the prediction model in the GeoLife project. Wesley proposed a hybrid model using the HMM, which clusters the available locations in the trajectory and different HMM models have been created for each cluster to train the system to predict the future location and measuring the similarity of the patterns. The previously analyzed trajectories are clustered according to their time of occurrence, and each is trained using different HMMs to predict the near future location. The geographical data are in the continuous location, which is converted into discrete specific for the regions which in turn used as the states. Each state will be associated with the probability distribution function for all the possible transitions. It helps in identifying the next location with higher probability. It provides an accuracy of 13 %. Osamma et al. proposed a template matching strategy, which compares the available patterns available in the prefix [7].

Hoyoung et al. proposed a novel method to predict the future location using the HMM. Hoyoung introduced a trajectory pattern model that describes an object's movement patterns based on hidden Markov process. It consists of a set of N frequent regions, each of which is associated with a set of M possible partitioned cells.

Cells are classified as observable states and hidden states. The discovered frequent regions are marked as hidden states and others as observable states. The probability of each observation sequence is calculated, which compares the current trend movement against historical data. It should be able to update the current trend along with the historical data of the available movement patterns identified. Accuracy of the prediction depends on the level of granularity of the information. To extract the frequent regions, a periodical mining method is used. The period of the mining data depends on the type of application, for example, animal movement is for a year, and human movement prediction can be for a day or an hour. The complete trajectory has to be divided to the equal number of periods, which is calculated and clustered together. Then, trajectory prediction model is constructed with the help of HMM. It identifies available states (N), observation Symbols in each state (M), initial state (π), state transition probabilities (A), and observation symbol probabilities (B). Using the state transition and observation symbol, probabilities and the initial state predict the future using the Baum–Welch algorithm [8].

4.2 Prediction Using String Matching Algorithms

In the trajectory mining using LCS, each reference point is considered as a character in a string. The prediction is done using the prefix string, i.e., the trajectory that has come across to reach the present state. Path similarity is measured in terms of similarity and importance. It should be non-overlapping and identical paths.

Banerjee and Ghosh [9] proposed a variation of LCS that is applied in Web usage mining which in turn can be applied to trajectory mining with slight variation. It uses the weighted LCS, which assigns weight value to the various reference points available on the trajectory based on the visits made to the reference point with respect to the available trajectory. Similarity graph can be constructed with the help of min-cut and balancing algorithm. Then, the cluster can be formed using the concept clustering algorithm. Ghosh specifies the tree construction start with the process of identifying first-level nodes, which constitutes to be the frequently visited reference points.

4.3 Semantic-Based Prediction Methods

Semantic trajectory mining is developed to improve the efficiency and using the meaningful movement of a human in order to trace and predict. It starts with identification of stay points and calculating the support and confidence value of each stay point with respect to another. Stay points are identified using the time spend on the particular location by human before making a decision to move or divert, etc. Stay points may serve as decision-making points. The trajectories can be presented with as a sequence of stay points. Each stay point is named with a

semantic name for representation and improves the mining process to gain the knowledge about the user prediction. It represents the path like $\langle \text{home}, \{\text{bank}, \text{park}\} \dots \rangle$, it represents the trajectory that follows the path of home, bank, and park. It specifies the information that bank and park go together which has the high support/confidence value. Support/confidence value is calculated using the conditional probability and data mining principles. Josh Jia-Ching Ying introduced a framework *Seman Predict* to evaluate the users' next location which may be either online or off-line. Off-line uses the notion of the stay location information to represent the user movement behavior. From the stay point information extracted the individual user's trajectory information can be identified (i.e) *Semantic Trajectory similarity measure* [10].

4.4 Pattern Matching Algorithms-Based Prediction Models

Monreale et al. [11] proposed a *WhereNext* predictor algorithm to predict the next location in the trajectory using the previously identified trajectory pattern (T-pattern). T-pattern is the common behavior of group users in space and time, which consists of node identifier, region identifier (spatial component), support value, and children for the node. They proposed a four step approach to predict the future location: (i) data selection, (ii) local models extraction, (iii) T-pattern tree building, and (iv) prediction. *WhereNext* represented the trajectory or spatiotemporal sequence is represented with the help of triples (x_i, y_i, t_i) , which corresponds to the location (x, y) in the time t . Trajectory pattern is an efficient algorithm to find out the frequent visiting location sequences using the threshold values including minimum support (σ) and temporal tolerance value (τ). Accuracy of this algorithm is analyzed using posterior analysis to calculate the average error rate, spatial and data coverage.

Prediction is done for three different scenarios

- (i) WhereNext_{r-1} intersects the region of the Node r .
- (ii) WhereNext_{r-1} enlarged by temporal tolerance t_h intersects the region of the node r .
- (iii) Does not intersect the region even after extending t_h .

Best matching path is identified by means of maximum path score and easy admissible prediction of future locations. *WhereNext* algorithm fails in the following occasions mainly when the length of the trajectory is lengthier than all other patterns and when it is distance from T-pattern (spatial and temporal distance).

Morzy [12] introduced a new approach by identifying the association rules using a modified apriori algorithm and uses *PrefixSpan* algorithm for predicting the future location. These models identify the frequent matching item set in the trajectories. Based on the frequent patterns, a classifier can be constructed. The construction creation involves three major steps mainly feature generation, feature selection, and modeling the classifier based on the features extracted in previous stages. This trajectories identified should be generalized but that cannot be done on a raw trajectory.

Morzy suggested dividing all the trajectories in equal-sized squares, which is named as cell and has four edges. Each cell is identified by the coordinates $\langle x, y \rangle$. When an object moves from one cell to the other, it crosses the edges. The edges can be either vertical (left/right) or horizontal (north or south).

Trajectories will be represented as the sequence of the edges it comes across, and the length of the trajectory is the number of edges it has come through. The trajectory can be maximal when the path does not match with any other patterns. Trajectories are divided into head and tail concatenation of both head and tail will lead to trajectory information. Support value for each trajectory is calculated and frequent trajectories in the trajectory. Frequent trajectories will be converted to a movement rule. If the movement is from $T1 \rightarrow T2$ where $T1$ and $T2$ are adjacent directories, then $T1 \oplus T2$ forms the frequent trajectory. $T2$ is named as head of the rule and $T1$ as the tail of the rule [12].

Morzy [12] proposed a way to decompose the location prediction problem into two sub problems:

- Discover movement rules with support and confidence greater than user-defined thresholds of min-sup and min-conf,
- Match the movement rules against the trajectory of a moving object for which the current location is to be determined.

5 Conclusion

This paper analyzes the various aspects of the trajectory mining algorithms. The trajectory-based algorithms are grouped based on the techniques; it uses to mine the trajectory information. Mainly they are based on probabilistic HMM models and string matching algorithms. By comparing the various techniques and algorithm, semantic-based algorithm provide better results for the trajectory with known and fixed reference points. The probabilistic and string matching algorithm-based prediction models can be used for unknown reference points, which provide a low accuracy. By analyzing the HMM and string matching algorithms, the string matching algorithms provide slightly better performance than the HMM-based models because most of the HMM models are designed specific for the local cluster, which changes dynamically.

References

1. J. Yang, M. Hu, TrajPattern, Mining Sequential Patterns from Imprecise Trajectories of Mobile Objects, EDBT 2006, pp. 664–681
2. L. Chen, M. Lv, Q. Ye, G. Chen, J. Woodward, A personal route prediction system based on trajectory data mining. *Inf. Sci.* **181**, 1264–1284 (2011)

3. J.G. Lee, J. Han, X. Li, H. Gonzalez, J.G. Lee, J. Han, X. Li, H. Gonzalez, TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. *J. Proc. VLDB Endowment* **1**(1), 1081–1094 (2008)
4. J.G. Lee, J. Han, K.Y. Whang, Trajectory clustering: a partition-and-group framework, in *Proceedings of the 2007 ACM SIGMOD International Conference on Management of data*, pp. 593–604
5. O. Ossama, H.M.O. Mokhtar, M.E. El-Sharkawi, An extended k-means technique for clustering moving objects. *Egypt. Inf. J.* **12**, 45–51 (2011)
6. F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, Trajectory Pattern Mining, *KDD'07* (2007) pp. 330–339
7. W. Mathew, R. Raposo, B. Martins, Predicting Future Locations With Hidden Markov Models *Ubicomp* (2012) pp. 911–918
8. H. Jeung, H.T. Shen, X. Zhou, *Mining Trajectory Patterns Using Hidden Markov Models* (Springer, Berlin, 2007), pp. 470–480
9. A. Banerjee, J. Ghosh, Characterizing visitors to a website across multiple sessions, in *Proceedings of the National Science Foundation (NSF) Workshop on Next Generation Data Mining*, pp. 218–227
10. J.J.C. Ying, W.C. Lee, T.C. Weng, V.S. Tseng, Semantic Trajectory Mining for Location Prediction GIS (2011), pp. 34–43
11. A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'09* (2009) pp. 637–646
12. M. Morzy, Mining Frequent Trajectories of Moving Objects for Location Prediction *MLDM'07*, pp. 667–680
13. Z. Li, M. Ji, J.G. Lee, L.A. Tang, Y. Yu, J. Han, R. Kays, MoveMine: Mining Moving Object Databases *SIGMOD10*
14. J.G. Lee, J. Han, K.Y. Whang, Trajectory clustering: a partition-and-group framework, in *SIGMOD'07*
15. J.G. Lee, J. Han, X. Li, Trajectory outlier detection: a partition-and-detect framework, in *ICDE'08*
16. F. Tung, J.S. Zelek, D.A. Clausi, Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance. *Image Vis. Comput.* **29**, 230–240 (2011)
17. Y. Wang, E.P. Lim, S.Y. Hwang, Efficient mining of group patterns from user movement data. *Data Knowl. Eng.* **57**, 240–282 (2006)
18. H.P. Tsai, D.N. Yang, M.S. Chen, Mining group movement patterns for tracking moving objects efficiently. *IEEE Trans. Knowl. Data Eng.* **23**(2), 266–281 (2011)
19. G. Yavas, D. Katsaros, O. Ulusoy, Y. Manolopoulos, A data mining approach for location prediction in mobile environments. *Data Knowl. Eng.* **54**, 121–146 (2005)
20. T.H. Cormen, C.E. Leiserson, R.L. Rivest, *Introduction to Algorithms* (MIT Press, Cambridge, 1990)
21. I.V. Cadez, S. Gaffney, P. Smyth, A general probabilistic framework for clustering individuals and objects, in *Proceedings of Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2000) pp. 140–149

Survey on Router Policies Providing Fairness and Service Differentiation Favoring Real-Time Data Transfers in Internet

Jyothish K. John and R.V. Siva Balan

Abstract Internet has IP-based routing mechanism. Forwarding and delivering in Internet is not guaranteed, and it is only a best-effort service. Situations of congestion make Internet difficult for real-time data transfers. Real-time data which have deadlines on quality may suffer from severe problems on throughput and delay. The term ‘fairness’ can be used to indicate a scenario where all flows get QOS approximately proportional to the data rate they contribute to the Internet. This can be achieved by regulating the high bandwidth consuming flows. But the real-time data flows which contribute only small data rate at regular intervals need service differentiation for achieving the deadlines on various QOS parameters. In modern Internet, which has more real-time data transfers, overall fairness can be ensured only if it has mechanisms to differentiate and support real-time transfers. The additional treatment they receive can be justified by such flows’ lower resource consumption. This paper is an extensive survey on router policies to ensure fairness to data flows and service differentiation that favors real-time traffic.

Keywords UDP protocol · Random early detection · RSVP · Per hop behavior · NCQ scheme

1 Introduction

Data transfers in the Internet can be classified into responsive and non-responsive. Responsive flows reduce their sending rate when congestion is sensed in the network. This happens when the network drops a few packets belonging to a flow. Hosts running their application using TCP have end-to-end congestion control

J.K. John (✉) · R.V. Siva Balan
Department of Computer Science and Engineering, Noorul Islam University,
Kumaracoil, India
e-mail: jyotiskj@rediffmail.com

R.V. Siva Balan
e-mail: rvsivan@gmail.com

mechanism. Non-responsive applications on the other side do not have a congestion detection mechanism, since they run over UDP protocol. Therefore, the growing rate of non-responsive flows in the Internet reduces the congestion adaptability of the Internet. Applications concentrating on performance rather than reliability fall on the latter category. Real-time data transfers which has high-delivery deadlines either on throughput or delay can be justified for their non-responsiveness because most of them uses small packet sizes [1], but long non-real-time flows being non-responsive reduces the stability of Internet. Therefore, fairness can be achieved by controlling the non-responsive flows which are non-real time. In this survey, data transfer from a particular source to a destination is mentioned as 'flow.'

2 Policies to Achieve Fairness

Random early detection (RED) gateway for congestion avoidance [2] was originally proposed in 1993 by Floyd and Jacobson and is now recommended for deployment in the Internet. RED allows a router to drop packets before its queue becomes saturated. Therefore, congestion responsive flows will back off early resulting in shorter average queue lengths which is good for interactive applications. Another advantage is that the packet drops will not occur in bursts. RED achieves this by dropping packets with a probability depending on the average queue length.

Floyd and Fall [3] introduce a router policy to restrict the unresponsive flows that does not reduce the sending rate even when packets are dropped. Such flows are termed as non-TCP friendly flows and they are identified from the drop history of RED. If the ratio of number of packets dropped per flow to the total number of packet dropped is higher, it signs an unresponsive flow. The flow level classification of data is needed in this approach and TCP friendliness of the flows is tested in the regular intervals. The non-TCP friendly flows are given much less priority while scheduling.

Pan et al. [4] propose a stateless active queue management scheme-CHOCe which deals with an alternate queue management scheme inspired from RED. The queue for incoming packet is FIFO which is having RED-like minimum and maximum thresholds. But, RED tries to maintain fairness only when the queue length become greater than minimum threshold, and by this time, misbehaving flows might have occupied the queue. Therefore, fairness in RED is granted only if the queue length is greater than minimum threshold. Compared to RED, CHOCe scheme differs in policy when the queue length falls between minimum and maximum thresholds. CHOCe algorithm tries to bring in more fairness. It assumes that the statistics of misbehaving flows are present in the occupancy before attaining minimum threshold. If the average queue size is less than minimum threshold, every incoming packet is queued into the FIFO buffer. If the average queue size is greater than maximum threshold, every arriving packet is dropped. When the average queue size is bigger than minimum threshold, each arriving packet is

compared with a randomly selected packet from the buffer, termed as drop candidate. If they have the same flow ID, they are both dropped. Otherwise, the randomly chosen packet is kept in the buffer and the arriving packet is dropped with a probability that depends on the average queue size. The drop probability is computed in the same way as RED.

Rangarajan [5] introduces an approach of regulating unresponsive flows. Flows need to be classified in this approach. The edge router takes over the control of rate adjustments of the unresponsive flows. The inner routers which drop the packet send back a source quench. The edge router analyses the source quenches for a particular flow and keeps a token bucket kind of controller for the flow. It adopts a multiplicative decrease and additive increase for the corresponding flow. Thus, the edge router is responsible for maintaining the correct rate of traffic to the Internet in the model.

Mahajan and Floyd [6] propose a mechanism to control the dropping of responsive flows. The technique has two parts: (1) Identifying responsive flows (2) Mechanism to prevent the dropping of responsive flows. Identification of the flows is performed by random sampling from the RED drop history. High bandwidth flows may have large number dropped packets. Such flows are termed as monitored flows. The monitored packets are dropped with high probability. The unmonitored flows are dropped only with normal priority of RED.

In order to stop the increasing packet loss rates caused by an exponential increase in network traffic, the IETF recommend deployment of active queue management techniques such as RED [2]. Active queue management can potentially reduce packet loss rates in the Internet. Feng et al. [7] proposes another Active Queue Management Algorithm, BLUE. The authors state that current techniques are ineffective in preventing high loss rates. The problem with the queue management algorithms (such as RED) is that they use queue lengths as the indicator of the severity of congestion. Instead, BLUE uses packet loss and link idle events to manage congestion.

All these are various methods to bring fairness in Internet by controlling the non-responsive flows. Real-time traffic which uses UDP as their delivery mechanism is non-responsive in nature. These methods certainly try to give a fair treatment to all traffic constrained to the equal-preference policy. However, most of these policies end up in punishing the real-time data flows rather than promoting them because of such flows' non-responsive nature.

3 Methods for Service Differentiation in Internet

The service differentiation methods in the Internet can be classified into three groups. The first two are general methods suited for all type of flows, while the third approach discusses policies that provide differentiation exclusive for real-time transfers (Fig. 1).

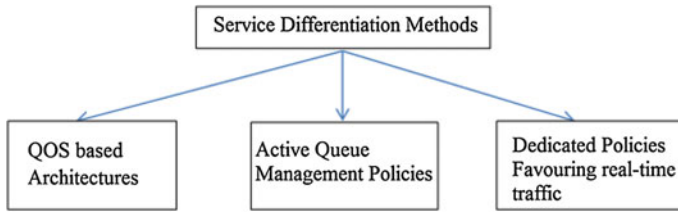


Fig. 1 Classification of service differentiation methods

3.1 QOS-based Architectures

Integrated services [8] are an application level architecture for ensuring QOS for any type of traffic. The model includes two sorts of service targeted toward real-time traffic: guaranteed and predictive service. The flow which utilizes this service has to reserve bandwidth using a signaling protocol, RSVP (Resource Reservation Protocol serves as an admission control and reservation protocol). The flows are sent to the network only if sufficient bandwidth is reserved for that particular flow. Resource reservation is done using the routers with respond to the RSVP signaling.

Differentiated services (DS) [9] are a scalable architecture when compared to the integrated services architecture. The DS domain consists of edge routers and core routers. The edge routers mark the packet with DS code point which is the classifier. The concept of marking is that the higher priority packet should get the priority proportional to the code point inside the DS domain. The core routers are responsible for conditioning the traffic and ensuring the quality. The core routers use the mechanism of metering, shaping, and dropping. The DS QOS architecture can classify the packet and use the IP TOS field for marking. Levels of quality assured in DS are indicated as per hop behavior (PHB).

3.2 Active Queue Management Policies

The general approach of these policies is that flows are graded and identified at the router level. The routers give preference for high priority flows at the time of scheduling and try to prevent dropping packets belonging to these flows at the time of congestion.

RED IN/OUT (RIO) [10] is an example of active queue management. When a packet reaches the router, it inspects the packet whether it is configured with 'in' or 'out.' If it is an 'in' packet, the router calculates the average queue length for the 'in' packets. The router calculates the average total queue size for all (both in and out) arriving packets in the case of an 'out' packet. The probability of dropping an 'in' packet depends on average 'in,' and the probability of dropping an 'out' packet depends on average total. Weighted random early detection (WRED) [11] is a

queuing discipline for a network scheduler suited for congestion avoidance. It is an extension to RED where a single queue may have several different queue thresholds. There is a queue threshold for each traffic class. Queue may have lower thresholds for lower priority packet. Queue buildup will cause the lower priority packets to be dropped, ultimately protecting the packet with higher priority in the same queue. This method allows quality of service prioritization possible for important packets from a pool of packets using the same buffer.

Lin and Morris [12] introduces a separate queue management scheme called Flow random early detection (FRED, modified version of RED) suitable for fragile and adaptive flows. FRED allows each flow to have a minimum and a maximum threshold and an average queue length. In the case where the queue length is less than maximum threshold and greater than minimum threshold, unlike RED to make a admit decision, it favors the packet with less average queue length than others. Therefore, FRED maintains better fairness by admitting the equal share of flows. Hence, it controls misbehaving flows from consuming more bandwidth.

3.3 Dedicated Policies Favoring Real-time Traffic

Mamatas and Tsaoussidis [13] propose a new service differentiation policy for real-time traffic. They suggest a new scheduling policy for non-congestive packets. Non-congestive packets are those packets which do not cause congestion in network (real-time packets). They are identified by their small packet sizes. The router captures the size of the packet and identifies whether it is real-time traffic or not. If real time, it is serviced faster. The limit of favor is controlled by a configurable threshold.

The same authors experiment the impact of non-congestive queuing, NCQ [14] for sensor traffic. For this, they simulate sensor traffic in NS2 environment. The size of the packet for prioritization is taken as 120 bytes. The experiment is done using a sensor access point (having a grid of 25 sensors) attached to the wired infrastructure having a dumbbell topology. They experimentally prove that throughput of sensor traffic is improved with NCQ implemented in router. The average delay is also reduced with NCQ.

Mamatas and Tsaoussidis [15] propose less impact better service (LIBS) philosophy. The packets are classified into congestive and non-congestive based on the size (size taken as 130 bytes). The prioritization based on this size will help sensor as well as VoIP traffic. The two types of traffic are highly sensitive to delay and need high throughput. The approach of prioritization is that for every packet, size is analyzed, and if it is <130, it is serviced using a non-preemptive priority queue. A threshold mentioned as 'ncqthreshold' is set to limit the prioritization. Simulations using NS2 by deploying NCQ in routers prove that the policy improves QOS for sensor and VoIP traffic.

Papastergiou et al. [16] suggest two levels of prioritization, tiny packets (e.g., sensor traffic) and small packets (e.g., VOIP traffic). The same NCQ Algorithm is

applied with modification and named as NCQ+. Tiny packets are less than 40 bytes in size, and small packets are having sizes between 40 and 120. Tiny packets are given more priority and scheduled faster. Then the priority is given to small packets. The prioritization is only up to a threshold level.

Many real-time applications use small packet sizes and have strong bandwidth requirements. Dimitriou and Tsaoussidis [17] justifies the selection of packet size as a service differentiation parameter. VoIP packets have size less than 140 bytes and sensor packets have less than 60 bytes. Bulk Data transfers having packet size greater than 1000 bytes cause congestion (e.g., FTP and Bit Torrent). But small amount of packet drops does not affect the performance of congestive traffic. A new scheme is proposed which is based on the axiom that 'different types of applications typically utilize different packet sizes.' RED which is ideal for maintaining QOS with its dropping policy treat all the packets alike while considering for drop. In size-based treatment, RED dropping policy is changed slightly in favor of small packets. The favoring of small packets should also be controlled and should not affect bulk flows.

The same authors propose size-oriented dropping policy (SDP) [18]. The policy makes sure that small packet is not dropped by the router in a congested network. The policy is inspired from the RED queue management scheme. The input queue is a RED queue. The RED queue drops the packet before the queue is full when the average queue size is in between minimum threshold and maximum threshold (to avoid tail drop). RED dropping policy is modified in SDP as when the queue size is between minimum and maximum thresholds as when a packet arrives, the average size is updated, and if the size of the packet is smaller than the average size and if the favoring has not exceeded the threshold, then the probability of dropping the packet is modified as SDP probability. Otherwise, the dropping probability is the probability of RED.

Dimitriou and Tsaoussidis [19] integrate the idea of SDP and NCQ in the router. This policy is termed as size-oriented queue management (SQM) Using SDP, small packets are favored by decreasing the dropping probabilities at the time of congestion. At the same time, the small packets are favored at the time of scheduling using NCQ scheme. SQM manages to satisfy broadly the quality constraints of real-time applications, without degrading the performance of bulk data applications. Making packet size as criterion, flows are identified and different dropping and scheduling policies are applied to favor time-sensitive traffic. Simple method to classify the packets into real time and non-real time is the specialty of size-oriented policy. This policy does not require flow state maintenance or complex packet inspection. Size is an easily extractable parameter of a packet. A comparison of the three service differentiation policies is given in Table 1.

Dedicated polices such as SDP try to promote the real-time transfers by altering their dropping plan. All these policies use RED as the queuing mechanism. But RED queue's admission control is limited between its minimum and maximum threshold. The initial queue occupancy before reaching the minimum threshold is not considered in all RED-based mechanism. Such evaluations help in computing the extent of promotion that can be given to real-time flows dynamically.

Table 1 Comparison of various service differentiation methods and their applicability to provide QOS to real-time traffic

	QOS-based architecture	Active queue management	Size-oriented policies
Features	Application level framework	Flows are categorized, assigned weightages and scheduled based on that	Identifies real-time packets and adopts scheduling and dropping schemes to favor real-time traffic
	Exclusive allocation of band-width to flows		
Advantages	Guaranteed QOS	Less implementation complexity (compared to the QOS-based architecture)	Less overhead when compared to the other two. No complex packet inspection
			State of flow need not be maintained
			Concentrates on packet Sizes rather than number of packets which is a better metric of band-width consumption
Disadvantages	High infrastructural needs	State of flow should be maintained	Less adaptability to future protocols
	Additional protocol overhead	Packet inspection is needed to analyze priority of each packet	Gateways implementing this policy may suffer from DOS attack
	Poor Scalability		RED-based policies do not consider initial queue occupancy
Applicability for achieving QOS for real-time traffic	General framework, but maybe used for QOS requirements for real-time traffic	General management policies suited for all sorts of traffic may be used for QOS requirements for real-time traffic	Policies are designed exclusively for real-time transfers

4 Conclusion

Policies concentrate in achieving fairness in Internet regulate the unresponsive flows. The fairness policies do not help real-time flows to maintain its QOS requirements, since such flows are treated as non-responsive because of their delivery mechanism. Therefore, real-time traffic needs an additional promotion at the time of scheduling and exception from dropping at the time of congestion to attain their delivery deadlines. Out of the three service differentiation approaches, Size-oriented policies are exclusively devised for promoting QOS in real-time transfers. In this approach, it is assumed that the real-time data transfers use comparatively smaller packet sizes. Packet size gives a good estimation of band-width consumption.

The shortcoming of SDP is that it is built over RED queuing mechanism which tries to maintain fairness in admission control only when the queue is nearing its capacity. It does not consider the initial occupancy of the queue for making the admission control as well. If the future protocols use bigger sizes, there is an issue in the classification of packets into real time and non-real time. Size-oriented policies concentrate on QOS parameters, throughput, and delay. There are some real-time data transfers which are jitter sensitive (interactive audio and video). The approach of size-oriented policies may be extended to the reduction of jitter for such kind of traffic. Also there may be applications trying to exploit the advantages of size-oriented gateways. The denial of service attack is possible by non-real-time applications purposefully reducing their size to get through the gateways. Methods to overcome such attacks are also the subject of future work.

References

1. S. Dimitriou, V. Tsaoussidis, A new service differentiation scheme: size based treatment, in *Proceedings of ICT 2008* (2008)
2. S. Floyd, V. Jacobson, Random early detection gateways for congestion avoidance. *IEEE/ACM Trans. Netw.* **1**(4), 397–413 (1993)
3. S. Floyd, K. Fall, Router Mechanisms to support end-to-end congestion control in the Internet. *IEEE/ACM Trans. Netw.* **7**(4), 458–472 (1999)
4. R. Pan, B. Prabhakar, K. Psounis, CHOKe: a stateless AQM scheme for approximating fair bandwidth allocation, in *Proceedings of INFOCOM 2000* (2000)
5. A. Rangarajan, Early Regulation of Unresponsive Flows. Technical Report TRCS99-26 (1999)
6. R. Mahajan, S. Floyd, Controlling high bandwidth flows at the congested router, in *Proceedings of ICNP 2001* (2001)
7. W. Feng, D. Kandlur, D. Saha, K.G. Shin, BLUE: A New Class of Active Queue Management Algorithms. Technical Report CSE-TR-387-99, University of Michigan (1999)
8. R. Braden, D. Clark, S. Shankar, RFC1633—integrated services in the internet architecture: an overview (1994)
9. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, RFC2475—an architecture for differentiated services (1998)
10. D. Clark, W. Fang, Explicit allocation of best-effort packet delivery service. *IEEE/ACM Trans. Netw.* **6**(4), 362–373 (1998)
11. www.cisco.com/en/US/docs/ios/12_2/qos_configuration/guide
12. D. Lin, R. Morris, Dynamics of random early detection, in *Proceedings of SIGCOMM 1997* (1997)
13. L. Mamatas, V. Tsaoussidis, Differentiating services with non-congestive queuing (NCQ). *IEEE Trans. Computers.* **58**(4), 591–604 (2009)
14. L. Mamatas, V. Tsaoussidis, Differentiating services for sensor internetworking, in *Proceedings of IFIP Fifth Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net '07)* (2007)
15. L. Mamatas, V. Tsaoussidis, Less Impact Better Service (LIBS): A Service Paradigm for Internet Telephony. Technical Report TR-DUTH-EE-2007-16, Democritus Univ. of Thrace (2007)

16. G. Papastergiou, C. Georgiou, L. Mamatas, V. Tsaoussidis, On short packets first: a delay-oriented prioritization policy. Technical Report TR: DUTH-EE-2008-8 (2009)
17. S. Dimitriou, V. Tsaoussidis, A new service differentiation scheme: size based treatment, in *Proceedings of ICT 2008* (2008)
18. S. Dimitriou, V. Tsaoussidis, Introducing size-oriented dropping policies as QoS-supportive functions. *IEEE Trans. Network Serv. Manage.* **7**(1), 14–27 (2010)
19. S. Dimitriou, V. Tsaoussidis, Promoting effective service differentiation with size-oriented Queue management. *Int. J. Comp. Telecommun. Netw.* **54**(18), 3360–3372 (2010)

Exploration of the Effect of Demographic and Clinical Confounding Variables on Results of Voxel-Based Morphometric Analysis in Schizophrenia

Anupa A. Vijayakumari, Priyadarshini Thirunavukkarasu, Ammu Lukose, Vikram Arunachalam, Jitender Saini, Sanjeev Jain, Bindu M. Kutty and John P. John

Abstract Brain morphometric abnormalities have been extensively reported in schizophrenia. In this research report, we used a voxel-based morphometry (VBM) approach to identify the effect of various confounding factors on gray matter (GM)

A.A. Vijayakumari (✉) · P. Thirunavukkarasu · A. Lukose · V. Arunachalam
Multimodal Brain Image Analysis Laboratory (MBIAL), Department of Psychiatry,
National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore 560029,
India

e-mail: av.anupa@gmail.com

P. Thirunavukkarasu

e-mail: galaxie2485@yahoo.co.in

A. Lukose

e-mail: ammu18lukose@gmail.com

V. Arunachalam

e-mail: drvikrm@gmail.com

J. Saini

Department of Neuroimaging and Interventional Radiology, National Institute
of Mental Health and Neurosciences (NIMHANS), Bangalore 560029, India

e-mail: jsaini76@gmail.com

S. Jain

Department of Psychiatry, National Institute of Mental Health
and Neurosciences (NIMHANS), Bangalore 560029, India

e-mail: sjain.nimhans@gmail.com

B.M. Kutty

Department of Neurophysiology, National Institute of Mental Health
and Neurosciences (NIMHANS), Bangalore 560029, India

e-mail: bindu.nimhans@gmail.com

J.P. John

Multimodal Brain Image Analysis Laboratory (MBIAL), Department of Psychiatry
and Department of Clinical Neuroscience, National Institute of Mental Health
and Neurosciences (NIMHANS), Bangalore 560029, India

e-mail: jppnimhans@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms
in Engineering Systems*, Advances in Intelligent Systems and Computing 324,
DOI 10.1007/978-81-322-2126-5_16

volume changes in patients with schizophrenia in comparison to healthy control subjects. Our findings highlight the importance of accounting for all possible confounding factors during study design and analyses, as well as setting appropriate statistical significance thresholds while reporting results in brain morphometric studies of schizophrenia.

Keywords Voxel-based morphometry · Computerized tomography · Structural magnetic resonance imaging · GLM ANCOVA analysis using VBM

1 Introduction

Structural brain abnormalities have been widely reported in schizophrenia, using various imaging modalities such as computerized tomography (CT), structural magnetic resonance imaging (sMRI), and diffusion tensor imaging (DTI) [1]. However, other than whole brain volume reductions, no regional brain morphometric abnormality has emerged from these studies to be considered pathognomonic of the disorder. Confounding variables such as age, age of onset [2], illness chronicity [3], medication exposure [4], substance abuse [5], unequal gender distribution [6], or handedness of study samples [7] could contribute to the inconsistent findings, especially when they are not controlled for in the analyses.

Various techniques have been employed to examine brain morphometry in schizophrenia. Voxel-based morphometry (VBM) implemented in the Statistical Parametric Mapping 8 (SPM8) software (Wellcome Department of Imaging Neuroscience, London; <http://www.fil.ion.ucl.ac.uk/spm>) is perhaps the most commonly used tool for identifying regional gray matter (GM) differences in schizophrenia by surveying the whole brain. It uses T1-weighted volumetric MRI scans, which are segmented into GM, white matter (WM), and cerebrospinal fluid (CSF) volumes. These tissue maps will be spatially normalized to a standard template. Then, statistical tests will be applied across all voxels in the image, to identify volumetric differences between diseased and control subjects [8]. Thus, VBM being an automated, rater-independent method provides an unbiased whole-brain approach for brain morphometric analysis.

Most studies that have reported significant regional GM reductions in the frontal, temporal, and parietal cortices in schizophrenia have controlled for confounding factors such as demographic characteristics (age and gender) and whole brain volume [total brain volume (TBV) or intracranial volume (ICV)] [9–12]. However, clinical variables such as duration of illness and medication exposure could also affect brain volumes as mentioned above. Thus, inclusion of these clinical variables could be essential for obtaining reliable results from VBM analysis. The aim of the present study is to demonstrate the importance of including the above clinical variables in VBM comparisons between patients with schizophrenia and healthy control subjects. We performed VBM analysis between a sample of patients with schizophrenia and matched healthy control subjects initially using only the

demographic (age, gender) and whole brain (TBV) confounding variables and then repeated the VBM analysis with further addition of the above 2 clinical variables (duration of illness, neuroleptic exposure in risperidone equivalents). We hypothesized that the above two VBM analyses would yield different results.

2 Methodology

2.1 Study Sample

The study was carried out at the National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India, with due approval from the NIMHANS Ethics Committee thus conforming to the ethical standards laid down in the 1964 Declaration of Helsinki. The study sample consists of 15 patients diagnosed with schizophrenia. The diagnosis for schizophrenia was ascertained with the Structured Clinical Interview for DSM-IV (Diagnostic Statistical Manual for Mental Disorders-Fourth edition) Axis I Disorders by a Psychiatrist. Eleven out of fifteen patients were not on neuroleptics. The remaining patients were on antipsychotics, the doses of which were converted to “risperidone equivalents” [13, 14]. Healthy control subjects ($N = 15$) with no neurological or psychiatric history matched for age, gender, and education with the schizophrenia sample were also recruited. Written informed consent was obtained from all participants prior to recruitment into the study.

2.2 Image Acquisition

MRI scans were performed on Siemens Magnetom Skyra 3.0 T scanner using a 20 channel head/neck coil. T1 anatomical images were acquired with Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence with a resolution of $1 \times 1 \times 1 \text{ mm}^3$. The image parameters were as follows: field of view = 200 mm, echo time, $TE = 2.44 \text{ ms}$, repetition time, $TR = 1,900 \text{ ms}$, flip angle = 9° , and bandwidth = 180 Hz/pixel. The GRAPPA technique with an acceleration factor of one was used for the experiment. The total acquisition time was 4 min 26 s. All scans were inspected for motion artifacts and gross pathology by an experienced neuroradiologist (J.S.).

2.3 Voxel-Based Morphometry (VBM)

VBM permits hypothesis-free whole brain, voxel-by-voxel between-group comparisons of GM volumes [15]. In our study, image processing and analysis was done using the VBM8 toolbox (Christian Gaser’s VBM8 toolbox; <http://dbm.neuro.uni-jena.de/>) of Statistical Parametric Mapping (SPM) version 8 algorithm

(www.fil.ion.ucl.ac.uk/spm) running under MATLAB R2012a. The primary step of VBM analysis of MR images involves spatial matching of MRI scans from different individuals to the same stereotactic space, a process known as spatial normalization. This preprocessing step confirms that location in one subject's MRI corresponds to the same location in other subject's MRI scan. Inter-subject registration of brain images was done by creating a sample-specific DARTEL template (Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra). A high-dimensional spatial normalization with DARTEL was used to normalize images to the DARTEL template. MR images were segmented to GM, WM, and CSF using prior tissue probability maps. Finally, images were spatially smoothed with a Gaussian kernel of 8 mm full-width-at-half-maximum which makes data conform to the Gaussian field model. The TBVs were calculated as sum of GM and WM volumes.

2.4 Statistical Analysis

For VBM analysis, analysis of covariance (ANCOVA) within the framework of general linear model (GLM) was used to compare between the two groups (schizophrenia and healthy control subjects). In the first step, only structural (TBV) and demographic (age, gender) characteristics were included as covariates, while, in the subsequent analysis, all of the above 3 variables along with the clinical confounding factors, viz., duration of illness and neuroleptic dosage (in risperidone equivalents), were included as covariates. A voxel-level peak threshold of family-wise error correction (FWE_v) $p < 0.05$ was set a priori to indicate significant volumetric differences between the two samples. However, in order to examine trend-level differences between the groups, VBM analyses were also performed using voxel-level peak threshold of false discovery rate correction (FDR_v) $p < 0.05$, cluster-extent-corrected family-wise error correction (FWE_c) $p < 0.05$, as well as at $p < 0.001$ (uncorrected, extent threshold (k) = 0 voxels) with 3 and 5 covariates, respectively.

3 Results

The mean age for schizophrenia and healthy participants was 29.06 ± 8.71 and 31.36 ± 12.24 , respectively. The gender (male/female) ratio for schizophrenia sample was 10:5, and for healthy control subjects, it was 9:6. GM volumetric differences in patients with schizophrenia when compared to healthy control subjects at the different statistical significance thresholds (FWE_v $p < 0.05$; FDR_v $p < 0.05$; FWE_c $p < 0.05$), using 3 (age, gender, TBV) and 5 covariates (age, gender, TBV, duration of illness, neuroleptic dosage), respectively, are given in Table 1.

Table 1 The effect of confounding variable on gray matter volume changes in the brain regions of patients with schizophrenia when compared to healthy control subjects

Type of confounding variable	Covariates	Correction for multiple comparisons*	Brain region	Coordinates of peak difference			P value
				X	Y	Z	
Demographic and structural	Age, gender, TBV	FWEv	Left inferior frontal gyrus	-35.54	16.23	-5.6	<0.05
			Right cerebellar declive	36.48	-67.6	-16.37	
		FDRv	Left inferior frontal gyrus	-35.54	16.23	-5.6	<0.05
			Right cerebellar declive	36.48	-67.6	-16.37	
			Right insula	39.27	12	6.08	
			Right claustrum	29.57	13.58	4.71	
			Right fusiform gyrus	18.33	-87.5	-18.51	
Demographic, structural and clinical	Age, gender, TBV, duration of illness, and cumulative neuroleptic exposure	FWEc	Left claustrum	-31.6	7.78	8.53	<0.05
			Left inferior frontal gyrus	-35.6	16.1	-4.26	

*FWEc cluster-wise family-wise error correction; FWEv voxel-level family-wise error; FDRv voxel-level false discovery rate

Voxel-based morphometric comparisons across groups at an uncorrected ($p < 0.001$) statistical threshold showed extensive volumetric reductions in the schizophrenia group when compared to the healthy control group with both 3 and 5 covariates (Figs. 3 and 5). When age, gender, and TBV were considered as covariates, GM volumes were significantly reduced in left inferior frontal gyrus (IFG) and right cerebellar posterior declive in patients with schizophrenia when compared to healthy control subjects after correcting for multiple comparisons using FWEv ($p < 0.05$) as shown in Fig. 1. At a threshold of FDRv ($p < 0.05$), significant regional GM volume reductions were observed in left IFG, right cerebellar posterior declive, right insula, right claustrum, and right fusiform gyrus in patients with schizophrenia compared to healthy control subjects as shown in Fig. 2. However, at an uncorrected threshold ($p < 0.001$; $k = 0$ voxels), diffuse GM volume deficits were observed in cortical and subcortical regions in schizophrenia patients when compared to healthy control subjects as shown in Fig. 3. When age, gender, TBV, duration of illness, and cumulative drug exposure were considered as covariates, no significant volume differences were observed at FWEv ($p < 0.05$) or FDRv ($p < 0.05$) thresholds. But at cluster-wise family-wise error (FWE)-corrected

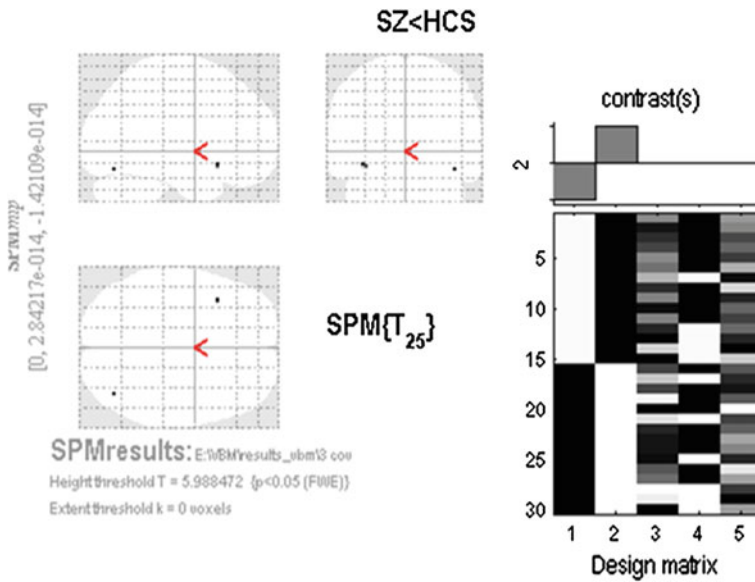


Fig. 1 Images depicting gray matter volume reductions in schizophrenia subjects, SZ ($n = 15$), when compared to healthy control subjects, HCS ($n = 15$), at a voxel-level FWE-corrected significance threshold of $p < 0.05$ and an extent threshold of 0 voxels, with age, gender, and TBV as covariates

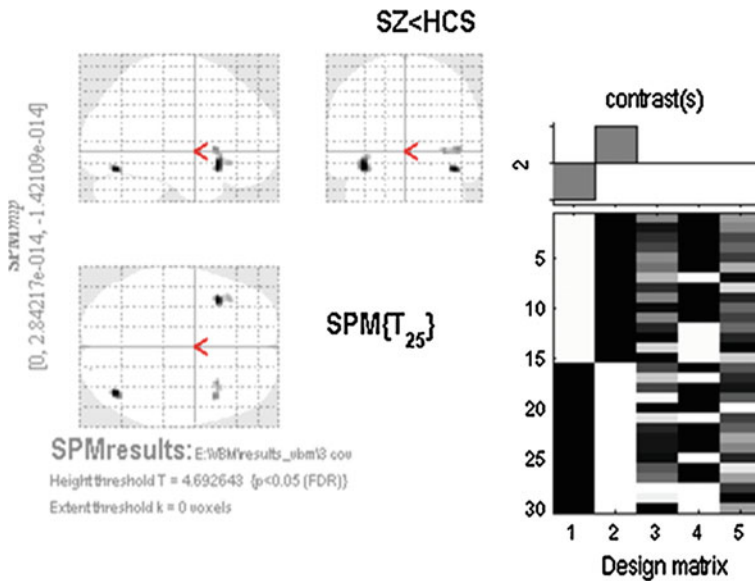


Fig. 2 Images depicting gray matter volume reductions in schizophrenia subjects, SZ ($n = 15$), when compared to healthy control subjects, HCS ($n = 15$), at a voxel-level FDR-corrected significance threshold of $p < 0.05$ and an extent threshold of 0 voxels, with age, gender, and TBV as covariates

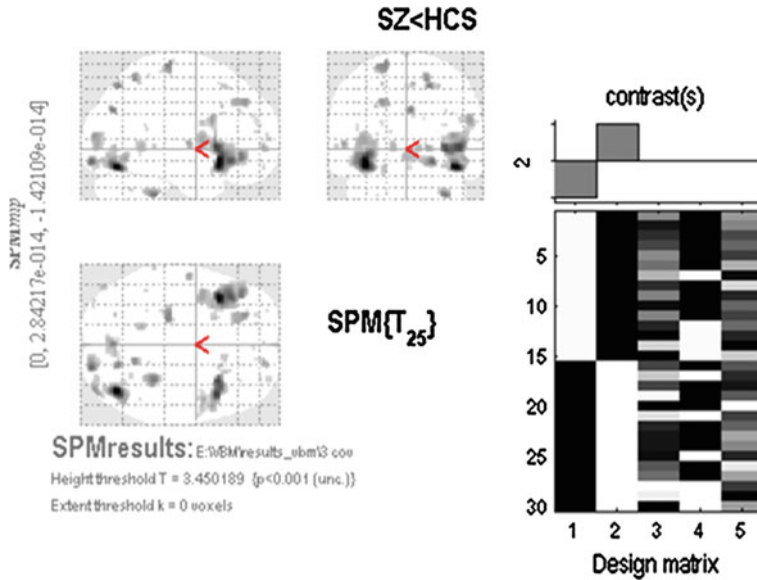


Fig. 3 Images depicting gray matter volume reductions in schizophrenia subjects, SZ ($n = 15$), when compared to healthy control subjects, HCS ($n = 15$), at an uncorrected significance threshold of $p < 0.001$ and an extent threshold of 0 voxels, with age, gender, and TBV as covariates

significance thresholding (FWEc; $p < 0.05$; $k = 648$ voxels), GM volume deficits in left claustrum and IFG were observed in patients with schizophrenia when compared to healthy control subjects as shown in Fig. 4. At an uncorrected threshold ($p < 0.001$; $k = 0$ voxels), widespread GM volume deficits were observed in cortical and subcortical regions in patients with schizophrenia in comparison to healthy control subjects as shown in Fig. 5.

4 Discussion

The findings of this exploratory study using VBM analyses highlight the variability of morphometric findings in schizophrenia at different statistical significance thresholds with or without inclusion of clinical confounding factors as covariates. When only demographic variables (age, gender) and whole brain volume (TBV) were included as covariates, significant morphometric reductions were noted in patients with schizophrenia even at the most stringent statistical significance threshold of FWEv $p < 0.05$. However, when clinical confounding factors (duration of illness and neuroleptic exposure) were included as covariates in addition to the above-mentioned covariates, volumetric reductions were noted only at more liberal statistical significance thresholds of FWEc ($p < 0.05$) and $p < 0.001$, uncorrected for multiple comparisons.

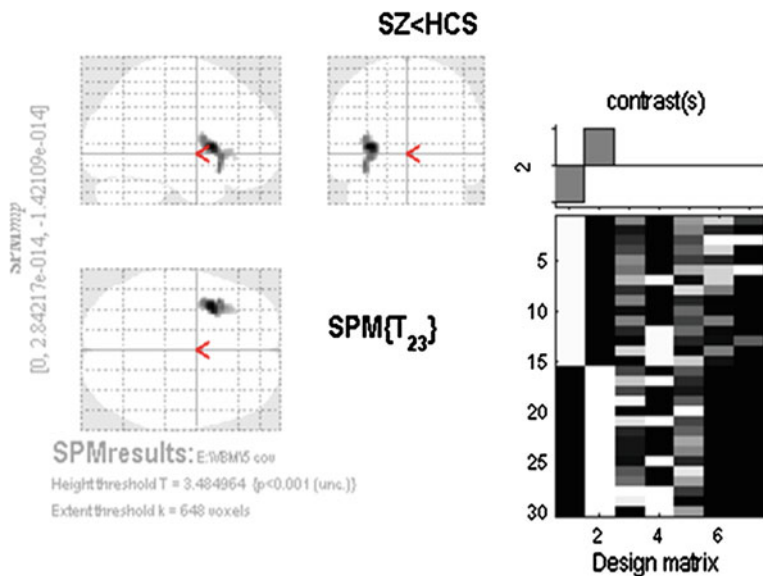


Fig. 4 Images depicting gray matter volume reductions in schizophrenia subjects, SZ ($n = 15$), when compared to healthy control subjects, HCS ($n = 15$), at a cluster-wise FWE-corrected significance threshold of $p < 0.05$ and an extent threshold of 648 voxels, with age, gender, TBV, duration of illness, and cumulative medication doses as covariates

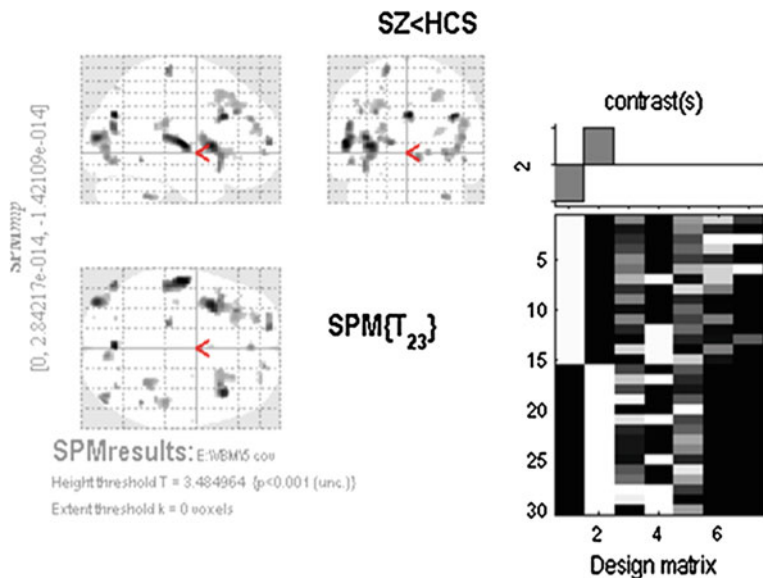


Fig. 5 Images depicting gray matter volume reductions in schizophrenia subjects, SZ ($n = 15$), when compared to healthy control subjects, HCS ($n = 15$), at an uncorrected significance threshold of $p < 0.001$ and an extent threshold of 0 voxels, with age, gender, total brain volume, duration of illness, and cumulative medication doses as covariates

When age, gender, and TBV were included as covariates in the GLM ANCOVA analysis using VBM, significant volumetric reductions were noted in the left inferior frontal gyrus (IFG) and right cerebellar declive in patients with schizophrenia in comparison to healthy control subjects at the stringent significance threshold of $FWE_v p < 0.05$. At a more liberal threshold of $FDR_v p < 0.05$, more brain regions showed volumetric reductions (Table 1). Even more brain regions showed volumetric reductions at an uncorrected ($p < 0.001$) significance threshold (Fig. 3). Our findings of volumetric reductions in schizophrenia are in agreement with a large volume of previous morphometric results [9, 10, 16, 17]. However, it is to be noted that the morphometric findings vary according to the statistical significance threshold employed. Moreover, only age, gender, and TBV were entered as covariates in the morphometric analyses.

When the clinical variables (duration of illness and neuroleptic exposure) were entered as covariates in the GLM ANCOVA in addition to age, gender, and TBV, there were no significant volumetric differences at the a priori decided significance threshold of $FWE_v p < 0.05$. Only a trend toward volumetric reductions was noted in the left claustrum and left IFG at $FWE_c p < 0.05$ (Table 1 and Fig. 4), with more brain regions showing trend toward volumetric reductions at a threshold of $p < 0.001$ (uncorrected) in patients with schizophrenia (Fig. 5).

In this exploratory analysis on a limited sample of patients with schizophrenia and healthy control subjects, the left inferior frontal gyrus seems to show robust volumetric reductions [$FWE_v p < 0.05$ with 3 covariates; $FWE_c p < 0.05$ with 5 covariates, *trend*] in patients with schizophrenia. Volumetric reduction in left IFG has been reported in many previous studies on patients with schizophrenia [18, 19]. Left IFG encompasses the Broca's area responsible for word generation, which is the most consistently replicated cognitive dysfunction in schizophrenia [20]. However, it is to be noted that even the left IFG only showed a trend for volumetric reduction at $FWE_c p < 0.05$, when 5 covariates were used in the GLM ANCOVA.

Overall, the results of the study highlight the importance of controlling for confounding variables in morphometric studies of schizophrenia using VBM. In order to obtain reliable results from morphometric studies, researchers should, therefore, control for all the various demographic, clinical, and other variables that might impact brain volumes, both at the sample recruitment stage and at the analysis stage. Another important issue that requires careful consideration is the statistical significance threshold that is used for making inferences regarding morphometric differences.

5 Conclusion

The present study demonstrates how the results of morphometric analyses vary according to the statistical significance threshold employed. Perhaps the inconsistency of previously reported morphometric findings in schizophrenia may reflect the variable extent to which the above methodological issues were given due

consideration [21, 22]. Therefore, future morphometric studies need to carefully control for various socio-demographic and clinical confounding variables that affect brain structure and employ appropriate statistical significance thresholds, in order to generate more reliable results.

Acknowledgments This study was supported by the Cognitive Science Research Initiative (CSI), Department of Science and Technology (DST), Government of India (Grant No. SR/CSI/79/2010).

References

1. M.E. Shenton, T.J. Whitford, M. Kubicki, Structural neuroimaging in schizophrenia: from methods to insights to treatments. *Dialogues Clin. Neurosci.* **12**, 317–332 (2010)
2. L. Burke, C. Androustos, J. Jogia, P. Byrne, S. Frangou, The Maudsley early onset schizophrenia study: the effect of age of onset and illness duration on fronto-parietal gray matter. *Eur. Psychiatry*. **23**, 233–236 (2008)
3. S.B. Schwarzkopf, S.C. Olson, J.A. Coffman, H.A. Nasrallah, Third and lateral ventricular volumes in schizophrenia: support for progressive enlargement of both structures. *Psychopharmacol. Bull.* **26**, 385–391 (1990)
4. M.S. Keshavan, W.W. Bagwell, G.L. Haas, J.A. Sweeney, N.R. Schooler, J.W. Pettegrew, Changes in caudate volume with neuroleptic treatment. *Lancet* **344**, 1434 (1994)
5. J. Borne, R. Riascos, H. Cuellar, D. Vargas, R. Rojas, Neuroimaging in drug and substance abuse part II: opioids and solvents. *Top. Magn. Reson. Imaging* **16**, 239–245 (2005)
6. J. Barnes, G.R. Ridgway, J. Bartlett, S.M. Henley, M. Lehmann, N. Hobbs, M.J. Clarkson, D. G. MacManus, S. Ourselin, N.C. Fox, Head size, age and gender adjustment in MRI studies: a necessary nuisance. *Neuroimage* **53**, 1244–1255 (2010)
7. C.D. Good, I. Johnsrude, J. Ashburner, R.N. Henson, K.J. Friston, R.S. Frackowiak, Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage* **14**, 685–700 (2001)
8. J.L. Whitwell, Voxel-based morphometry: an automated technique for assessing structural changes in the brain. *J. Neurosci.* **29**, 9661–9664 (2009)
9. R.A. Honea, A. Meyer-Lindenberg, K.B. Hobbs, L. Pezawas, V.S. Mattay, M.F. Egan, B. Verchinski, R.E. Passingham, D.R. Weinberger, H. Callicott, Is gray matter volume an intermediate phenotype for schizophrenia. A voxel-based morphometry study of patients with schizophrenia and their healthy siblings. *Biol. Psychiatry* **63**, 465–474 (2008)
10. E.M. Meisenzahl, N. Koutsouleris, R. Bottlender, J. Scheuerecker, M. Jäger, S.J. Teipel, S. Holzinger, T. Frodl, U. Preuss, G. Schmitt, B. Burgermeister, M. Reiser, C. Born, H.J. Möller, Structural brain alterations at different stages of schizophrenia: a voxel-based morphometric study. *Schizophr. Res.* **104**, 44–60 (2008)
11. V. Molina, G. Galindo, B. Cortés, A.G. de Herrera, A. Ledo, J. Sanz, C. Montes, J.A. Hernández-Tamames, Different gray matter patterns in chronic schizophrenia and chronic bipolar disorder patients identified using voxel-based morphometry. *Eur. Arch. Psychiatry Clin. Neurosci.* **261**, 313–322 (2011)
12. D.R. Watson, J.M. Anderson, F. Bai, S.L. Barrett, T.M. McGinnity, C.C. Mulholland, T.M. Rushe, S.J. Cooper, A voxel based morphometry study investigating brain structural changes in first episode psychosis. *Behav. Brain Res.* **227**, 91–99 (2012)
13. S.W. Woods, Chlorpromazine equivalent doses for the newer atypical antipsychotics. *J. Clin. Psychiatry* **64**, 663–667 (2003)
14. R.A. Kroken, E. Johnsen, T. Ruud, T. Wentzel-Larsen, H.A. Jorgensen, Treatment of schizophrenia with antipsychotics in Norwegian emergency wards, a cross-sectional national study. *BMC Psychiatry* **9**, 24 (2009)

15. J. Ashburner, K.J. Friston, Voxel-based morphometry-the methods. *Neuroimage* **11**, 805–821 (2000)
16. I. Harvey, M.A. Ron, G. Du Boulay, D. Wicks, S.W. Lewis, R.M. Murray, Reduction of cortical volume in schizophrenia on magnetic resonance imaging. *Psychol. Med.* **23**, 591–604 (1993)
17. K.O. Lim, W. Tew, M. Kushner, K. Chow, B. Matsumoto, L.E. DeLisi, Cortical gray matter volume deficit in patients with first-episode schizophrenia. *Am. J. Psychiatry* **153**, 1548–1553 (1996)
18. T. Ohtani, J.J. Levitt, P.G. Nestor, T. Kawashima, T. Asami, M.E. Shenton, M. Niznikiewicz, R.W. McCarley, Prefrontal cortex volume deficit in schizophrenia: A new look using 3T MRI with manual parcellation. *Schizophr. Res.* **152**, 184–190 (2014)
19. U.S. Torres, E. Portela-Oliveira, S. Borgwardt, G.F. Busatto, Structural brain changes associated with antipsychotic treatment in schizophrenia as revealed by voxel-based morphometric MRI: an activation likelihood estimation meta-analysis. *BMC Psychiatry* **13**, 342 (2013)
20. A. Szöke, F. Schürhoff, F. Mathieu, A. Meary, S. Ionescu, M. Leboyer, Tests of executive functions in first degree relatives of schizophrenia patients: a meta-analysis. *Psychol. Med.* **35**, 771–782 (2005)
21. C.M. Bennett, G.L. Wolford, M.B. Miller, The principled control of false positives in neuroimaging. *Soc Cogn Affect Neurosci.* **4**, 417–422 (2009)
22. J.P. Ioannidis, Excess significance bias in the literature on brain volume abnormalities. *Arch. Gen. Psychiatry* **68**, 773–780 (2011)

A Novel Method for Secure Image Steganography

S. Anjana and P.P. Amritha

Abstract Steganography is the science that involves communicating secret data in an appropriate multimedia carrier. The secret message is hidden in such a way that no significant degradation can be detected in the quality of the original image. In this paper, a new technique for embedding messages inside images is proposed. The pixels for message embedding are chosen such that the distortion introduced after embedding will be minimum. A distortion function is designed to calculate the cost of embedding for each pixel. The function evaluates the cost of changing an image element from directional residuals obtained using a wavelet filter bank. The intuition is to limit the embedding changes only to those parts of the cover that are difficult to model in multiple directions, avoiding smooth regions and clean edges. A technique that introduces less distortion to the carrier image will generally cause changes that are more difficult to detect, therefore providing more security.

Keywords Steganography · Steganalysis · Wavelets · Filter banks

1 Introduction

In recent years, steganography has emerged as an increasingly active research area, with information being imperceptibly hidden in images, video, and audio among others. With the wide availability of digital images and the high degree of redundancy present in them despite compression, there has been an increased interest in using digital images as cover-objects for the purpose of steganography. We use three main terminologies in steganography: the cover image, secret message, and the embedding algorithm. The cover image corresponds to the medium in which the

S. Anjana (✉) · P.P. Amritha

TIFAC CORE in Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: anjanaossery9@gmail.com

P.P. Amritha

e-mail: ammuviyu@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_17

151

message is hidden. Embedding algorithm is the method by which message is hidden within the cover medium. The cover image with the message hidden inside is known as the stego-image.

There are two main challenges in information hiding systems: high payload capacity and high robustness to modification. In steganography, robustness means the embedded data should be as immune as possible to modifications from attacks, and capacity refers to the amount of information that can be hidden within a given image. There is always a tradeoff between capacity and robustness. When the size of a secret message increases, there is always a chance for an attack to happen. So, the challenge for the designer is to develop an algorithm which would embed messages of large size with minimum possible embedding artifacts introduced [1]. Detectability of a steganographic system is defined as the relative entropy between the probability distribution of cover image and the stego-image. Any steganography system is called ϵ -secure if the relative entropy of the system is at most [2].

In this paper, an algorithm is proposed which will embed with minimum embedding artifacts while maximizing the payload. A set of wavelet filter banks are constructed to measure the cost of embedding for each pixel. Wavelet filter banks are constructed using daubechies 8-tap filters. We conducted experiments with different filters, and db filters gave the better result. We use filters with the assumption that edges and noisy regions have higher wavelet coefficients, and when we embed in those regions, the chance of detectability will be minimum. Filters are used to get the regions with high wavelet coefficients. And the algorithm which we use here will embed in those regions with high value for wavelet coefficients, such that detectability will be minimum.

1.1 Preliminaries

Currently, many practical steganographic algorithms [2] use LSB hiding techniques to hide the message. LSB hiding techniques hide the secret message into the least significant positions of the image pixels that affect the image resolution, which will reduce the image quality and make the image easy to attack.

1.2 LSB Embedding

The most common method used in steganography is LSB embedding. In this method, message is hidden by taking the image pixel and replacing the least significant bit of this pixel by the message bit. LSB replacement is the simplest type of embedding. If the LSB bit of the pixel and the message bit to be hidden are same, then leave the pixel as it is, whereas if the LSB bit and the message bit are different, then replace the LSB bit of the pixel with the message bit.

1.3 Attacks on the Existing Systems

There are three types of attacks on stego systems: Visual attacks, statistical attacks, and structural attacks. The following sections give a brief idea of these attacks.

1.3.1 Visual Attacks

The majority of steganographic algorithms embed messages replacing carefully selected bits by message bits. The idea of visual attacks is to remove all parts of the image covering the message. The human eye can now distinguish whether there is a potential message or still image content [2].

1.3.2 Statistical Attacks

The idea of the statistical attack is to compare the theoretically expected frequency distribution in steganograms with some sample distribution observed in the possibly changed carrier medium. The degree of similarity of the observed sample distribution and the theoretically expected frequency distribution is a measure of the probability that some embedding has taken place. The degree of similarity is determined using the chi-square test.

1.3.3 Structural Attacks

For structural attacks, consider palette-based steganography for palette images. Here, before embedding data, we reduce the number of colors so that the number of pixel color difference is very less [3]. This is done by changing the palette of the image. When this type of change in characteristic structure can be identified in the stego-image, then structural attacks occur.

2 Proposed System

In this paper, a new embedding technique is proposed which will embed in those pixels, which when altered gives minimum distortion. In this technique, an algorithm to calculate the cost of embedding for each pixel is developed and the embedding is done in such a way that the cost is minimum.

Wavelets and Wavelet Filter banks: In this method, we are using a set of wavelet filter banks to measure embedding distortion. Before constructing wavelet filter banks, we should know about low-pass and high-pass filters. A high-pass filter is an electronic filter that passes high-frequency signals and attenuates low-frequency

components. It is also called a low-cut-filter or bass-cut-filter. Whereas a low-pass filter passes low-frequency components and attenuates high-frequency components. Here, a directional filter bank is used to detect edges in local neighborhoods of each pixel. Then the changes in residuals caused by embedding are weighted and aggregated using a specially designed rule such that we get a low embedding cost only when the content is not smooth in any direction [4].

Before embedding, we have to calculate the cost of embedding for each pixel. For this, we construct a set of filter banks using daubechies 8-tap filters. It is constructed with low-pass and high-pass filters. FB (1), FB (2), and FB (3) are the set of filter banks we construct.

$$\begin{aligned} \text{FB}(1) &= h \cdot g^t. \\ \text{FB}(2) &= g \cdot h^t. \\ \text{FB}(3) &= g \cdot g^t \end{aligned} \quad (1)$$

The filter banks consists of low–high, high–low, and high–high decomposition filters, respectively, in Eq. (1). The support of each one-dimensional filter is 16, which gives each filter bank, a size of 16×16 . We define the k th directional residual as follows:

$$R(k) = \text{FB}(k) * C. \quad (2)$$

where $*$ is the mirror padded convolution, and C is the cover image. Mirror padding is used to prevent embedding artifacts at the boundary.

For each pixel, cost of changing is calculated by using a set of filter banks. When we apply high-pass filter to an image, the high-frequency coefficients are filtered out. That is, we get the pixels corresponding to edges and noisy regions. When we embed in these regions, chance for detection is less.

Now given a cover image C and stego-image S , we define the distortion between both the images as the sum of relative changes of the wavelet coefficients w.r.t the cover image and distortion is given as follows:

$$D(C, S) = \sum_K \sum_{uv} \frac{W_{u,v}^k(C) - W_{u,v}^k(S)}{e + W_{u,v}^k(C)}. \quad (3)$$

where $W(C)$ and $W(S)$ correspond to the wavelet coefficients in the k th decomposition obtained using the Eq. (2), for the cover image and the stego-image, respectively. From the Eq. (3), it is clear that the ratio is smaller when a large cover wavelet coefficient is changed, which corresponds to the edges and noisy regions [5]. When pixels in these regions are changed, the chance of detection is less. We develop our embedding algorithm in such a way that the pixels with the small value for the distortion function are taken first for embedding. It is clear that the embedding algorithm discourages making changes in areas where the content is smooth in at least one direction [6] (Fig. 1).

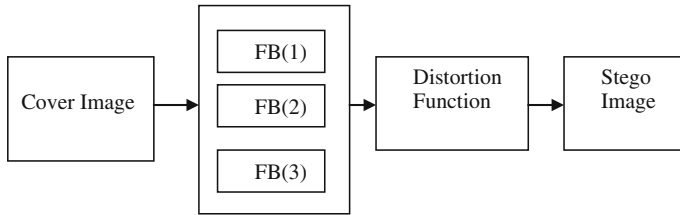


Fig. 1 Computing cost matrix

Embedding algorithm

The procedure of data hiding in the embedding algorithm works is as follows:

Input: Image file and text file.

Output: Text embedded image.

Procedure:

- Step 1: Take the input image and calculate the cost of embedding.
- Step 2: Find the length of input message.
- Step 3: Sort the cost array in increasing order.
- Step 4: Take each pixel from the sorted array.
- Step 5: Change the LSB of the pixel until the length of message is over.
- Step 6: Stop.

We start with the input image, and the output will be the stego-image with message hidden within it. Input images are taken from BOSS database [7]. Cover image is taken and the cost of embedding for each pixel is calculated using filter banks. After calculating the cost matrix, the values inside it are sorted, preserving the actual positions. Next, find the length of the message to be hidden. Then take each pixel value from the sorted array and replace the LSB of the pixel by looking at the message bit. If the LSB of the cover image and the message bit to be hidden match, then take the next pixel. Otherwise, change the least significant bit of the cover image.

Extraction Algorithm

The procedure for extracting messages inside images is as follows:

Input: Stego-image, Message length.

Output: Message.

Procedure:

- Step 1: Calculate the cost matrix for the image.
- Step 2: Sort the cost array.
- Step 3: Find the LSB of each image pixel from the cost array until the length of the message.
- Step 4: Concatenate the LSB's.
- Step 5: Return the message after concatenating.

3 Experiments and Results

The proposed technique has been simulated using the MATLAB-07 platform. A set of 8-bit grayscale images of size 512×512 are used as the cover image to form the stego-image.

Experiments were conducted with images from BOSS database [8]. The strength of the stego system is checked with statistical steganalysis tools. Chi-square test and RS steganalysis were conducted on the results and the strength of the stego system is verified [9].

RS steganalysis was conducted on images using virtual steganographic laboratory and the outputs proved the resistance of the stego system against RS steganalysis [10] (Fig. 2).

Chi-square test was also conducted on the output images. The results of chi-square test were compared with the results which used various other methods for embedding. The test was conducted on maximal length embedded images (i.e., all the pixels were embedded with message bits). A plot of probability of embedding with percentage of pixels embedded was obtained from chi-square test. Even though maximal length embedding was done, only a small percentage of pixels were detected to contain embedded bits. Also the values of chi-square statistics were large, which correspond to cover images [1]. False positiveness was comparatively small.

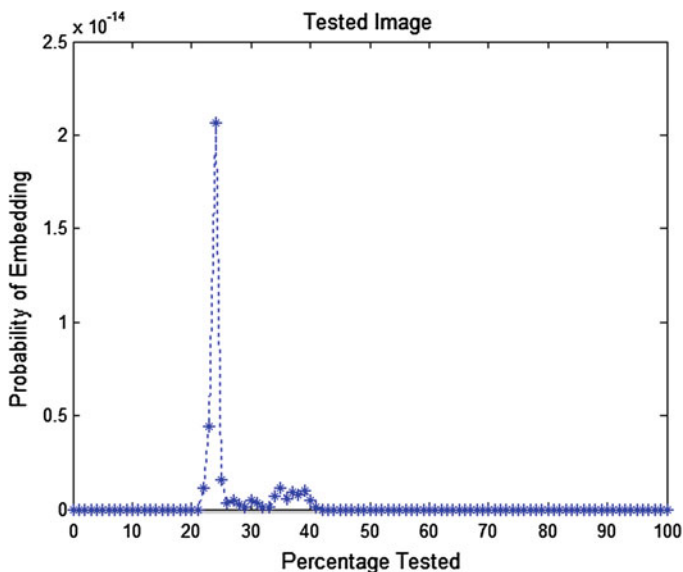


Fig. 2 Result of chi-square test on a stego-image obtained using the proposed method

Table 1 Performance evaluation of algorithm

Embedding rate (%)	Distortion	Coding loss	PSNR
50	12,402.7	9.98	50.97
60	13,356.9	9.91	55.67
70	17,257.9	10.58	45.98
75	12,829.37	9.67	59.34

3.1 Performance Evaluation

The above table gives the amount of distortion, PSNR, and the coding loss obtained on three different test images. Total distortion is a sum over the embedding costs where the pixel is changed. Coding loss is found as the ratio of actual payload with theoretically best possible payload (Table 1).

4 Conclusions and Future Work

This paper proves that embedding distortion can be minimized by restricting embedding changes to textures while avoiding smooth areas. Wavelet filter banks measure the embedding distortion in an effective way. The smoothness of the image is evaluated in multiple directions using the filter banks. Hence, cost matrix which we get from the distortion function is more accurate. The strength of the steganographic system is verified by different steganalysis tools. Due to the novel design of distortion function, we obtained good results.

Future works include using better directional filter banks to get a more effective design of the distortion.

References

1. R. Bobme, *Advanced Statistical Steganalysis* (Springer, Berlin, 2010)
2. A. Westfield, A. Pfitzmann, *Attacks on Steganographic Systems* (Dresden University Of Technology, 1999)
3. T. Filler, J. Fridrich, Design of adaptive steganographic schemes for digital images. In *Information Hiding, 9th International Workshop* (2007)
4. M. Siffuzzman, M.R. Islam, M.S. Ali, Wavelet transform and its advantages compared to fourier transform. *Recent trends and developments. J. Phys. Sci.* 13 (2009)
5. J. Fridrich, V. Holub, *Digital image steganography using universal distortion* (2013)
6. J. Fridrich, V. Holub, *Designing steganographic distortion using directional filters*. Air force Office of Scientific Research (2013)
7. M. Vetterli, Wavelets and filter banks: theory and design. *IEEE Trans. Sig. Proces.* 40 (1992)

8. T. Filler, T. Pevny, T. Bass, Break our steganography systems: the ins and outs of organizing BOSS. In *Information Hiding*, pp. 59–70 (2010)
9. J. Fridrich, M. Goljan, Practical steganalysis of digital images-state of the art. In *Proceedings of SPIE*, vol. 4675, pp. 1–13 (2002)
10. J. Fridrich, M. Goljan, R. Du, Reliable detection of LSB steganography in color and gray scale images. *IEEE Multimedia* 8 (2001)

Techniques for Enhancing the Performance of TCP in Wireless Networks

MD. Sirajuddin, Ch. Rupa and A. Prasad

Abstract TCP is a connection-oriented and reliable transport layer protocol. Currently, it is most widely used protocol in the Internet. All the reliable Internet applications make use of this protocol. This protocol works well in the wired network, but it does not produce satisfactory results if it is used in the wireless networks. Usage of TCP in wireless networks leads to performance degradation, because it considers all the packets losses as due to congestion and reduces the packet sending rate, and at the same time, it diminishes the network throughput. This feature is suitable for wired networks in which packet loss mainly occurs due to congestion. However, this is inappropriate in wireless networks where packet loss occurs due to signals fading, high bit error rate, hand-off, etc. It misinterprets all packet losses as due to congestion and reduces its congestion window. This misinterpretation of packet loss mainly decreases the throughput. This paper depicts the performance of all the efficient mechanisms which have been developed to improve the performance of TCP.

Keywords Wireless networks · Congestion control · Improved mechanisms · RTT

MD. Sirajuddin (✉)

Department of C.S.E, S.M.C.E, Guntur, Andhra Pradesh, India
e-mail: siraj.cs@gmail.com

Ch. Rupa

Department of C.S.E, VRSEC, Vijayawada, Andhra Pradesh, India
e-mail: rupamtech@gmail.com

A. Prasad

Department of C.S, Vikrama Simhapuri University, Nellore, Andhra Pradesh, India
e-mail: prasadjkc@yahoo.co.in

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_18

1 Introduction

TCP [1] is most widely used protocol in the Internet. All the Internet applications which provide reliability make use of TCP. When it came into an existence all the networks were made up of wired links. It was designed for wired networks. In such networks, the packet loss triggers TCP congestion control algorithm at sender and hence reduces the congestion window. The congestion control mechanism of TCP is very effective in the wired networks because the packet loss rate is very low. In such networks, packet loss occurs mainly due to the congestion. In effect, the TCP congestion algorithm cannot differentiate the packet losses caused by congestion from the one caused by error. Once a loss occurs, TCP considers it as a loss occurred due to congestion and reduces its congestion window. This leads to performance degradation and underutilization of channel bandwidth. This behavior is acceptable in wired networks because 99 % packet loss occurs due to congestion. However, this behavior is not acceptable in wireless networks because packet loss in such networks cannot be designated as due to congestion; there are also many reasons for packet losses such as interferences, signal fading, hand-off, and other radio effects.

TCP is unable to differentiate between losses due to congestion and due to corruption; it became the main cause of performance degradation of TCP on wireless links. This problem is popularly known as the TCP performance problem over wireless network [2]. In this paper, we study the performance of various techniques to improve the performance of TCP over wireless links. This paper will be useful for analyzing all the existing methods and for implementing innovative congestion control schemes.

The rest of paper is structured as follows. Section 2 describes the performance issues of TCP in wireless networks. Section 3 represents the existing techniques which addresses the performance issues. Section 4 consists of proposed solution. Section 5 represents the conclusion.

2 Obstacles in TCP

TCP is responsible for end-to-end delivery of messages. In wireless network, the performance of TCP is more important, because in such networks possibility of error rate is very high. For reliable data transmission, TCP uses congestion window. This congestion window specifies the number of packets that can be sent by the sender without worrying about acknowledgment. When packet loss occurs, the sender TCP decreases the congestion window directly without investigating the reason for packet loss. There are two reasons for packet loss, first due to congestion and second is due to link failure or noise in wireless links. If packet loss occurs due to link errors, the reducing congestion window leads to the reduction in data transmission rate. This feature affects the throughput of the network. The

congestion window should be reduced only in the presence of congestion. So, we need to design an intelligent congestion control scheme which will identify the reason for packet loss and then decide whether to decrease the congestion window or not.

2.1 TCP Issues in Wireless Networks

For reliable and effective data transmission, the performance of the TCP needs to be improved. The conventional TCP which was developed for wired networks is to be modified to support existing wireless network features with minimum overhead and able to address all the below listed issues.

- How to detect congestion and reduce data rate?
- How to detect multiple packet loss within the same window?
- Time required to detect each packet loss should be minimized.
- How to determine the condition of wireless link?
- TCP should only retransmit the lost packet caused by bit error and do not reduce the size of congestion window.
- All the transmission error caused by bit error should be transparent to the sender and it can be resumed quickly.

3 Existing Approaches

There are many approaches that have been developed to solve this misinterpretation of packet loss in TCP. Whenever packet loss occurs, TCP misinterprets that this packet loss was due to congestion and reduces its congestion window. In this section, we have presented various schemes that solve above-mentioned problems. Each existing approach is explained briefly along with its strengths and weaknesses by the following ways.

3.1 Fast Retransmission and Fast Recovery

In order to detect congestion, TCP Tahoe [3] makes use of slow start, congestion avoidance, and fast retransmission. This algorithm gives better result initially. As soon as the packet is lost, it reduces the data rate. This algorithm reduces congestion window without identifying the reason for packet loss. It has a drawback of drastic reduction in congestion window.

This limitation of TCP Tahoe is overcome by TCP Reno [4]. It based on TCP Tahoe, but uses the concept of duplicate acknowledgments to trigger fast

retransmission. When TCP Reno receives 3 duplicate acknowledgments, it enters fast recovery. In fast recovery, congestion window is set to half the value of current window. But, this technique cannot detect multiple packet loss within the same window.

The limitation of TCP Reno is overcome by TCP New-Reno [5, 6]. In this protocol, fast recovery scheme is modified to deal with multiple packet losses in a single window of data. It is inefficient in terms of bandwidth. TCP new-Reno suffers from the fact that it takes one RTT to detect each packet loss. All these three techniques will address three issues of Sect. 2, but misinterpretation of packet loss still exists.

3.2 Determining the Cause of Packet Loss

The cause of packet loss can be determined by considering the network parameters. Xiao et al. [7] proposed a mechanism in which change trend of RTT is used to determine the cause of packet loss and the status of the network. When the RTT between consecutive packets increases continuously, then it signals the occurrence of congestion. In this approach, certain parameters are used to detect the condition to link that are as follows: t_i denotes the time of the packet received, Receive Packet Time Interval (RPTI) denotes the interval time of two consecutive received packets of the receiver, and λ denotes the difference between two consecutive RPTI. That is,

$$\text{RPTI}_i = t_i - t_{i-1}; \text{RPTI}_{i-1} = t_{i-1} - t_{i-2}; \lambda = \text{RPTI}_i - \text{RPTI}_{i-1}$$

Wireless link condition can be determined according to the λ on the receiver. This technique is compatible with IPSec.

3.3 Congestion Control Mechanism for Wireless Networks and Detection of Multiple Packet Loss

An alternative congestion control mechanism which is specially designed for wireless networks is TCP Westwood [8]. It uses TCP Reno operation with exponential growth during slow start and linear growth during congestion avoidance phase. Based on the packet sizes and RTT estimates, it uses a series of equations to estimate the bandwidth usage of the link. When a packet is lost, the window is reset to the bandwidth estimate rather than reducing it to half. It is having a drawback that sometimes it overestimates the available bandwidth which affects the throughput of the network.

To detect multiple packet loss within the same window of data, Selective Acknowledgements scheme (SACK) [6] was developed. SACK scheme supports

recovery of multiple packet loss within the same RTT window. SACK can specify which blocks of data, following the loss have been successfully received. By informing the sender which packets have been received, and which packets must be resent, multiple lost packets can be recovered in one RTT. SACK is proven to be beneficial [5] for error recovery in wireless networks.

To improve the performance, modified version of SACK TCP was proposed by Metha and Vithalani which is called as SACK_OK [9]. They used a flag SACK_OK to differentiate between random loss and congestion loss. It avoids the reduction of congestion window when the random loss is identified. This mechanism uses distance as another parameter to handle the fast retransmission. This technique does not require changes on receiver side and in header also. The process of modified SACK TCP is shown in Fig. 1.

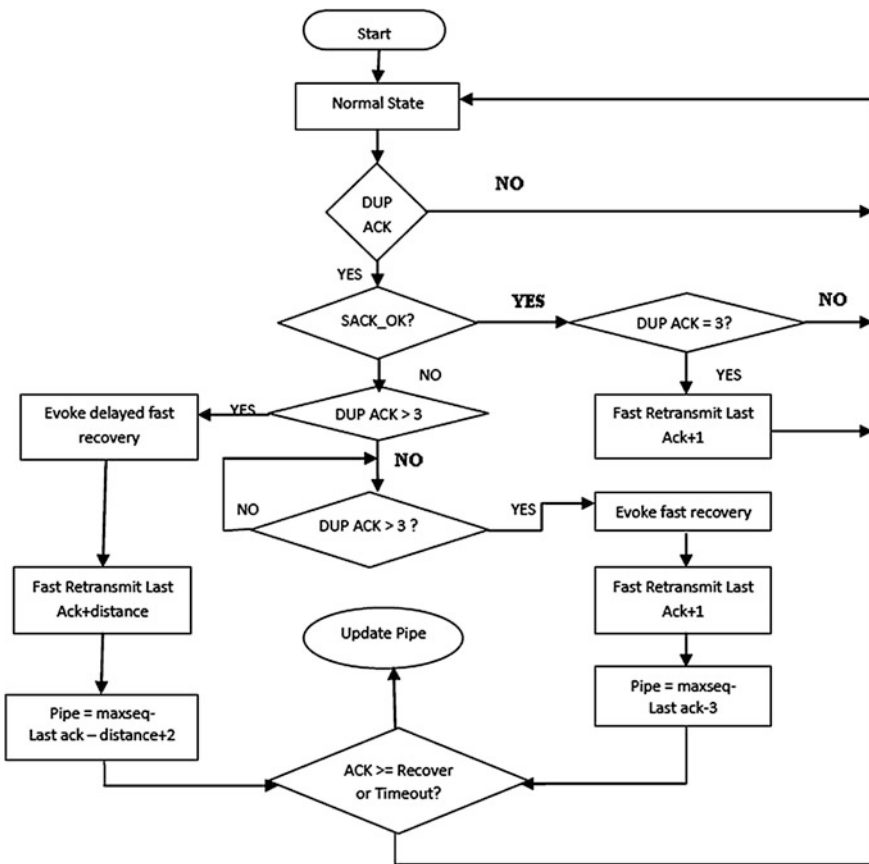


Fig. 1 Flow chart of modified SACK TCP

3.4 Explicit Congestion Notification and Usage of PEP

It is also possible to use certain bits in the header to signal the occurrence of congestion. Explicit congestion notification [10] uses two bits in the IP header and two bits in the TCP header to record the status of the network.

To improve the throughput of TCP, performance enhancing proxy (PEP) can also be used. PEP provides reliability in wireless networks with minimum overhead [11–14]. Split TCP and Snoop are commonly used PEPs. These two approaches suffer from the number of problems which are listed below.

- This scheme breaks TCP end-to-end semantics.
- Path failure between PEP and receiver TCP will lead to unexpected problems.
- Does not support route changes between the sender TCP and the PEP.
- Many security protocols are not compatible with these approaches.
- Snoop proxy does not specify what should be done if the retransmitted packet is also lost.
- Cannot able to hide TCP losses due to interoperability problems.

Above-stated limitations are overcome by innovative proactive distributed TCP proxy called D-proxy [2, 8]. D-proxy is distributed because it uses a proxy either side of the wireless link. It is proactive. The basic concept of D-proxy is relatively simple, but the implementation is complex. D-proxy is able to maintain and recover losses in very high loss situations. It can perform this task using negative acknowledgments. The key benefit is that D-proxy is a negatively acknowledging proxy; sending messages only when a packet is missing. This technique is dependent on D-proxy. Problem occurs when the D-proxy fails.

3.5 Using Reserved Bits of TCP Header and SNR Ratio

Another congestion control scheme was proposed by Bassil [2]. It uses one of the reserved bits of the TCP header and SNR ratio to detect the reliability of the link and decide whether to reduce the packet burst or to retransmit the lost packet. The reserved bit (RB) specifies the type of the link over which the TCP connection is established. For wired link, the RB = 0 and for wireless, RB = 1. In wired mode, any timeout is considered a congestion loss. If packet loss occurs in wireless link, then SNR ratio of the link is measured. Based on SNR value, the decision is made whether to reduce the congestion window or to retransmit the packet. This process is shown in Fig. 2. This scheme maintains true end-to-end semantics, without involving any proxies between the sender and the receiver. This feature eliminates extra processing overhead and the need for extra buffer space. It does not require modifying the source and receiver code.

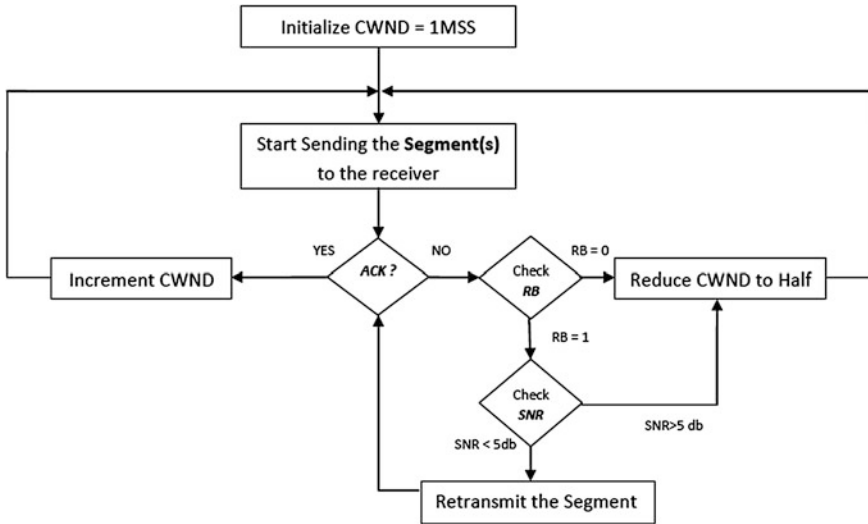


Fig. 2 Flow chart for detecting the cause for packet loss

3.6 WiTracer

Recently proposed approach to improve the performance of TCP is WiTracer [15]. It addresses various issues of TCP Reno and gives solutions for them. TCP Reno suffers from various problems such as Hungry RTO, RTX-DATA Loss, and spurious DATA. WiTracer solves all these problems by using congestion identifier and opportunistic-recovery scheme. Congestion identifier checks for each duplicate acknowledgment packet and implement local recovery rate, while opportunistic recovery provides a faster recovery rate of DupAck and overcomes several TCP drawbacks. WiTracer can be installed on mobile devices for various applications.

4 Proposed Solution

The proposed scheme uses certain network parameters to determine the status of the network and the reason for the packet loss. The performance of TCP can be improved by considering the parameters such as RTT, throughput, and by setting the timers carefully. By setting the timers carefully, we can reduce the delay in wireless networks.

The best technique to solve TCP congestion control problem is to use RTT and throughput. In proposed scheme, for every packet, TCP records the RTT and it also estimates the throughput. By using these two parameters, the TCP can determine the status of the network and can able to detect the real cause of packet loss.

The RTT between consecutive packets is compared. Increasing RTT and decreasing throughput signals the occurrence of congestion. Whenever the packet is lost, the TCP estimates the throughput and also compares the latest RTT with the previous RTT's in a record and also with the threshold value let it be RTT_{Thresh} . If the latest RTT reaches the RTT_{Thresh} , then the TCP considers the packet loss as due to congestion and hence reduces the congestion window based on the capacity of the network. If RTT between packets is uniform and less than the RTT_{Thresh} , then the packet loss is interpreted as due to link error. In this case, TCP only retransmits the lost packet without reducing the congestion window. But, when the packet is retransmitted, the RTO [16] value is doubled. For efficiency, the RTO value should be doubled only when the packet loss is due to congestion, otherwise it will be same. The following algorithm can be used to set the RTO value whenever the packet needs to retransmitted.

4.1 Algorithm

When timeout occurs, throughput and RTT values are considered to know the status of congestion. If occurrence of congestion is confirmed, then the RTO value is doubled and the lost packet is retransmitted, otherwise RTO value is kept same. This process is shown by the following steps.

- Step 1: When timeout occurs
- Step 2: if(Throughput * RTT > congestion_window)
- Step 3: then no congestion
- Step 4: Retransmit the lost packet with the same RTO value.
- Step 5: else RTO = 2 * RTO
- Step 6: End-if

By considering above-stated parameters, the TCP congestion control problem can be solved. Alternatively, we can also use cross-layer approach in which the below layer will provide the information to the above layer. Generally, the network layer monitors the operation of subnet, and it also controls the congestion; it would be better to provide the status of the network by the network layer to the transport layer so that the transport layer can know the real cause of packet loss.

5 Conclusion

In this paper, we have explained various causes that results in TCP performance degradation and proposed some solutions to this problem. This paper presents various techniques that can be used for improving the performance of TCP both in wired as well as in wireless networks. This paper would be helpful for the researchers to better analyze the TCP performance problems and to invent better solutions for it.

References

1. J. Postel, Transmission control protocol specification. RFC 793 (1981)
2. Y. Bassil, TCP congestion control scheme for wireless networks based on tcp reserved field and snr ratio. *IJRRIS* **2**(2) (2012)
3. M. Tayade, S. Sharma, Performance improvement of TCP in MANET. *IJEST* **3**(3) (2011)
4. B. Moraru, F. Copaciu, G. Lazar, V. Dobrota, Practical analysis of TCP implementations: Tahoe, Reno, New Reno
5. S. Floyd, T. Henderson, A. Gurtov, The new reno modification to TCP's fast recovery algorithm. RFC 3782 (2004)
6. M. Mathis, J. Mahadevi, S. Floyd, A. Romanw, TCP selective acknowledgement options (1996)
7. L. Xiao, Z. Li, N. Zhao, A TCP performance improved mechanism for wireless network. UIC-ATC. In *Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*. IEEE, pp. 212–216 (2009)
8. D. Murray, T. Koziniec, M. Dixon, D-Proxy: reliability in wireless networks. In *2010 16th Asia Pacific Conference on Communications (APCC)*, pp. 129–134
9. R.D. Metha, C.H. Vithalani (2012) Distinguishing congestion loss from random loss on wireless erroneous links to improve performance of wireless TCP-SACK. In *CSNT '12 Proceedings of the 2012 International Conference on Communication Systems and Network Technologies*. IEEE, pp. 383–387 (2012)
10. K. Ramakrishnan, S. Floyd, A proposal to add explicit congestion notification to IP. Internet Draft (1999)
11. J. Border, M. Kojo, J. Griner, G. Montenegro, Z. Shelby, Performance enhancing proxies intended to mitigate link-related degradations. IETF RFC 3135 (2001)
12. A. Bakre, B.R. Badrinath, I-TCP: indirect TCP for mobile hosts. In *15th International Conference on Distributed Computing Systems. ICDCS'95 Proceedings of the 15th International Conference on Distributed Computing Systems*. IEEE, pp. 136–143 (1995)
13. H. Balakrishnan, S. Seshan, E. Amir, R.H. Katz, Improving TCP/IP performance over wireless networks. In *MobiCom'95 Proceedings of the 1st annual international conference on Mobile computing and networking*, ACM, pp. 2–11 (1995)
14. B. Francis, V. Narsimhan, A. Nayak, I. Stojmenovic, Techniques for enhancing TCP performance in wireless networks. In *2012 32nd International Conference on Distributed Computing Systems Workshops*, pp. 222–230 (2012)
15. C. Hu, X. Yang, M. Fan, P. Zhao, WiTracer: a novel solution to improve TCP Performance over Wireless Network. In *9th IEEE International Wireless Communications and Mobile Computing Conference*. IEEE, pp. 450–455 (2013)
16. H. Touati, I. Lengliz, F. Kamoun, TCP adaptive RTO to improve TCP performance in mobile ad-hoc networks, pp. 176–180 (2007)

Analysis of Classification Models Using Image Statistics and Data Miner for Grade Prediction of Astrocytoma

M. Monica Subashini, Sarat Kumar Sahoo,
S. Prabhakar Karthikeyan and I. Jacob Raglend

Abstract Astrocytoma is the most common primary tumor which develops from glial cells of brain. They are generally classified as low grade (Grade I and Grade II) and high grade (Grade III and Grade IV), and these classifications are very important in clinical practice which signifies the rate of growth. Grading of astrocytoma relies on magnetic resonant images, and pathological information is also available in clinical settings. In this proposed method, we introduce a novel approach to grade the tumor using first- and second-order image statistical parameters combined with a tool termed as 'XLMiner.' The actual grade of astrocytoma and the predicted grade by the classifiers are compared and the accuracy of the classifiers is summarized based on the classifier-predicted output. Experimental results demonstrate the effectiveness of the method. The accuracy of Naives Bayes, discriminant analysis, regression tree, and classification tree classifiers for the prediction of grades from lower (I, II) to higher (III, IV) are 100, 81, 76, and 78 % for all the views, respectively.

Keywords Brain MR images · GLCM · Discriminate analysis · Classification tree · Regression tree · Naives Bayes classifier

M. Monica Subashini (✉) · S.K. Sahoo · S. Prabhakar Karthikeyan
School of Electrical Engineering, VIT University, Vellore, Tamil Nadu, India
e-mail: monicasubashini.m@vit.ac.in

S.K. Sahoo
e-mail: sksahoo@vit.ac.in

S. Prabhakar Karthikeyan
e-mail: spk25in@yahoo.co.in

I. Jacob Raglend
NI University, Kumaracoil, Thuckalay, Kanyakumari, Tamil Nadu, India
e-mail: jacobraglend@rediffmail.com

1 Introduction

1.1 Medical Issue

There are almost 120 different types of glioma. A glioma (tumor) is named based on the cell type, origin, and their location. The World Health Organisation (WHO) classifies the tumor by origin and cell behavior from benign to malignant [1]. Primary brain tumors originate in the brain, whereas secondary tumor begins as a cancer in any part of the body and spreads to brain by blood or adjacent tissues. The rate of growth of a tumor is determined by the grade and the course of treatment begins only after diagnosing the grade of a tumor. Astrocytoma is a primary brain tumor derived from astrocytes, star-shaped glial cells. Astrocytoma tumor types by grade as follows:

- Grade I: Pilocytic astrocytoma is a most common type of glioma found in children.
- Grade II: Low-grade astrocytoma typically occurs in men and women of ages 20–60.
- Grade III: Anaplastic astrocytoma typically occurs in adults ages 30–60 and is more common among men than women.
- Grade IV: Glioblastoma multiforme (GBM) accounts for almost 50 % of all astrocytomas. These are highly malignant aggressive tumors and common in older adults (50s–70s), particularly men.

1.2 Classification and Data Mining

Artificial intelligence plays an important role in classification. Classification deals identification of a category to which a new observation belongs from a group of data. A program or an algorithm that implements classification is a classifier. Classification involves supervised and unsupervised in terms of machine learning. In supervised learning, the output is classified from a training set where correct observations are available. Unsupervised learning fix the output based on few similarities among the input features. They actually group data into categories. This refers to ‘clustering.’ The input patterns are recognized and assigned a ‘class’ or a ‘cluster.’ The available classifiers are as follows: Linear classifiers, support vector machines, quadratic classifiers, Kernel estimation, boosting, decision trees, neural networks, gene expression programming, Bayesian networks, hidden Markov models, and learning vector quantization.

The proposed method is a data mining procedure which implemented few classifiers mentioned. They are decision trees, discriminant analysis, regression trees, and naïve Bayes. Data mining predicts or discovers the grade of the diseased MR image from large amounts of input data. The input data sources are the

statistical parameters extracted from the MR images. We extract six parameters from the MR images (normal/abnormal). The classifiers are suitably selected to recognize the features and predict the grade. XLMiner from analytic solver platform is a powerful and simulation tool for Excel. The extracted features in Excel sheets are the input data to the tool. Clustering and classification are performed by the tool on the data (features) on the instructions given on classification models. The output of the model is prediction of class/grade. In order to suggest the best classifier for the grade prediction, four models in data miner were selected and the experiment was performed using the image statistics. The XLMiner provided automatic partitioning of dataset into training, validation, and test samples. The developed classification model is suitably trained for any new data. Therefore, a fast and direct output would improve the medical decision.

1.3 Related Research Work

Many schemes are available in literature-related tumor identification and grade prediction including data mining and machine learning techniques. Specifically mentioning, Refs. [2, 3] includes works on predictive models for data mining and on tree structure for efficient data mining.

A tumor in brain is recognized using least squared support vector machines based on feature selection in Refs. [4, 5] deals with generation of prior probabilities for classifiers. Reference [6] developed a supervised pattern recognition method for the prediction of contrast-enhancement in brain tumors. Reference paper [7] classified the tumor type and grade using machine learning techniques. In [8, 9], the authors came up with a semi-supervised graph-based tumor classification method. A pattern recognition system for brain tumor grade prediction was built based on histopathological material and were successful in grade prediction [10]. Reference [11] refers the work identified for malignant transformations in low-grade gliomas and tumor-specific analysis. All of these references motivated to analyze the classifiers utilized in data miners for the specific grade prediction of astrocytoma.

2 Proposed Method

2.1 Image Acquisition

The proposed method is highlighted in Fig. 1. The brain MR images for the process are stored as a dataset. It consists of seventy-five normal and abnormal (astrocytoma) images. T1-weighted sequence images in three orthogonal views of normal and astrocytoma (Grade I, Grade II, Grade III, and Grade IV) are loaded for analysis. T1-weighted scans provide better contrast between white and gray matter.

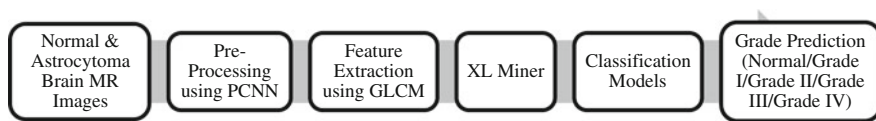


Fig. 1 Proposed methodology outline

In 75 images, 15 were normal brain images comprising of five axial view images, five Saggital view images, and five coronal view images. Similarly, we collected 15 images of Grade I (5 (Axial); 5 (Saggital); 5 (Coronal) from 5 patients. Hence, a total of 60 diseased images from twenty patients in all three orthogonal views are saved in the dataset.

2.2 Preprocessing

The MR images are preprocessed to remove noise due to motion artifacts and magnetic field. Median filter is applied, since it removes noise as well as preserve edges. Contrast-enhancement results in better differentiation between tissues. The images are subjected to pulse-coupled neural network for enhancement [12].

2.3 First- and Second-Order Image Statistics for Image Information

The overall information about the image is obtained by first- and second-order statistics. The statistical features are extracted using Gray Level Co Occurrence Matrix. The extracted parameters for normal, grade I, grade II, grade III, and grade IV vary from each other in their values (variance, skewness, entropy, energy, contrast, and homogeneity). Seventy-five images of three views have been subjected to feature extraction.

2.4 Data Mining Approach

Data mining is a process of extraction, exploration, and analysis by semi-automatic or automatic methods [7]. A process flow of data mining principle is shown in Fig. 2. Some common tasks such as classification, prediction, data analysis, data reduction, data exploration are performed in data mining.

Classification The given data is examined and grouped into a class (Known/Unknown).

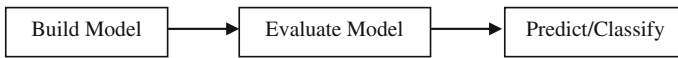


Fig. 2 The three blocks represent the XLMiners’ process flow

Prediction	The value of a numerical variable is predicted.
Association Rules	The data is associated with few set of rules to map the output.
Data Reduction	Complex data is distilled into simpler data, i.e., a large number of variables are consolidated into smaller groups.
Data Exploration	The data must be completely explored to analyze each variable individually. Similar variables are grouped together under a single variable. This is the utmost objective of data exploration.
Data Visualization	Analyzing the variables individually and with respect to neighbors is done visually. Graphical analyses aid this task through bar charts, pie charts, histograms, and box plots.

Data mining was carried out using a tool, ‘XLMiner’ to distinguish the normal/abnormal (Grade I, Grade II, Grade III, and Grade IV) brain MR images with the help of first- and second-order image statistics defined above. The procedure followed in grade prediction is described in detailed steps.

- Step 1: The purpose of this data mining project is to predict the grade of astrocytoma based on feature variables. The dataset is shown in Table 1. The features presented are extracted from axial view brain images (normal/abnormal). The ‘type’ corresponds 0—Normal; 1—Grade I; 2—Grade II; 3—Grade III, and 4—Grade IV.
- Step 2: Data Obtaining
The dataset is small, and hence, it is used completely.
- Step 3: Explore, Clean, and Preprocess the data
The variables in the dataset should be analyzed thoroughly. The ‘variance’ in the dataset of the MR images, predominantly shows the difference among them for each grade. So, variance should be included in the model. The purpose of the inquiry is to verify the variables and useful variables has to be applied to the model. Variance is a useful predictor. Similarly, the remaining 5 variables are analyzed in the same manner. Six independent predictor variables are allowed to be used in the model.
- Step 4: Partitioning
The data is partitioned into training, validation, and test partitions. Training set would build the model. Validation set would see how well the model works with the new data. XLMiner has an option ‘standard data partition’ under which the data is partitioned as 60 % training data, 40 % validation data, and 0 % test data. A model is trained and tested with multiple data sets once the successful completion of building a model.

Table 1 Data set containing the extracted features from Saggital plane MR images

Type	Variance	Skewness	Entropy	Energy	Contrast	Area
0	128.00	4.20	7.16	0.18	0.13	447836.88
1	2450.00	2.42	7.05	0.25	0.17	538959.50
2	5000.00	2.22	6.93	0.31	0.16	362004.75
3	3042.00	3.38	7.28	0.19	0.17	190155.38
4	242.00	2.16	7.03	0.25	0.20	305807.50
0	1058.00	5.36	6.95	0.22	0.21	511563.63
1	2178.00	2.06	7.20	0.22	0.19	535979.50
2	882.00	1.70	7.03	0.21	0.18	283816.63
3	288.00	2.22	7.13	0.22	0.21	188410.63
4	392.00	2.37	7.20	0.22	0.22	300795.25
0	2.00	3.64	7.16	0.20	0.14	477847.75
1	800.00	2.21	7.15	0.25	0.16	518288.13
2	800.00	6.25	7.01	0.19	0.20	207742.38
3	648.00	3.64	7.07	0.23	0.23	214567.13
4	12.50	2.98	6.32	0.27	0.20	388898.25
0	162.00	3.52	7.17	0.23	0.11	488765.38
1	242.00	3.00	6.87	0.24	0.35	303519.25
2	882.00	4.98	7.06	0.21	0.20	218182.50
3	200.00	0.94	7.10	0.18	0.23	155704.38
4	40.50	2.98	6.95	0.25	0.24	279627.75
0	882.00	4.92	7.10	0.22	0.24	269524.50
1	1922.00	3.67	7.11	0.22	0.19	404352.38
2	1568.00	2.49	7.11	0.21	0.22	215903.38
3	882.00	2.39	7.30	0.14	0.32	218329.00
4	2244.50	2.88	6.95	0.25	0.20	318645.75

Step 5: Task

The task to be mined is fixed. Here, the task is to predict the grade of astrocytoma using the 6 predictor variables.

Step 6: Choose the Technique

XLMiner is used to build the classification models. Four models were selected from the tool for the grade prediction process.

1. Discriminant Analysis
2. Regression Tree
3. Classification Tree
4. Naive Bayes

Step 7: Use of algorithm

XLMiner uses the algorithm of prediction and classification. The techniques utilized in prediction and classification are regression trees classification trees, Naive Bayes, and discriminant analysis. The fitted values on the training data and predicted values on the validation are obtained after simulation. The software shows the standard output for all the specified models as shown in figure. The prediction error (average error, total sum of squared error, and RMS error) for the training and validation data is compared. The RMS error for the validation data is larger than the training data.

Step 8: Interpret the results

Four prediction algorithms were tried with the dataset for the grade prediction problem. The model with the lowest error on the validation data seems to be the best model. The selected classification model can be used to predict the output variable with fresh input data.

Step 9: Deploy the model

Usage of selected classification model is deployed for the prediction problems.

3 Classifiers

3.1 Discriminant Analysis

Discriminant analysis is a statistical technique which is applied for profiling and classification. Continuous variable measurements are utilized for the purpose of classification. The statistical distance or Mahalanobis distance method is applied for classification.

Mathematically,

$$D_{\text{statistical}}(X, \bar{X}) = [X - \bar{X}]' S^{-1} [X - \bar{X}] \quad (1)$$

S is the covariance matrix between the feature variables to be classified. S^{-1} is the inverse matrix of S . The distance between an observation and a class is computed based on the centroid and covariances between each pair of variables. A separating hyper plane is constructed using classification functions for allocating an observation to the closest class. These classification functions compute scores which measures proximity of an observation to each of the grades. Table 2 gives a summary report comprising the classification confusion matrix and the time elapsed for the classification process. Discriminant analysis is a statistical tool utilized for separating the classes with predictor's optimal weights. These predictors are from multivariate normal distribution. The computational is simple and it is useful for small datasets.

Table 2 Training data scoring—summary report

Classification confusion matrix					
Actual class	Predicted class				
	0	1	2	3	4
0	4	0	0	0	1
1	0	5	0	0	0
2	0	0	3	2	0
3	0	0	1	4	0
4	0	0	1	0	4
Elapsed time					
Overall (s)		2.00			

3.2 Classification Tree

Recursive partitioning of the variables (features) is the key idea behind any classification tree. Recursive partitioning divides the p dimensional space of x variables into non-overlapping multi-dimensional rectangles. The X variables here are considered to be features extracted. The tree splits are recursively performed. One of the variables is selected from validation set and compared with variable in p dimensional space to split into two parts, either greater or less than. This procedure is repeated to check all the variables in the space resulting in small rectangular regions. The entire space is divided into rectangles and they are assigned a class.

- Splits selection into decision nodes.
- A terminal node is designated based on decisions made in decision nodes.
- The terminal node is assigned a class.

The tree developed after completing the grade prediction process is shown in Fig. 3. The terminals predict the grades in numerical format. ‘Zero’ represents ‘normal image,’ ‘1’ represents ‘Pilocytic astrocytoma,’ ‘2’ represents ‘low-grade astrocytoma,’ ‘3’ represents ‘Anaplastic Astrocytoma,’ ‘4’ represents ‘Glioblastoma Multiforme.’ Area and Variance are the two features that showed variation and hence the classification is based on these feature values.

3.3 Regression Tree

Regression trees are similar to classification trees. The end point is a predicted function instead of predicted class as in classification tree. The results obtained for five images from each type based on regression is shown in Table 3.

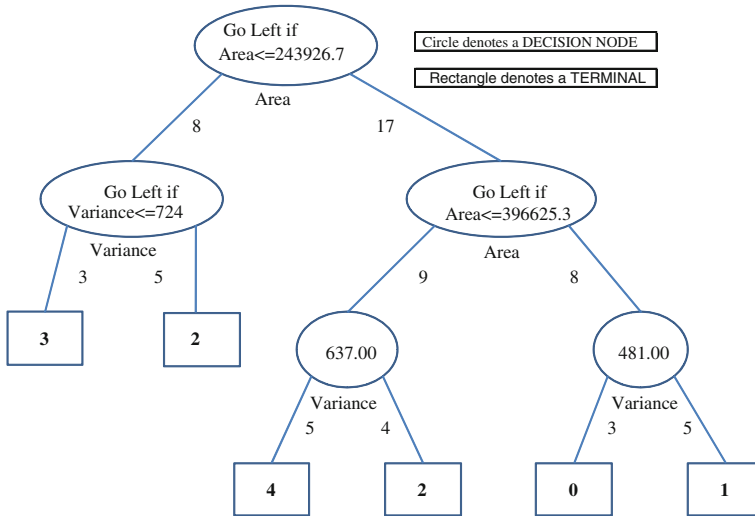


Fig. 3 Classification tree developed by the Miner for the grade prediction

Table 3 Regression tree predicted result

Predicted value	Actual value
0	0
1	1
2.25	2
3	3
2.25	4

3.4 Naives Bayes

Naive Bayes classifier works only with predictors that are categorical. This classifier is very much useful when the dataset or database is large. The information given in the set of predictors is integrated into the Naive rule to classify accurately. A fundamental theorem based on conditional probabilities called Bayes theorem that plays the major role in classifications. In the theorem, if given training data is D , hypothesis h , the posteriori probability $P(h|D)$ could be stated by the following Eq. 2

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \tag{2}$$

Table 4 Naives Bayes predicted results

Predicted class	Actual class
0	0
1	1
2	2
3	3
4	4

The actual and predicted class based on the feature inputs and posteriori probability is tabulated. Twenty-five images chosen as training data for training the classifier and sample output is shown for five images in Table 4. The actual class is known from medical experts and the predicted class by the classifiers is benchmarked against the actual class to determine the accuracy of the classifiers.

4 Discussion and Results

The classifiers developed by the XLMiner predicted the grade in terms of class. Class 0, Class 1, Class 2, Class 3, and Class 4 represents normal, Grade I, Grade II, Grade III, and Grade IV, respectively. The actual class and the predicted class of the images contribute to the accuracy findings of the models. The classification models were chosen studying their advantages and disadvantages for this specific grade prediction problem. Classification and regression tree models could cope up with any data structures. The models are so simple to construct with the effective usage of conditional information. Naïve Bayes is selected since the classifier is incremental and supports probabilistic learning. The performance of each classifier for this chosen astrocytoma grade prediction problem is analyzed. The deviation from the actual class to the predicted class is necessary to define the accuracy. The predicted class of discriminant analysis and Naïve Bayes almost reaches the actual class. But, the classification and regression tree methods show much deviation. Cent percent accuracy is achieved when the predicted class equals the actual class. Table 5 highlights the accuracy of classifiers with respect to all the three views of MR brain images.

Table 5 A comparison on classifier results

Brain MR images (normal/grade I/grade II/grade III/grade IV)	Classifier	Accuracy (%)
Axial view	Discriminant analysis	92
	Regression tree	80
	Classification tree	78
	Naives Bayes	100
Coronal view	Discriminant analysis	72
	Regression tree	76
	Classification tree	80
	Naives Bayes	100
Saggital view	Discriminant analysis	80
	Regression tree	72
	Classification tree	76
	Naives Bayes	100

5 Conclusion

The grade prediction of astrocytoma (tumor) was the problem chosen and a suitable classifier has to be suggested for the problem identified. The scheme comprises of several steps including preprocessing (enhancement), feature extraction, and classification model selection and classification. The proposed method was applied on a population of 75 brain images which has normal and tumor images diagnosed as Pilocytic astrocytoma, low-grade astrocytoma, Anaplastic astrocytoma, and GBM as per World Health Organization grading. The MR images are subjected to GLCM, a first-order statistical feature extracting technique. The extracted features such as variance, skewness, entropy, energy, contrast, and homogeneity are tabulated for every view namely, Axial, Saggital, and Coronal. ‘XLMiner’ is a tool in data mining which supports classification. The classification models are discriminant analysis, regression tree, classification tree, and Naives Bayes. The results predict the grade of the tumor and based on the actual and predicted class, a comparison on classifiers is thus performed. Since the features overlap in few cases, the accuracy of the constructed classifiers is less. In all the three views, Naives Bayes classifier showed satisfying results. To improve the accuracy, the proposed work is to be extended with histopathological brain astrocytoma images.

Acknowledgments The brain MR images were collected from Medpix online and Krishna Scan Centre, Vellore. This work has been supported by School of Electrical Engineering, VIT University.

References

1. D.N. Louis, H. Ohgaki, O.D. Wiestler, W.K. Cavenee, P.C. Burger, A. Jouvet, B.W. Scheithauer, P. Kleihues, The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97–109 (2007)
2. V.S. Ananthanarayana, M. Narasimha Murty, D.K. Subramanian, Tree structure for efficient data mining using rough sets. *Pattern Recogn. Lett.* **24**, 851–862 (2003)
3. Se June Hong, Sholom M. Weiss, Advances in predictive models for data mining. *Pattern Recogn. Lett.* **22**, 55–61 (2001)
4. J. Luts, A. Heerschap, J. Suykens, S. Van Huffel, A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. *Artif. Intell. Med.* **40**, 87–102 (2007)
5. M. Greg Reynolds, C. Andrew Peet, N. Theodoros Arvanitis, Generating prior probabilities for classifiers of brain tumours using belief networks. *BMC Med. Inf. Decis. Making* **7** (2007)
6. M.C. Lee, S.J. Nelson, Supervised pattern recognition for the prediction of contrast-enhancement appearance in brain tumors from multivariate magnetic resonance imaging and spectroscopy. *Artif. Intell. Med.* **43**, 61–74 (2008)
7. E.I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E.R. Melhem, C. Davatzikos, Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* **62**, 1609–1618 (2009)
8. J. Gui, S.L. Wang, Y.K. Lei, Multi-step dimensionality reduction and semi-supervised graph-based tumor classification using gene expression data. *Artif. Intell. Med.* **50**, 181–191 (2010)
9. S. Ortega-Martorell, H. Ruiz, A. Vellido, I. Olier, E. Romero, A novel semi-supervised methodology for extracting tumor type-specific mrs sources in human brain data. *Plos One* **8**, 12(2013)
10. C. Konstantinou, E. Maneas, G. Dimitris, K. Spiros, R. Panagiota, D. Cavouras, A pattern recognition system for brain tumour grade prediction based on histopathological material and features extracted at different optical magnifications. In *Workshop on Bio-Medical Instrumentation and related Engineering and Physical Sciences, e-Journal of Science and Technology (e-JST)* (2012)
11. A. Constantin, A. Elkhaled, L. Jalbert, R. Srinivasan, S. Cha, S.M. Chang, R. Bajcsy, S.J. Nelson, Identifying malignant transformations in recurrent low grade gliomas using high resolution magic angle spinning spectroscopy. *Artif. Intell. Med.* **55**, 61–70 (2012)
12. M. Monica Subashini, S.K. Sahoo, Pulse coupled neural networks and its applications. *Expert Syst. Appl.* **41**, 3965–3974 (2014)

Object Detection in Cluttered Environment Using 3D Map

Deepesh Jain, Renuka Ramachandran, Anuhya Vunnam
and P. Vignesh

Abstract Autonomous mobile robot must act intelligently without external control by definition and require fundamental capabilities such as the awareness of its environment and of its location within the environment. These two problems are known, respectively, as mapping and localization. The ability to detect and identify mobile and fixed obstacles also plays an important role for achieving robots autonomy. The project is concerned with the problem of designing and implementing a robot system to recognize objects in cluttered environment using a 3D map generated by the system using efficient algorithms. For building dense 3D maps of the environment and to recognize objects, use RGB-D camera which accurately identifies objects as they take into consideration the shape and three-dimensional characteristics of the object.

Keywords Kinect · SURF · RANSAC

1 Introduction

Robots are present in our daily life, not only in industry but also at home as service robots such as vacuum cleaners, luggage transfer, disabled people assistance, or intelligent device switching. They can be either controlled by the human or

D. Jain (✉) · R. Ramachandran · A. Vunnam · P. Vignesh
Department of Computer Science and Engineering, Amrita School of Engineering,
Amritanagar (P.O.), Ettimadai, Coimbatore 641112, Tamilnadu, India
e-mail: jain_deepesh94@yahoo.com

R. Ramachandran
e-mail: renukaram06@gmail.com

A. Vunnam
e-mail: anuhya.vunnam@gmail.com

P. Vignesh
e-mail: vigneshprakash92@gmail.com

autonomously perform a predefined task. In recent years, robotics research has focused on the problem of planning and executing various tasks autonomously, i.e., without human guidance. Building 3D maps of environments is one such important task for mobile robotics, with applications in navigation, manipulation, semantic mapping, etc. Such a facility is essential in the emerging field of service robotics.

One of the main shortcomings in conventional mobile robotics is perception that mobile robots can travel across much of earth's man-made surface, but they cannot perceive the world nearly as well as human beings. The success of mobile robotics platforms depends on their ability to perform effective and efficient autonomous perception and interaction with objects in various environments. We can overcome this by improving the efficiency with which robots can track and recognize objects in their environment.

In this paper, we propose a novel idea of building a 3D map of a cluttered indoor environment and use this map to detect objects present in the environment.

The rest of this paper is organized as follows: Sect. 2 describes the available system architectures. Section 3 derives the components of object tracking system and implementation of the above system. Section 4 describes the conclusion and future enhancements.

2 Related Works

One methodology is to use dual sensors, i.e., thermal sensor and optical sensor, for detection. Here, the concept of sensor handover can be implemented. The concept of sensor handover is used specifically to address the issues of extreme changes in illumination over a long timescale where the advantages of thermal sensing are under certain twilight/night illumination conditions, while optical sensing remains for brighter illumination periods [1]. This approach is not applicable to differentiate between two objects of different shapes but with same temperature.

Object recognition and tracking can be implemented by utilizing the depth information from a low-cost depth sensor such as Kinect. Conventional object recognition methods that utilize RGB cameras are unable to accurately identify objects in the real world since they do not take into consideration the shape and three-dimensional characteristics of the object. Another major factor determining the accuracy of recognition is the lighting conditions and object pose at the time of recognition. We discuss an approach making use of the depth information and 3D properties of objects in order to accurately identify them independent of lighting conditions [2]. In this approach, only object recognition is achieved, and no object is tracked in the environment.

Another method to detect object is using color and depth segmentation. Usually, object segmentation from an image is achieved using color segmentation. This segmentation can be achieved by processing the R, G, and B chromatic components. However, this method has the disadvantage of being very sensitive to the changes on lighting. Converting the RGB image to the CIE-Lab color space avoids

the lack of sensitivity by increasing the accuracy of the color segmentation. Unfortunately, if multiple objects of the same color are presented in the scene, it is not possible to identify one of these objects using only this color space. Therefore, we need to consider an additional data source, in this case, the depth, in order to discriminate objects that are not in the same plane as the object of interest [3].

Another method for detection of object is based on integrated global template and local feature-based recognition. Here, for local feature path matching, they are using local feature models such as SIFT, speeded up robust features (SURF), and bag-of-words matching schemes. For the global path, they rely on a contemporary histogram-of-gradients descriptor that includes latent part components; this method extends the histogram-of-gradients (HOG) object template model with a deformable high-resolution part structure [4].

3 Proposed Model

In our model, we will track an object in 3D map of environment. We have divided object tracking into two modules:

1. Building a 3D map of cluttered environment
2. Detection of object in 3D Map

Above two modules are not interdependent. So we can use 3D map for other applications.

3.1 *Building a 3D Map of Cluttered Environment* [5]

To build a 3D map of environment, we have to follow the following steps:

- *Getting a 3D color image of environment:* To get 3D image of environment, we are using Microsoft Kinect Xbox which has inbuilt 3D sensors and cameras to capture the image. To view the image, we are using Microsoft SDK 1.7 which has already 3D viewing code available.
- *Feature extraction from 3D image:* To extract the feature from 3D image, we are using SURF which is based on finding key frames which are stable and has periodic properties.
- *Implementation of ICP and RANSAC Algorithm:* To match the key frames features and to find the best alignment between them, we are using iterative closest point (ICP) and random sample consequences (RANSAC). Use of ICP algorithm is to check the features periodically, putting a threshold to check feature alignment. It is also used to put a maximum threshold over number of iterations or alignment angle. RANSAC is used to find out alignment between two consecutives frames.

- *Loop closure detection*: A loop closure detection algorithm is run in parallel to check whether attachment of a frame will cause a loop closure or not because when we are attaching a frame in map, there is a possibility for a loop closure, so we will match every n th frame for loop closure; here, n value depends upon environment.
- *Pose graph optimization*: It is a global optimization algorithm used to optimize graph. Here, complete graph is represented in form of nodes and if there is any constraint between them, then there will be an edge between both nodes, so without any loop closure, graph will be a straight chain, and TORO can be used to optimize it.

3.2 Object Detection in 3D Map [6]

To detect an object in map, we need to use following steps:

- *Selection of appropriate dataset*: Selection of dataset depends upon the object which we are interested to detect in 3D map.
- *Smoothing the image from database*: To detect an object in 3D map, we have to smoothen the image because we are only interested in shape of object.
- *Object detection*: To detect the object, we are using sliding window protocol to match each and every segment of image by sliding an even-sized window.
- *Object evaluation*: To verify the presence of the object, we can use Pascal formula where detection is considered correct if $\frac{\text{area}(B \cap G)}{\text{area}(B \cup G)} > 0.5$ where B is the bounding box of the detection and G is the ground truth bounding box of the same class. For a given ground truth bounding box, only a single detection is considered to be correct, with the rest considered as false positives.

3.3 System Architecture

See Figs. 1 and 2.

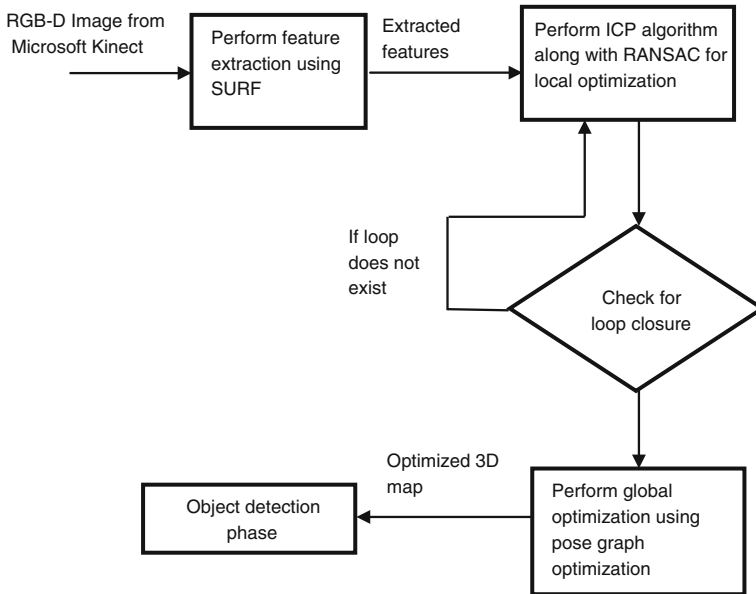


Fig. 1 Building 3D map

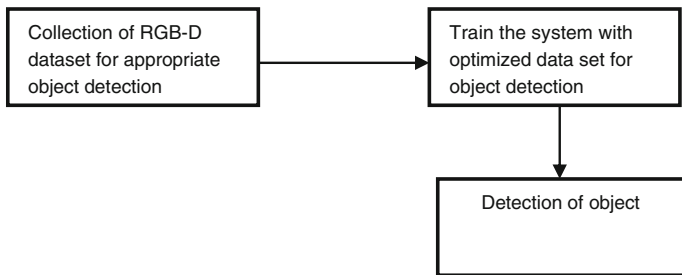


Fig. 2 Object detection phase

4 Conclusions

The above approach can be deployed to detect objects in a cluttered indoor environment effectively. Here, 3D map building is independent of the object detection phase and thus can be used for semiautonomous robotic applications.

References

1. M. Magnabosco et al., Cross-spectral visual simultaneous localization and mapping (SLAM) with sensor handover. *Robot. Auton. Syst.* **61**(2), 195–208 (2012)
2. P. Bongale et al., Implementation of 3D object recognition and tracking, in *International Conference on Recent Advances in Computer and Software Systems* (2012), pp. 77–79
3. J.J. Hernández-López et al., Detecting objects using color and depth segmentation, in *IberoAmerican conference on electronics engineering and computer science* (Elsevier, Amsterdam, 2012), pp. 196–204
4. K. Sanko et al, Practical 3D object detection using category and instance-level appearance model, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2011), pp. 793–800
5. P. Henry et al., RGB-D Mapping, using Kinect-style depth cameras for dense 3-D modeling of an indoor environment. *Int. J. Rob. Res.*, 1–17 (2012)
6. A. Janoch et al., A category-level 3D object dataset: putting the kinect to work, in *IEEE International Conference on Computer Vision Workshops* (2011), pp. 1168–1174

Smart Energy Meter with Instant Billing and Payment

Dhananjayan Ravi, J. Shibu and E. Shanthi

Abstract This paper mainly focuses on the measurement of energy consumption and providing data for billing and a system for payment at your place. In this paper, we present a simple design for automatic energy meter reading with payment facility with the help of ZigBee communication technology. By this technology, we can communicate at faster rate without any data loss and it provides high security in serial communication. In this system, the energy is measured in units and the data are fed to a remote computer server where a software solution is provided to generate bill for energy consumption and the data are send back by using same communication method, the consumer can pay the bill at home by using a keypad system.

Keywords Energy meter · Wireless data · Data management · Billing and payment

1 Introduction

World without electricity is unimaginable. Countries development depends on per-capita consumptions. In India, there are many sectors which have attained a rapid development but only few developments are made in electricity board sector. Traditional electro-mechanical meters, still widely used today, are prone to drift over temperature and time as a result of the analog and mechanical nature of the components in these meters. Collection of meter readings is also inefficient, because a meter reader has to physically be on-site to take the readings. This method of

D. Ravi (✉) · J. Shibu · E. Shanthi
Department of Electronics and Communication Engineering,
Bannari Amman Institute of Technology, Sathyamangalam, India
e-mail: dhananjayanravi@gmail.com

J. Shibu
e-mail: shibu.097@gmail.com

collecting of meter readings becomes more problematic and costly when readings have to be collected from vast, and often scattered rural areas. Meter readers are reluctant to make the effort to travel to such areas and will often submit inaccurate estimations of the amount of electricity consumed. For households at the top of high buildings and luxury housing plots, traditional meter reading is highly inefficient. There exists chance for missing bills, absence of consumer, etc. [1]. Digital meters were placed in certain places, which indicates voltage, current, power, and time and date in liquid-crystal display (LCD). Automatic meter reading (AMR) is popular because of its remote nature of data collection. There are different technologies being used to capture and transfer data remotely, but the accuracy, speed, efficiency, reliability and cost-effectiveness are the usual benefits properly achieved in this system, but an additional facility for payment was also introduced. AMR system consists of measuring sensors, microcontroller, and wireless communication network. The meter reading and management of data are free from human error [3]. After the measurement of readings, the data are fed to remote location server which consists of software solution which generates bill and it will send back to the same protocol so the consumer can collect his bill in the meter display. By using a recharge card, consumer can pay the bill with the help of keypad in the system.

1.1 Existing Technologies

Energy consumption is measured using various technologies. Bill for usage is generated and provided to the customer using certain methods. Payment is collected in electricity board from the customer. Recently, research into the field of AMR system has continued to receive much attention in academia.

- (i) Traditional electro-mechanical meters were used to measure the energy consumption. It is an analog meter where readings are noted in the card by a person and the reading was taken to electricity board station where the bill is generated for the consumption; consumer has to pay the bill for the usage in electricity board station. Human error is main disadvantage of this method.
- (ii) Electronic meters were introduced which has replaced the analog into digital system, but the procedure was same as electro-mechanical meter. Here, the consumer can note down voltage, current, power, time and date; this is the only advantage over electro-mechanical meter.
- (iii) Developed a Bluetooth based system, were a method was introduced to retrieve data by means of wireless communication known as AMR. AMR is a mechanism whereby the energy meter sends the recorded power consumption of a household in the certain interval of time to a “wirelessly” connected reader, which could be a personal computer (PC) [3]. The reading was noted in a database, and bill will be generated.

- (iv) Introduced a GSM based system, where Bluetooth technology was replaced by GSM technology, here the readings were send to remote location by using SMS and bill is generated and send to consumer mobile using GSM modem.
Disadvantage of this method is the cost for SMS and jam of network.
- (v) Design of an electric energy meter for long-distance data information transfers which is based upon GPRS is proposed, where the data are transferred by means of GPRS.

1.2 Proposed System

The system consists of AMR facility with the help of voltage and current sensor, and the value is fed to controller to calculate the usage of power. Then, the data are fed to remote station server with the help of ZigBee. A software is created using .net in the server which will generate the bill for our usage according to the tariff. A database is maintained in the server which consists of customer details and his consumption. Amount for usage is again fed to energy meter display in home using same ZigBee network. Thus, we can avoid human errors in measuring the readings.

- AMR for measuring consumption
- Microcontroller to calculate energy consumption
- ZigBee to transfer data
- Display for readings and information
- Keypad for payment

EB card was introduced in this system which will resemble as top-up cards for mobile phones. According to usage of power, we can purchase the card. Then, we can enter the secret number in the card by using a keypad and send to server by using ZigBee network. Amount for usage will detected by the server, and it will credit the amount if there is balance. Thus, we be able pay the bill in our place.

2 System Architecture

This section describes the conceptual design of a low-cost energy meter with payment facility (see Fig. 1). As depicted in Fig. 1, the proposed system has the facility for AMR and payment of bill. Customer can pay the bill for the consumption using EB card which is introduced for this design which is like recharge card for mobile phone usage.

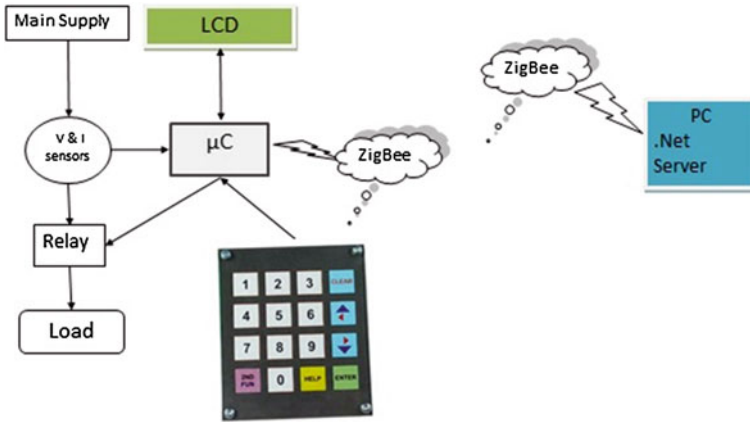


Fig. 1 Conceptual architecture overview

2.1 Microcontroller

This unit contains the software components such as the automatic energy meter system through which the consumption is calculated and fed to remote station. Here, we used Atmel's AT89S52 microcontroller which is having four ports consisting of 32 input/output pins. Here, the microcontroller acts as the central processing unit, i.e., brain of the system. The microcontroller is able to communicate when there is a need to access the network for sending or receiving data. Since the microcontroller is an electronic device, the signals that enter it must be electrical. In order to make an electrical or mechanical machine communicate with the microcontroller, an interface has to be used. The microcontroller takes the data from the interface and makes some calculations if needed then translates the data into some commands so the module can understand it. The controller also takes the data from the module then translates the data to the interface to produce the bill. Then, after producing the bill, it will send it to microcontroller with the network communication and it will be displayed in LCD.

2.2 ZigBee Technology

ZigBee is a radio frequency (RF) communications standard based on IEEE 802.15.4. ZigBee is new wireless communication technology, representing a wireless sensor network which is highly reliable, secure, low data rate, low power consumption, low cost and fast reaction. The ZigBee coordinator is responsible for creating and maintaining the network. All communication between devices propagates through the coordinator to the destination device. The wireless nature of ZigBee helps to overcome the intrusive installation problem with the existing

systems identified earlier. The ZigBee standard theoretically provides 250 kbps data rate, and as 40 kbps can meet the requirements of most control systems, it is sufficient for controlling the system. The low installation and running cost offered by ZigBee helps to tackle the expensive and complex architecture problems with existing systems, as identified earlier.

2.3 Personal Computer (PC)

PC consists of necessary operating system which can control the software created, and it acts as a server to receive and provide data for billing. This will generate bill according to tariff. This will contain all details of a customer.

2.4 Keypad

A keypad is a set of buttons arranged in a block or “pad” that usually bear digits which is known as numeric keypad. This keypad helps the customer to enter the secret code for the payment.

2.5 Liquid-Crystal Display (LCD)

A LCD is a flat panel display, electronic visual display, or video display that uses the light modulating properties of liquid crystals. Liquid crystals do not emit light directly. LCDs are available to display arbitrary images such as preset words and digits. Thus, we can able to get information by the display.

2.6 Voltage and Current Sensors

Voltage and Current sensor are used to detect the flow of energy, and it helps to calculate the power consumption.

3 System Implementation

3.1 Hardware Implementation

Voltage divider circuit and current divider circuit are used to note down the voltage and current value of consumption, the phase angle is also noted and they are fed to microcontroller to calculate the power consumption using our program which is

burn into the microcontroller, the reading is also displayed in the LCD, where we can be able to collect information on voltage, current, power, readings, etc., thus we can measure energy consumption.

By using ZigBee network, the reading is send to remote location server, where a software program is created using GUI. Here, the tariff for our usage is given which helps to generate bill for the usage. Then, it is sent back to energy meter using ZigBee network, and it is displayed in LCD.

After getting the bill in LCD, we can user EB card to settle our bill which is introduced for this system; here, the card contains some secret numbers which possess certain amount according to our bill we can pay the amount using the card. For example, if our consumption is about Rs. 200, we can buy a card for Rs. 200, which will contain a secret code number. Then, we should type that number using keypad and press a button to send to the server. By receiving the code, it will decode the code in amount, and thus, it will clear our account on payment.

3.2 Software Implementation

The software developed here is with GUI, which makes easier to work on it. It is created using .NET. It has database management system, which helps to maintain the data. Customer can directly pay the amount in electricity board station where this software helps to complete the payment.

4 Conclusion

This paper presents the design and the implementation of an interactive ZigBee-based energy meter. As the mobility in the world increases, the need for communicating to remote locations also increases. The use of ZigBee communications technology helps lower the expense of the system and the intrusiveness of the respective system installation. This system helps to pay our electricity bill in our own place.

4.1 Simulation Results

Simulation is done in Proteus software. Simulation output represents the voltage, current, unit consumption, and cost for consumption. By means of keypad, we can enter the secret code for the amount to be paid, where programming is done in MikroC Pro and hex file is generated (Figs. 2, 3, 4 and 5).

The key number can be pressed by using keypad, and then, the key will be sent to remote location station where the key number will be decoded and the amount

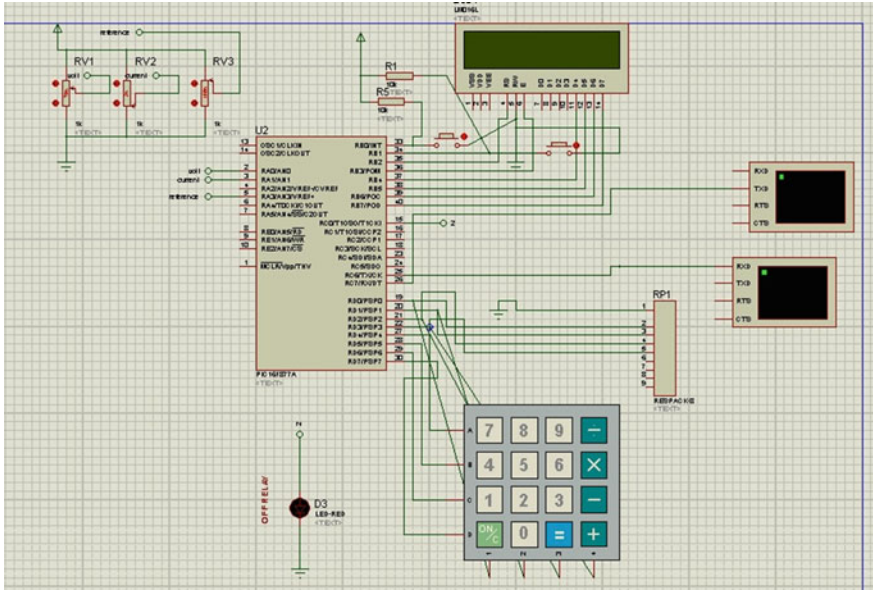


Fig. 2 Simulation circuit

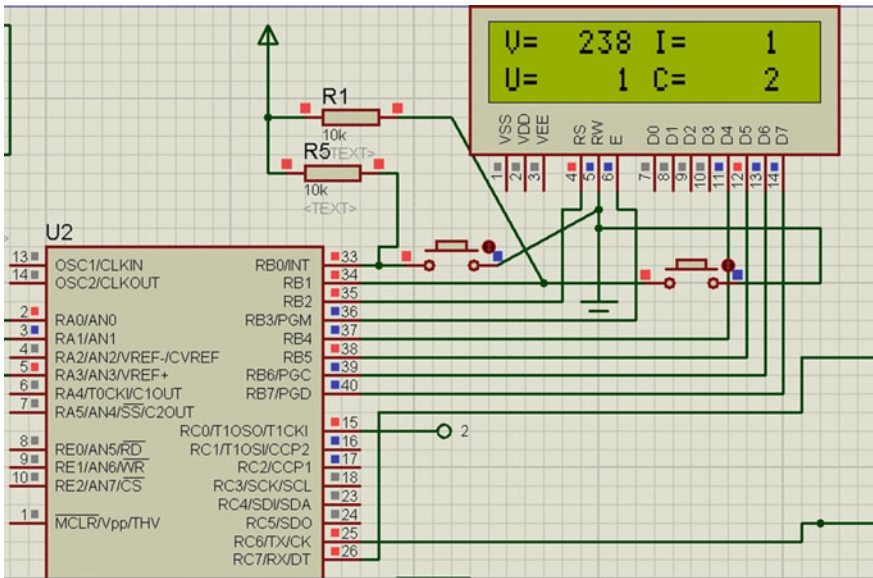


Fig. 3 Output representing voltage, current, units, and cost

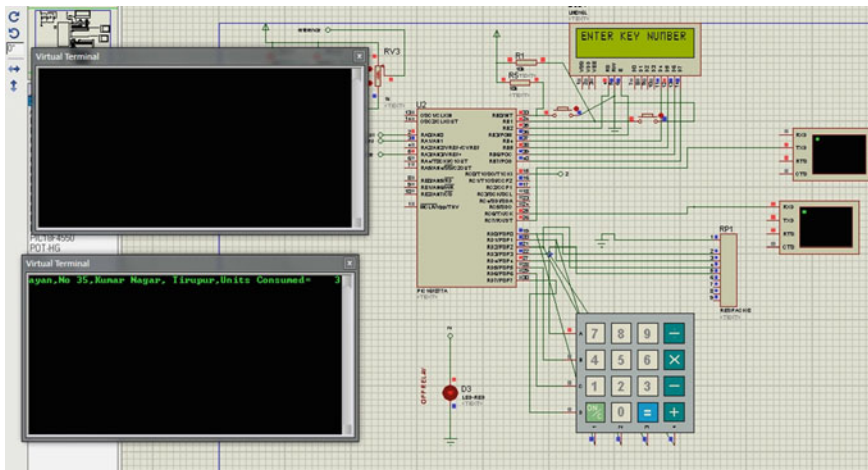


Fig. 4 Using keypad we can enter the key number

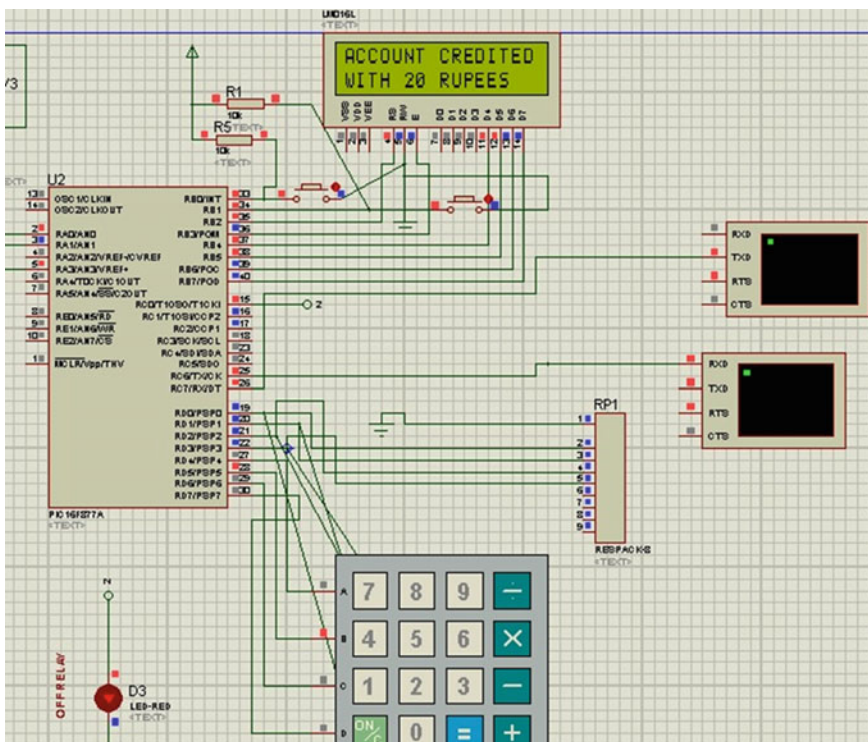


Fig. 5 Amount credited to your account

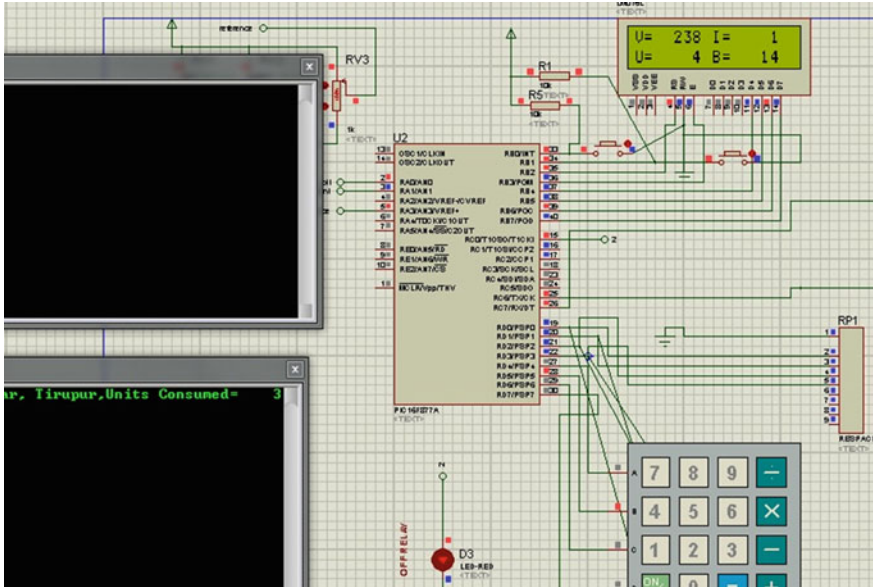


Fig. 6 Represent the balance amount after credit

for the key number will be credited to your account. If there is any balance in your account, it will be detected for next month charge and it will add the cost to your account if there is no balance (Fig. 6).

Simulation figures represent the model of hardware which provides the result that deserve for this project.

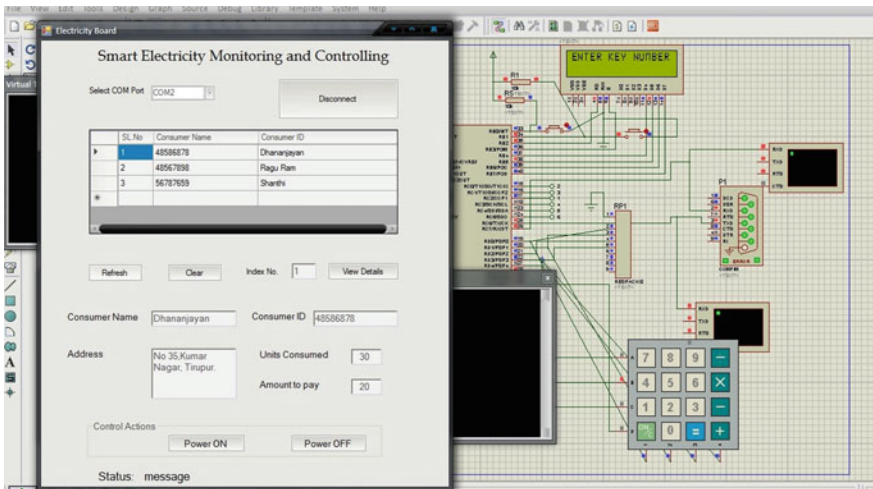


Fig. 7 .Net work frame to receive data

Visual studio has been used to frame an executable application which is used to receive data and store data in database; here, for simulation, we have used Eltima software, which is virtual serial port connector which helps us to communicate with the executable file and hardware. Controlling buttons are provided in the framework, which used to control the hardware. The power on and power off button will provide the control action. 1 is given as logic for power on button, and 0 is given as logic for power of button. Customer name and customer ID had been registered in the database using servers (Fig. 7).

Acknowledgments We wish to thank our Professor E. Santhi Bannari Amman institute of Technology for his support and encouragement.

References

1. Y. Bao, X. Jiang, Design of electric energy meter for long-distance data information transfers which based upon GPRS ISA2009, in International Workshop on Intelligent Systems and Applications (2009)
2. K. Ashna, S.N. George, GSM based automatic energy meter reading system. IEEE (2013)
3. T. Maity, P.S. Das, *Intelligent Online Measurement and Management of Energy Meter Data Through Advanced Wireless Network*. IEEE (2011)
4. B.S. Koay, S.S. Cheah, Y.H. Sng, P.H.J. Chong, H.W. Kuek, *Design and Implementation of Bluetooth Energy Meter*. IEEE (2003)
5. V.K. Sehgal, N. Panda, N.R. Handa, Electronic energy meter with instant billing, in UKSim Fourth European Modelling Symposium on Computer Modelling and Simulation

GA-Based Compiler Parameter Set Tuning

N.A.B Sankar Chebolu, Rajeev Wankar
and Raghavendra Rao Chillarige

Abstract Determining nearly optimal optimization options for modern-day compilers is a combinatorial problem. Added to this, specific to a given application, platform and optimization objective, fine-tuning the parameter set being used by various optimization passes, enhance the complexity further. In this paper, we apply genetic algorithm (GA) to tune compiler parameter set and investigate the impact of fine-tuning the parameter set on the code size. The effectiveness of GA-based parameter tuning mechanism is demonstrated with the benchmark programs from SPEC2006 benchmark suite that there is a significant impact of tuning the parameter values on the code size. Results obtained by the proposed GA-based parameter tuning technique are compared with existing methods and that shows significant performance gains.

Keywords Compiler optimization · Genetic algorithms · Parameter tuning

1 Introduction

Today's compilers provide a vast spectrum of optimizations to improve the code aggressively. These optimizations include local optimizations, global optimizations, inter-procedural optimizations, feedback-directed optimizations, link-time optimizations, etc., with impact on execution time, code size, and power. The best

N.A.B.S. Chebolu (✉)
ANURAG, Hyderabad, India
e-mail: chnabsankar@yahoo.com

N.A.B.S. Chebolu · R. Wankar · R.R. Chillarige
School of Computer and Information Sciences,
University of Hyderabad, Hyderabad, India
e-mail: rajeev.wankar@gmail.com

R.R. Chillarige
e-mail: crres@uohyd.ernet.in

optimization sequence will depend on the application, optimization objective, and the target architecture. Determining the best set of optimization options and their ordering is most difficult task as the interactions between different optimizations is nonlinear. Choosing the right set of optimizations can make a significant difference in the code size, power usage, and also on the execution time of a program. Study shows that standard optimization levels result in poor performance [1, 2], and tuning the compiler optimization settings will result in maximal performance [3]. Sophisticated auto-tuning strategies for exploring the optimization sequences is considered as one of the major sources of unexploited performance improvements with existing compiler technology [4]. Due to the sheer number of optimizations available and the range of parameters that they could take, identifying the best sequence by hand is impossible [5]. The literature survey shows many attempts by various researchers to tune the optimization options using various approaches including the statistical tuning [3], genetic algorithms (GAs) based [2], and also machine learning based [1]. Tuning based on multi-objective compiler optimization [6, 10] was also proved effective. Rather applying these techniques at program level, attempts to obtain the best set of optimizations at method level [11] were also proposed.

It is intuitive that the parameter set also plays an important role in achieving better performance. Fine-tuning of the parameter set is not considered by earlier researchers. In our previous study [7], parameter tuning was done using the greedy-based iterative approach. Greedy approach works sequentially with a restricted parameter space and may lead to a solution which may not be globally best. As the GA has no such subjectivity, we are using this technique in the current study of tuning the parameter set. The effectiveness of GA-based tuning technique is compared with our previous tuning approach. We have considered the code size as our optimization objective as it is important factor in embedded systems design. Many traditional compiler optimizations are designed to reduce the execution time of compiled code, but not necessarily the code size. For embedded system developers, minimizing the code size and power usage are more vital than the performance in terms of average execution time.

The rest of the paper is structured as following. Section 2 explores the compiler optimization space and also illustrates the experimental setup. Section 3 describes the GA-based tuning strategy, and Sect. 4 discusses the effectiveness of tuning. This section also compares this tuning technique with other techniques. Finally, in Sect. 5, the results are summarized and a case is made for future research in this area.

2 Compiler Optimization Space and Experimental Setup

Modern-day compilers, both commercial and open source, are generally equipped with vast number of optimizations of various kinds implemented in several passes and provide many optimization options and parameters, so that user can invoke them to produce the best result as per their optimization objective. We have considered GCC 4.6.4 for our study.

2.1 Compiler Optimization Options

The optimization space of widely known and used open-source GCC compiler encompasses optimization transformations in more than 240 passes on both GIMPLE and RTL. GCC offers architecture-independent optimization options of ‘-f’ type and architecture specific options of ‘-m’ type. There are around 176 optimization options of ‘-f’ type supported by the GCC 4.6.4 compiler [8], which can be made on or off by the user. GCC implements the notion of standard optimization levels, and these levels from O0, O1, O2, O3, and Ofast are meant for gradually increasing their stress on average execution time, and -Os will reduce the code size. However, these levels may have side effect on other objectives like execution time/size and also on compilation time, debugging information, etc. But these dependencies are not exactly known.

2.2 Parameter Set

The optimization options that correspond to various classes of optimizations like alignment-related, branch and loop-related, local and global optimizations, inter-procedural analysis-based, register allocation-related, link-time optimizations, etc., are further supported by around 120 parameters. These parameters correspond to various optimization pass implementations as provided by the compiler writers. Each parameter takes a discrete value from its unique allowed range with default, minimum, and maximum values as provided in the GCC internal documentation. For example, user can invoke the loop peeling optimization using the option ‘-fpeel-loops’ and further tweak this optimization using related parameters like ‘max-peeled-insns’ and ‘max-peel-times.’ The default values of these parameters are 400 and 16, respectively, and each parameter has its own allowed range. However, GCC provides an option of the form ‘-param name = value,’ which can be used to change these parameters explicitly. The ‘name’ refers to the parameter name, and the ‘value’ is the allowed value from its range. Tuning of these parameters will have impact provided the corresponding optimizations are invoked through enabling relevant optimization options. Unlike the options, neither reordering nor repetition of these parameter settings will have any impact. We have considered 104 parameters for our study, and rest of the parameters not considered, as their values are fixed as per host processor architecture, or they are not relevant from the code size objective, or enough documentation is not available.

2.3 Test Platform

Intel Xeon E540-based 4-core system, in which each core is operating at 2.66 GHZ with 6 MB of L1 cache and with Fedora release 17 having Linux Kernel 3.6.9, is used for the experimentation. Test setup is developed using MATLAB and bash scripting. Test cases are selected from CINT2006 (integer component of SPEC CPU2006) and CFP2006 (floating point component of SPEC CPU 2006) benchmark suites [9]. Compute-intensive benchmark programs used in our experimentation are *Perlbench*, *bzip2*, *gcc*, *mcf*, *gobmk*, *hmmmer*, *sjeng*, *libquantum*, *h264ref*, and *lbm*.

3 Parameter Set Tuning Using Genetic Algorithm

GAs are non-traditional optimization methods that simulate the natural evolutionary process of survival of the fittest. For the current problem, the coded values of compiler runtime parameters are referred as genes. We have considered 104 compiler parameters $p_1, p_2 \dots p_{104}$ of GCC 4.6.4, and these parameters can be supplied to the compiler using the command line switch of type ‘*-param name = value*,’ where name refers p_i and value is a discrete value taken from its corresponding range using the encoding function. The range of each variable is mapped with the gene length of 8 bits using the encoding function. String of such genes is called chromosome, representing the set of all compiler switches of type ‘*-param*,’ which in turn change the values of runtime parameters. Each chromosome represents possible solution to the problem, and set of all such chromosomes will constitute the population. The chromosome will be applied to the cost function, which in turn invoke the compiler under study and provides the code size as fitness value. As the objective under consideration is code space, the fitness function will in turn invoke the compiler with the corresponding chromosome and provide the code size of the executable. The code size is the sum of text and data section sizes obtained using the size command. We desire the chromosome with low fitness value, as we are keen in minimizing the code size. As the values that all parameters can take are only discrete values from its respective range, we have used the binary GA for experimentation. The initial population was selected randomly; that is, the initial values of all genes (parameters) in each chromosome (compilation sequence) of the initial population were taken randomly. The evolutionary process involves cycles of generating and testing new offspring solution using one of the three operations: reproduction, crossover, and mutation. Using crossover operation, we combine part of one chromosome with part of another chromosome and create a new chromosome. Similarly, mutation operator is applied to change the genes randomly with mutation rate probability. The GA-based evolution is repeated for a fixed number of generations, and the best chromosome is selected. Values of different GA parameters used are given in the Table 1.

Table 1 Lists of GA properties and their values considered

GA property	Value	GA property	Value
Number of generations	100	Selection method	Weighted random pairing
No. of chromosomes in each generation	16	Crossover method	Single point
No. of genes in each chromosome	104	Crossover probability	0.5
Gene length	8 bit	Mutation rate	0.15

4 Analysis of the Results

As stated earlier, 10 benchmark programs from SPEC 2006 are considered for our experimentation. ‘-O2’ is taken as our default optimization switch. We have applied the GA-based parameter tuning on top of this. This tuning is carried out for 100 generations, and it is observed that for many cases parameter set is getting converged to optimum level and local minima is reached. Figure 1 will depict the minimum cost and mean cost in each generation for the program ‘lbn.’ Figures 2 and 3 illustrate the impact of the GA-based tuning parameter tuning on the code size. Empirically, it is observed that with GA tuning, code size has come down up to 6.7 %.

Parameter set was evolved starting from the initial random values. As mentioned earlier, we have compared the effectiveness of GA approach with our previous study of greedy-based iterative approach. Table 2 illustrates these results, wherein CS1 (compiler setting 1) refers to compilation with O2, CS2 refers to compilation

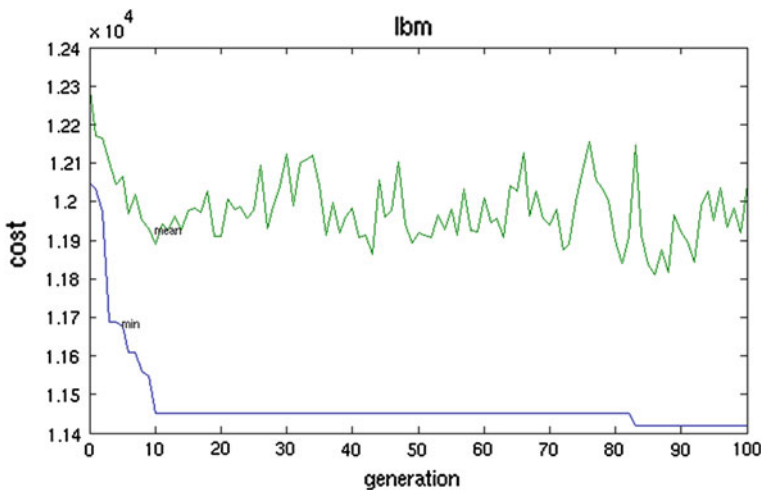


Fig. 1 Best and average costs in each generation for ‘lbn’ benchmark program

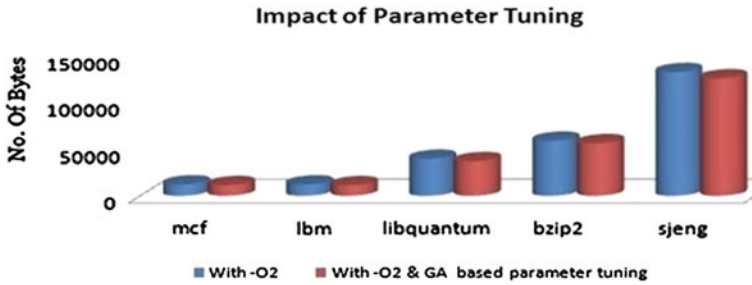


Fig. 2 Impact of the GA-based parameter set tuning on code size (in terms no of bytes) on selected benchmark programs

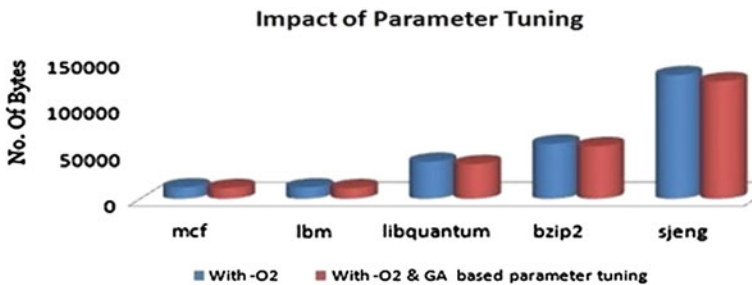


Fig. 3 Impact of the GA-based parameter set tuning on code size (in terms no of bytes) on selected benchmark programs

Table 2 Code size of benchmark programs (in terms of number of bytes) with different tuning techniques and start points

Benchmark program	CS1	CS2	CS3	CS4
hmmmer	39,748	39,732	37,164	36,844
Libquantum	132,036	132,184	125,356	126,112
Sjeng	280,548	282,680	269,532	269,492

with O2 and tuning using greedy approach, CS3 refers to GA-based tuning from random start point, and CS4 refers to GA-based tuning with initial start from the results of CS2.

Analysis of this Table 2 shows that GA-based tuning has greater impact than the other tuning technique by 6.5, 5.2, and 4.7 % for the benchmark programs ‘hmmmer,’ ‘libquantum,’ and ‘sjeng,’ respectively. It is also evident that initial parameter set taken from local minima of greedy-based result has a positive impact on couple of programs than starting with random initial values. Further to this, a statistical analysis based on the analysis of variance (ANOVA) is carried out on the code size values of all ten selected benchmark programs at standard optimization level -O2

with default parameter set and also with the fitness values obtained through this GA-assisted fine-tuning exercise. It is evident from the results of statistical analysis with 5 % level of significance that tuning the parameter set will play significant role in optimizing the code size.

5 Conclusions and Future Work

This study brings out fact that tuning the parameter set apart from the optimization options is necessary to obtain the best results. It is also evident that GA-based tuning strategy outperforms our previous approach. These tuning techniques can further be applied to other optimization objectives. Optimizing for code size has influence on other optimization objectives like average execution time, power usage, etc. Hence, it is proposed to study these multi-objective optimization problems.

References

1. F. Agakov, E. Bonilla, J. Cavazos et al., Using machine learning to focus iterative optimization, in *Proceedings of CGO* (2006)
2. K.D. Cooper, P.J. Schielke, D. Subramanian, Optimizing for reduced code space using genetic algorithms. *SIGPLAN Not.* **34**(7), 1–9 (1999)
3. M. Haneda, P.M.W. Knijnenburg, H.A.G. Wijshoff, Automatic selection of compiler options using non-parametric inferential statistics. 14th International Conference on Parallel Architectures and Compilation Techniques (PACT'05)
4. V. Adve, The next generation of compilers, in *Proceedings of CGO* (2009)
5. M. Duranton, D. Black-Schaffer, S. Yehia, K. De Bosschere, Computing Systems: Research Challenges Ahead the HiPEAC Vision (2011/2012)
6. J. Cavazos, M.F.P. O'Boyle, Method-specific dynamic compilation using logistic regression, in *Proceedings of OOPSLA'06*
7. P. Lokuciejewski, S. Plazar, H. Falk, P. Marwedel, L. Thiele, Multi-objective exploration of compiler optimizations for real-time systems, in *Proceedings of ISORC* (2010)
8. N.A.B.S. Chebolu, R. Wankar, R.R. Chillarige, Tuning the optimization parameter set for code size, in *Proceedings of MIWAI* (2012)
9. A. Martinez-Alvarez, J. Calvo-Zaragoza, S. Cuenca-Asensi, A. Ortiz, A. Jimeno-Morenilla, Multi-objective adaptive evolutionary strategy for tuning compilations. *Neurocomputing* **123**, 381–389 (2014)

An Intelligent Intrusion Detection System Using Average Manhattan Distance-based Decision Tree

R. Selvi, S. Saravan Kumar and A. Suresh

Abstract Recently, security is an important challenge in Internet-based communication. In such a scenario, intrusion detection systems help to secure the information through the identification of normal and abnormal behaviors. In order to model these behaviors accurately and to improve the performance of the intrusion detection system, intelligent decision tree algorithm based on average Manhattan distance algorithm (IDTAMD) is proposed in this paper. In this proposed new classification algorithm for effective decision making in the network data set. Moreover, an attribute selection algorithm called modified heuristic greedy algorithm [1] is used to select itemsets from redundant data. The experimental results obtained in this work show high detection rates and reduce the false alarm rate. This system has been tested using the tenfold cross-validations on the KDD'99 Cup data set. The results have been tested with tenfold cross-validation.

Keywords Intrusion detection system (IDS) • Attribute selection • Modified heuristic greedy • Itemsets

1 Introduction

Computer networks used to transfer a lot of secret information shared between the different types of computer devices from huge servers. Network security algorithms are introduced every day based on the various security concepts such as encryption,

R. Selvi (✉)
St. Peter's University, Chennai, India
e-mail: ss12.selvi@gmail.com

S. Saravan Kumar
Panimalar Institute of Technology, Chennai, India
e-mail: saravanakumars81@gmail.com

A. Suresh
SMK Fomra Institute of Technology, Chennai, India
e-mail: asuresz@yahoo.com

firewalls, machine learning, and access control. Urgent need of the current world is to develop an effective intrusion detection system (IDS) to detect known and unknown attacks dynamically. There are two major intrusion detection methods, namely misuse detection and anomaly detection. Misuse detection compares network activities with pre-defined signatures taken from typical features that represent an exact attack. Anomaly detection discovers attacks by identifying deviations from normal network activities [2].

Current intrusion detection systems are rule-based systems that depend on a set of rules representing abnormal or normal which are identified by security experts. Manual rule encoding is a very costly and lengthy process [3]. Moreover, it depends on the efficiency of human experts in analyzing a large amount of network activities to discover intrusion patterns. However, these drawbacks are overcome by introducing many data mining techniques in IDSs [2, 4–6]. Data mining is the analysis of observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [4]. The data mining techniques have been used to classify network connections into abnormal and normal data based on labeled training data in misuse detection [7].

In this paper, we propose a new intelligent classification algorithm called intelligent decision tree algorithm which is based on average Manhattan distance (IDTAMD) for effective decision making on the data. Moreover, we use an attribute selection algorithm called modified heuristic greedy algorithm of itemset [1] for eliminating the redundant data. We have focused mainly on the effective detection of DDoS attacks by using SVM algorithm, since DDoS attacks are more serious than other attacks.

The subsequent sections are organized as follows: Sect. 2 presents a general survey in field of misuse detection, anomaly detection, and data reduction. Section 3 describes the architecture of the system proposed in this paper for implementing a new attribute selection algorithm. Section 4 gives a brief explanation about the proposed algorithm and implementation. Section 5 discusses the results and its possible implications. Conclusions and future works are given in Sect. 6.

2 Literature Survey

In the past, a number of feature selection techniques have been proposed in the literature. Among them, a simplified decision table was proposed for improving the efficiency of attributes reduction to obtain the minimal attribute reduction in [8]. Geng et al. [9] proposed a novel feature selection method for network intrusion based on fast attribute reduction algorithm of rough set. Onpans et al. [1] proposed a new method for feature selection using the modified heuristic greedy algorithm of itemset (MHGIS) by implementing the apriori algorithm to improve the detection rate, accuracy rate, and false alarm rate. Om and Kundu [10] proposed a hybrid intrusion detection system that combines the merits of anomaly and misuse detection principles.

Farid et al. [11] proposed a new machine learning approach for network intrusion detection that performs data reduction by selecting important subset of attributes. It has reduced the false positives when compared to other classifiers such as ID3 algorithm and naïve Bayesian classifiers. There are many classification algorithms that are found in the literature [4–6]. Among them, Mulay et al. [4] proposed an effective classification algorithm called tree-structured multiclass SVM for classifying data effectively. They proposed decision tree-based algorithms to construct multiclass IDS which were used to improve the training time, testing time, and accuracy of IDS.

Madzarov et al. [5] proposed a new architecture for support vector machine classifiers that utilized binary decision tree (SVM-BDT) for solving the multiclass problems. This architecture provides techniques for achieving better classification accuracy. Redundant and irrelevant attributes of intrusion detection data set have led to a complex intrusion detection model which enhances the detection accuracy [6]. Sannasi et al. [7] proposed two new feature selection algorithms, namely an intelligent rule-based attribute selection algorithm and an intelligent rule-based enhanced multiclass support vector machine for effective classification of intrusion data.

Zhu et al. [3] proposed a novel method to reduce the scale of training set for one-class classification. This method only preserves the sample locating near the data distribution which may become support vector. Zhang and Wang [12] proposed an effective wrapper feature selection approach based on Bayesian networks for network intrusion detection.

3 System Architecture

This proposed intrusion detection system consists of five components, namely KDD Cup data set, user interface module, feature selection module, classification module, and decision-making module. The system architecture is shown in Fig. 1.

User interface module collects the data from KDD Cup data set. Feature selection module selects the important features. The classification module classifies the data by using the proposed classification algorithm called IDTAMD. The decision-making module decides whether the particular data are “normal” or “abnormal”.

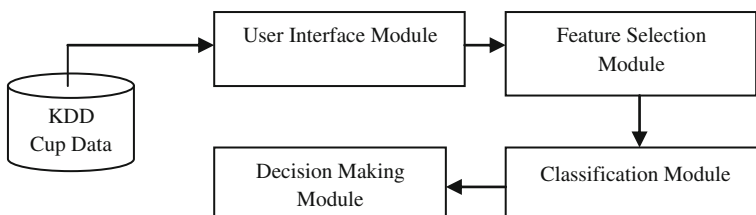


Fig. 1 System architecture

4 Implementation Details

4.1 Feature Selection

In this paper, we used an attribute selection algorithm called modified heuristic greedy algorithm of Itemset [1] for effective preprocessing. In this technique, select only the valuable attributes from the data set using projection. Moreover, data cleaning, data integration, and data transformation are carried out for performing effective preprocessing.

4.2 Intelligent Decision Tree Algorithm Based on Average Manhattan Distance Algorithm (IDTAMD)

The main idea of IDTAMD is to calculate the average Manhattan distance of every testing attribute. It selects the attribute with greatest Manhattan distance as the root node of the decision tree. Based on the values of attribute, the sample data are set into some subsets, and each subset corresponds to a sub-tree. All the steps of the algorithm have been repeated until all samples belong to the same class are obtained in every sub-tree. Moreover, leaf nodes in their respective classes were created. All the attributes of parent nodes no longer involve in the generation of sub-nodes, when building the sub-trees.

Algorithm: Intelligent decision tree algorithm based on average Manhattan distance AMD (D, A, C)

Input: Training data set D, testing attribute set A, class attribute C

Output: a decision tree

1. Agent creates a root node for the decision tree.
2. Every sample belongs to the same class c and the node is a leaf node, with label as c , return root.
3. The testing attributes are set A is empty and also returns a single node root.
4. Otherwise, the agent calculates the average Manhattan distance of every attribute in testing attribute set A.
5. Agent selects the testing attribute a with the highest average Manhattan distance as the expanding attribute.
 - (i) $\text{Root} \leftarrow a$ // set the attribute a as the root node of the decision attribute
 - (ii) For every possible value of attribute a , create a new branch Branch (a_i), where the value of attribute a is a_i in the sample data set.

6. For each node in a branch Branch (a_i):
 - (i) Every sample belongs to the same class and below this new branch add a leaf node with the corresponding to the class.
 - (ii) The sample set is empty and below this new branch add a leaf node with the most common class in the sample set.
 - (iii) Otherwise, call the AMD (D, A-(a), C) recursively.
7. End

The process of the IDTAMD algorithm is the similar process of the ID3 algorithm, and the main difference is the selection process of the testing attribute using intelligent agent. In this IDTAMD algorithm, find the average Manhattan distance when uses the information gain in ID3 algorithm. Moreover, the information gain and entropy hold a number of logarithmic operations. The proposed algorithm involves a great deal of logarithmic operations in all the selection of testing node. When the sample set is larger, this logarithmic symbol affects the efficiency of decision tree. The improved AMD algorithm uses the average Manhattan distance which has not enormous logarithmic operations, so it takes less amount of time for calculation and constructing the decision tree.

5 Experimental Results

5.1 Training and Test Data

The data set used in the experiment was taken from the Third International Knowledge Discovery and Data Mining Tools Competition (KDD Cup 99) [7]. Each connection record is described by 41 attributes. The list of attributes consists of both continuous-type and discrete-type variables, with statistical distributions varying drastically from each other, which makes the intrusion detection a very challenging task.

5.2 Experimental Results

Table 1 shows the list of selected features from 41 features in the KDD'99 Cup data set using the modified heuristic greedy algorithm of itemset [1].

Table 2 shows the overall results of three different classifiers, namely ID3, naive Bayes, and the proposed classification algorithm with 41 features. This classifier has achieved the highest detection rates for old DOS, PROBE, and R2L attacks. As to the new attacks, it also has the highest detection rates for DOS and U2R attacks.

However, a weakness observed is that it did not perform so well for detecting R2L attacks. This is the trade off for the high detection rate of other attack types.

Table 1 List of selected features

protocol_type, src_byte, wrong_fragment, hot, root_shell, su_attempted, num_access_shells, error_rate, diff_srv_rate, srv_serror_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count,dst_host_same_srv_count, dst_host_diff_srv_count, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_error_rate

Table 2 Performance analysis of the proposed algorithm with 41 features

Classes	ID3 algorithm [12]	NB classifier [3]	IDTAMD
Normal	99.71	99.65	99.73
Probe	98.22	99.35	99.65
DoS	99.63	99.71	99.74
U2R	86.11	64.84	88.12
R2L	97.79	99.15	99.21

Furthermore, the low number of instances for R2L connections in the training and testing data makes the detection rate of R2L negligible compared to other attacks.

Figure 2 shows the performance analysis of ID3 algorithm [12] and the proposed IDTAMD with reduced features. From Fig. 2, it can be observed that the detection accuracy of the proposed algorithm is better when it is compared with the existing ID3 algorithm in the detection of various attacks.

Table 3 shows the false-positive analysis. From Table 3, it can be observed that the proposed system performs better when it is compared with the existing ID3 due to use of the agent and the effective distance measurement formula.

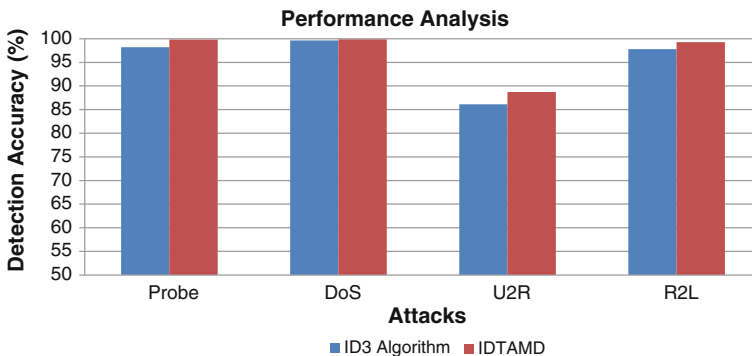


Fig. 2 Performance analysis between ID3 algorithm and IDTAMD

Table 3 False-positive rate analyses

Classes	ID3 algorithm [12]	IDTAMD
Normal	0.10	0.06
Probe	0.55	0.40
DoS	0.04	0.03
U2R	0.14	0.11
R2L	10.03	7.80

The tables and figures show the performance of the proposed algorithm with full features and reduced features. From the above results, it can be observed that the significant improvements were achieved in this proposed intrusion detection model.

6 Conclusion

In this paper, a new intrusion detection system is proposed which is the combination of modified heuristic greedy algorithm of Itemset and the proposed intelligent decision tree algorithm based on average Manhattan distance for effective decision making. The experimental results of the proposed algorithm achieved the higher detection accuracy and reduced the false alarm rate. Further work in this direction is possible to improve the performance of the intrusion detection model further using soft computing techniques.

References

1. J. Onpans, S. Rasmeguan, B. Jantarakongkul, K. Chinnasarn, A. Rodtook, Intrusion feature selection using modified heuristic greedy algorithm of itemset. 13th International Symposium on Communications and Information Technologies (ISCIT) (2013), pp. 627–632
2. D.E. Denning, An intrusion detection model. *IEEE Trans. Softw. Eng.* **51**(8), 12–26 (2003)
3. F. Zhu, N. Ye, W. Yu, S. Xu, G. Li, Boundary detection and sample reduction for one-class support vector machines. *Neurocomputing* **123**, 166–173 (2014)
4. S.A. Mulay, P.R. Devale, G.V. Garje, Intrusion detection system using support vector machine and decision tree. *Int. J. Comput. Appl.* **3**(3), 0975–8887 (2010)
5. G. Madzarov, D. Gjorgjevikj, I. Chorbev, A multiclass SVM classifier utilizing binary decision tree. *Informatica* **33**, 233–241 (2009)
6. W.K. Lee, S.J. Stolfo, A data mining framework for building intrusion detection models, in *Proceedings of the IEEE Symposium on Security and Privacy* (1999), pp. 120–132
7. S. Ganapathy, K. Kulothungan, S. Muthurajkumar, M. Vijayalakshmi, P. Yogesh, A. Kannan, Intelligent feature selection and classification techniques for intrusion detection in networks: a survey. *J. Wirel. Commun. Networking* **271**, 1–16 (2013)
8. C. Yang, H. Ge, G. Yao, L. Ma, Quick complete attribute reduction algorithm. 2009 6th International Conference on Fuzzy Systems and Knowledge Discovery, IEEE (2010), pp. 576–580

9. G. Geng, N. Li, S. Gong, Feature selection method for network intrusion based on fast attribute reduction of fuzzy rough set. 2012 International Conference on Industrial Control and Electronics Engineering (2012), pp. 530–534
10. H. Om, A. Kundu, A hybrid system for reducing the false alarm rate of anomaly intrusion detection system. 1st International Conference on Recent Advances in Information Technology (2012), pp. 1–6
11. D.M. Farid, J. Dormant, N. Harbi, H.H. Nguyen, M.Z. Rahman, Adaptive network intrusion detection learning: attribute selection and classification. International Conference on Computers Systems Engineering (2009)
12. F. Zhang, D. Wang, An effective feature selection approach for network intrusion detection. 8th International IEEE Conference on Networking, Architecture and Storage (2013), pp. 307–311
13. C. Jin, L. De-lin, M. Fen-xiang, An improved ID3 decision tree algorithm, in *Proceedings of 4th International Conference on Computer Science and Education* (2009), pp. 33–38

Intelligent Medical Diagnosis System Using Weighted Genetic and New Weighted Fuzzy C-Means Clustering Algorithm

P.S. Jeetha Lakshmi, S. Saravan Kumar and A. Suresh

Abstract In this paper, we propose a new intelligent prediction system to predict more accurately the presence of heart diseases effectively from feature-selected medical dataset. For this purpose, a new weighted genetic algorithm is proposed for selecting very important features from the dataset for improving the prediction accuracy of the disease. In this proposed intelligent prediction system, the data are preprocessed using the new weighted genetic algorithm and the new weighted fuzzy C-means clustering algorithm is used for effective fragmentation. Finally, we have used the ID3 algorithm for classification which is useful for making effective decision.

Keywords Weighted genetic algorithm · New weighted fuzzy C-means · Decision tree

1 Introduction

Machine learning techniques have been mostly used to help the medical experts in analyzing medical data [1–3]. However, with the emergence of significant medical data, many machine learning techniques suffer from the intractability problem. This is mainly due to the presence of large number of attributes, leading to higher dimensionality in the medical data. This high dimensionality influences the decision

P.S. Jeetha Lakshmi (✉)
St. Peter's University, Chennai, India
e-mail: jeethaps@gmail.com

S. Saravan Kumar
Panimalar Institute of Technology, Chennai, India
e-mail: saravanakumars81@gmail.com

A. Suresh
SMK Fomra Institute of Technology, Chennai, India
e-mail: asuresz@yahoo.com

in many aspects. Moreover, it may increase the time on both training and runtime phases of a learning system. In addition, it may cause problem in which is to be applied on such data handling the high dimensional medical data and hence feature reduction is an important technique [4–6].

The feature selection is one of the major topics during the past 10 years. It focuses on the selection of only the important features that are most predictive on a given outcome. Moreover, feature selection problems are found in most machine learning algorithms namely classification, prediction, and clustering. It is a difficult task since it has to select more attractive features for a classification problem, especially from a large dataset in which the feature space is very large [3]. Feature selection tries to achieve three main goals, namely reduction of cost extracting features, improving the classification accuracy, and enhancing the reliability of the estimated performance [4].

The existing techniques such as adaptive resonance theory (ART) algorithm [2] are used to perform clustering and to test all existing clusters for suitability when classifying a new input pattern after clusters are formed. For many applications, this results in the inability to use many useful clustering methods. Hence, it is necessary to use a suitable clustering algorithm for grouping the dataset effectively.

In this paper, a new genetic algorithm-based feature selection technique is proposed. In this process, n number of chromosomes is selected randomly for a given n . These chromosomes are decoded, and the fitness value is computed. The genetic operations namely selection, crossover, and mutation are used to generate a new population, and their fitness values are recomputed. The main contribution of this paper is the proposal of a new fitness function for the evaluation of new populations.

Remainder of this paper is organized as follows: Sect. 2 describes the related work. Section 4 explains the system architecture. Section 4 discusses the proposed work. Section 5 poses the results and discussion. Section 6 gives the conclusion and future enhancements.

2 Related Work

Benkaci et al. [2] established a computational architecture and techniques to deal with complex classification systems for feature selection. Their approach works are realized in two stages. In the first stage, they classified the faults from the data using the fuzzy ARTMAP classification, and in the second stage, it accounts between features of test data based on the hyper-cubes resulted in the first stage. Guan et al. [4] proposed a new RSFS method which can learn from both diagnosed and undiagnosed samples. Their method is called undiagnosed samples aided rough set feature selection. The main benefit of the proposed method is that it reduces the requirement on diagnosed samples by the help of undiagnosed ones. Finally, the promising performance of their work is validated through experiments on medical datasets.

Iyakaremye et al. [3] studied the fuzzy entropy and similarity-based feature selection in which they used entropy measures to discard redundant and irrelevant features in datasets. They reduced the number of features as well as computational time and also achieved higher mean classification accuracy. Their results show that their similarity measure based on Yu's norms could effectively perform feature selection when it is combined with fuzzy entropy measures. They achieved 98.83 % classification accuracy which is a prominent result in this area. Morgado et al. [5] proposed a new feature selection algorithm based on mutual information which is able to avoid the redundancy that exists between brain voxels that are highly dependent. Their approach is able to join higher amounts of relevant information from a feature vector of fixed dimension. Moreover, their method improves the classification performance when it is compared with the other feature selection techniques.

Turhali et al. [6] implemented and tested different ensemble methods on a dataset for feature reduction. They evaluated the performance of the classification methods by using the kappa, accuracy, and MCC values. Paja and Wrzesie [7] explained the results of a feature selection method which was used to find the most important or all important features that characterize melanocytic spots on the skin. Maulik et al. [8] show the effectiveness of the proposed transductive SVM scheme in the framework of transductive inference learning, using two feature selection methods. Sethukkarasi et al. [9] proposed an intelligent knowledge representation model for mining temporal patterns called fuzzy temporal cognitive map (FTCM) that is proposed for medical diagnosis and decision support system. Song et al. [10] suggested an enhancement in the learning process by incorporating the inference mechanism of fuzzy cognitive maps with automatic identification of membership functions and quantification of causalities. Their system does not provide the sufficient decisions made on critical applications in medical diagnosis.

3 System Architecture

The architecture of the prediction system proposed in this paper is shown in Fig. 1. The proposed prediction system consists of five modules, namely heart disease database, data preprocessing module, clustering module, classification module, and prediction.

The data preprocessing module collects the data from heart disease database. It also selects the necessary features based on the genetic operation. The clustering module is used to split the original dataset into sub-datasets. The classification module is used to classify the data by using the decision tree classifier.

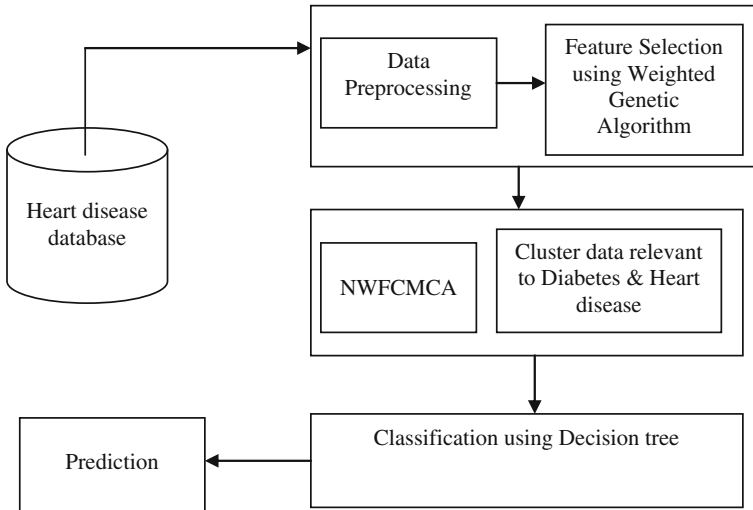


Fig. 1 System architecture

4 Proposed Work

The proposed system consists of four steps, namely data preprocessing, feature subset selection using weighted genetic algorithm, fragmentation of heart disease data using new weighted fuzzy C-means clustering algorithm, and decision tree representation for classification.

4.1 Data Preprocessing

The knowledge discovery during the training phase becomes more difficult if the database contains noisy or unreliable data. Hence, the data preprocessing is considered as initial and essential step in data mining projects. The various actions performed in preprocessing of dataset are removal of duplicate records, normalizing the values, handling the missing values, and removal of unneeded data fields.

4.2 Feature Subset Selection Using Weighted Genetic Algorithm

Feature selection is used to reduce the number of features before applying the classification algorithm. Irrelevant features may lead to take negative decision on the data in prediction task. Moreover, the computational complexity of a

classification algorithm may suffer from the curse of dimensionality caused by several features. Thus, the goal of feature selection is to find the minimal subset of attributes such that the resulting probability distribution of data classes is close to original distribution obtained using all attributes. In this proposed feature selection method, use weighted genetic algorithm for the feature selection process. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution such as inheritance, mutation, selection, and crossover. In this paper, we use a weighted genetic algorithm to select top “N” *best* features for classification out of total “M” features. The steps in simple genetic algorithm are as follows:

- Step 1: Select the initial population of individuals. These individuals are the chromosomes.
- Step 2: The fitness function evaluates the fitness value of each chromosome in that population.
- Step 3: Rank the chromosomes using fitness values.
- Step 4: Assign a weight for all chromosomes based on the ranking.
- Step 5: Choose the best-fit chromosomes for reproduction.
- Step 6: Use selected chromosomes to generate new chromosomes through crossover and mutation and the application of the fitness function

$$f(n) = (f(n-1) + f(n-2))/2. \quad (1)$$

- Step 7: Repeat steps 2–6 until the termination condition is reached. The termination conditions may be time limit, attributes limit, sufficient fitness achieved, etc.
- Step 8: If any feature repeats, eliminate the copies.
- Step 9: Stop the process when the desired output is met.

4.3 Decision Tree Representation

Decision tree is a popular classifier which is simple and easy to implement. It requires no domain knowledge or parameter setting and can handle high-dimensional data. Hence, it is more appropriate for exploratory knowledge discovery. An advantage of decision tree methods is that decision trees can be converted into understandable rules. The performance of decision trees can be enhanced with suitable attribute selection. Correct selection of attributes partitions the dataset into distinct classes. This proposed work uses ID3 decision tree [11] for classification.

5 Results and Discussion

In order to evaluate the proposed system, it is tested on a benchmark dataset, the medical data from the UCI machine learning repository [12]. Table 1 provides brief information about three datasets with reduced features and number of instances. Reduced features were obtained after applied the proposed feature selection algorithm.

Table 2 shows the experimental results. From Table 2, it can be observed that the proposed system classification accuracy is better than the existing system.

Figure 2 shows the comparison of the classification accuracy between USA-QR and the proposed system. From Fig. 2, it can be observed that the proposed system classification accuracy is better than the existing USA-QR.

Figure 3 shows the comparison of the classification accuracy between fuzzy kernel K-means clustering algorithm with immune genetic algorithm [10] and the proposed system.

From Fig. 3, it can be observed that the proposed system classification accuracy is better than the existing FKKMCA with GA due to the weighted features applied in the clustering and feature selection process.

Table 1 Reduced features and instances considered

ID	Medical data name	Number of features	Reduced features	Number of instances
1	Diabetes dataset	9	6	769
2	Heart disease dataset	14	9	270
3	Promoter gene sequence	58	52	106

Table 2 Experiment results

Data	Percentage of data (%)	Classification accuracy			
		Full features	Reduced features	Full features	Reduced features
1	20	39.5	41.7	42.3	43.4
	30	50.6	50.6	52.8	53.6
	40	64.5	64.1	65.6	66.3
	50	49.8	49.1	53.2	54.5
2	20	71.9	75.1	73.4	76.2
	30	76.3	78.1	77.3	79.3
	40	73.1	72.9	74.2	75.9
	50	75.8	76.6	76.9	77.8

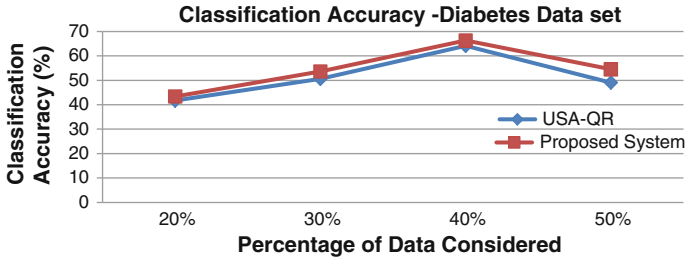


Fig. 2 Classification accuracy in diabetes dataset

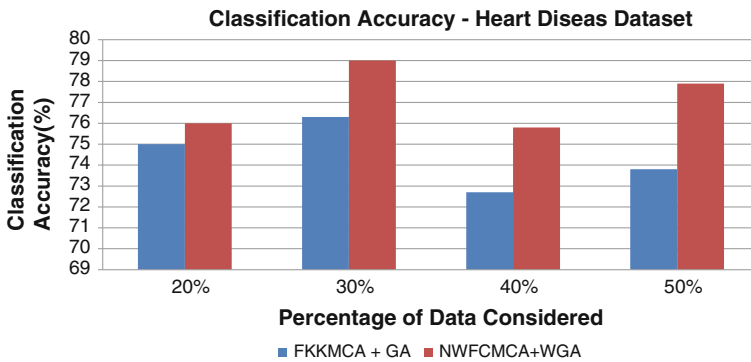


Fig. 3 Classification accuracy in heart disease dataset

6 Conclusions and Future Works

In this paper, a new system for medical diagnosis is proposed. This proposed system consists of a new weighted genetic algorithm, new weighted fuzzy C-means clustering algorithm, and decision tree classification algorithm. This system has been designed to achieve the better classification accuracy using the proposed weighted genetic-based feature selection algorithm. The main advantage of this method is that it reduces the false-positive rates. Future works in this direction could be the use of soft computing techniques for enhancing the power of the decision manager.

References

1. C.C. Hung, S. Kulkarni, B.C. Kuo, A new weighted fuzzy C-means clustering algorithm for remotely sensed image classification. *IEEE J. Sel. Top. Sig. Proc.* 5(3), 543–553 (2011)
2. M. Benkaci, B. Jammes, A. Doncescu, Feature selection for medical diagnosis using fuzzy ARTMAP classification and intersection conflict, in *IEEE 24th International Conference on Advanced Information Networking and Applications Workshops.* (2010), pp. 790–795

3. C. Iyakaremye, P. Luukka, D. Koloseni, Feature selection using Yu's similarity measure and fuzzy entropy measures, in *IEEE World Congress on Computational Intelligence (2012)*, pp. 1–5
4. D. Guan, W. Yuan, Z. Jin, S. Lee, Undiagnosed Samples Aided Rough Set Feature Selection for Medical Data, in *2nd IEEE International Conference on Parallel, Distributed and Grid Computing (2012)*, pp. 639–644
5. P.M. Morgado, M. Silveira, J.S. Marques, Efficient Selection of Non-Redundant Features for the Diagnosis of Alzheimer's Disease, in *IEEE 10th International Symposium on Biomedical Imaging (2013)*, pp. 640–643
6. U. Turhali, S. Babur, C. Avci, A. Akba, Performance Improvement for Diagnosis of Colon Cancer by Using Ensemble Classification Methods, in *IEEE Conference (2013)*, pp. 271–275
7. W. Paja, M. Wrzesie, Melanoma Important Features Selection Using Random Forest Approach, in *HIS (2013)*, pp. 415–418
8. Ujjwal Maulik, Senior Member, Anirban Mukhopadhyay, Senior Member, Debasis Chakraborty: gene-expression-based cancer subtypes prediction through feature selection and transductive SVM. *IEEE Trans. Biomed. Eng.* **60**(4), 1111–1117 (2013)
9. R. Sethukkarasi, S. Ganapathy, P. Yogesh, A. Kannan, An intelligent neuro fuzzy temporal knowledge representation model for mining temporal patterns. *J. Int. Fuzzy Syst.* **26**, 1167–1178 (2014)
10. G. Chengjie, S. Zhang, K. Liu, H. Huang, Fuzzy Kernel K-means clustering method based on immune genetic algorithm. *J. Comput. Inf. Syst.* **7**(1), 221–231 (2011)
11. C. Jin, L. De-lin, M. Fen-xiang, An improved ID3 decision tree algorithm, in *Proceedings of 4th International Conference on Computer Science and Education (2009)*, pp. 33–38
12. D.J. Newman, S. Hettich, C.L. Blake, C.J. Merz, UCI repository of machine learning databases

Security Analysis in Cloud Environment

M.S. Akshay, Ashina Kakkar, K. Jayasree, P. Prudhvi
and Prathibha Shridhara Metgal

Abstract Cloud computing is a new environment similar to distributed systems and is limited to its use of networking. Therefore, security issues are prevalent and cannot be ignored. Intrusion detection systems (IDS) are used to detect malicious behavior in network communication and hosts in real time. An open-source IDS widely in use is Snort, powerful IDS that can be configured by writing simple rules to detect a wide variety of hostile or suspicious network traffic. But IDS alone cannot effectively analyze the security threats as they have high rates of false alerts. Several machine learning and neural network algorithms have been tested on available datasets, and it has been proved that these algorithms help reduce false alerts up to a large extent. In this paper, we propose a way to bridge the gap between intrusion detection system benchmarking and real-world attacks by making use of effective and efficient algorithms.

Keywords Cloud · Intrusion detection system · Cloud fusion unit · Snort · Genetic algorithm · AdaBoost · Self-organizing map · Assessment units · Decision units · Virtual machines

M.S. Akshay (✉) · A. Kakkar · K. Jayasree · P. Prudhvi · P.S. Metgal
Department of IT, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Ettimadai,
Coimbatore 641112, India
e-mail: akshaymudumbi1992@gmail.com

A. Kakkar
e-mail: kakkardashina@gmail.com

K. Jayasree
e-mail: katarijayasree@gmail.com

P. Prudhvi
e-mail: prudhvi.koti@gmail.com

P.S. Metgal
e-mail: pratzmetgal272@gmail.com

1 Introduction

The cloud environment is relatively a new concept and is constantly being explored. Cloud service users rent the services and pay for only the services used. With the advent of Internet, numerous ways to compromise stability and security have been created. Hence, there has been an evolution of a cloud computing security discipline with ongoing efforts to cope with the requirements and capabilities regarding security issues. There should be a close watch on a very technical level, focusing primarily on attacks and hacking attempts related to cloud computing providers and systems [1].

Mechanisms such as intrusion detection systems (IDSs) are utilized for better security. An intrusion detection system like Snort requires expensive human input in the form of configuration of rules to create attack signatures or to determine effective models for normal behavior [2]. The high false alert rates discourage the use of the IDS alone. Effective algorithms like supervised algorithmic approaches have been considered as a potential alternative to human input. The main task of these algorithms is to analyze new data and characterize into normal or attack behavior. By improving the security issues using the effective algorithms and investigating the relevance of it with respect to dataset labels for normal behavior and each type of attack, we propose a more efficient approach [3].

The rest of this paper is organized as follows: Sect. 2 describes the related works. Section 3 discusses the threats faced by cloud environment. Section 4 discusses about Snort intrusion detection mechanisms. Section 5 discusses about the proposed system in the paper, and Sect. 6 describes the conclusion.

2 Related Works

Sondhiya et al. proposed an efficient IDS using multilayer perceptron algorithm and neural network for detecting unknown intrusion in an intrusion detection system.

Gunes et al. analyzed two machine learning algorithms, namely a clustering algorithm and neural network algorithm. The main objective of the paper is to analyze different sources and determine the differences between synthetic and real-world traffic.

Shrivastava et al. provided an overview of different attacks possible in a cloud environment and the attack surfaces possible stating the attacks for each of these surfaces.

Amirreza et al. proposed an Internet intrusion detection system CIDSS which is developed based on cloud computing and make up for the deficiency of traditional intrusion detection system.

Roesch et al. give a detailed study on snort, its components, and working. High-performance rules are written to identify intrusion behavior.

Ms. Parag K et al. give a detailed study on the different type of algorithmic approaches for the security in cloud computing and also the techniques already in existence.

3 Threats Faced by the Cloud Environment

Cloud service providers should ensure users with connectivity and availability. The attacks are usually featured as four categories, namely user to root, remote to local, denial of service, and probe. The different interfaces and attacks are as follows:

1. Server-to-client interface: buffer overflow and SQL injection.
2. Client-to-server interface: browser-based attacks and phishing attacks on mail.
3. Cloud-to-service interface: Attack which is possible is DOS attack.
4. Service-to-cloud interface: availability reductions and malicious interferences.
5. Cloud-to-user interface: change of instance of user or granting privileges to unprivileged users, etc.
6. User-to-cloud interface: phishing-like attacks [4].

There are different interfaces in which attacks are possible. And different attacks are possible in each of these interfaces. It is not possible for a human effort to monitor the attacks. So there is defense mechanism like intrusion detection mechanism which is needed to reduce the human effort and to match to the real-time situations.

4 Snort Intrusion Detection Mechanisms

(IDSs) are systems that automate the process of monitoring the event sourcing in a computer system or network, analyzing them for malicious activities, and produce reports to a management station. Intrusion detection can be a host-based intrusion detection system (HIDS) or a network-based intrusion detection system (NIDS). The approach used by them can be anomaly detection or misuse detection. An IDS is generally composed of several components:

1. Sensors which generate security events.
2. Console to monitor events and alerts and control the sensors.
3. Central engine that records events logged by the sensors in a database [5].

Snort is one such IDS that can be configured as a packet sniffer, packet logger, and NIDS. Snort rules are simple to write, yet powerful enough to detect a wide variety of hostile or merely suspicious network traffic. There are three base action directives that Snort can use when a packet matches a specified rule pattern, namely pass, log, or alert [6].

Even though Snort can be used for intrusion detection, Snort rules are to be manually written. So it requires human input, and they have high false alarm rates. Hence, to have efficiency in detecting attack and its cause, we require a more efficient technique like algorithmic approaches as proposed in the paper.

5 Proposed System

The proposed design and stepwise working is described below:

1. Deployment of private cloud like Eucalyptus. Topology is as follows: front-end cloud fusion unit (cloud controller (CLC), cluster controller, walrus, and storage controller) and back end (node controllers).
2. IDS are created by installing and configuring Snort on each virtual machines (VM) of the cloud systems. The advantage of this is to split the network traffic to all IDS and hence prevent overloading; to reduce impact of possible attacks; and prevent system from single-point failure attack [7].
3. Alerts yielded by the IDS will be stored in MySQL database placed within cloud fusion unit of front-end server. Single database is suggested for lower risk of loss of data and centralization of alerts for admin.
4. The cloud fusion unit consists of the database, two levels of assessment, and a decision unit. The first level of assessment will be using the multi-class AdaBoost algorithm, and the second level will be using the genetic and hierarchy of SOM algorithm. The decision unit compares the two outcomes before deciding whether the behavior is that of an intrusion or not.
5. The decision unit helps maximize the true positive rates and minimize false positives.

5.1 Cloud Controller

The cloud fusion unit is on the CLC which is also the cluster controller, storage controller, and walrus. The CLC is the front end to the entire cloud infrastructure. CLC provides web service interfaces to the client tools on one side and interacts with rest of the components of the cloud infrastructure on the other side. Functions of the fusion unit are as follows:

1. Monitor the availability of resources of components of the cloud infrastructure.
2. Deciding which cluster will be used for provisioning the instances.
3. Monitoring the running instances.

The cloud fusion unit consists of the MySQL database. Alerts from all VM are stored in the database. It also consists of the assessment units. The first level implements the supervised AdaBoost algorithm and gives the output to the decision

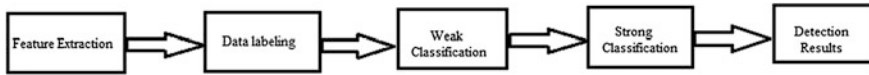


Fig. 1 AdaBoost algorithm implemented for the first assessment in intrusion detection with different stages involved

unit. The second level implements the hybrid of genetic as well as SOM algorithm and sends the output to the decision unit. The decision unit compares the output of the two assessment units and gives the final decision whether it is an attack or not when there is a conflict happening between the two assessment units.

5.2 Analysis Using AdaBoost Algorithm

The AdaBoost algorithm is a supervised learning algorithm which uses weak classifiers or learner. A weak learner is defined to be a classifier which is only slightly correlated with the true classification. In contrast, a strong learner is a classifier that is arbitrarily well correlated with the true classification. By adaptive boosting, the algorithm helps to build strong classifiers from weak classifiers [8]. It initially chooses the learner that classifies more data correctly. In the next step, the data are reweighted to increase the importance of misclassified samples. This process continues, and at each step, the weight of each weak learner among other learners is determined. A multi-class AdaBoost algorithm, unlike the traditional two-class AdaBoost, can be implemented for detecting normal behavior and different attack behavior like denial of service, cloud malware injection, and SQL injection based on the packet features. The system first extracts the features considered for classification. Since it is a supervised algorithm, the features to be considered are chosen beforehand [9].

Data labeling is then done to initialize the system with the probability distribution for each feature before classification. Classification then takes place with the classifier trying to attain minimum error. By taking a number of such weak classifiers where the error rates are considerably minimized each time, the strong classifier is attained. This is done by converging the result to minimize exponential loss. The detection results are then passed to the decision unit for comparison with results from the second assessment unit. The diagrammatic approach for the first assessment unit is shown in Fig. 1.

5.3 Analysis Using Genetic Algorithm and Self-Organizing Map (SOM)

The second stage of analysis implements the hybrid of genetic and two-level hierarchy of SOM algorithm. The genetic algorithm reduces the number of features

that are used for inspecting. A two-level SOM hierarchy was developed for intrusion detection. The system utilizes the basic features of a connection, which can be derived from packet headers without inspecting the packet payload. Genetic algorithm optimizes the number of features to be assessed and reduces the complexity. Then, these packets with corresponding features are sent to SOM hierarchy.

SOM is an artificial neural network algorithm, which employs unsupervised learning for training [10]. At the first level, SOMs are trained, one for each feature where the general objective is to encode temporal relationships within the features. The second level combines the information from the first-level SOMs.

Since real-world datasets are unlabeled, hit histograms are employed to analyze datasets with training datasets. Hit histograms summarize how often the neurons are excited. As a neuron is excited for more patterns in a dataset, the area of the hexagon being colored becomes larger. If the training and the test datasets have similar statistical properties such as similar range, probability distribution, and dispersion, then the test dataset would populate a considerable portion of the SOM. After performing a two-layer hierarchy of SOM, we can comparatively distinguish different patterns in a real-world dataset. The results are then passed to the decision unit for comparison with results from the first assessment unit.

5.4 Decision Unit

The decision unit is a console which is needed when a conflict arises between the decisions taken by the two assessment units. When the results arising from both the assessment units are conflicting, the decision unit raises the alarm where the system administrator can then control the final decision. Also the administrator can grant access rights to users by updating the database and rules by ignoring the decision taken by the unit.

The flow diagram of the system has been indicated in the Fig. 2, and the architectural diagram has been indicated in the Fig. 3.

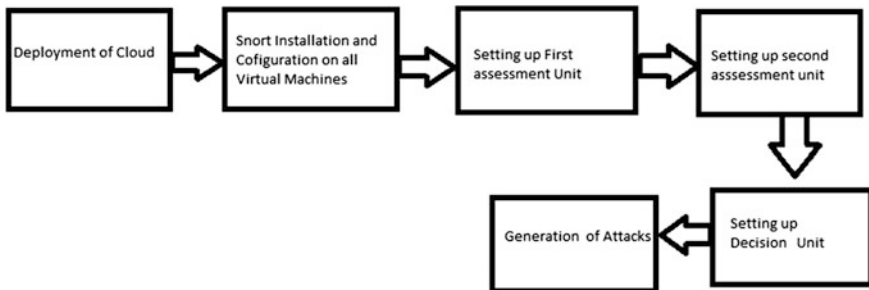


Fig. 2 Flow diagram of the proposed system with the intrusion detection system, assessment units, and decision unit

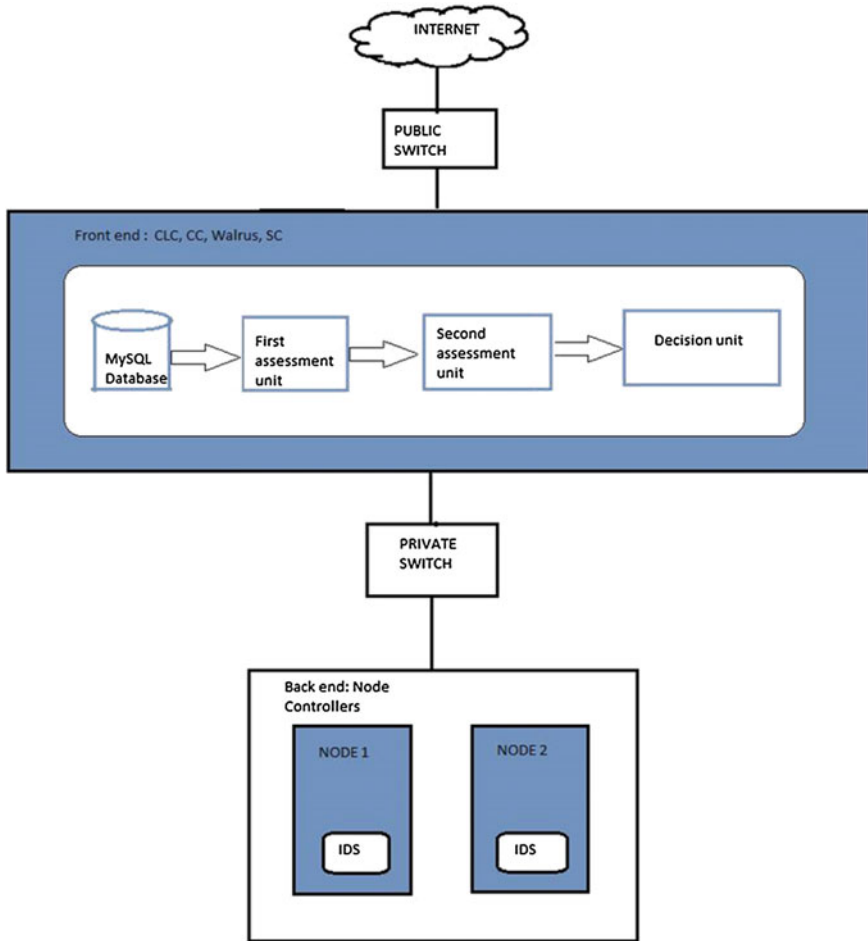


Fig. 3 Architecture of the proposed system with cloud fusion unit and the back end as node controllers

6 Conclusions

An approach using intrusion detection system with the help of AdaBoost as well as the hybrid of genetic and SOM algorithm has been proposed to detect and analyze the attacks. The proposed system fills in the gap that was present while using the intrusion detection system alone. It reduces the human effort required as well as the false alarms generated by the intrusion detection system. It can also be applied to the large network as the network traffic is split among all virtual machine by which one point failure does not happen. By this approach, we can bridge the gap between

intrusion detection system benchmarking and real-world attacks by making use of effective and efficient algorithms.

References

1. A. Zarrabi, A. Zarrabi, Internet intrusion detection system service in a cloud. *Int. J. Comput. Sci.* **9**(5) (2012)
2. S. Mukkamala, G. Janoski, A. Sung, Intrusion detection using neural networks and support vector machines. *IEEE-IJCNN* (2002)
3. A. Singh, M. Shrivastava, Overview of attacks on cloud computing. *Int. J. Eng. Innovative Technol.* **1**(4) (2012)
4. M. Roesch, Snort-lightweight intrusion detection for network, in *Proceedings of Lisa '99* (1999)
5. R. Sondhiya, M. Shreevastav, M. Mishra, To improve security in cloud computing with intrusion detection system using neural network. *Int. J. Soft Comput. Eng.* **3**(2) (2013)
6. K. Patel, R. Srivastava, Classification of cloud data using bayesian classification. *Int. J. Sci. Res.* **2**(6) (2013)
7. M.D. Holtz, B.M David, R. Timoteo, Building scalable distributed intrusion detection systems based on map reduce framework. *Revista Telecommun.* **1** (2011)
8. H.G. Kayacık, N. Zincir-Heywood, Analysis of three intrusion detection system benchmark dataset using machine learning algorithm. *Int. J. Eng. Innovative Technol.* (2009)

A Bi-level Clustering Analysis for Studying About the Sources of Vehicular Pollution in Chennai

Gunaselvi Manohar, S. Prasanna Devi and K. Suryaprakasa Rao

Abstract The aim of this paper is to study about the awareness among the people in Chennai city, Tamil Nadu, about the causes of pollution. Initially, the k-means clustering method was applied to group variables rather than observations in the design of questionnaires. The first draft of a questionnaire contained more questions than is prudent to ensure a good response rate. When the draft questionnaire is tested on a smaller number of respondents (75 samples), it was observed that the responses to certain groups of questions are highly correlated. Hence, clustering analysis was applied to identify groups of questions that are most predominant in contributing to the reduction in air pollution in Chennai. Thus, the selected questions were used for survey purpose to study the acceptability among different sectors of people. Primary data were collected from 110 people belonging to different sectors of Chennai using questionnaire method. In the second level of cluster analysis, the cluster analysis was carried out to assign observations to groups. These results were further applied to identify the recommendation of suitable transport policies to mitigate vehicular pollution. This method of applying clustering techniques in two levels of the questionnaire analysis has been newly proposed in this paper.

Keywords Pollution · Clustering · Questionnaire survey · Transport policies · Data mining

G. Manohar (✉)

Department of Industrial Engineering, College of Engineering Guindy, Chennai, India
e-mail: gunaselvim@rediffmail.com

S. Prasanna Devi

Department of CSE, Apollo Engineering College, Chennai 602105, India

K. Suryaprakasa Rao

Department of Industrial Engineering, Anna University, Chennai 600025, India

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_26

229

1 Introduction

It is a well-known fact that due to tremendous increase in the use of private, public, and personal vehicles in mega cities, air pollution has been up by many folds causing damage to urban public health and properties. Due to this sudden increase in vehicular population in the city of Chennai, South India, there is an increase in vehicular emissions caused by the vehicular effluents such as carbon monoxide, oxides of nitrogen, oxide of sulfur, and particulate matter. Thus, it has become necessary to conduct a study about public awareness of vehicular pollution in order to develop new transport policies to mitigate vehicular pollution.

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cut costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Cluster analysis is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics and policy decision process.

2 Literature Survey

As highlighted by Faiz and Sturm [5], one of the major environmental problems causing urban air pollution is road traffic. As per areas Xia and Leslie [10], traffic exhaust emission is one of the most important air pollution sources in urban. According to Fenger [6], air pollutant concentrations are dominated by the exhaust emissions of carbon monoxide, oxides of nitrogen, and suspended particles from the usage of vehicles in mega cities. Goyal et al. [7] estimated the contribution of transport sector as 72 % while understanding the problem of vehicular pollution vis-a-vis ambient air quality of a highly traffic-affected mega city, Delhi. A number of policies have been activated in India in order to control the level of air pollutants such as particulate matter, oxides of sulfur, and oxides of nitrogen Chelani and Devotta, [3]. According to Huai et al. [8], vehicular emission is the major contributor of air pollution in the cities. Aziz and Ihsan [1], and Ojo and Awokola [9] stated that discharge of motor vehicular carbon monoxide in Lahore is due to mass transit system with frequent stoppages, entering, and exit in flow of traffic.

The goal of cluster analysis, broadly stated, is to find the arrangement of observations and clusters that maximizes both within-group homogeneity and

between-group heterogeneity Borden, [2]; Within-group homogeneity refers to the extent to which observations that are assigned to a given cluster share similar attributes on the variables included in the cluster analysis. Between-group heterogeneity refers to the extent to which each cluster is dissimilar in the aggregate from other clusters with respect to the variables included in the analysis. The cluster analysis technique known as k-means is an iterative algorithm that attempts to generate the most appropriate fit of observations to clusters, given the number of clusters (k) selected by the researcher prior to the execution of the cluster analysis (Everitt Landau, Leese, and Stahl [4]).

In other words, the researcher selects the number of clusters in advance, which is the k in “k-means.” The algorithm generates the assignment of samples to clusters that minimizes differences between observations within a cluster and maximizes the differences between clusters.

3 Survey Methodology

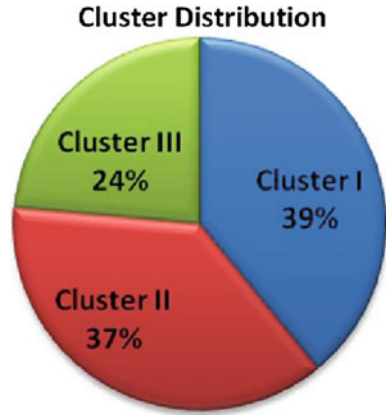
Aiming toward the implementation of suitable transport policies for mitigating the vehicular pollution data collected from different sources were integrated. The study focused on engineering colleges and public sector offices in Chennai, Tamil Nadu, which is the target population on which the findings were generalized. Data collection was through a questionnaire, which was interviewer administered. The study is among the people from different engineering colleges (post graduate students), hospitals, and major public sector offices. Twenty questions are framed to know about the public degree of awareness toward causes of increase of vehicular pollution, their knowledge about standards of emission norms, vehicular taxes, transport rules and regulations, and willingness to follow new transport policies.

The questionnaire used for the survey is shown in Appendix 1. The data were entered using the specially prepared software in database MS Access. In order to ensure quality control, the software was programmed to check the internal consistency of data entered. The data collection and processing were started in May and completed in July 2013. The SPSS-10 statistical package, being very suitable for this kind of analysis, was used for data tabulation and analysis.

4 Results and Discussions

The collected data from a sample of 75 respondents are grouped into three clusters by the k-means algorithm ($k = 3$) by the clustering technique. From the pie chart shown in Fig. 1, it was observed that cluster C1 has got maximum percentage contribution (39 %) among the three clusters, followed by C2 (37 %) and finally C3 (24 %). Hence, C1 has got more number of people and the result obtained for each question will be the central value of each cluster. The corresponding result will be

Fig. 1 Percentage distribution of clusters



the willingness or decision of all people in that respective cluster. So based on the result, it is possible to recommend suitable transport policy for implementation in future.

Clustering methods are applied to group variables rather than observations. Cluster 1, cluster 2, and cluster 3 are characterized by the high, average, and low mean value of the observations. The results of k-means clustering ($k = 3$) applied to group questionnaire data are given in Table 1.

Cluster analysis has been used to group variables into homogeneous and distinct groups. This approach is used, for example, in revising a questionnaire on the basis of responses received to a draft of the questionnaire. The grouping of the questions by means of cluster analysis helps to identify redundant questions and reduce their number, thus improving the chances of a good response rate to the final version of the questionnaire. From the results shown in Table 1, the questions whose centroids values are more than 4 in two or more clusters were grouped as the most distinctive

Table 1 Results of cluster analysis. **a** Questions 1 through 10. **b** Questions 11 through 20

Clusters	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
(a)										
C1	3.37	3.535	3.46	4.00	4.37	4.53	4.25	4.418	3.63	3.72
C2	3.05	2.92	3.36	3.29	3.94	4.31	4.29	4.09	2.97	1.66
C3	2.61	3.07	2.307	2.884	2.846	2.846	2.92	2.46	2.769	2.34
Clusters	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
(b)										
C1	3.83	3.37	4.32	3.604	3.163	3.558	3.79	4.16	3.69	4.62
C2	2.43	2.56	3.39	3.58	1.82	2.73	3.14	3.87	3.21	4.73
C3	2.69	3.5	2.42	2.69	3.9	2.96	3.03	2.65	3.307	3.115

Table 2 Description of clusters

Cluster	Percentage (%)	Classification	Description of clusters
C1	63	Self-oriented, Less ready to take personal constraints	Searches satisfaction by doing good thing
			Is not altruistic
			High level of formal education
			Believes policy decisions will be efficient for developing countries
C2	37	Value-oriented	Highly involved in social/political action
			Highest household income among all clusters
			Own security is the basis for solitary action

questions, which were strongly agreed by most of the respondents. It is observed that questions Q6, Q7, Q8, and Q20 were found to be most receptive among the respondents to achieve reduction in air pollution in Chennai.

In the second level of analysis, these four questions were distributed to 110 respondents. The cluster analysis was performed on the respondents to identify their typologies. Two clusters were formed. The respondent groups were classified as given in Table 2.

Table 3 a Initial clustering results. **b** Final clustering results

Mean	Cluster	
	1	2
(a)		
Q6: The main source of air pollution in Chennai is due to transport vehicle emissions	4.30	1.8
Q7: The problem of vehicular air pollution is worst in urban area	3.66	2.6
Q8: "Installation of traffic information boards about level of air pollution" will be useful	3.9	2.3
Q20: Vehicular pollution affects the public health and creates global warming problem	3.7	2.3
Mean	Cluster	
	C1	C2
(b)		
Q6	4.1	2.61
Q7	3.59	3.08
Q8	3.1	2.88
Q20	3.88	3.1

Table 4 The difference between the clusters

Mean	Cluster		Error		F	Signal
	MS	DF	MS	DF		
Q6	9.74	2	0.24	2	40.61	0.000
Q7	9.26	2	0.282	2	32.8	0.000
Q8	11.43	2	0.093	2	122.8	0.000
Q20	10.72	2	0.153	2	69	0.000

With k-means cluster analysis, the four predominant questions selected (cases) were clustered into 2 homogeneous groups based on the respondents characteristics. Hence, individual policies suitable to each group can be thought to enforce to reduce the level of air pollution in Chennai (Table 3a).

After the initial cluster centers have been selected, each case is assigned to the closest cluster, based on its distance from the cluster centers. After all of the cases have been assigned to clusters, the cluster centers are recomputed, based on all of the cases in the cluster. Case assignment is done again, using these updated cluster centers (10 iterations in this case) (Table 3b).

It is inferred from the above table that majority of the respondents are aware that vehicular pollution is the major source of pollution in Chennai, followed by the fact/awareness that vehicular pollution leads to health hazards. The F ratios have been computed to describe the difference between the clusters as shown in Table 4.

5 Conclusion

There has been a rapid increase in the number of vehicles, as a result of urbanization, economic growth, and easy availability of finance. Apart from new vehicles, old vehicles also exist often with outdated technology and nonobservance of emission norms. The quality of fuel supplied has also compounded the problem of vehicular pollution. In this regard, the study undertaken by this research to study the awareness among the people about the sources of air pollution clearly concludes that the people are very much aware about the fact that vehicle pollution leads to health hazards. Research is warranted, and the measures discussed in the questionnaire study are effective or showing promise in order to further identify cumulative measures that together can assure sufficient exposure reduction and health protection for those living near busy roadways.

6 Appendix 1: Questionnaire: Survey of Pollution Awareness Among People of Chennai

6.1 Questions

1. New motor vehicle may be subjected to general VAT or special vehicle sales tax may be levied based on either value or other attributes such as weight or engine capacity.
2. Annual fees for the registration or use of motor vehicle may take the form of an annual fixed amount for all vehicle of certain type (private cars or may be more finely according to vehicle characteristics).
3. Higher motor fuel taxes will lead to reduction of vehicle use.
4. Environmental Protection Agency (EPA) helps to control emission standards.
5. Biodiesel fuel is better to use than regular diesel fuel.
6. The main source of air pollution in Chennai is due to transport vehicle emissions.
7. The problem of vehicular air pollution is worst in urban area.
8. "Installation of traffic information boards about level of air pollution" will be useful.
9. People will re-route their trip on available alternative re-route if traffic information board displays higher level of air pollution.
10. If the alternate route is longer than the one with pollution hazard, will the people to sacrifice some time and fuels to prefer that long route?
11. Will the people be willing to pay for the information regarding pollution hazard on roads if they are available on payment only?
12. Will the people prefer public transport to travel?
13. The increase in fuel price will affect the usage of private vehicle.
14. The reason for preferring private vehicle is less frequency of operation of public transport.
15. Present conditions of the public transport system are good.
16. The reason for preferring public transport is safety in travel.
17. People are willing to share their vehicle with their colleagues.
18. Less awareness about vehicular pollution is the reason for not cooperating with the government to get emission check certificate for their own vehicle periodically.
19. Vehicles should be scrapped after a certain period.
20. Vehicular pollution affects the public health and contributes to global warming problem.

References

1. A. Aziz, I.U. Bajwa, Minimizing human health effects of urban air pollution through quantification and control of motor vehicular carbon monoxide (CO) in Lahore. *Environ. Monit. Assess.* **135**, 459–464 (2007)
2. V.M.H. Borden, Identifying and Analyzing Group Differences, in *Intermediate/Advanced Statistics in Institutional Research*, ed. by M.A. Coughlin (2005), pp. 132–168
3. A.B. Chelani, S. Devotta, Air quality assessment in Delhi: before and after CNG as fuel. *Environ. Monit. Assess.* **125**, 257–263 (2007)
4. P. Ewell, M. Boeke, *Critical Connections: Linking States' Unit Record Systems To Track Student Progress* (Lumina Foundation for Education, Indianapolis, 2011)
5. A. Faiz, P.J. Sturm, Air pollution and road traffic in developing countries. *Atmos. Environ.* **34**, 4745–4746 (2000)
6. J. Fenger, Urban air quality. *Atmos. Environ.* **33**, 4877–4900 (1999)
7. S.K. Goyal, S.V. Ghatge, P. Nema, M. Tamhane, Understanding urban vehicular pollution problem vis-à-vis ambient air quality-case study of a megacity (Delhi, India). *Environ. Monit. Assess.* **119**, 557–569 (2005)
8. T. Huai, S.D. Shah, W.J. Millera, T.Y. Loved, D.J. Chernichb, A. Ayalab, Analysis of heavy-duty diesel truck activity and emissions data. *Atmos. Environ.* **40**, 2333–2344 (2006)
9. O.O.S. Ojo, O.S. Awokola, Investigation of air pollution from automobiles at intersections on some selected major roads in Ogbomoso. *Int. Organ.Sci. Res. J. Mech. Civil Eng.* **1**, 31–35 (2012)
10. L. Xia, L.M. Leslie, A GIS framework for traffic emission information system. *Meteorol. Atmos. Phys.* **87**, 153–160 (2004)

Investigation of Fault Detection Techniques for an Industrial Pneumatic Actuator Using Neural Network: DAMADICS Case Study

V. Elakkiya, K. Ram Kumar, V. Gomathi and S. Rakesh Kumar

Abstract The objective of this work was to develop the novel approach for fault detection using neural network in industrial actuator for DAMADICS benchmark case. This neural network model has the ability to produce effective result for fault detection. In this paper, a model-based technique is proposed for the residual generation which results from the deviation of fault-free behavior of actuator from faulty behavior on actuator. The actuator is the multi-input–multi-output (MIMO) system which is designed using four kinds of neural network architectures (NNARX, NNARMAX, NNRARX, and feed forward), and best structure is chosen based on performance indices. The actuator faults can be grouped using k-means clustering technique. This technique is applied to the DAMADICS benchmark case.

Keywords Fault detection · Neural network · k-means · DAMADICS

1 Introduction

Control systems are implemented everywhere in the industry. They are used to control various types of chemical processes in industries. The supervision of chemical processes and the quality control of products are mainly based on monitoring the present state indicating any undesired states present in the process, and

V. Elakkiya (✉) · K. Ram Kumar · V. Gomathi · S. Rakesh Kumar
Department of Electronics and Instrumentation Engineering, SASTRA University,
Tirumalaisamudram, Thanjavur, India
e-mail: elakkiyatamil56@gmail.com

K. Ram Kumar
e-mail: ramkumar@eie.sastra.edu

V. Gomathi
e-mail: gomathi@eie.sastra.edu

S. Rakesh Kumar
e-mail: srakesh@eie.sastra.edu

taking appropriate actions to avoid damage or accidents. The deviations from normal process behavior result between faults and errors. If faults could timely be detected and diagnosed in many cases, it is possible to concurrently reconfigure the control system so that it can safely continue its operation [1].

Fault diagnosis is performed using these three categories: model-based method, model-free method, and process history-based method. In model-based method, parameter estimation, state estimation, observer design, and soft computing technique are the most applied techniques. The model-based fault detection and isolation (FDI) is based on state space model of the process. If the process is linear process, then it is quite easy to find the linear state space model; if not linear, then it is somewhat difficult to get the nonlinear state space model. Most of the research works in nonlinear FDI focused on the development of models, which can accurately approximate the dynamics of the nonlinear system, to be used for generating residuals. Many techniques were developed based on universal approximators, such as FSs, NNs, and neuro-fuzzy networks (NFN) [2]. Based on the accurately generated residuals, faults can then be isolated by classifiers derived from statistics, fuzzy logic, NN, and NFN [3, 4]. Since the actuator system is highly nonlinear and many parameters are involved in this valve mathematical modeling, it is difficult to find the accurate model for valve. To avoid this kind of problem, artificial neural network (ANN)-based system identification procedure is developed [5]. The neural network model is developed for fault-free data.

The residual generation via neural network is implemented by comparing the output of the fault-free model with faulty model. This paper focuses on finding the best method for generating residuals among these neural network models and diagnoses the faults into groups by k-means clustering method. This scheme is designed and applied to the benchmark DAMADICS.

2 Model-Based Fault Detection Method

The model based fault diagnosis can be defined as the identification of the faults present in a system. It consists of two stages. The generation of residues and decision-making. The difference between the system output and fault-free model output is called residual. Based on this residue value, the fault is detected. The block diagram of FDI is shown below in Fig. 1 [6]. The problem that will arise when using a mathematical model for the given system is that it cannot model the monitoring

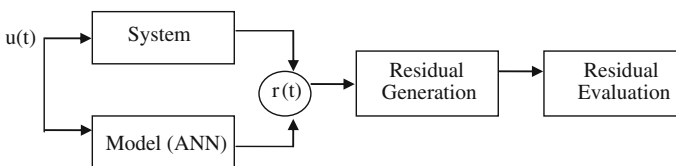


Fig. 1 General block diagram of FDI

system accurately due to disturbances and noise. This results in differences between the analytical model output and the system output due to non-modeled dynamics of the system and other uncertainties. Faults are detected by setting fixed or variable thresholds on residual signals generated from the difference between actual measurements and their estimates obtained by using the process model.

2.1 Residual Generation Using Neural Network Model

The basic idea in fault detection system consists of two stages to generate signals that reflect residual between the normal operating condition and faulty operating conditions. That kind of signals are known as residuals and are usually calculated using a variety of analytical methods [7]. Another stage is making the decision. The generation of residuals is done using various types of neural network; among this, best network is chosen based on the performance indices. Neural network is able to detect the faults present in dynamic system and able to model the multi-input–multi-output system [8].

Artificial neural networks are widely used for developing the data-based model. The objective is to design fault-free and faulty models that will be used for the residual generation [9, 10]. ANNs are used to differentiate various faults from the normal condition to abnormal condition, according to different fault patterns presented in the measured input–output system data. ANN is trained with data collected during the normal functioning (fault free); then, the ANN models are validated with another set of data.

The modeling equation of valve using neural network is shown below

$$\begin{aligned} X' &= \text{net } X(\text{CV}, \text{P1}, \text{P2}, \text{T}), \\ F' &= \text{net } F(X, \text{P1}, \text{P2}, \text{T}). \end{aligned}$$

Identify a neural network model of a dynamic system by using network architecture. The toolbox provides different model structures as below.

Neural network autoregressive with external input (NNARX)

$$\hat{y}(t-1) = y(t-1) \dots y(t-n)u(t-p) \dots u(t-q-p+1). \quad (27.1)$$

n, q number of past inputs and outputs used for determining the prediction
 p Time delay

Neural network recursive autoregressive with external input (NNRARX)

Identify a neural network model of a dynamic system by using a recursive algorithm.

Neural network autoregressive moving average with external input (NNARMAX1)

$$\hat{y}(t - 1) = y(t - 1) \dots y(t - n) \dots u(t - p) \dots u(t - q - p + 1) + c(q - 1)e(t). \tag{27.2}$$

In order to get the best architecture, performance measure of each network is evaluated based on sum of squared error.

3 Case Study (DAMADICS Actuator)

This procedure is studied for DAMADICS actuator which is the benchmark system mainly developed to analyze the various kinds of fault detection and isolation schemes. This DAMADICS benchmark is based on the electropneumatic valve actuator in the Lublin sugar factory in Poland [11]. The actuator consists of a control valve, pneumatic servomotor, and positioner shown in Fig. 2. Totally, five available measurements and 1 control value signal have been considered for

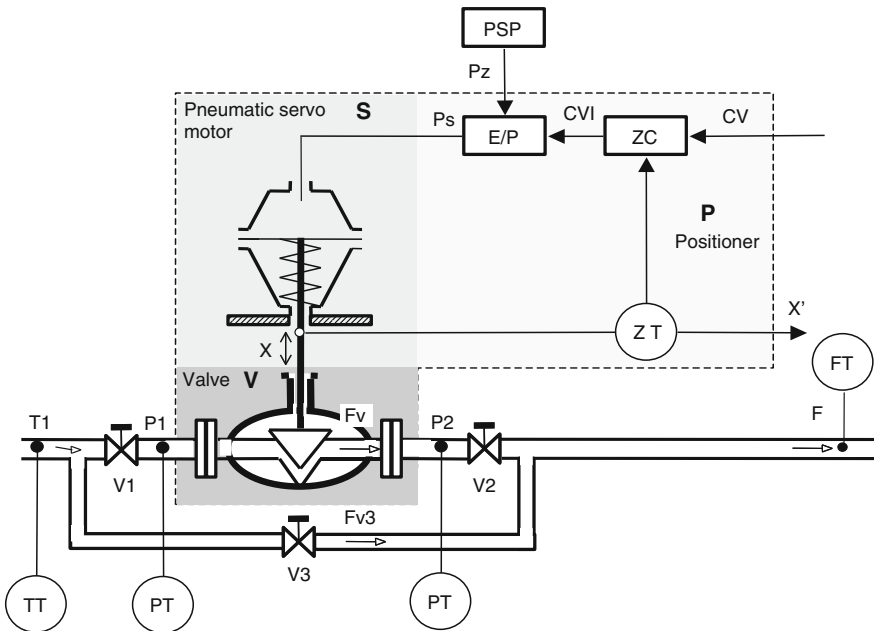


Fig. 2 Actuator structure, V1 hand-driven bypass valve, V2 hand-driven bypass valve, V3 hand-driven bypass valve, V control valve, P1 pressure sensor (valve inlet), P2 pressure sensor (valve outlet), F process media flow meter, X valve displacement, x positioner feedback signal, p, pneumatic servomotor supply, and CV control signal from external PI Controller

benchmarking purposes. The input to the actuator is as follows: process control external signal CV, values of liquid pressure on the valve inlet P1 and outlet P2, and liquid temperature T1. The output of the actuator is liquid flow rate F and stem displacement X. The data set is collected from the DAMADICS Web site during various operating conditions which means with fault and fault-free operating conditions [12]. The actuator consists of 4 inputs and 2 outputs.

3.1 Residual Design

The residual generation can be done by comparing the neural network output (fault free) with actual system measurement [13]. In this work, the residual generation was done by using various kinds of neural network [14]. The neural network model can be represented as multi-input–single output (MISO) for pneumatic actuator.

Using “Neural Network Based System Identification Toolbox” (NNSYSID), various structures like NNARX, NNRARX, and NNARMAX are designed and they are shown in Figs. 3, 4, 5 and 6. The network is designed with tansig transfer function for hidden layer and purelin for output layer. The network is trained with data collected during normal operation and validated with another set of data. The model which is designed using neural network is able to map the actuator system output. So this model will behave like the system (actuator).

The best architecture performance measure of each network is evaluated based on sum of squared error. The performance measure is shown in Table 1.

Fig. 3 Output of the model and system

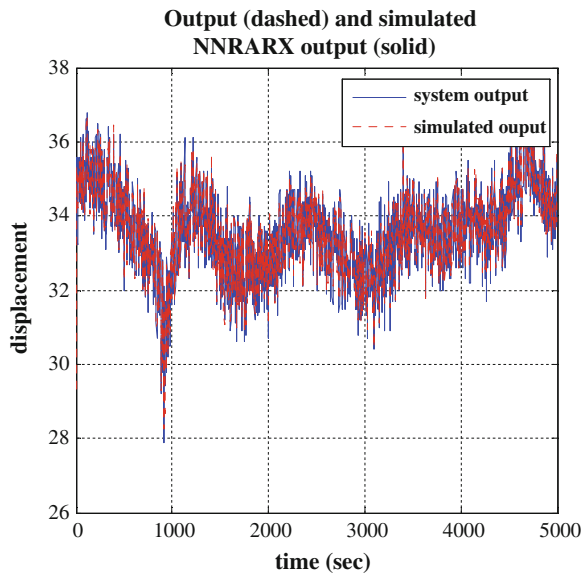


Fig. 4 Output of the model and system

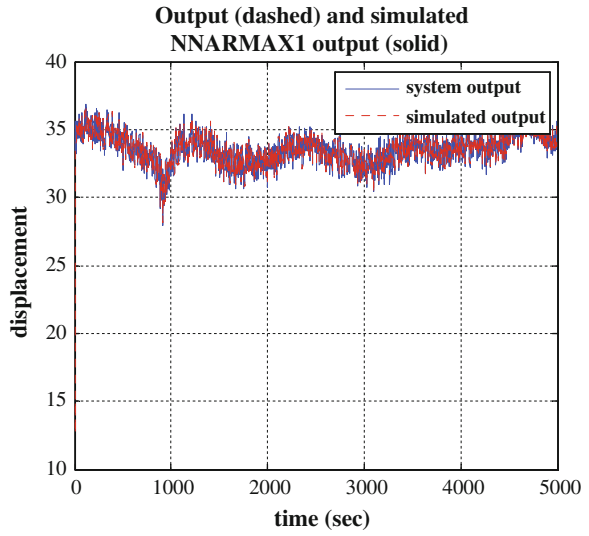


Fig. 5 Output of the model and system

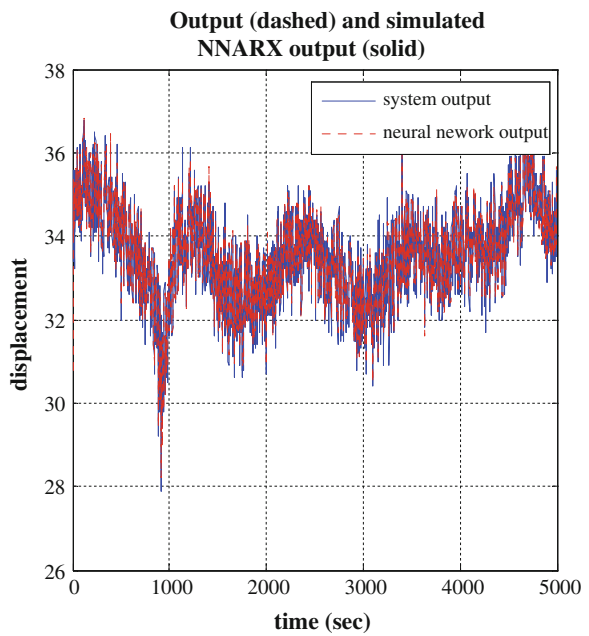


Fig. 6 Output of the model and system (after training)

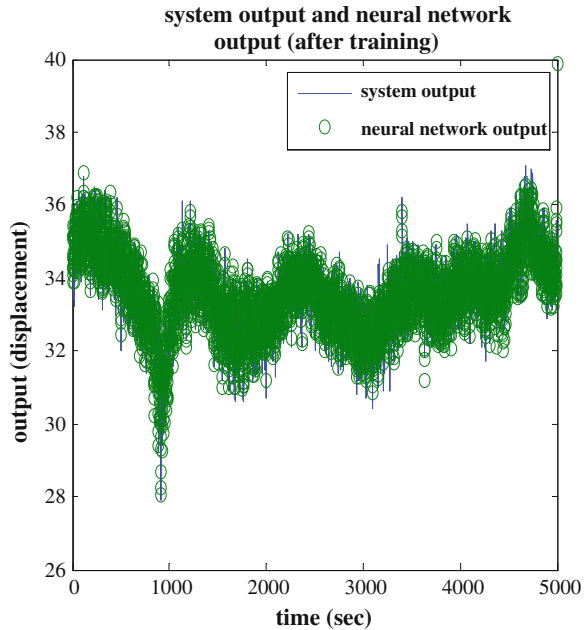


Table 1 Validation of different models using ANN

Network type	Performance measure (sum of squared error)
Feed-forward type	2.604
NNARX	0.5571
NNRARX	0.6328
NNARMAX1	16.64

4 Actuator Fault Detection and Isolation

In the proposed FDI system, 19 classes of system behaviors, normal operating condition f_0 , and nineteen faults f_1 – f_{19} are modeled by a bank of dynamic neural networks. But the result is shown only for partially opened bypass valve (f_{19}). To evaluate residuals and to obtain information about faults, simple thresholding can be applied. If residuals are smaller than the threshold value, a process is considered to be healthy; otherwise, it is faulty. In practice, due to modeling uncertainty and measurement noise, it is necessary to assign thresholds larger than zero in order to avoid false alarms [5]. The fault (f_{19}) is detected by setting the fixed threshold value (1.5) as shown in Fig. 7. Since the isolation task can be performed using some statistical classification and fault signature matrix, but the isolation is not effective, it is not possible to segregate the type of fault based on magnitude because some

Fig. 7 Residual between system output and neural network output for f19

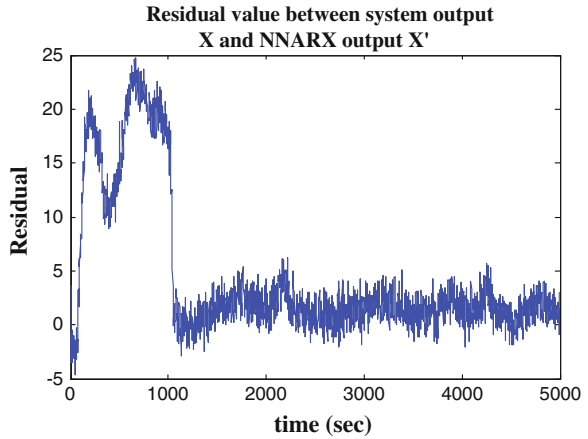


Table 2 Types of faults

Fault	Description
<i>Control valve faults</i>	
f1	Valve clogging
f2	Valve or valve seat sedimentation
f3	Valve or valve seat erosion
f4	Bushing friction
f5	External leakage
f6	Internal leakage (valve tightness)
f7	Medium cavity or critical flow
<i>Pneumatic servomotor faults</i>	
f8	Twisted servomotor's rod
f9	Terminals tightness
f10	Servomotor's diaphragm perforation
f11	Servomotor's spring fault
<i>Positioner faults</i>	
f12	Electropneumatic transducer
f13	Rod displacement sensor fault
f14	Pressure sensor fault
f15	Positioner spring fault
f16	Positioner lever fault
f17	Positioner supply pressure drop
f18	Unexpected change of pressure difference
f19	Fully or partly opened bypass valves

faults are having same magnitude (f1, f7, f10). So all the faults are grouped into two based on k-means clustering technique using squared Euclidean distance. The list of faults is shown in Table 2.

4.1 Cluster Analysis

The aim of cluster analysis is the classification of objects according to similarities. Clustering techniques are among the unsupervised (learning) methods, since they do not use prior class identifiers. Most clustering algorithms also do not rely on assumptions common to conventional statistical methods, such as the underlying statistical distribution of data, and therefore, they are useful in situations where little prior knowledge exists.

The most well-known clustering algorithm is the so-called k -means algorithm or Lloyd's method which attempts to address the following objective [15]: Given a set of points in a Euclidean space and a positive integer k (the number of clusters), split the points into k clusters so that the total sum of the (squared Euclidean) distances of each point to its nearest cluster center is minimized. The clustering task can be done based on the error value, sum of squared error, and mean squared error. The faults are clustered into two groups. The group 1 consists of f1, f7, f10, f12, f15. The remaining faults will fall in group 2 (f2, f3, f4, f5, f6, f8, f9, f11, f13, f14, f16, f17, f18, f19).

5 Conclusion

Diagnostics of industrial processes has been extensively developed in recent years. We can conclude that by using artificial neural networks composed of dynamic neurons, one can design an effective fault diagnosis system. A group of neural models are used to model for normal operation conditions. The faults are grouped into two types based on k -means clustering technique. The limitation of this model-based FDI is we can use this for only known working condition. The unknown faults can be detected but not isolated. Another limitation is the faults cannot be isolated using the magnitude, and ability to detect and isolate multiple faults is very difficult.

Future activities include addressing issues related to isolation of each faults and to design fault-tolerant control system.

References

1. C.D. Bocaniala, Sa da Costa.: application of a novel fuzzy classifier to fault detection and isolation of DAMADICS benchmark problem. *Control Eng. Pract.* **14**(6), 653–669 (2006)
2. J.J. Gertler, *Fault Detection and Diagnosis in Engineering Systems* (Marcel Dekker Inc, New York, 1998)
3. Y. Kourd, Guersi, N., Lefebvre, D.: Neuro-fuzzy approach for fault diagnosis: application to the DAMADICS, in *Proceedings of ICINCO 2010* (2010)
4. R. Sundarmahesh, B. Kannapiran, Fault diagnosis of pneumatic valve with DAMADICS simulator using ANN based classifier approach. *Int. J. Comput. Appl.* 11–17 (2013)

5. K. Prabhakaran, U. Mageshwari, D. Prakash, A. Suguna, Fault diagnosis in process control valve using artificial neural network. *Int. J. Innov. Appl. Stud.* **3**(1), 138–144 (2013)
6. J. Chen, R.J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems* (Kluwer Academic Publishers, Dordrecht, 1999)
7. S. Haykin, *Neural networks. A comprehensive foundation* (1999)
8. Y. Kourad, D. Lefebvre, N. Guersi, Early FDI based on residuals design according to the analysis of models of faults: application to DAMADIC. *Hindawi Publishing Corporation Advances in Artificial Neural* (2011), pp. 1–10
9. R. Isermann, Process fault detection based on modeling and estimation method-a survey. *Automatica* **20**, 387–404 (1984)
10. Y. Kourad, D. Lefebvre, N. Guersi.: Fault diagnosis based on neural networks and decision trees: application to Samadics. *Int. J. Innov. Comput. Inf. Control* **9**(8) (2013)
11. M. Bartys, R. Patton, M. Syfert, S. De las Heras, J. Quevedo, Introduction to the DAMADICS actuator FDI benchmark study. *Control Eng. Pract.* **14**, 577–596 (2006)
12. M. Syfert, M. Bartys, J. Quevedo, Benchmark definition. <http://diag.mchtr.pw.edu.pl/damadics> (2002)
13. R. Isermann, *Fault Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance* (Springer, New York, 2006)
14. H. Demuth, M. Beale, *Neural network toolbox for use with MATLAB. User's Guide Version 3*
15. C. Boutsidis, M.W. Mahoney, Unsupervised Feature Selection for the k-means Clustering Problem

Design of Portable Security System Using Face Recognition with Back-Propagation Algorithm and MATLAB

M. Hareesh Babu, M. Bala Naga Bhushanamu, B. Benarji
and M. Purnachandra Rao

Abstract In our globally connected world, threats from various aspects are going at an alarming rate. These are controlling with different security systems such as metal detector, closed circuit cameras, and scanning systems. All these aids are meant to recognize and identify the explosives and others weapons. Here, it is more important to identify the particular person or persons, who were planning to distract the society or particular event. This paper is aimed to design that to control the threats by identifying the suspected people by a simple face recognition technique using simple PC or laptop with the help of scientific software MATLAB and its neural network tool box. In general, all major events are fully securitized with well-developed protection systems but only problem with non-major and small events, where security systems are matter of financial issues. So, militants and other destroyers are taking advantage of these situations and creating a panic and terror situations. This paper is also designed like that a PC or laptop with camera can be a face recognition system to identify the suspected peoples and most wanted criminal. By recognizing the people, we can mostly avoid the threats from these people and dangerous situations. Neural network is a science that has been extensively applied to numerous pattern recognition problems such as character recognition, object recognition, and face recognition, where this paper has programmed for face recognition with the back-propagation algorithm and simulated with the software MATLAB and its neural network tool box. Here, the back propagation plays the central operation role to get the key features were extracted from the picture for training the network. Since the major role of the project is mainly focusing on the

M. Hareesh Babu (✉) · M. Bala Naga Bhushanamu · M. Purnachandra Rao
Department of Systems Design, Andhra University, Visakhapatnam, India
e-mail: hareesh.makesu@gmail.com

M. Bala Naga Bhushanamu
e-mail: balanagabhushanamu@gmail.com

M. Purnachandra Rao
e-mail: raomp17@gmail.com

B. Benarji
Department of Electronics and Communication, Andhra University, Visakhapatnam, India
e-mail: benarjiec@gmail.com

training of the neural network, already extracted key features of the person's image from the database were taken for training the back-propagation network. Here, we have taken 7 input units, 6 hidden units, and 4 output units contained back-propagation network. The output unit, 4 output units, generates the 4-bit output which gives the person identity.

Keywords Security · Back-propagation · Face recognition · Neural network · Images · MATLAB · Data learning rate · Weight · Train

1 Introduction

In this globalized world, security of a person or country is a major issue. An annual budget for the country and their people security has raising drastically by every nation. Even though the security system in root levels are failing at small and non-recognitions events, which leads to financial and death loses. So, security at these events need to be enhance to control the various loses. All these situations can be overcome by utilizing the technology such as Artificial Neural Network, Face recognition, Digital Image Processing, and MATLAB software. The main aim of the paper was to identify the persons from the face recognition technique and back-propagation algorithm, which is a part of neural network. In this paper, we have designed as that will collect the still photograph of the person, especially face part of the of the person, by the personal computer or laptop's built-in camera and then fed to the MATLAB software, where captured image has processed according to back-propagation algorithm using neural network tool box. This image is stored in the database and also compared with the previous images from database system, which has taken previously or collected from internet database system. If images found to be matched, then the system immediately intimating that particular person has already listed or identified in this database. Here, the data base system will readily stores the suspected people and criminals images. In this manner, we can identify the criminals, militant, and suspected persons at a small and non-recognized events at rural areas by simple using personal computer and laptops.

MATLAB is a high-level language and interactive environment for computations, visualizations, and programming. Meanwhile, you have a chance to analyze data, develop algorithms, and create models and applications. This language tool and built-in mathematical functions will enable you to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C or C++ or VC++. Main Key features of the MATLAB are as follows: (1) Mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, numerical integration, and solving ordinary differential equations. (2) Built-in graphics for visualizing data and tools for creating custom plots. (3) Development tools for improving code quality and maintainability and maximizing performance. (4) Tools for building applications with custom graphical

interfaces. (5) Functions for integrating MATLAB-based algorithms with external applications and languages such as C, Java, .NET, and Microsoft Excel. (6) Creating Apps with graphical user interfaces in MATLAB.

Back-propagation is a common method of training artificial neural networks and an abbreviated as “backward propagation of errors.” From a desired output, the network learns from many inputs, similar to the way a child learns to identify a dog from examples of dogs. It is a supervised learning method and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks. Back-propagation requires that the activation function used by the artificial neurons or “nodes” be differentiable. The back-propagation learning algorithm can be classified into two phases: propagation and weight update. Each propagation involves the following steps: (1) Forward propagation of a training pattern’s input through the neural network in order to generate the propagation’s output activations. (2) Backward propagation of the propagation’s output activations through the neural network using the training pattern target in order to generate the deltas of all output and hidden neurons. For each weight-synapse follow the following steps: (1) Multiply its output delta and input activation to get the gradient of the weight. (2) Subtract a ratio (percentage) of the gradient from the weight. This ratio (percentage) influences the speed and quality of learning; it is called the *learning rate*. The greater the ratio, the faster the neuron trains; the lower the ratio, the more accurate the training is. The sign of the gradient of a weight indicates where the error is increasing; this is why the weight must be updated in the opposite direction. Phase 1 and 2 repeats until the performance of the network is satisfactory.

2 Design Technique

The face recognition task is a challenging because of variability in the pose, orientation, location, and scale. It is also depends on the facial expressions, lighting conditions, and age excreta. In this paper, we have mostly considered about the three conditions as follows: (1) Variations in image plane and pose (2) Facial expressions, and (3) Lighting conditions and background. Section one due to the faces in the image vary due to rotation, translation, and scaling of the camera-pose or the face itself. Second sections due to the appearance of a face are largely affected by the expression on the face. The presence or absence of additional features such as glasses, breads, and mustaches further add to this variability. The final section describes about the lighting conditions that depend on the object and light source properties, and affect the appearance of the face. The back ground which defines the profile of the face is important and cannot be ignored. The whole process involves training the network with a set of pictures and after the training process is over, an arbitrary image is given as input. The output is a set of binary values of yes or no, indicating whether the image resembles any trained picture or not.

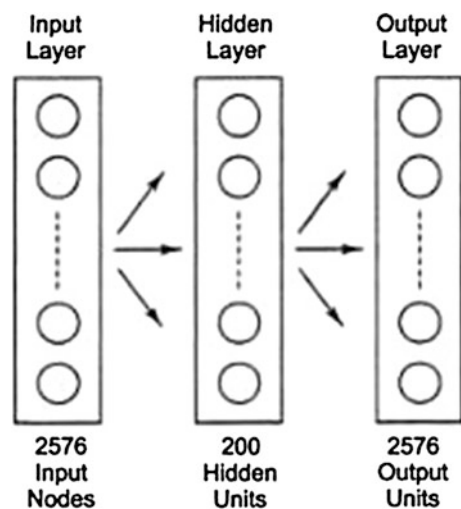
3 Training Approach

The training approach has taken into three stages. In first stage, the back-propagation network is used to extract various unique features from every image, and learning vector quantization is used for training the network. Here, the back-propagation network is also used as auto-associative network and each image is resized to 56×46 pixels then fed to back-propagation network as an input, which leads to be 2,576 input units, 2,576 output units, and 200 hidden units. Further, each pixel value is converted into a bipolar to accept the values as '-1' and if the pixel value is greater than 0.5 as '+1' (Figs. 1 and 2).

In second stage, the back-propagation network is used as a string network to act as a database system with a 2,576 input units, maximum of 200 hidden units, and 6 output units. The output is converted from binary to hexadecimal to address the memory location of the database, where the information about a person is stored. This 6-bit output gives the person identity (Fig. 3).

In final stage, the key features of the image were only extracted for the training the network because if we give the picture as input to the network is not feasible due to the size of the weight matrix is huge. To train the network, we have extracted key features from the database for train the back-propagation network with 7 input units, 6 hidden units, and 4 output units has used. The seven attributes of the image features are 1. Frontal or central, 2. Upper frontal, 3. Lower frontal, 4. Top right frontal, 5. Top left frontal, 6. Bottom right frontal, and 7. Bottom left frontal. So, the dimensions of the image are considerably reduced to 7-6-4 from 2576-50-6. Finally, the 6-bit output units of the image converted into 4-bit output, which gives the person identity. Here also the output is converted from binary to hexadecimal to address the memory location where the information of a person has stored.

Fig. 1 BPN architecture



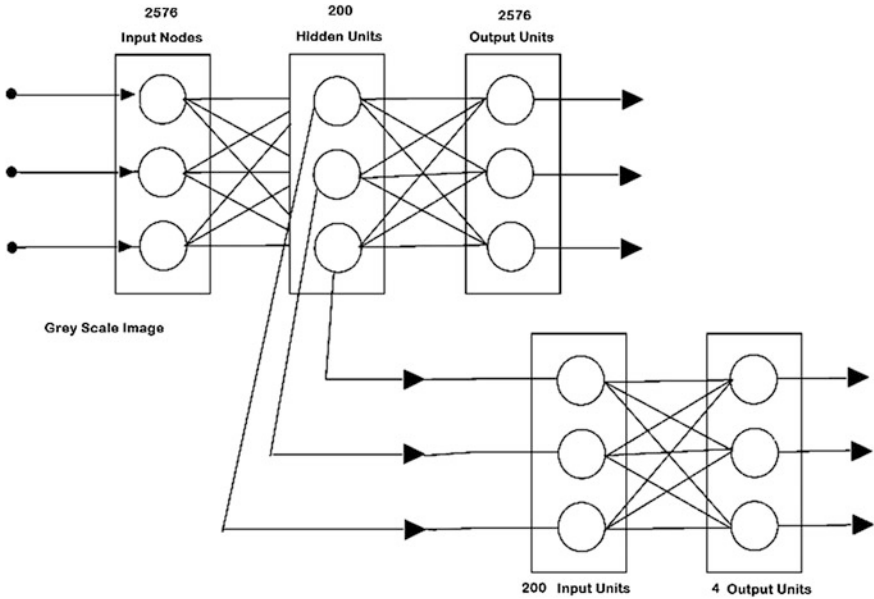
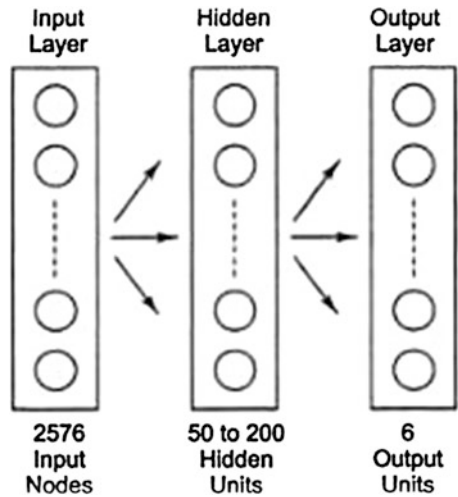


Fig. 2 Hybrid BPN network

Fig. 3 BPN network for process the original image



To analyze the efficiency of the back-propagation network for varying inputs to identify the correct person, we trained the network in both continuous valued and discrete value features (Figs. 4 and 5).

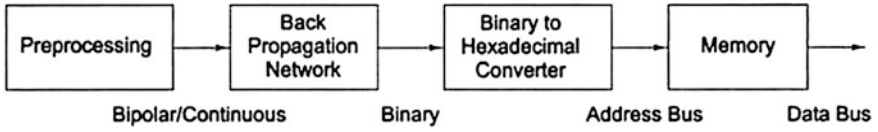


Fig. 4 Block diagram for the entitled paper

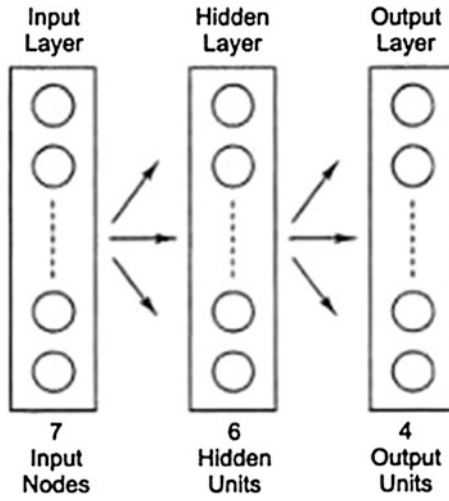


Fig. 5 BPN network after process the original image



Fig. 6 A set of training images

4 Simulation Result

See Figs. 6, 7, 8, 9 and 10.

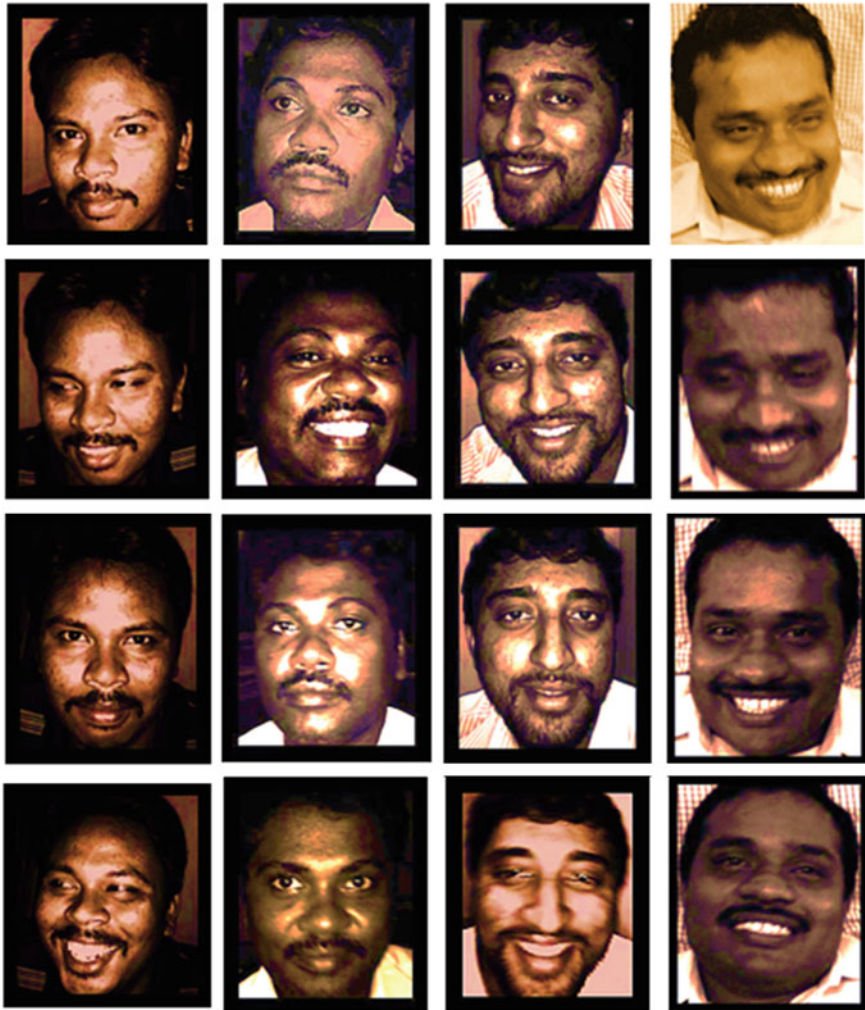


Fig. 7 A set of testing images

5 Conclusion

Advantages in technology such as face recognition and Image Processing has leads to overcome many sensitive problems in security aspects such as identifying the suspected people and enemies by capturing the image and identify the validity of a person by using the extracted features available in the database. For practical implementation, more attention should be given to the preprocessing stage to extract the features directly from the image. Finally, the intention of the paper

Frontal	Upper Frontal	Lower Frontal	Top Left Frontal	Top Right Frontal	Bottom Left Frontal	Bottom Right Frontal
0.5800	0.4500	0.1850	0.9955	0.3945	0.2720	0.2850
0.5600	0.4400	0.1400	0.9285	0.3825	0.1880	0.3000
0.6150	0.4800	0.1650	0.8385	0.5130	0.3010	0.3050
0.5500	0.4150	0.1350	0.7635	0.3180	0.2100	0.2000
0.5650	0.4400	0.1550	0.9395	0.4275	0.2140	0.2700
0.3800	0.2750	0.1000	0.2255	0.0800	0.0490	0.0850
0.3550	0.2800	0.0950	0.2455	0.0955	0.0620	0.0750
0.4500	0.3200	0.1000	0.3810	0.1705	0.0750	0.1150
0.3650	0.2950	0.0800	0.2555	0.0970	0.0430	0.1000
0.4400	0.3400	0.1000	0.4510	0.1880	0.0870	0.1300
0.3550	0.2800	0.0850	0.2905	0.0950	0.0395	0.1150
0.5000	0.4000	0.1300	0.6645	0.2580	0.1330	0.2400
0.4700	0.3550	0.1000	0.4755	0.1675	0.0805	0.1850
0.5350	0.4050	0.1450	0.6845	0.2725	0.1710	0.2050
0.4900	0.3800	0.1350	0.5415	0.2175	0.0950	0.1900

Fig. 8 Sample dataset for continuous input

Frontal	Upper Frontal	Lower Frontal	Top Left Frontal	Top Right Frontal	Bottom Left Frontal	Bottom Right Frontal
1	1	-1	1	1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
1	-1	-1	1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1

Fig. 9 Sample dataset for discrete input

“Design of Portable Security System Using Face Recognition with Back-propagation Algorithm and MATLAB” has simulated in PC and laptop for very low cost and advantage is that it can mobilized to anywhere.

Fig. 10 Binary data for persons' identity

1	1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
1	1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1

Acknowledgments We thank to our beloved Guide and Head of the Department Prof. M. Purnachanra Rao, who gave us such an opportunity to work on NEURAL NETWORK, ARTIFICIAL INTELLIGENCE, and MATLAB SOFTWARE and his assistance on our presentation.

References

1. J.R. Anderon, *Cognitive psychology and its implications*, 2nd edn. (Freeman, San Francisco, 1985)
2. D.H. Ballard, Parameter nets. *Artif. Intell* **22**(3), 235–267 (1984)
3. E.A. Feigenbaum, A. Barr, P.R. Cohen (eds) *The Handbook of Artificial Intelligence* (Addison-Wesley, New York, 1989)
4. M.A. Eshera, K.S. Fu, A graph distance measure for image analysis. *IEEE Trans. syst. Man Cybern. SMC* **14**(3)
5. K.S. Fu, Sequential methods in pattern recognition and matching of pictorial structures. *IEEE Trans. Comput* (1968)
6. J.L. Kolodener, Reconstructive memory: a computer model. *Cogn. Sci.* **7**(4), 281–328 (1983)

Model-Based Control for Moisture in Paper Making Process

C. Karthik, K. Suresh, K. Valarmathi and R. Jacob Rajesh

Abstract This project deals with the performance evaluation on the comparison of model-based control for drying process of paper industry. The dryer section is the last part of the paper machine and consists of a large number of rotating steam-heated cast iron cylinders by adjusting the set point of the stream pressure controller to the cylinders. In the design of model reference adaptive control, schema is used, in which the adaptive law has been developed by MIT rule. Similarly, design of PID and MRAC controller is used. This paper presents a nonlinear dynamic control, based on heat and mass balance for steam, cylinder, and paper. The control was performed to the combined drying process system using both the adaptive control algorithm and MPC controller method and its results were analyzed. A simulation is carried out using MATLAB. Simulation results reveal clear benefits of the model reference adaptive control over traditional controller and MPC controller methods. Thus, by controlling, this process proves real incentives for industrial implementation.

Keywords Drying section · System identification · PID · MPC · MRAC

1 Introduction

The function of a paper machine is to form the paper sheet and remove the water from the sheet. A paper machine is divided into three main parts as wire section, press section, and drying section. Nonlinear modeling of moisture control of drying

C. Karthik (✉) · K. Suresh · R. Jacob Rajesh
Kalasalingam University, Krishnankoil 626126, Tamil Nadu, India
e-mail: karthikmtech86@gmail.com

K. Suresh
e-mail: suresharulraj@gmail.com

K. Valarmathi
P.S.R. Engineering College, Sivakasi 626140, Tamil Nadu, India

process in paper machine has been proposed an approach to define that the paper machine is modeled for designing moisture content control loop using DCS which is available in the paper plant. A transfer function to validate the moisture control process is obtained with the real-time data [1]. MPC as control strategy for pasta drying processes has been proposed an approach to MPC that produces high performance and accuracy, with relatively small computational rate and gives better results than PID [2, 3]. Model predictive control of an industrial dryer has been proposed an approach to its high performance due to the use of the direct control of the product moisture content based on a state observer, to updating the model of the process on which MPC relies [4, 5]. Direct model reference adaptive control of linear systems with input/output delays has been proposed an approach to Direct Model Reference Adaptive Tracking Controller for linear systems with unknown time varying input delays [8]. MRAC using observers with unknown inputs has been proposed a new solution for MRAC, based on the design of a state observer with unknown inputs has been proposed [8].

A multivariable MRAC scheme with sensor uncertainty compensation has been proposed a crucial step, the derivation of a properly parameterized error model in terms of the system and sensor parameter errors and the output tracking errors. Based on the developed error model, stable adaptive laws have been derived for updating the parameter of the compensator and feedback controller [10].

2 System Identification

System identification is a procedure to build a mathematical model of dynamics of a system from measured data. System identification is a process of obtaining models based on a data set collected from experimental setup as well as the real-time models.

2.1 Identification

The design of a control system requires a mathematical model of the dynamics of the process. Different types of identification model structure based on early principles model parameters are estimated from measured data. If the physical laws governing the behavior of the system are known, we can use these to construct so-called *white-box* models of the system. In a white-box model, all parameters and variables can be interpreted in terms of physical entities and all constants are known a priori. At the other end of the modeling scale, we have so-called *black-box* model or identification.

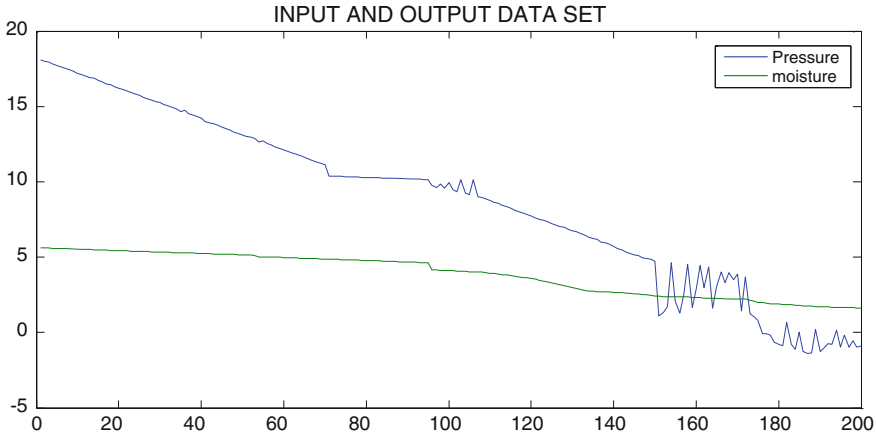


Fig. 1 Data set of input and output

2.1.1 Experiment Design

Collecting data is a very essential step. The data set Z^N should be as informative as possible to fully identify the model. Their pressure and moisture data were collected by using ABB DCS in TNPL. A total of 2,000 data were collected that are shown in Fig. 1.

Model sets or model structures are families of models with adjustable parameters. Parameter estimation amounts to conclude the “best” values of these parameters. The system identification complication amounts to find both the good models. Model Validation is the process of gaining confidence in a model. Crucial this is achieved by “twisting and turning” the model to scrutinize all attitude of it. Of particular importance is the model’s ability to reproduce the behavior of the validation data sets. Thus, it is important to review the properties of the residuals from the model when applied to the validation data.

3 PID Controller

A PID controller calculates an “error” value as the difference between a measured process variable and a desired set point. The controller experiments to minimize the error by adjusting the process control inputs. Be able to use common methods of analysis for a system with a PID controller in order to predict the behavior of the system and controller, and to be able to choose PID parameters. Defining $u(t)$ as the controller output, the final model of the PID algorithm is

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + k_d \frac{d}{dt} e(t) \quad (1)$$

3.1 Ziegler–Nichols Tuning

This procedure is only valid for open loop stable plants, and it is carried out through the following steps. Set the true plant under proportional control, with a very small gain, and increase the gain until the loop starts oscillating. Note that linear oscillation is needed and it should be detected at the controller output. Record the controller critical gain $K_p = K_c$ and the oscillation period of the controller output P_c .

4 Model Predictive Control

Future values of output variables are predicted using a dynamic model of the process. The control calculations are based on both future predictions and current measurement. Inequality, equality constraints, and measured disturbances obtain including the control calculations. The calculated manipulated variables obtain implemented set point for lower level control loops. A discrete-time implementation of model-based control algorithm is called as model predictive control.

4.1 MPC Design

The first step in the design is to load a plant model. Its dimensions and signal specifically set the context for the remaining steps. The model can be loaded directly or indirectly by importing a controller or a saved design. To import from MATLAB workspace, radio button should be selected by default. The dialog section labeled in the workspace lists the LTI models. They select the state space model for the process. The dialog section labeled the properties and then displays the number of input and output—their names, signal types, etc (Fig. 2).

5 Model Reference Adaptive Control

The MRAC is one of the main adaptive control approaches. When the system specifications are in terms of a reference model, it tells how the process output should ideally respond to command signals. It is then possible to use MRAC. Model reference adaptive system is to create a closed loop controller with parameters that can be updated to change the response of the system. The output of

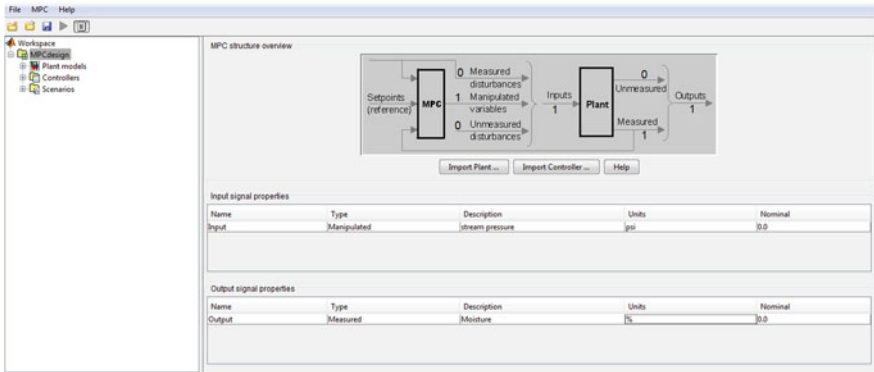


Fig. 2 MPC control and estimation tools manager

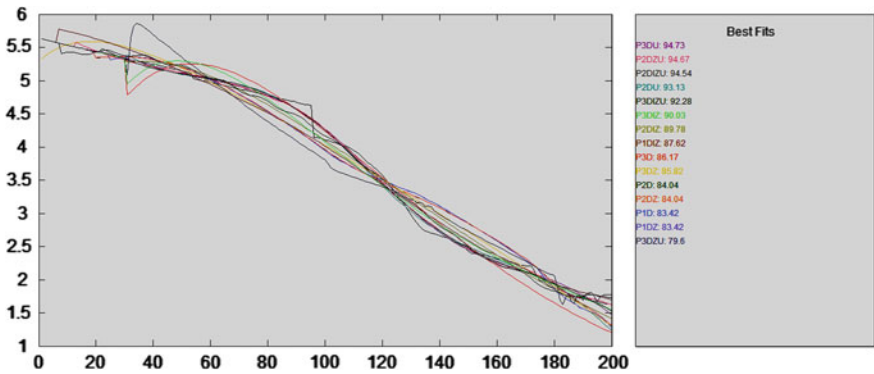


Fig. 3 Process model validation fitness output

the system is compared toward a desired response from a reference model. The control parameters will update based on this error. By adjusting, the mechanism parameters in a model reference adaptive system can be obtained using gradient method (Figs. 3, 4, 5 and 6) (Tables 1 and 2).

5.1 Gradient Method—MIT Rule

$$\text{Capture error : } e = y_{\text{plant output}} - y_{\text{model output}} \tag{2}$$

$$\text{Regarding cost function : } J(\theta) = \frac{1}{2} e^2(\theta) \tag{3}$$

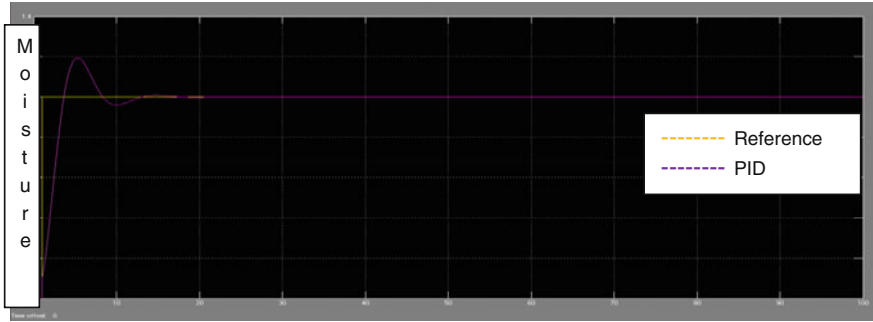


Fig. 4 PID simulation scenario

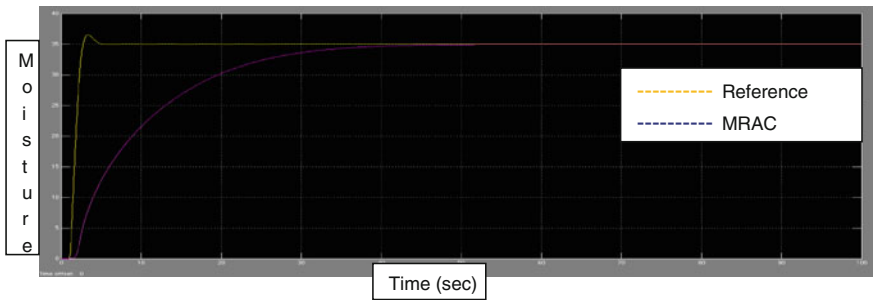


Fig. 5 MRAC simulation scenario

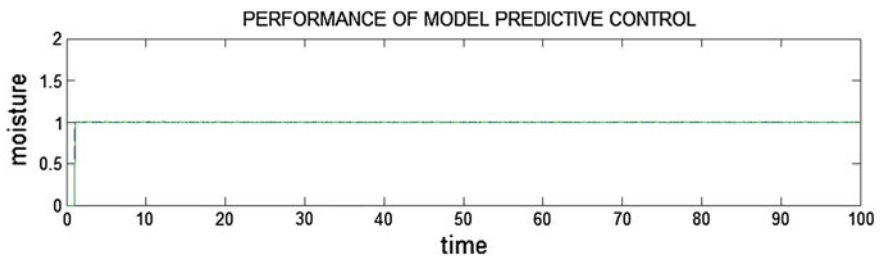


Fig. 6 MPC simulation scenario

$$\text{MIT standard : } \frac{d\theta}{dx} = -\gamma \frac{\delta J}{\delta \theta} = -\gamma e \frac{\delta e}{\delta \theta} \tag{4}$$

$$\frac{d\theta 1}{dt} = -\gamma(a_{muc}|s^3 + a_m)e \tag{5}$$

Table 1 Comparison of process model

Models	MSE for real-time data	Fitness (%)
PID	0.0508	83.42
P2D	0.0471	84.04
P3D	0.0353	86.17
P2DU	0.00872	93.13
P3DU	0.00512	94.74
P1DZ	0.0508	83.42
P2DZ	0.0471	84.04
P3DZ	0.0371	85.82
P2DZU	0.0471	86.48
P3DZU	0.0371	87.45
P1DIZ	0.0285	87.62
P2DIZ	0.0194	89.76
P3DIZ	0.0184	90.03
P2D1ZU	0.00551	94.54
P3DIZU	0.011	92.28

Table 2 Comparison between PID, MPC, and MRAC controller response

Controller performance	PID	MPC	MRAC
Rise time	3 s	2.2 s	13 s
Settling time	22 s	4.6 s	44 s
Peak overshoot	1.118	0	35

$$\frac{d\theta_2}{dt} = -\gamma(a_{myplant}|s^3 + a_m)e \tag{6}$$

From $\frac{d\theta_1}{dt}$ and $\frac{d\theta_2}{dt}$ we get updating controller parameter θ_1 and θ_2 are

$$\theta_1 = \frac{-0.0001}{s} (9s^2 + 7s + 1|s^3 + 9s^2 + 7s + 1) \tag{7}$$

$$\theta_2 = \frac{0.0001}{s} (9s^2 + 7s + 1|s^3 + 9s^2 + 7s + 1) \tag{8}$$

6 Result and Discussion

From system process model identification, we get third-order transfer function for the model P3DU which is given by,

$$Tf = 1/(s^3 + 9s^2 + 7s + 1) \quad (9)$$

7 Conclusion

The proposed controllers are tested by using MATLAB simulinkprogram. The simulation shows that MPC provides better performance than MRAC and PID controller. The proposed model-based control system increases its efficiency and quality of the product. This will reduce the production cost by controlling the moisture.

References

1. C. Karthik, K. Valarmathi, M. Rajalakshmi, Non linear modeling of moisture control of drying process in paper machine. *Sci. Direct Trans. Procedia Eng.* **38**, 1104–1111 (2012)
2. J. De Temmerman, P. Dufourb, B. Nicolaia, H. Ramona, MPC as control strategy for pasta drying processes. *Sci. Direct Trans. Comput. Chem. Eng.* **33**, 50–57 (2011)
3. L. Obregon, L. Quinones, C. Velazquez, Model predictive control of a fluidized bed dryer with an inline NIR as moisture sensor. *Sci. Direct Trans. Control Eng. Pract.* **21**, 509–517 (2012)
4. V.M. Cristea, M. Baldea, P. Agachi, Model predictive control of an industrial Dryer. *Science direct symposium transactions on computer aided process engineering*, **10** (2012)
5. A. Cortinovis, M. Mercang, T. Mathur, J. Poland, M. Blaumann, Nonlinear coal mill modeling and its application to model predictive control. *Sci. Direct Trans. Dept. control Eng.* **21**, 308–320 (2013)
6. A.J. Gallego, E.F. Camacho, Adaptive state-space model predictive control of a parabolic-trough field. *Sci. Direct Trans. Dept. Control Eng.* **20**, 904–911 (2012)
7. M. Morari, U. Maede, Nonlinear offset-free model predictive control. *Sci. Direct Trans. Dept. Autom.* **48**, 2059–2067 (2012)
8. P. James, M.J. Balas, Direct model reference adaptive control of linear systems with input/output delays. *Sci. Direct Trans. Dept. Electr. Comput. Eng.* **3**, 445–462 (2013)
9. M. Duarte-Mermoud, P. La Rosa, MRAC using observers with unknown inputs. *Dept. Electr. Eng.* (2007)
10. J. Guo, G. Tao, A multivariable MRAC scheme with sensor uncertainty. *IEEE Trans. Dept. Electr. Comput. Eng.* **22904**, 6632–6637 (2009)
11. C. Karthik, M. Rajalakshmi, Nonlinear identification of pH process using NNARX model. *CIIT Int. J. Artif. Intell. Syst. Mach. Learn.* **4**(8), 502–506 (2012)
12. C. Karthik, M. Rajalakshmi, On linear structure identification of pH process, in *IEEE - ICAESM* (2012), p. 45
13. C. Karthik, M. Rajalakshmi, K. Valarmathi, Nonlinear modeling of moisture control of drying process in paper machine. *Elsevier Procedia Eng.* **38**, 1104–1111 (2012)

Elimination of Harmonics in Seven-Level Cascaded Multilevel Inverter Using Particle Swarm Optimization Technique

W. Razia Sultana, Sarat Kumar Sahoo, S. Prabhakar Karthikeyan, I. Jacob Raglend, Pasam Harsha Vardhan Reddy and Gangireddy Taraka Rajasekhar Reddy

Abstract This paper presents a robust stochastic search algorithm called particle swarm optimization (PSO) to resolve the optimum switching angles for 7-level cascaded multilevel inverter (MLI) to eliminate selected harmonics in order to get low total harmonic distortion (THD) by taking the resultant equations of the output voltage THD of an inverter as an objective function. The simulation is carried out on a 7-level cascaded inverter to ensure the accuracy of the projected technique. Results show that 5th and 7th harmonics are effectively suppressed at faster converging rate and show how the THD varies with modulation index.

Keywords Multilevel inverter · THD · PSO · Selective harmonic elimination

1 Introduction

Due to the requirement of high-power apparatus in numerous industrial applications, multilevel inverter (MLI) has been introduced as an alternative in high-power high-voltage situations. It can be operated at both fundamental and high switching frequency (pulse width modulation (PWM)), and in addition to that, it also enables the use of renewable energy sources. Renewable energy sources such as photovoltaic and wind can be easily interfaced to a multilevel converter system for a

W. Razia Sultana (✉) · S.K. Sahoo · P. Harsha Vardhan Reddy · G.T. Rajasekhar Reddy
VIT University, Vellore 632014, Tamil Nadu, India
e-mail: wraziasultana@vit.ac.in

S.K. Sahoo
e-mail: sarata1@rediffmail.com

S. Prabhakar Karthikeyan · I. Jacob Raglend
NI University, Kumaracoil, Thuckalay, Kanyakumari Tamil nadu, India
e-mail: spk25in@yahoo.co.in

I. Jacob Raglend
e-mail: jacobraglend@rediffmail.com

high-power application [1]. The higher power can be achieved by linking a series of power semiconductor switches with several lower-voltage DC sources to perform the power conversion by synthesizing a staircase waveform. Batteries, capacitors, and renewable energy voltage sources can be used as the multiple dc voltage sources [2]. Three different major multilevel converter structures have been reported in the literature: cascaded H-bridge converter with separate dc sources, diode clamped (neutral clamped), and flying capacitors (capacitor clamped). Abundant modulation techniques and control paradigms have been developed for multilevel converters such as sinusoidal pulse width modulation (SPWM), selective harmonic elimination (SHE), and others [3]. SHE refers to the choice of switching angles designed to eliminate specific harmonics by effectively eliminating certain harmonics by DC link voltages of an inverter on and off via power switches at pre-determined points [4]. By using different combinations of switches, three different output voltage levels can be obtained (+Vdc, 0, -Vdc) [1]. The complete elimination of the lower-order harmonics using PWM techniques is not possible [5]. So selected lower-order harmonics can be eliminated by considering the output waveform's Fourier series which gives the $(N + 1)$ nonlinear equations. There is a possibility of multiple solutions in these equations, and practical method of solving these equations is by trial and error method.

Iterative methods such as Newton–Raphson are also used to solve these equations, but right choice of the initial guess is required for converging. The other choice is to use resultant theory where these equations are converted into polynomials. The complexity of the problem increases as the number of DC scores increases, making the degree of the polynomial as well as computational burden quiet high.

The important measure in the voltage non-equality is its THD, and this needs to be as small as possible for many applications on MLI [6].

2 Cascaded Multilevel Inverter

Cascaded MLI is the cascade connection of the n H-bridge inverters. The value of n depends upon the level of inverter; for $2n + 1$ -level inverter, we require n number of H-bridges. Figure 1 shows cascaded 7-level inverter with 3 H-bridges [7].

Here, $V_{dc1} = V_{dc2} = V_{dc3} = V_{dc}$. Each inverter level can generate +Vdc, 0, -Vdc. Switching sequence for different voltage levels is shown in Table 1. It shows that by turning on the switches S1 and S4 in each bridge we can get +3Vdc. Similarly, turning on S2 and S3 switches in each bridge yields -3 Vdc. By turning off all the switches, we get the voltage to be equal to be 0 V.

Fig. 1 Single-phase cascaded 7-level inverter

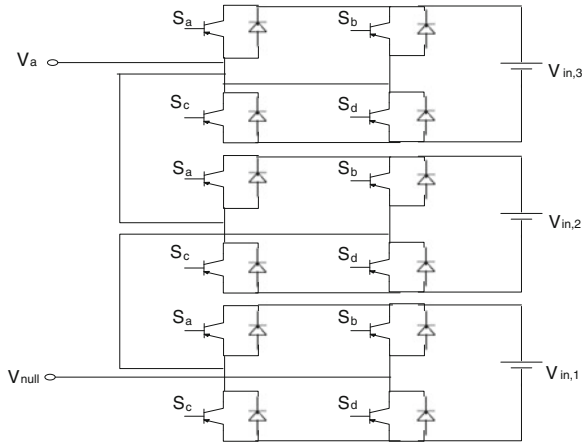


Table 1 Switching sequence of 7-level cascaded multilevel inverter

Voltage level	1st H-bridge	2nd H-bridge	3rd H-bridge
$3 V_{dc}$	S_1, S_4	S_1, S_4	S_1, S_4
$2 V_{dc}$	S_1, S_4	S_1, S_4	S_4, D_3
V_{dc}	S_1, S_4	S_4, D_3	S_4, D_3
0	Off	Off	Off
$-V_{dc}$	S_3, D_4	S_3, D_4	S_2, S_3
$-2V_{dc}$	S_3, D_4	S_2, S_3	S_2, S_3
$-3V_{dc}$	S_2, S_3	S_2, S_3	S_2, S_3

3 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a robust stochastic search algorithm for optimization problem. PSO technique is developed by Kennedy and Eberhart in 1995 [8]. PSO combines social psychology principles in sociocognition human agents and evolutionary operations [9]. Inspired by social behavior of bird flocking or fish schooling, searching process in PSO is based on population of particles. A particle represents a potential solution to the problem under investigation. Each particle in a given population adjusts its position by flying in a multidimensional search space to find the optimal solution to the given problem. PSO is very simple in concept, easy in implementation, and efficient algorithm when compared to other algorithms such as genetic algorithm (GA) and other iterative algorithms [10]. The system is initialized randomly under certain limits and searches for optimal solution by updating particle position. However, unlike GA, PSO has no crossover. In PSO, particles are flown through the problem space by following the current optimum particles. Each particle keeps on updating itself by comparing it to the best position it has attained so far and the global best position attained and keeps following it.

All these particles are to find the best solution (fitness). This fitness of the particle will decide the global and the particle best values. The best value of particle is called global best. In PSO, the global best position of particle (gbest) leads the other individual particles [11]. PSO takes less iteration than GA in order to find the optimal solution [11]. PSO modeling can be done mathematically by using velocity and position vectors, respectively [10]. The equations are shown below.

$$V_{i+1} = W \times V_i + C_1 \times r_1 \times (Pb_i - X_i) + C_2 \times r_2 (gb_i - X_i) \quad (1)$$

$$X_{i+1} = X_i + V_{i+1} \quad (2)$$

where i is the iteration and C_1 and C_2 are the cognitive and social parameters, respectively

X is the position vector, and V is the velocity vector

$$W = (W_{\max} - W_{\min}) \times \frac{\text{iter}_{\max} - \text{iter}}{\text{iter}_{\max}} + W_{\min} \quad (3)$$

where iter_{\max} is maximum iterations taken and iter is the present iteration, and the flowchart representing the basic PSO algorithm is shown in Fig. 2.

4 Optimization of Switching Angles Using PSO

PSO-based optimization [12] is to find the best switching angles for the MLI. To find those switching angles, this must be randomly initialized [11]. As the output has quarter wave symmetry, initialization must be in the order given $0 \leq \theta_1 \leq \theta_2 \leq \theta_3 \leq \frac{\pi}{2}$. The output voltage of the 7-level cascaded MLI is shown in Fig. 3.

The step-by-step procedure to solve SHE problem using PSO is as follows

- Initialize the iteration and the iteration count is set to 1.
- Initialize each particle in the population between 0 and $\pi/2$, also the velocity vector of each particle between $+V_{\max}$ and $-V_{\max}$. Evaluate each particle fitness by using the objective function

$$F_n(\theta_1, \theta_2, \theta_3) = \frac{\sqrt{\sum_{n=3,5,7\dots}^n (V_n^2)}}{V_1} \quad (4)$$

- By the fitness of the particle, evaluate the particle best and the global best values P_{best} and G_{best} , respectively. These are used to update the vectors.
- Update velocity and position vectors through (1) and (2), respectively.

Fig. 2 Flowchart of general PSO

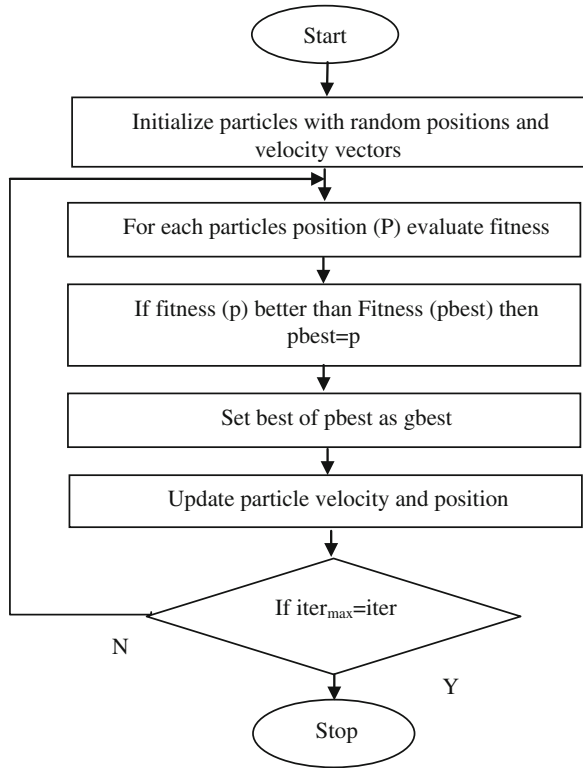
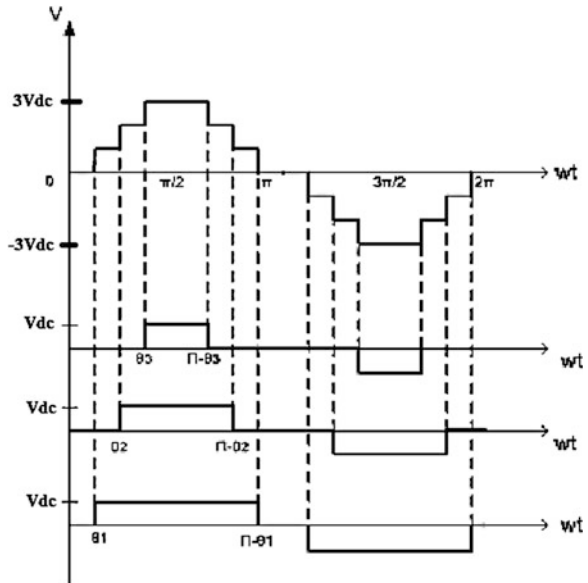


Fig. 3 Output voltage waveform of a 7-level MLI



- If the iteration count reaches the maximum iterations, then it will terminate; if it is less than the maximum iterations, then the loop continues. If the solution is not the desired one, then increase the number of iterations in the first step.

The output voltage Foulter equation with respect to the three switching angles is

$$V(\omega t) = \sum_{n=1,3,5\dots}^{\infty} \frac{4V_{dc}}{n\pi} (\cos(n\theta_1) + \cos(n\theta_2) + \cos(n\theta_3) + \dots \cos(n\theta_n)) \times \sin(n\omega t) \quad (5)$$

The main objective is to eliminate the 5th and 7th harmonic content [3] by satisfying the equations below

$$\cos(7\theta_1) + \cos(7\theta_2) + \cos(7\theta_3) = 0 \quad (6)$$

$$\cos(5\theta_1) + \cos(5\theta_2) + \cos(5\theta_3) = 0 \quad (7)$$

$$\cos(\theta_1) + \cos(\theta_2) + \cos(\theta_3) = m \quad (8)$$

where the modulation index M is defined from m . $M = m/s$, where s is the number of switching angles; in this case, $s = 3$. Number of harmonics that can be eliminated by using SHE depends upon the number of switching angles. For s switching angles, $s-1$ number of harmonics can be eliminated. If the value of s is even, then maximum harmonic that it can eliminate is $2s + 3$, whereas for odd s , it is $2s + 1$.

In the objective function of the 7-level inverter, we need to consider the harmonics up to 7th. The 5th and 7th harmonic magnitudes can be suppressed by this algorithm, so V_5 and V_7 are the main constrains of the problem; minimizing this percentage with respect to the fundamental voltage is our main objective, so these values are substituted in objective function equation to calculate the fitness of the particle.

5 Simulation Results

The results that were carried out in MATLAB/SIMULINK software for a seven-level H-bridge inverter and the validated outcome intended for switching angles are shown in Table 2.

It presents the THD, and $\%V_5$, $\%V_7$ for the various modulation indexes. The computation of switching angles by the PSO is accomplished in MATLAB (M-File) using MATLAB coding. Figures 4 and 5 show the line voltage and phase voltage of the load voltage, respectively.

The FFT analysis shown in Fig. 6 shows that 5th harmonics and 7th harmonics are eliminated. The results are presented for modulation index ranging from 0.5 to 1. As we observe from the Fig. 7, $\%V_5$ and $\%V_7$ are suppressed for modulation

Table 2 Results for various modulation index (phase voltage)

Switching angles in deg			Modulation index (M)	%V ₅	%V ₇	THD
θ_1	θ_2	θ_3				
12.124	33.025	59.652	0.77	0	0.5	13.79
13.170	22.698	53.860	0.82	0.45	0.4	14.45
12.464	42.531	85.617	0.59	0.08	0.58	18.62
16.431	41.502	63.757	0.71	0.16	0.11	19.61
5.813	16.562	35.819	0.91	0.12	0.63	17.68
13.589	36.643	61.604	0.75	0.52	0.24	15.90

Fig. 4 V_{ab} line voltage waveform

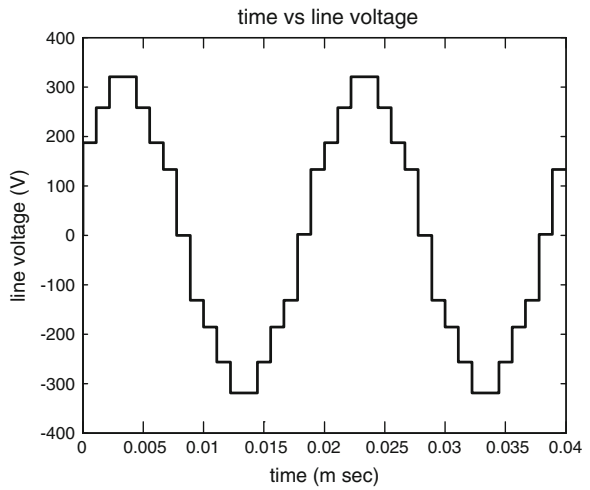
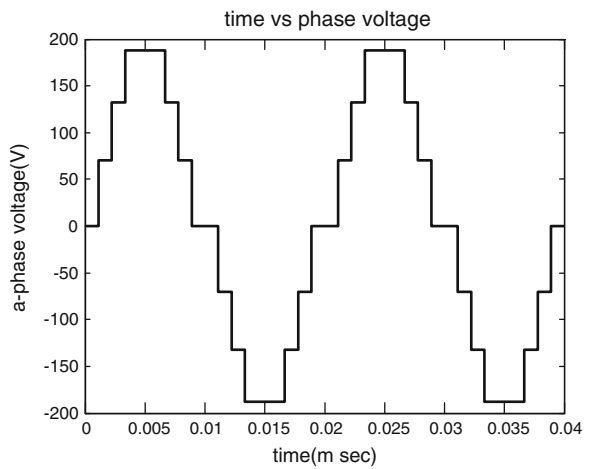


Fig. 5 Phase-a voltage output wave form



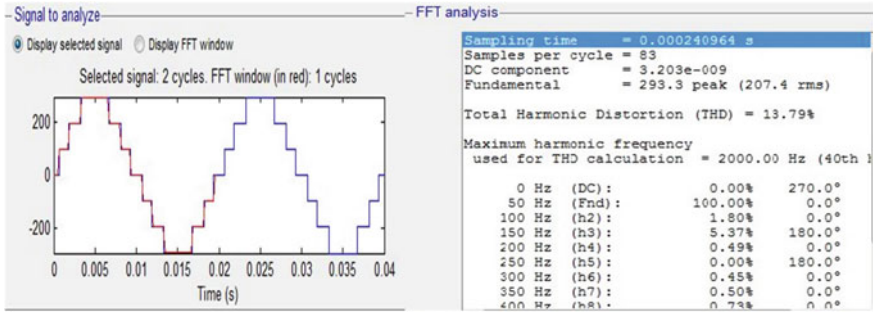


Fig. 6 FFT analysis of output voltage

Fig. 7 Modulation index versus %V₅ and %V₇

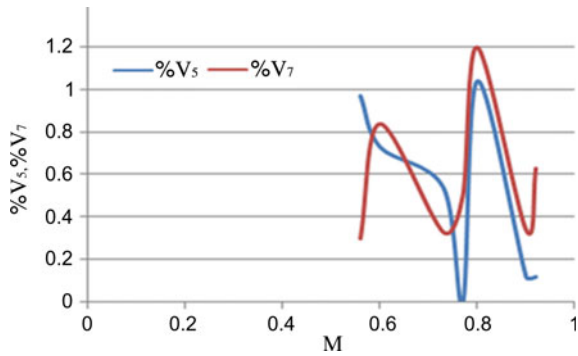
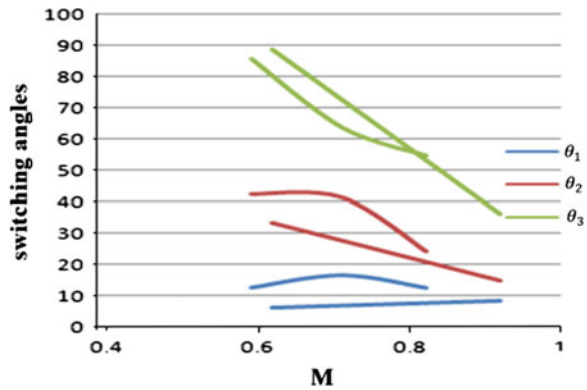
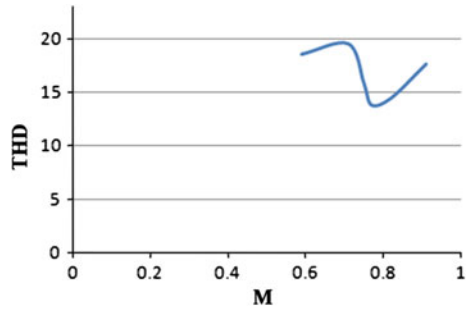


Fig. 8 Modulation index versus switching angles



index 0.7–0.8 and the higher-order harmonics are optimized to contribute minimum voltage THD.

Fig. 9 Modulation index versus THD



To validate the optimal switching angles, a MATLAB code is written and the simulation is carried out for 100 iterations. The result for switching patterns with different modulation indexes is presented in Fig. 8.

From the simulation, we can see that for different modulation index, our aim to reduce the lower-order harmonics and produce minimum THD is satisfied. From Fig. 9, it is clear that for the modulation index ranging from 0.7 to 0.8, the THD is minimum.

6 Conclusion

This paper applies the SHE technique to a 3-phase, 7-level cascaded MLI to assess the ability of the technique to eliminate specific harmonics. The paper shows that when mitigating specific sets of harmonics, there is a reduction in THD. The simulation is carried out for 7-level cascaded MLI with three switching angles to eliminate 5th and 7th harmonics. From the results, we can observe that between the modulation index in the range of 0.7–0.8, the magnitudes of 5th and 7th harmonics are less and THD is also much reduced. Since our main objective is satisfied, the optimal switching angles will be in between the range of 0.7 and 0.8. This work can be further extended by increasing the level of MLI, thereby increasing the number of switching angles. As switching angles are increased, more number of harmonics can be reduced.

References

1. L.G. Franquelo, J. Rodriguez, J.I. Leon, S. Kouro, R. Portillo, M.A.M. Prats, The age of multilevel converters arrives. *IEEE Ind. Electron. Mag.* 2(2), 28–39 (2008)
2. L.M. Tolbert, J.N. Chiasson, K. McKenzie, Z. Du, Elimination of harmonics in a multilevel converter with non-equal DC sources, in *Proceedings of IEEE Applied Power Electronics Conference* (2003), pp. 589–595

3. J.N. Chiasson, L.M. Tolbert, K.J. McKenzie, Z. Du, A complete solution to the harmonic elimination problem. *IEEE Trans. Power Electron.* **19**(2), 491–499 (2004)
4. J. Rodríguez, J. Lai, F. Peng, Multilevel inverters: A survey of topologies, controls and applications. *IEEE Trans. Ind. Electron.* **49**(4), 724–738 (2002)
5. W. Fei, X. Du, B. Wu, A generalized half-wave symmetry SHE-PWM formulation for multilevel voltage inverters. *IEEE Trans. Ind. Electron.* **57**, 3030–3038 (2010)
6. H. Taghizadeh, M.T. Hagh, Harmonic elimination of cascade multilevel inverters with non-equal DC sources using particle swarm optimization. *IEEE Trans. Ind. Electron.* **57**, 3684–3687 (2010)
7. K. Xiaomin, K. Corzine, M. Wielebski, Over detention operation of cascaded multilevel inverters. *IEEE Trans. Ind. Appl.* **42**(3), 817–824 (2006)
8. J. Kennedy, R. Eberhart, Particle swarm optimization, in *International Conference of Neural Networks* (1995)
9. J. Kennedy, R.C. Eberhart, Particle swarm optimization, in *Proceedings of IEEE International of Neural Networks* (1995), pp. 1942–1948
10. B. Khokhar, K.P.S. Parmar, A novel weight-improved particle swarm optimization for combined economic and emission dispatch problems. *Int. J. Eng. Sci. Tech (IJEST)* **4**, 2015–2021 (2012)
11. R. Eberhart, Y. Shi, Comparison between genetic algorithms and particle swarm optimization, in *Proceedings of International Conference on Evolutionary Computation* (1998), pp. 611–616
12. R.C. Eberhart, Y. Shi, Guest editorial. *IEEE Trans. Evol. Comput. (Spec. Issue Part. Swarm Optim.)* **8**(3), 201–203 (2004)

Optimization and Quality-of-Service Protocols in VANETs: A Review

K.R. Jothi and A. Ebenezer Jeyakumar

Abstract Optimization and quality-of-service (QoS) protocols are very important for safety, emergency, and multimedia applications in vehicular ad hoc network (VANET). VANET requires real-time message propagation that is able to deliver data in a timely and accurate manner. For example, considering the case of safety applications, any delay in the message delivery may entrain risky and fatal accidents. Similarly, sending multimedia files and video streams requires a high level of QoS. This paper gives the brief review of various optimization techniques and QoS protocols in VANETs which will help in better understanding of these protocols in VANETs and pave their way to develop a new protocol to overcome the drawbacks of existing protocols.

Keywords Vehicular ad hoc networks • Optimization • Quality of service

1 Introduction

Vehicular ad hoc network (VANET) is becoming nowadays an attractive topic for the research community with the increase in the number of traffic accidents and the complexity of the roads infrastructure. VANET is a new class of wireless networks that allows vehicle-to-infrastructure communication and vehicle-to-vehicle communications. This technology offers a wide set of applications and services ranging from safety applications and traffic management systems to commercial and marketing services. Safety, emergency, and multimedia applications of VANET require

K.R. Jothi (✉)

Department of Computer Science and Engineering, Sri Ramakrishna Institute of Technology, Coimbatore, India
e-mail: prof.krj@gmail.com

A. Ebenezer Jeyakumar

Sri Ramakrishna Institute of Technology, Coimbatore, India
e-mail: ebeyjkumar@rediffmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_31

275

to assure a high level of quality of service (QoS) through the network. Practically, VANET requires real-time message propagation that is able to deliver data in a timely and accurate manner. For example, considering the case of safety applications, any delay in the message delivery may entrain dangerous and mortal accidents. Similarly, exchanging multimedia services such as files and video streams requires a high level of QoS [1].

2 Optimization Techniques

2.1 Ant Colony Optimization

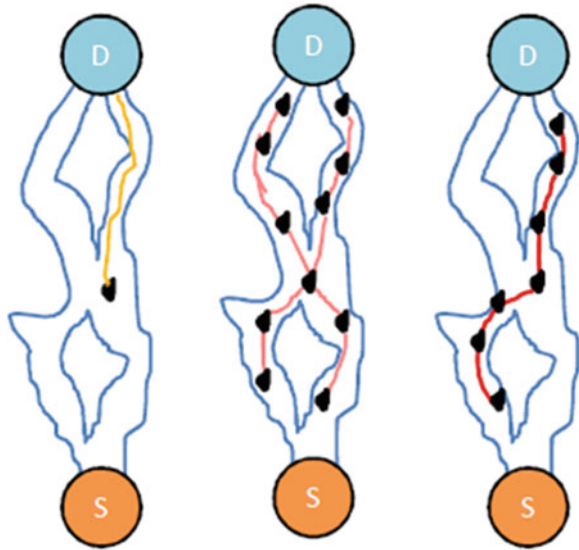
Ant colony optimization is a probabilistic approach that is used to solve several discrete optimization problems. This approach inherits the normal behavior of ants that tend to find the shortest route while searching for food. In fact, the ants that move randomly toward the food indicate the other ants the shortest path to follow by depositing chemical substance called pheromone. Thereafter, paths with higher pheromone values are chosen to be followed.

Thus, the shortest path will be continuously reinforced by more pheromone values since it will get marched repeatedly by the ants [2]. In contrary, the pheromone trails of the not marched paths will decrease due to the evaporation process. The evaporation is important to avoid the convergence for local optimal solution since without evaporation, the routes marched by the first ants will be extremely attractive for the following ants. The behavior of ants has attracted the researchers due to its dynamic nature that makes it adaptive to changes in real-time applications. Therefore, it has been widely used to solve many problems such as vehicle routing, traveling salesman problem (TSP), machine scheduling, telecommunication networks, ad hoc networks routing, and personnel placement in airline companies.

To illustrate how the ant colony optimization algorithm works, we show in the following how ACO can be applied to solve the TSP, which is a routing problem. Given a set of cities and the distances separating each pair of cities, the problem of TSP is concerned with finding the shortest path that visits each city just once and turns back to the original city. This problem could be modeled as a construction graph where the cities represent the vertices and the distances separating the cities are the edges. The ACO solution works as follows: Initially, every ant begins from a randomly selected vertex (city). Next, ants choose their next vertex to get in a probabilistic manner according to the highest pheromone value.

This process continues until each ant has visited all the vertices on the graph only once. Now, pheromone values are updated on all the edges according to the quality of solution to which they belong so that the pheromone values for shorter routes would be greater than other routes. Thereafter, the pheromone values begin to evaporate and only the short tours will be reinforced by more pheromone values. This process is repeated for a specified number of iterations, and the best discovered tour is maintained as final solution. This process is illustrated in Fig. 1.

Fig. 1 Ant colony optimization



Therefore, in VANET, to guarantee choosing the optimal paths in terms of QoS, mobility, and delay, an ant colony optimization algorithm is used where proactive discovery by ant agents is initiated the find the best paths.

2.2 Dempster–Shafer

Dempster-Shafer is a mathematical theory elaborated by Arthur P. Dempster and Glenn Shafer [3]. This theory combines evidences from independent sources to come up with a degree of belief (belief function). It relies on two main ideas: (1) acquiring degrees of belief from subjective probabilities and (2) combining these beliefs.

To illustrate how these two ideas work, as an example, let us assume that the subjective probabilities for the trustworthiness of Doctor John are known. The probability that John is trustworthy is 0.9, while the probability that he is untrustworthy is 0.1. Suppose that John says that a patient, Bob, suffers from diabetes. This allegation must be true if John is trustworthy but not necessarily false if he is untrustworthy. Thus, his testimony alone justifies a 0.9 degree of belief that Bob suffers from diabetes, but only a zero degree of belief (not a 0.1) that Bob is healthy. This zero does not imply that we are sure that Bob does not suffer from diabetes, but simply implies that John’s testimony gives no reason to believe that Bob is healthy. The 0.9 and the zero together form a belief function.

In order to explain how the combination rule for degrees of belief works, we suppose that we know another doctor, called Alice, and that Alice is trustworthy with probability of 0.9 and untrustworthy with probability of 0.1. Assume that Alice

witnesses as well that Bob suffers from diabetes. Since the trustworthiness of John is independent from the trustworthiness of Alice, we may multiply the probabilities of these two events. The probability that both are trustworthy is $0.9 \times 0.9 = 0.81$. The probability that neither John nor Alice is trustworthy will be $0.1 \times 0.1 = 0.01$. Finally, the likelihood that at least one is trustworthy is $1 - 0.01 = 0.99$. Since both doctors said that Bob suffers from diabetes, at least of them being trustworthy means that Bob really suffers from diabetes, and we may hence assign this event a degree of belief of 0.99 as explained before.

On the other hand, if John's and Alice's testimonies were contradictory in the sense that John says that Bob suffers from diabetes, while Alice says that he does not. Now, both cannot be right and hence cannot be both trustworthy. The prior probabilities that only John is trustworthy, that only Alice is trustworthy, and that neither is trustworthy are 0.09, 0.09, and 0.01, respectively, and the posterior probabilities (given that not both are trustworthy) are $9/19$, $9/19$, and $1/19$, respectively. Hence, we have a $9/19^\circ$ of belief that Bob suffers from diabetes (because John is trustworthy) and a $9/19^\circ$ of belief that Bob is sound and healthy (because Alice is trustworthy). Thus, Dempster-Shafer gives a weight for each evidence according to the trustworthiness level of the person giving the evidence and is necessary hence to discount evidences from untrustworthy observers upon aggregating the different testimonies. This appealing feature motivated us to use Dempster-Shafer while detecting the misbehaving vehicles since we are proposing a cooperative detection mechanism where evidences from different sources need to be aggregated. Therefore, Dempster-Shafer can be a solution to come up with reliable and credible decisions.

Another important characteristic of Dempster-Shafer is that it supports uncertain evidences. Suppose, for example, that Alice and John witness that a thief got in Bob's home. However, they might both heard the voice of a noise coming from a cat and thought it is coming from a thief. To express this uncertainty, Bob can consider three evidences: (1) evidence for Alice's trustworthiness, (2) evidence for John's trustworthiness, and (3) evidence for the possibility of the presence of a cat. Now, Bob can combine these three items of evidences through Dempster's rule of combination to come up with the final decision taking into consideration that both observers may be unreliable. This feature can be exploited to improve the detection for misbehaving vehicles in VANET and overcome the problem of ambiguity caused by the high mobility of the vehicles and the channel collisions.

2.3 Repeated Game Theory

Game theory is a formal study of conflict and cooperation that applies whenever the actions of several peers are interdependent in the sense that the strategy of one game's component depends on the action of another game component [4]. The motivation behind using game theory is arriving at optimal decision. Consider, for example, that a company decides to reduce prices in order to augment its profit.

Without considering the other players' (companies) actions, this action may be counterproductive and the company will lose money if the other companies apply a policy of price cuts. Here lies the importance of considering the different parties' strategies upon building any strategy. Game theoretical concepts have been widely used to solve problems in the fields of economy, biology, military, and computer science. To be clear and meaningful, each game should describe seven principal elements: players, actions, information, strategies, outcomes, payoff, and equilibrium. The players are the game parties that are responsible for making decisions. The actions are the set of options from which the players have to choose. The information represents the learning of the player upon making decision. The strategies describe the set of principles that control the decisions of the player at each stage of the game. The outcomes are the expected or desired output of the game such as increase in profits. The payoffs describe the utilities yielded by the player in a specific outcome. Finally, the equilibrium represents a stable solution in which no player has an interest to take unilateral decisions and change his strategy.

Repeated game is a type of game theory in which players repeat their actions over and over again. To show the importance of using repeated game models, we present a motivating example based on the Prisoner's Dilemma [5]. The Prisoner's Dilemma models the investigation in a crime where two prisoners suspected to be committed a crime together are arrested. The investigator isolates them and suggests a deal saying: (1) if one of them confesses against the other one, the confessor will get free (payoff: 0) and the offender will spend 4 years in prison (payoff: 4), (2) if they both confess, they will bear a less cruel punishment by being jailed 3 years (payoff: 3), and (3) if they both decline to confess, they will both bear a reduced sentence lack of evidences (payoff: 1). This deal can be summarized in the following bimatrix (Table 1):

The question is how should the prisoners behave in such game? Each prisoner will have the following thinking:

- If the other prisoner confesses, I have to confess (since 3 years are less than 4 years).
- If the other prisoner refuses to confess, I have to confess (since getting free is better than 1 year in jail).

If the game is played one shot, the best strategy for both players is to confess whatever the opponent did, thus staying 3 years in jail. However, if the game is played repeatedly, then the previous actions of each other become observable and they will know each other's decisions. Then, they may get a better result by not confessing together (1 year in jail). Here lies the dilemma of the prisoners.

Table 1 Payoff matrix of the prisoner's dilemma

	Confess	Do not confess
Confess	(3, 3)	(0, 4)
Do not confess	(4, 0)	(1, 1)

In VANET, the vehicles have to make decisions about cooperating with each other. In making these decisions, nodes may behave selfishly, seeking exclusively for their own interests. This makes the objectives of the different nodes conflicting (some nodes need to be served and others consider that their interests lie in being uncooperative). Thus, the application of game theory may be appropriate, as game theory analyzes situations in which player objectives are in conflict. Moreover, the vehicles' decision depends on the other vehicles' decisions. Therefore, the repeated games are the best to model such situation.

3 Clustering in VANET

The communication in VANETs entrains a high level of overhead, collision, and contention. In order to ensure efficient communications and mitigate the channel collision, overhead, and contention, there should be wireless backbone architecture able to elect some nodes to assume the network responsibilities. One solution is to gather the nodes into clusters and elect for each cluster a specified node to serve as cluster head. The function of the cluster head is to achieve both intra-cluster coordination, and inter-cluster communication. The intra-cluster coordination involves the coordination among the nodes within each cluster. In the inter-cluster communication, the cluster members charge the cluster head to communicate with the other cluster heads on behalf on them. The clustering imposes several challenges that should be taken into consideration such as which node has to be elected as cluster head? How the election procedure is done? What are the requirements of the cluster heads? How to increase and maintain the clusters lifetime? Based on these challenges, several clustering algorithms for VANET have been proposed trying to answer these questions. In the following, we present an overview on the main contributions in this context.

APROVE [6] uses the affinity propagation algorithm to perform a clustering that minimizes the distance and the mobility between cluster heads and members. The affinity metric is composed of responsibility and availability factors. Responsibility signals how compatible is one node to become exemplar while availability signals the willingness of the node to become exemplar.

Modified DMAC [7] was proposed on top of the original Basagni's distributed and mobility-adaptive clustering algorithm. Its basic idea is to increase the stability and avoid re-clustering of the group of vehicles moving in different directions using a freshness parameter. In this algorithm, each node has to know its moving direction, current position, and velocity. The authors in [8] propose a multi-hop clustering that uses the relative mobility between multi-hop away nodes. The beacon delay is used to calculate this metric. The cluster head is elected according to the smallest aggregate mobility value. This approach considers also the problem of re-clustering by postponing it for some time. In [9], the authors use complex metric composed of traffic conditions, connection graph, and link quality. Before

assigning a node to a cluster, a check on the node's reliability is done using the membership lifetime counter. This has the advantage of avoiding needless re-clustering.

Presented clustering algorithms are proposed for different purposes such as clusters stability and overhead minimization. However, these algorithms ignore the QoS which is important for safety, emergency, and multimedia services in VANET [10]. The QoS relies primarily on connectivity, reliability, and end-to-end delay. Thus, any clustering protocol in VANET is to maintain the stability of the vehicular network while achieving a trade-off between QoS requirements and mobility constraints.

3.1 Routing in VANET

After the clusters are formed, it is important to develop a routing algorithm able to achieve the communications among clusters. Routing refers to the process of carrying a packet from source to destination. In ad hoc networks, this process encounters several challenges. These challenges range from the dynamic topology of the network, scalability, and limited physical security, to bandwidth and energy constraints. Based on these challenges, several routing protocols are presented MANETs and VANETs.

Routing based on ant colony optimization routing algorithm using ant agents for MANETs (RAAM) [11] was proposed to reduce the end-to-end delay. This can be done by creating multiple ant colonies that will travel through different paths to select the optimal one. Nevertheless, the overhead is the shortcoming that encounters this algorithm.

Ant-colony-based routing algorithm (ARA) [12] gets several paths from source to destination to transfer the packets. The drawback of ARA is that it cannot respond directly to topology change because of its passive nature. Probabilistic emergent routing algorithm (PERA) [13] is, in contrary, an active method that periodically broadcasts ants so as to avoid the local best solution. However, the overhead of the routing table and the periodic broadcasts is a drawback that faces PERA.

The idea of AntHocNet [14] is to achieve a dynamic traffic loading balance for the whole network in order to reveal the importance of the quality-of-service issue. Nevertheless, AntHocNet suffers from several limitations such as the long search time and the early convergence for large scales.

3.2 Routing Based on Multi-point Relay Nodes

The classical OLSR [15] protocol has been modeled to cope with mobile ad hoc networks (MANETs). Its basic idea is to elect a cluster head for each group of neighbor nodes and divides hence the network into clusters. These heads then select

a set of specialized nodes called MultiPoints relay (MPRs). The function of the MPR nodes is to reduce the overhead of flooding messages by minimizing the duplicate transmissions within the same zone.

QOLSR [16] was design on top of OLSR to consider the QoS of the nodes during the election of heads and the selection of MPRs. In fact, QOLSR focuses on choosing optimal paths satisfying the QoS constraints. Though the QOLSR is unable to deal with VANETs, it considers exclusively the nodes' bandwidth ignoring thus some other important metrics such as mobility.

Then came QoS-OLSR [17], a cluster-based protocol that aims to prolong the network lifetime. When electing heads and choosing MPRs, this protocol considers, in addition to the bandwidth, some metrics that may affect the network lifetime such as the residual energy. Nevertheless, the QoS-OLSR has many limitations that make it inadequate to achieve the VANET requirements since it ignores the mobility of nodes while computing the QoS.

In summary, VANETs have some characteristics that make them unique among other types of networks. Practically, VANET is characterized by the very high mobility of its nodes and the frequent disconnections. Numerous routing protocols have been proposed for MANETs, and some of them could be applied to VANETs. Nevertheless, simulation results proved that they suffer from bad performances due to the specific features of VANET such as rapid vehicles movement, dynamic packets exchange, and high speed of nodes. Thus, finding and maintaining routes is a very challenging task in VANETs.

4 Conclusion

Various optimization and QoS protocols have been developed that works in IEEE 802.11p which is suitable for the VANETs. This paper gives the review of various optimization techniques and QoS protocols in VANETs. Every protocol has its own advantages and disadvantages. Moreover, the existing routing algorithms are unable to select and maintain the optimal paths in terms of QoS, stability, delay, and overhead. Concerning the misbehaving vehicles, the existing approaches have several limitations that make them inefficient to deal with the selfish nodes such as lack for scalability and centralization, ambiguous collisions, and false alarms. We hope that this concise work will help in better understanding of QoS protocols in VANETs and pave their way to develop a new protocol to overcome the drawbacks of existing protocols.

References

1. O.M. Abdel Wahab, Cooperative clustering models for vehicular ad hoc networks. Thesis, Lebanese American University (2013)
2. M. Dorigo, G. Di Caro, L.M. Gambardella, Ant algorithms for discrete optimization. *Artif Life* **5**, 137–172 (1999)
3. J. Schubert, Conflict management in Dempster-Shafer theory using the degree of falsity. *Int. J. Approximate Reasoning* **52**(3), 449–460 (2011)
4. J.M. Robert, H. Otrok, A.N. Quttoum, R.A. Boukhris, Distributed resource management model for virtual private networks: tit-for-tat strategies. *Elsevier—Comput. Netw.* **56**(2), 927–939 (2012)
5. J. Andreoni, H. Varian, Preplay contracting in the prisoners dilemma. *PNAS* **96**(19), 10933–10938 (1999)
6. C. Shea, B. Hassanabad, S. Valaee, Mobility-based clustering in VANETs using affinity propagation, in *Proceedings of 28th IEEE Conference on Global Telecommunications* (2009)
7. G. Wolny, Modified DMAC clustering algorithm for VANETs, in *Proceedings of 2008 Third International Conference on Systems and Networks Communications* (2008)
8. Z. Zhang, A. Boukerche, R. Pazzi, A Novel multi-hop clustering scheme for vehicular ad-hoc network. in *Proceedings of 9th ACM International Symposium on Mobility Manage and Wireless Access* (2011)
9. S. Kuklinski, G. Wolny, Density based clustering algorithm for vehicular ad hoc networks. *Int. J. Internet Protoc. Technol.* **4**(3), 149–157 (2009)
10. S. Vodopivec, J. Bester, A. Kos, A survey on clustering algorithms for vehicular ad-hoc networks. in *Proceedings of 35th International Conference on Telecommunications and Signal Process* (2012)
11. K.R. Ramkumar, M. Ravichandran, N. Hemachandar, D. Manoj Prasad, M. Ganesh Kumar, RAAM: routing algorithm using ant agents for MANETS. in *Proceedings of 11th International Conference on Advanced Communications Technology* (2009)
12. M. Gunes, U. Sorges, I. Bouazizi, ARA—the ant-colony based routing algorithm for MANETS, in *Proceedings of 2002 International Conference on Parallel Process* (2002)
13. J.S. Baras, H. Mehta, A probabilistic emergent routing algorithm for mobile ad hoc networks, in *Proceedings of WiOpt'03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks* (2003)
14. G. Di Caro, M. Dorigo, Ant net: distributed stigmergetic control for communications networks. *J. Artif. Intell. Res.*, **9** (1998)
15. T. Clausen, P. Jacquet, P. Muhlethaler, A. Laouiti, A. Qayyum, L. Viennot, Optimized link state routing protocol for ad hoc networks, in *Proceedings of IEEE International Multi Topic Conference* (2001)
16. H. Badis, K.A. Agha, QOLSR, QoS routing for ad hoc wireless networks using OLSR. *European Trans. Telecommun.* **16**(5), 427–442 (2005)
17. H. Otrok, A. Mourad, J.M. Robert, N. Moati, H. Sanadiki, A cluster-based model for QoS-OLSR protocol, in *Proceedings of 7th International Wireless Communications and Mobile Computing Conference* (2011)

Author Biographies



Jothi K.R. received his M.E. degree and B.E. degree in Computer Science and Engineering from Anna University and Bharadhidasan University, respectively. He is currently working as Assistant Professor (Selection Grade) in the Department of Computer Science and Engineering at Sri Ramakrishna Institute of Technology, Coimbatore. His research interests accumulate in the area of vehicular ad hoc networks.



Dr. Ebenezer Jeyakumar A. received his Ph.D. degree, M.E. degree, and B.E. degree from Anna University Chennai, University of Madras, and Annamalai University, respectively. He is currently working as Director Academics, Sri Ramakrishna Engineering College, Coimbatore. He has contributed richly to augmentation of knowledge through his research activities by guiding various research scholars in the area of mobile networks, computer networks, power systems, VLSI circuit systems, energy, renewable resources—wind, software engineering, and high energy batteries.

Object Detection Using Robust Image Features

Khande Bharath Kumar and D. Venkataraman

Abstract Object detection is a challenging field of research in computer vision. Research approaches have become increasingly popular in overcoming the challenges of object detection like occlusions, changes in scale, rotation, and illumination. Object detection methods that utilize RGB cameras are used to accurately identify objects in the real world, but they do not consider shape and three-dimensional characteristics of the object. Recognizing the objects in 3D is not an easy task for computers, like as in humans. Robust features like shape, color, size, etc., are necessary for 3D object detection for ensuring accuracy.

Keywords Feature extraction · Color model · SIFT · Object detection

1 Introduction

Object detection is a well-established field of research in computer vision, and feature-based approaches have become more popular due to their robustness to clutter, occlusions, and changes in scale, rotation, and illumination. Once the model is obtained off-line, online recognition and pose estimation is performed by matching the image features against the model features and solving the perspective-n-point problem for the 2D–3D correspondences. Given a set of correct matches, pose estimation is a well-solved problem, and various solutions have been devised. Feature-based methods can be grouped on the basis of the feature used, e.g., edges, shape, patches, and interest points. Recent approaches based on this paradigm rely on feature discriminability for correct matches and on the robust estimator capabilities for outlier

K.B. Kumar (✉) · D. Venkataraman
Department of Computer Science and Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: khandebharathkumar@gmail.com

D. Venkataraman
e-mail: d_venkat@cb.amrita.edu

rejection. The basic objective is robust object detection using multiple methods and selecting the appropriate one depending on the characteristics of the object.

The paper is organized as follows. Introduction is given in Sect. 1. Section 2 gives an idea about previous work in object detection. Proposed work is carried out in Sect. 3, Feature Matching in Sect. 4, Object Detection in Sect. 5, Discussion and Results in Sect. 6, and Conclusion and Future work in Sect. 7.

2 Previous Work

Object detection is a task of finding 3-dimensional (3D) objects from two-dimensional (2D) images and classifying them into one of the many known object types. It is important for computer vision because it is closely related to the success of many computer vision applications such as robotics, surveillance, registration, and manipulation, etc. Object detection is to extract set of features from given images. Features for object detection can be grouped into global features and local features. Global features usually describe an object as a whole, while local features are extracted from the parts of the object, such as robust regions. Global features usually require an exhaustive search of the whole image over various scales and sizes to localize the target object. Thus, it is more time-consuming when compared to a search using local features, which are typically scale and rotation invariant. The extraction of these local features usually consists of two steps. First, the salient regions are detected in a way robust to geometric transformations. Second, a region description for each of the detected region is generated to make them distinguishable. Reliable region detectors are robust to illumination and viewpoint changes.

In particular, 3D object detection is an indispensable process for robot manipulation. Many researchers have proposed many 3D object recognition approaches. Among them, model-based recognition method is one of the most general one, which generates the hypothesized model posed by finding correspondences between the model features and image features, and then, the final pose can be verified with additional image features. Three-dimensional (3D) object detection is a process that matches the input stimulus to stored object representations in memory. The search for a match, when viewpoint invariant features (e.g., color or material) are absent, must be based on the object shape. A major challenge for object recognition is to understand how potential matches are verified despite shape variations in the image due to rotations in viewpoint. Empirical evidence has shown that human object recognition strongly depends on familiar views, a result particularly pronounced for structurally similar objects.

The basic 2D object detection refers to identify a location that identify and register components of a particular object class at various levels of detail. Color of an image is formed by mixture of different lights. Decomposition of an image into relevant information helps to localize detection. The goal of Basic 2D Object Detection system is to identify the basic geometric shape of objects present in image. Shape of an object is based on geometry of object parts that are distributed

all over the object rather than object boundary. Key points like corners, vertices in an image are tolerance to clutter and occlusion. Detecting object in a clutter scene by extracting features of small segments helps to find object's interest regions. Segmentation helps to locate objects and boundaries in an image. Pixels in the segmented region share certain visual characteristics of object.

Model-based recognition method fails in areas where no interest regions are extracted by the detector. Classification of objects can be done with the help of textures. Texture patches in an image describes better histogram features. Texture helps in image categorization at different boundary regions. Edge detection reduces the processing data and filters out the unnecessary information. Different edge detectors are Sobel, Prewitt, Canny, etc.

The challenges for real-world images are the variety of object appearances caused by view changes, intravariation, and occlusion. In object detection, segmentation plays an important role in identifying object. Object segmentation is achieved by processing red, green, and blue chromatic components. But the disadvantage of this method is sensitive to changes in lighting and if the scene having multiple objects of same color is not possible to detect object using only color information [1].

The region-based features are inspired by segmentation approaches and are mostly used in algorithms whose goal is to combine localization, segmentation, and categorization. Region-based features use various rectangular or square local regions of the same size in order to derive the object templates [2]. Such features cannot deal with multi-scaling (appearance of the object in various sizes) effectively. A fixed patch size may not be suitable. If the patch size is small, it may not cover a large but important local feature. Information of such feature may be lost in the smaller patch. On the other hand, if the patch size is large, it may cover more than one independent feature, which may not be present simultaneously in other images. Intensity is the most commonly used cue for generating region-based features, texture, color, and minimum entropy and [3] has also been used for generating these features. These features are very sensitive to lighting conditions and are generally difficult from the perspective of scale and rotation invariance.

SIFT consists of four major stages: scale-space extrema detection, key point localization, orientation assignment, and key point descriptor. The first stage used difference of Gaussian function to identify potential interest points [4], which were invariant to scale and orientation. Difference of Gaussian was used instead of Gaussian to improve the computation speed [4]. In the key point localization step, they rejected the low-contrast points and eliminated the edge response. Hessian matrix was used to compute the principal curvatures and eliminate the key points that have a ratio between the principal curvatures greater than the ratio. An orientation histogram was formed from the gradient orientations of sample points within a region around the key point.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, \sigma) - L(x, y, \sigma) \quad (1)$$

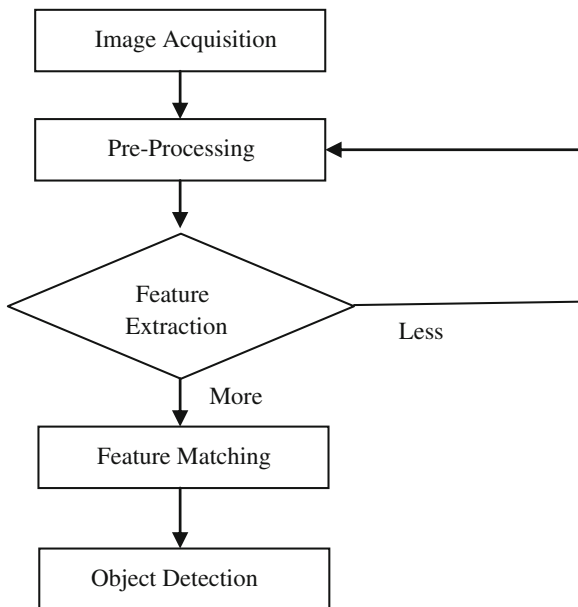
3 Proposed Work

The work aims to detect objects for arbitrary images. For achieving this, several object detection techniques have been done for 2D images out of which no single feature alone is sufficient for object detection due to illumination problem and object characteristics. Preprocessing plays a crucial role in detection and recognition of object. Histogram equalization plays a very crucial role in enhancing the contrast of an image. Variation in illumination is difficult to detect object with color feature for detecting, so for effective segmentation, we need color models to differentiate between objects. For achieving this, we are using HSV color model, canny edge detection technique, contour-based feature are used for object detection. Feature-based extraction methods classify situations depending on the characteristics of the object of interest. Different feature extraction methods can be used to extract features for area of interest. The main objective of any feature extraction method is to get features with maximum information. The suitable object detection method is different for different environments. Here, our goal is to find best suitable recognition methods for different environments. SIFT is more robust to noise and scale variations. For specific object recognition, SIFT would be a good choice. Viewing angle-based histogram technique, in which the matching needs much less computation time. Viewing angle uses range images for classification. Segmentation of range images is done by using region, edge base technique. Region-based segmentation is done in range level instead of pixel level produces gaps between boundaries. Object detection by using range images depends on illusions and suffers from noise and optical occlusions [5]. In order to achieve good quality, the information that can be used in an image are color, texture, depth, etc., SIFT will work better for different structure object detections (Fig. 1).

3.1 Color Detection Algorithm

Color provides more powerful information for object detection. In RGB image, each pixel has three color components: red, green, and blue. Amount of mixing of these three colors determines value of pixel. Image is a collection of pixels; each pixel is a combination of red, green, and blue colors. So it is difficult to process each pixel; hence, we need color model for robust detection. In HSV image, each pixel has only one color which is represented by Hue component. Saturation and value components determine how much amount of black and white color is added into that color; it helps to differentiate object with other color so we use HSV color model for object detection.

The R, G, B values are divided by 255 to change the range from 0–255 to 0–1.

Fig. 1 Overview of object detection

$$\mathbf{R}' = \mathbf{R}/255 \quad (2)$$

$$\mathbf{G}' = \mathbf{G}/255 \quad (3)$$

$$\mathbf{B}' = \mathbf{B}/255 \quad (4)$$

$$\mathbf{C}_{\max} = \max(\mathbf{R}', \mathbf{G}', \mathbf{B}') \quad (5)$$

$$\mathbf{C}_{\min} = \min(\mathbf{R}', \mathbf{G}', \mathbf{B}') \quad (6)$$

$$\Delta = \mathbf{C}_{\max} - \mathbf{C}_{\min} \quad (7)$$

Hue Calculation

$$\mathbf{H} = \begin{cases} 60^\circ \times \left(\frac{\mathbf{G}' - \mathbf{B}'}{\Delta} \bmod 6 \right), & \mathbf{C}_{\max} = \mathbf{R}' \\ 60^\circ \times \left(\frac{\mathbf{B}' - \mathbf{R}'}{\Delta} + 2 \right), & \mathbf{C}_{\max} = \mathbf{G}' \\ 60^\circ \times \left(\frac{\mathbf{R}' - \mathbf{G}'}{\Delta} + 4 \right), & \mathbf{C}_{\max} = \mathbf{B}' \end{cases} \quad (8)$$

Saturation Calculation

$$\mathbf{S} = \begin{cases} 0, & \Delta = 0 \\ \frac{\Delta}{\mathbf{C}_{\max}}, & \Delta < > 0 \end{cases} \quad (9)$$

Value Calculation

$$V = C_{\max} \quad (10)$$

3.2 Edge Detection

In this phase, it detects edges of objects present in input image using edge detection technique. The Edge Detection technique detects the edges of the objects. So, we can easily separate the objects in given image. Then, the second phase is image segmentation, in which each object is separated by labeling each region and finally the third phase is shape recognition that identifies the shape of each object or region of an image and recognizes the object. Advantage of object detection through edge detection is changes in lighting and color usually do not have much effect on image edges. Different edge detectors are Sobel, Prewitt, Canny, etc. The disadvantages of Sobel and Prewitt are sensitivity to the noise, in the detection of the edges and their orientations. The increase in the noise to the image will eventually degrade the magnitude of the edges. Equation (11) shows masks used in Sobel and Prewitt edge detectors. Canny edge detector is able to produce single pixel thick, continuous edges, ability to detect strong and weak edges and its insusceptibility to noise interference. Canny edge detection algorithm has a better performance, but it is costly when compared to Sobel, Prewitt and Robert's operator and is cost-effective [6, 7, 8].

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (11)$$

Masks used in Sobel Kernel

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Masks used in Prewitt Kernel.

4 Feature Matching

In RGB image, each pixel has three color components: red, green, and blue. Amount of mixing of these three colors determines value of pixel. In HSV image, each pixel has only one color, which is represented by hue, saturation, and value components determining how much amount of black and white color is added into that color. We need to convert HSV image to binary image for specific object detection. Contours

are white areas in binary image. We will find each contour and calculate their area based on the area we will detect object. SIFT features are used to increase the detection speed. The scale invariant feature transform (SIFT) algorithm for image features generation which are invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine projection. Features of an image are interesting key points. Key points histogram values are used to index the image by using supervised learning techniques such as k-mean clustering [9].

5 Object Detection

By using multiple features helps to detect object accurately. Recognition rate for some objects is not satisfactory, and use of random feature results in poor recognition. To achieve a good recognition performance on particular classes of objects, it is not wise to randomly select intensity and color. We need to categorize depending on their characteristics. Texture is not similar for same kind of objects, so we need to split as different categories. Using SIFT, any specific textured object of this category can be recognized. By using contour and shape feature, the object is detected accurately. SIFT features can be increased by deformable part model and k-d tree [10].

6 Discussion and Results

From the observations for specific object recognition, SIFT would be a good choice. There is no single object recognition method that can work equally well on various types of objects and backgrounds. Rather, it must rely on multiple methods and should be able to select the appropriate one depending on the characteristics of the object and the surroundings. Color plays an important role in segmenting the objects from other objects. Figure 4 shows the segmenting object using color feature. Categorization is done on the basis of their characteristics. Edges and corners are robust in nature in different illumination conditions. Harris corner detection and SIFT are good feature extraction algorithms. Harris corner detector extracts the informative points in the scene; these features are enough for structured environment. In unstructured environment, we cannot expect productive feature points. SIFT can be effective feature in such environments. So, SIFT can be used as a feature detectors for our work. Figure 6 shows the feature matching using SIFT feature. Contour can be matched invariant to lighting condition and object color. All the objects in the image show similar feature then we can consider additional features like shape, size, etc., to classify object. Statistical measurements like mean, variance, skewness help in object detection [11]. Figure 7 shows object detection using contour feature. If objects in an image are showing similar features, then we need features to classify the objects of different kinds by using classifiers (Figs. 2, 3 and 5).

Fig. 2 Input image



Fig. 3 Target object



Fig. 4 RGB to HSV color conversion





Fig. 5 Binary image



Fig. 6 Feature matching

Fig. 7 Object detection



7 Conclusion and Future Work

In this work, we proposed object detection by using color, shape, and edge features. The proposed method is capable of detecting under different lighting conditions. Advantage of using organized dataset helps in knowing the relationship between the adjacent pixels and also nearest neighbor operations can perform effectively. For efficient object detection, features need to be strong for accurate object detection. In some cases, if the object is of small size so extracting features from the small object is difficult hence the processing time is increased. So small object feature extraction is difficult compared to larger objects. For accurate object detection, object must be of good size for feature extraction and processing.

Our method detects particular object from multiple objects we plan further to detect object using depth feature which helps to process under different illumination conditions. Advancement in technology made the consumer cameras such as Microsoft Kinect produce not only color image but also depth images. Microsoft Kinect is economically reliable compared to other sensing device. This work can be carried forward to estimate shape, color features of the objects using Kinect. Inclusion of these features in the system can lead to deployment of the system in real time.

References

1. J. Shotton, J. Winn, C. Rother, A. Criminisi, Texton Boost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comput. Vision* **81**, 2–23 (2009)
2. M. Varma, A. Zisserman, A statistical approach to material classification using image patch exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 2032–2047 (2009)
3. H. Wang, J. Oliensis, Rigid shape matching by segmentation averaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 619–635 (2010)
4. D. Lowe, Distinctive image features from scale-invariant key points. *IJCV* **60**(2), 91–110 (2004)
5. L.-C. Chen, X.-L. Nguyen, S.-T. Lin, Automated object detection employing viewing angle histogram for range images, in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics* (2012), pp. 196–201
6. B. Ommer, J. Buhmann, Learning the compositional nature of visual object categories for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(3), 501–516 (2010)
7. P. Carbonetto, G. Dorko, C. Schmid, H. Kuck, N. De Freitas, Learning to recognize objects with little supervision. *Int. J. Comput. Vision* **77**, 219–237 (2008)
8. Z. Si, H. Gong, Y.N. Wu, S.C. Zhu, Learning mixed templates for object recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 272–279 (2009)
9. K.K. Thyagarajan, R.I. Minu, prevalent color extraction and indexing. *Int. J. Eng. Technol.* **5** (6), (2013–2014)
10. J. Dou, J. Li, J. Li, Robust object detection based on deformable part model and improved scale invariant feature transform. *Int. J. Light Electron. Opt.* **124**, 6485–6492 (2013)

11. C. Richao, Y. Gaobo, Z. Ningbo, Detection of object-based manipulation by the statistical features of object contour. *Forensic Sci. Int.* **236**, 164–169 (2014)
12. J. Canny, A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–714 (1986)
13. C. Ma et al., An improved Sobel algorithm based on median filter, in *Institute of Electrical and Electronics Engineers, 2nd International IEEE Conference* **1**, pp. 88–93 (2010)
14. A. Seif et al., A hardware architecture of Prewitt edge detection, in *Sustainable Utilization and Development in Engineering and Technology, 2010 IEEE Conference*, pp. 99–101 (2010)
15. A.-L. Quintanilla, J.-L. Lopez-Ramirez, M.A. Ibarra-Manzano, Detecting objects using color and depth segmentation with kinectsensor, in *Iberoamerican Conference on Electronics Engineering and Computer Science*, pp. 196–204 (2012)
16. B. Leibe, A. Leonardis, B. Schiele, Robust object detection with interleaved categorization and segmentation. *Int. J. Comput. Vision* **77**, 259–289 (2008)
17. S. Tangruamsub, K. Takada, O. Hasegawa, 3D object recognition using voting algorithm in a real-world environment, in *2011 IEEE Conference on Applications of Computer Vision*, pp. 153–158 (2011)
18. A. Mansur, Y. Kuno, Integration of multiple methods for Robust object recognition, in *SICE Annual Conference*, pp. 1990–1995 (2007)
19. D. Lowe, Distinctive image features from scale invariant keypoints. *Int. J. Comput. Vision* **60** (2), 91–110 (2004)
20. T. Serre, L. Wolf, T. Poggio, A new biologically motivated framework for robust object recognition. *Ai memo* 2004–026 (2004)
21. C. Harris, M. Stephens, A combined corner and edge detector, in *Presented at the Alvey Vision Conference* (1988)
22. A. Opelt, A. Pinz, A. Zisserman, Learning an alphabet of shape and appearance for multi-class object detection. *Int. J. Comput. Vision* **80**, 16–44 (2008)
23. Z. Si, H. Gong, Y.N. Wu, S.C. Zhu, Learning mixed templates for object recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 272–279 (2009)

Service-Adaptive Fuzzy Multi Criteria Based Intelligent Vertical Handover Decision Algorithm for Heterogeneous Wireless Networks

V. Anantha Narayanan, A. Rajeswari and V. Sureshkumar

Abstract The next-generation wireless networks should be interoperable with other communication technologies to offer the best connectivity to the users.

But providing such interworking is a key challenge. Currently, existing decision engines are simple, proprietary, and have no support for profile-based service, and handover is only based on the received signal strength which is not intelligible enough to make handoff decision. The proposed decision algorithm gains intelligence by combining fuzzy logic system, multiple attribute decision making, and context-aware strategies. It synchronizes the user schedule and usage pattern through software agent for making handover at appropriate time. The performance analysis shows that the proposed algorithm efficiently uses the network resources by switching between 3G and Wi-Fi based on the user schedule. It is observed that average handover delay for the experiment is 25–35 ms under good RF conditions, running time of algorithm is around 1 ms, and it reduces the call dropping rate (<0.006), blocking probability (<0.00607), as well as unnecessary handover in heterogeneous networks.

Keywords Vertical handover · Fuzzy logic · Multiple attribute decision making · Context awareness · Software agent · User schedule

V.A. Narayanan (✉) · V. Sureshkumar
Department of CSE, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: veluananthu@gmail.com

V. Sureshkumar
e-mail: sureshkumarvcse@gmail.com

A. Rajeswari
Department of ECE, Coimbatore Institute of Technology, Coimbatore, India
e-mail: rajeswari.ece.cit@gmail.com

1 Introduction

A number of proposals have been made for vertical handover decision algorithms. This section reviews most of the algorithms with their significance. Qualcomm [1] designed a connectivity engine that uses dual-stack mobile IP for sending selected IP traffic to particular interface with support of simultaneous 3G and Wi-Fi access. But the decision engine is based on received signal strength which has already been proved as a not an intelligent solution. This bottleneck is addressed in proposed approach using multiple attribute decision-making algorithms. Akyildiz et al. [2] made a survey on mobility management techniques for next-generation IP-based wireless systems. The survey has all protocols to provide mobility management at different layers. Its review deals mostly with the algorithms used for decision making and its significance. This significance of each algorithm helps to fine-tune the proposed algorithm to obtain intelligence to the algorithm. Ning et al. [3] proposed fuzzy clustering-based group vertical handover decision by combining traffic of users and the status of available networks. Fuzzy clustering method assists user to choose best network with reduced handover-blocking probability. Grishaeva and Voropayeva [4] proposed fuzzy logic-based vertical handover decision algorithm to increase the quality of service and transparent mobility.

Even though these methods choose appropriate time for handover, it fails to use context awareness for choosing appropriate network. Buburuzan [5] presented a new handover model derived from the IEEE 802.21 standard which allows the seamless integration of broadcast technologies in a wireless heterogeneous environment. But it leads to major changes in the core architecture of cellular-WLAN architecture. The proposed algorithm applies decision algorithm over the cellular-UMTS integration architecture which does not require any change in the core architecture. Chen et al. [6] proposed a scheme based on bandwidth, dropping probability, and cost parameters as the metrics for the network selection function. These values are placed in target-visiting network to reduce the processing delay. The metrics collected are not stable in nature due to rapidly changing RF conditions. It requires frequent distribution of the distribution of the metrics collected, which increases the network traffic. The proposed algorithm handles this problem by processing the collected information locally in mobile terminal itself.

Kirsal et al. [7] proposed a Markov model for cellular/WLAN integration based on the policies. This model clearly differentiates requests originating in the cellular system, from requests being handed over from WLAN to cellular system. This ensures that calls handed over from WLAN to cellular are not handed over back to the WLAN. But the prediction of user movement of this algorithm makes it complicated to deploy in mobile terminal. The proposed algorithm reduces unnecessary handover and call dropping probability by using multiple attribute decision-making algorithms. Kassar et al. [8] and Mehbodniya et al. [9] presented an idea to eliminate the imprecision of data in the work of Ling et al. [10] by combining fuzzy logic with multiple attribute decision making. It includes received signal strength, QoS parameters, and mobile velocity attributes with analytic

hierarchy process as a weighting scheme. This algorithm works fine in indoor environmental conditions, but not for outdoor due to rapidly changing RF conditions. The proposed algorithm handles this problem by taking context-aware strategies in decision-making algorithm.

Ekiz et al. [11] proposed fuzzy logic–analytic hierarchy process-based handover decision algorithm with received signal strength as the handover parameter. But it is proven that the decision only based on the received signal strength is not intelligent enough. This triggers unwanted handover which causes poor quality of service faced by the user. Ekiz et al. [12] proposed vertical handover decision based on the fuzzy inference system and clustering method, and they have shown that their proposed method enables fast handover between heterogeneous network users. But it fails to use context awareness for choosing appropriate network, and it does not adopt based on the service.

The proposed algorithm combines fuzzy logic system with multiple attribute decision making (FMADM) and context-aware strategy for including the current status of the target network and the application network resource requirements into consideration. It synchronizes the user schedule and usage pattern for making handover at appropriate time without any processing to reduce the load on the decision engine. The multiple attributes for making handoff and application network resource requirements are collected using the software agent to reduce the processing time of decision engine.

2 Service-Adaptive FMADM-Based Decision Algorithm

The proposed algorithm (Fig. 1) uses multiple attributes from the network for intelligent vertical handover decision making. These multiple attributes are collected using software agents for reducing the workload on the decision engine. The software agent is a program that acts on behalf of another software program (decision engine) to retrieve and to process information. Once the software agent identifies the new network, it starts to collect the information that is required to make handover decision. Sometimes, the discovered network may offer poor service quality. At that case, processing of information from the discovered network is useless. To avoid such a problem, the software agent checks the received signal strength, service quality, and service/network load on the discovered network. This process helps to avoid the processing of poor-connectivity network. The software agent also collects the usage pattern, preferred network interface of user through learning to make an intelligent solution. Sometimes, the discovered network may have good received signal strength, and service quality.

But there may be a chance that all channels are occupied. If the decision engine makes the handover to that network, then it leads to poor connectivity to the user. This can be avoided by using a software agent that transmits load on the network at every beacon of network. The collected information using software agents, user schedule, and usage pattern are assigned with predefined weights and then passed

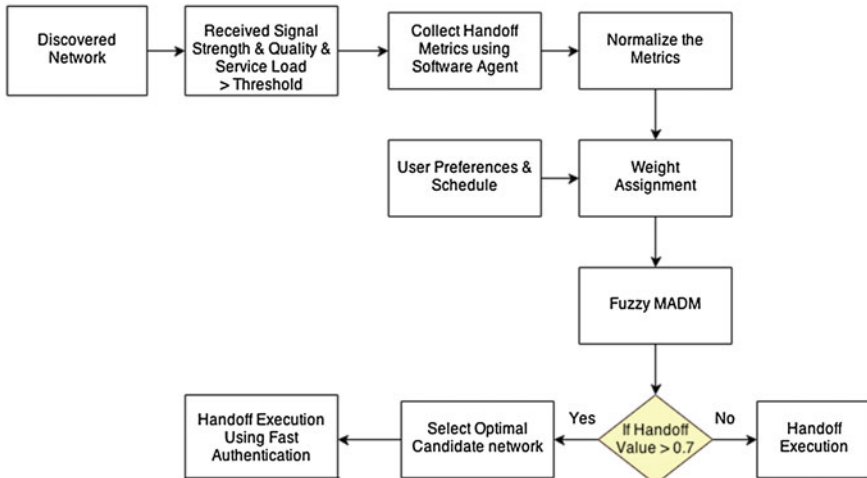


Fig. 1 Working flow of intelligent vertical handoff decision algorithm

into fuzzy-based multiple attribute decision-making algorithm. The collected information from different network is converted into common scale (out of 10 scales), and these collected values are analog in nature. But the analog values cannot be processed by the fuzzy logic systems, and it should be converted into linguistic values (fuzzification/aggregation) on the basis of predetermined membership functions. Then, the linguistic values are fed into a fuzzy inference engine. The inference engine applies IF-THEN rules to obtain fuzzy decision sets (activation). The output fuzzy decision sets include linguistic values like strongly yes, yes, uncertain, no, and strongly no. This output fuzzy sets are combined into a single fuzzy set (accumulation) and defuzzified into analog values using defuzzifier by applying center of gravity method (COG).

COG: $U = \frac{\int_{Min}^{Max} u \mu(u) du}{\int_{Min}^{Max} \mu(u) du}$, where U , u , p , μ , Min , and Max refers the result of defuzzification, output variable, number of singletons, membership function after accumulation, lower limit for defuzzification, and upper limit for defuzzification, respectively. Handover decision is based on the defuzzified value to choose the appropriate time and the most suitable access network according to user preferences. If the candidate network gets handoff probability greater than 0.7 (optimal value), then decision algorithm applies “make before break handover” and does the regular fast authentication and mobile IP registration. Once the selected network interface becomes active, the decision algorithm seamlessly routes the traffic to the selected interface with minimum packet loss.

3 Results and Analysis

The proposed approach has been implemented in Ubuntu-based machine for measuring accurate efficiency and effectiveness of the proposed algorithm. The decision algorithm uses common parameters to support integration of all high-data rate technologies. But currently, most of the mobile phones and laptop machines have the support for only Wi-Fi and 3G connectivity. So for implementation, only these two technologies have been taken into account.

3.1 Weight Assignment

The proposed algorithm uses signal strength, quality of service, service cost, power requirements, mobile velocity, location information, data rate, network latency, user preferences, and security. This collected information using software agents, user schedule, and usage pattern has to be assigned with predefined weights before passing into fuzzy-based multiple attribute decision-making algorithm. The service from the cellular and WLAN operators is classified into real-time and non-real-time services, and optimal weights are assigned using trial and error method.

Handoff value (UMTS_Non-Real time) = $0.148 * (\text{Remaining battery capacity}) + 0.095 * (\text{Received signal strength}) + 0.098 * (\text{Link quality Indication}) + 0.092 * (\text{Data rate supported}) + 0.082 * (\text{Network latency}) + 0.088 * (\text{Service cost}) + 0.213 * (\text{Mobile velocity}) + 0.222 * (\text{Network coverage}) + 0.026 * (\text{Security})$.

Handoff value (WLAN_Real time) = $0.154 * (\text{Remaining battery capacity}) + 0.085 * (\text{Received signal strength}) + 0.168 * (\text{Link quality Indication}) + 0.126 * (\text{Data rate supported}) + 0.092 * (\text{Network latency}) + 0.108 * (\text{Service cost}) + 0.153 * (\text{Mobile velocity}) + 0.082 * (\text{Network coverage}) + 0.032 * (\text{Security})$.

Figure 2 shows the user profile for the student and user preference and probability to make handover to the particular candidate network. In decision engine text area, it clearly shows that running time of this algorithm to make decision is around 1 ms. This running time is relatively less when compared to other approaches.

3.2 Device Interoperability

The proposed approach uses software-defined networking to handle the device interoperability. The decision engine manages the network services through lower level of functionality by decoupling the wireless access technology into control plane and data plane.

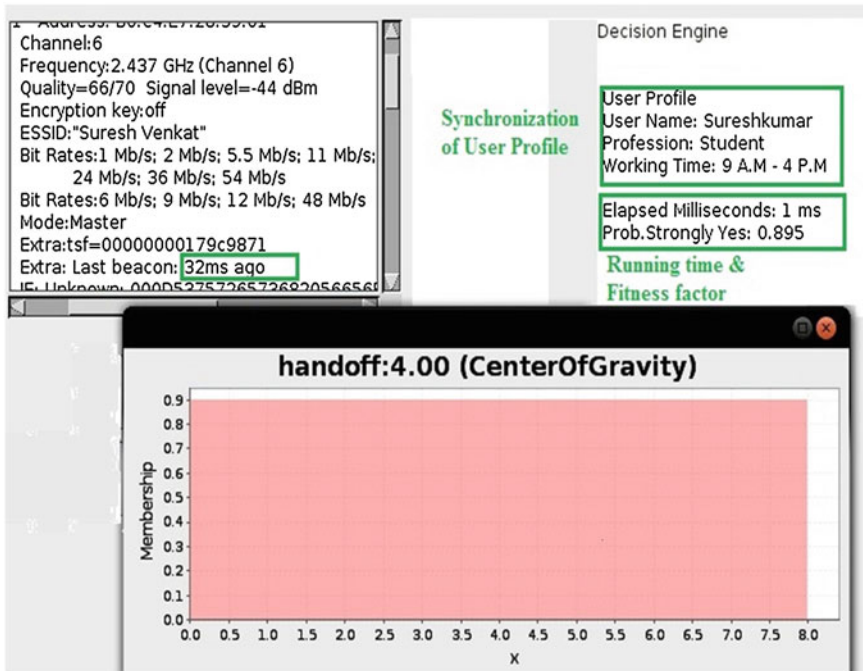


Fig. 2 User profile synchronization, running time, and fitness factor

The control plane selects the interface to route the traffic, and the data plane will route the traffic into the selected interface. To support interoperability between devices, most of the data card has been studied and common dialing, redialing, and disconnect features are extracted. These features are coded using scripts and used in the decision engine to obtain handoff metrics independent of devices. This novel approach in the proposed approach helps to support different data cards including Huawei E303C, Huawei E303U, Huawei E3121, Huawei E355S, Huawei E173, Huawei E 303, Huawei E303S, and Reliance CDMA data cards on most of the carriers.

Figure 3 shows that when number of handover increases, the packet loss (<0.022 %) and round trip time (deviation (<2.5 ms)) also increase due to high signaling loads. Call dropping and blocking probability are the most important factor in traffic usage during the handover. The experiment was conducted with total number of 500 calls (Table 1); the calling rate was 20/h, and call holding time was 120 s. It was observed that the total number of blocked call was 3 and dropped call was 2.

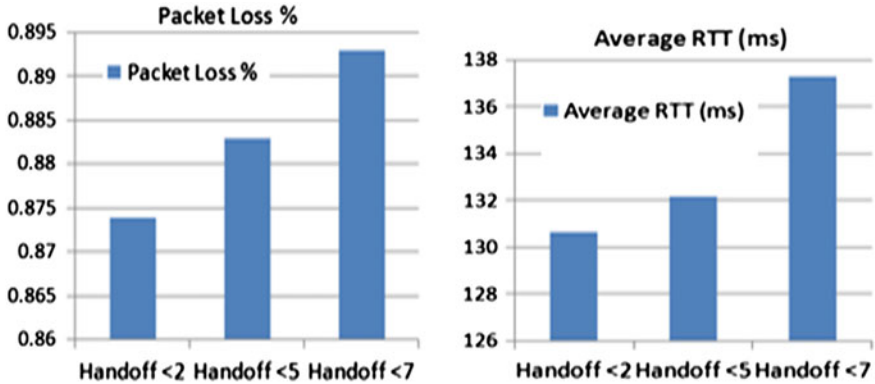


Fig. 3 Packet loss and average round trip time (RTT) analysis

Table 1 Call blocking probability and drop call rate

Number of handoffs	Blocking probability in busy hours with handover	Drop call rate in busy hours with handover
Handoff <3	3/494	3/500
Handoff <5	3/494	3/500
Handoff <7	2/500	4/500

Blocking Probability = (Number of lost calls)/(Total number of offered calls) and Drop call rate = (Number of dropped calls)/(No of call attempts), Blocking Probability in busy hours = (3/494) = 0.00607 and Drop call rate in busy hour = (2/500) = 0.004

4 Conclusion

The proposed intelligent vertical handoff decision algorithm uses a software agent to collect and process handoff information to reduce the load on the decision engine, and the user profile is given as the input to the decision engine to reduce the processing time. The reduction in the computational time reduces the handover delay and power required to compute handover decision. The proposed approach uses novel scheme to handle device interoperability by splitting control plane from the data plane. The performance analysis shows that the effectiveness of proposed algorithm in terms of minimal packet loss (<1 %), running time (1 ms), average round trip time (<137 ms), and efficient resource utilization is based on the application requirements, and the proposed approach fulfills QoS requirements of audio, video, and data in terms of packet loss and handover delay during the handover as recommended by Cisco Systems. It is observed that average handover delay for the experiment is 25–35 ms, and the proposed intelligent decision algorithm reduces the call dropping rate (<0.006), call blocking probability (<0.00607), as well as unnecessary handover in heterogeneous networks. The proposed algorithm uses predefined weights to choose the target network. If unsupervised

learning algorithm is used to adapt automated weighting, then decision will perform better, and this work has been taken for the future work.

Acknowledgments We are highly indebted to the authorities of Mobile and Wireless Networks Research Laboratory of CSE department of Amrita Vishwa Vidyapeetham for providing necessary hardware resources and test bed for carrying out this research work.

References

1. Qualcomm Technology, 3G/LTE Wi-Fi Offload Framework: Connectivity Engine (CnE) to Manage Inter-System Radio Connections and Applications (2012)
2. I.F. Akyildiz, J. Xie, S. Mohanty, A survey of mobility management in next-generation all-IP-based wireless systems. *Wirel. Commun.* **11**(4), 16–28 (2004)
3. L. Ning, Z. Wang, Q. Guo, K. Jiang, Fuzzy clustering based group vertical handover decision for heterogeneous wireless networks, in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, vol. 7(10) (IEEE, 2013), pp. 1231–1236
4. A.D. Grishaeva, V.Y. Voropayeva, Development of the vertical handover algorithm for heterogeneous wireless networks, in *23rd International Crimean Conference on Microwave and Telecommunication Technology*, vol. 8(14) (2013), pp. 480–481
5. T. Buburuzan, Performance evaluation of a handover model for integrating mobile broadcast technologies within heterogeneous networks, in *IEEE 13th International Symposium* (2009), pp. 603–607
6. J. Chen, Z. Wei, Y. Wang, L. Sang, D. Yang, A service-adaptive multi-criteria vertical handover algorithm in heterogeneous wireless networks, in *Personal Indoor and Mobile Radio Communications* (2012), pp. 899–904
7. Y. Kirsal, E. Ever, G. Mapp, O. Gemikonakli, Enhancing the modeling of vertical handover in integrated cellular/WLAN environments, in *Advanced Information Networking and Applications (AINA)* (2013), pp. 924–930
8. M. Kassar, B. Kervella, G. Pujolle, An overview of vertical handover decision strategies in heterogeneous wireless networks. *Comput. Commun.* **31**(10), 2607–2620 (2008)
9. A. Mehbodniya, F. Kaleem, K.K. Yen, F. Adachi, A fuzzy MADM ranking approach for vertical mobility in next generation hybrid networks, in *Ultra Modern Telecommunications and Control Systems and Workshops* (2012), pp. 262–267
10. Y. Ling, B. Yi, Q. Zhu, An improved vertical handover decision algorithm for heterogeneous wireless networks. *Wirel. Commun. Netw. Mob. Comput.* **1**(3), 12–14 Oct 2008
11. L. Ekiz, C. Lottermann, D. Ohmann, T. Tran, O. Klemp, C. Wietfeld, C.F. Mecklenbrauker, Potential of cooperative information for vertical handover decision algorithms, in *Intelligent Transportation Systems. 16th International IEEE Conference*, vol. 6(9) (2013), pp. 455–460
12. M.T. Ekiz, M.L. Hossain, M.A. Kabir, M.T. Rahman, S. Salekin, S.S. Alam, A.F. Mitul, Vertical handover decision using fuzzy logic in a heterogeneous environment, in *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, vol. 17(18) (2013), pp. 1–3

Automatic Traffic Classification Using Machine Learning Algorithm for Policy-Based Routing in UMTS–WLAN Interworking

V. Anantha Narayanan, V. Sureshkumar and A. Rajeswari

Abstract The future mobile terminal will be dependent on the multiple wireless access technology simultaneously for accessing Internet to offer best Internet connectivity to the user. But providing such interworking among wireless heterogeneous networks and routing the selected traffic to particular wireless interface is a key challenge. Currently, existing algorithms are simple and proprietary, and there is no support to route the specific application traffic automatically. The proposed decision algorithm finds the optimal network by combining fuzzy logic system with multiple-attribute decision-making and uses naïve Bayes classifier to classify the application traffic to route into appropriate interface to reduce the service cost. The performance analysis shows that the proposed algorithm efficiently uses the network resources by maintaining active connection simultaneously with 3G and Wi-Fi. It routes 71.99 % of application traffic using Wi-Fi network and 28.008 % of application traffic using UMTS network to reduce the service cost and to reduce network load on the cellular operator.

Keywords Internet traffic classification · 3G and Wi-Fi · UMTS network · Fuzzy logic multiple-attribute decision-making

V. Anantha Narayanan (✉) · V. Sureshkumar
Department of CSE, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: veluananthu@gmail.com

V. Sureshkumar
e-mail: sureshkumarvcse@gmail.com

A. Rajeswari
Department of ECE, Coimbatore Institute of Technology, Coimbatore, India
e-mail: rajeswari.ece.cit@gmail.com

1 Introduction

A number of proposals have been made for traffic identification in Internet traffic classification. This section reviews most of the algorithms with their significance. While traffic classification techniques are improving in accuracy and efficiency, the behavior of Internet application changes rapidly. It makes the classification algorithm to exhibit high false positives. To route the specific application traffic without manual intervention requires more clear identification of the traffic actually originating from the application. Traffic classification in vertical handover problem does not require high accuracy, but demands minimum time to classify the traffic to route in specific wireless interface. Hu and Shen [1], Nguyen and Armitage [2], and Dainotti and Pescapè [3] presented a survey about constructive analysis of the supervised and unsupervised traffic classification algorithms, its achievements, and obstacles to progress. Erman et al. [4] presented an expectation maximization-algorithm-based clustering algorithm for Internet traffic classification and shown that their approach achieves an accuracy up to 91 %. But it does not perform well for unknown traffic. If the number of clusters is large, then EM algorithm creates outfitting problem. Kirsal et al. [5] proposed a Markov model for cellular/WLAN integration based on the policies. But the policy is defined for avoiding frequent handover during single voice call. The proposed algorithm reduces call-dropping probability by maintaining active connection in both interfaces. Ning et al. [6] proposed fuzzy clustering-based vertical handover decision by combining traffic of users and the status of available networks. Fuzzy clustering method assists user to choose best network with reduced handover blocking probability. Grishaeva and Voropayeva [7] proposed fuzzy logic-based vertical handover decision algorithm to increase the quality of service and transparent mobility. Even though these methods choose appropriate time for handover, it fails to use context awareness for choosing appropriate network.

Kassar et al. [8] and Mehboodniya et al. [9] proposed a fuzzy logic-based multiple-attribute decision-making which includes received signal strength, QoS parameters, and mobile velocity attributes with analytic hierarchy process as a weighting scheme. This algorithm works fine in indoor environmental conditions. But in outdoor, it cannot perform well due to rapidly changing RF conditions. The proposed algorithm handles this problem by taking context-aware strategies in decision-making algorithm. Wang et al. [10] presented an application-based traffic classification for unknown traffic by deriving its signature. It combines the traffic clustering algorithm with statistical flow properties using signature construction from payload and proven that it is highly effective. Wang et al. [11] proposed a classification algorithm using machine learning by pattern matching with common substrings in the payloads and review about the classification algorithms. Zander et al. [12] proposed a novel method for ML-based flow classification and application identification based on statistical flow properties using unsupervised machine learning. The average accuracy across all traces is 86.5 %. But all these presented algorithms are taking considerably maximum classification when it is applied to the network with gigabits of traffic.

This problem has formed as fuzzy multiple-attribute decision-making (FMADM) for identifying the quality of service the network can offer. Automatic traffic classifier identifies well-known application traffic based on the application identifier, and unknown application traffic is analyzed using naïve Bayes algorithm. The proposed approach identifies the traffic without relying on destination IP and port by using throughput threshold limitation, by setting download size threshold for interface, and by identification using application unique identifier.

2 Proposed Automatic Traffic Classification for Policy-Based Internet Traffic Routing

2.1 Identification Best Network to Route High Demand Traffic

The proposed algorithm (Fig. 1) uses software agent which is generally a program used to collect information about the available networks, application requirements, and network conditions. The network condition includes available bandwidth, received signal strength, link quality, signal-to-noise ratio, security, and remaining battery capacity. These collected multiple attributes using software agent help to reduce work load on the decision engine. The software agent runs periodically to collect the information. Sometimes, the discovered network may offer poor service quality. At that case, processing information from that network is useless. To avoid such a problem, the software agent checks the received signal strength, service quality, and service/network load on the discovered network. This process helps to avoid the processing of poor connectivity network. The collected information using software agents along with the usage pattern is assigned with predefined weights and then passed into fuzzy-based multiple-attribute decision-making algorithms.

The collected analog information from different networks is normalized into common scale using the fuzzifier through the membership values. These membership values help to reduce the impact of error during the experiment. These membership values are assigned with preferred weight based on the type of

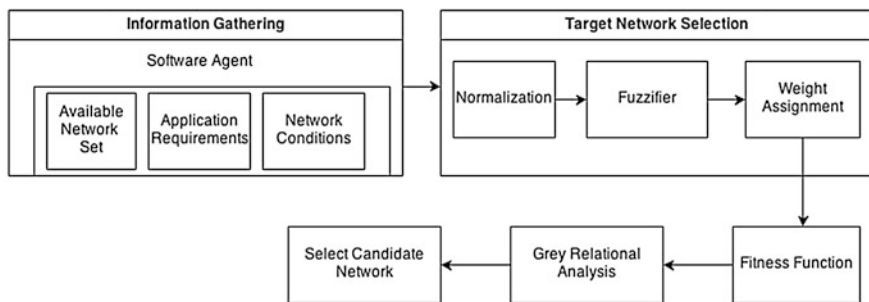


Fig. 1 Working flow in the identification of optimal network

application. For example, network latency plays a major role in voice interactive services than network bandwidth. Fitness function value is calculated by sum of the product of attributes and their associated weights. The fitness values and the user preferences are given to the gray relational analysis to rank the network. The network with best rank will be selected to route high network-resource-demanding application's traffic. Once the selected network interface becomes active, the decision algorithm seamlessly routes the selected traffic to the selected interface with the minimum packet loss.

2.2 Classification of Application Traffic

The proposed algorithm uses naïve Bayes supervised learning algorithm to automatically classify and identify network applications based on selected features. The naïve Bayes algorithm automatically builds a classifier by learning the inherent structure of application traffic by using features in the application traffic.

The supervised naïve Bayes algorithm (Fig. 2) has two steps: First one is to build model using training data set and second is to use the model to classify the test data set by the classifier. The naïve Bayes algorithm estimates the Gaussian distribution of the attributes for each class based on labeled training data for building classifier model. This classifier model is used to classify the real-time application traffic based on the conditional probability of the traffic belongs to a class based on the selected features.

2.3 Routing the Application Traffic Based on User Policy

Figure 3 shows the automatic traffic classifier that identifies well-known application traffic based on the application identifier. The unknown application traffic is analyzed using naïve Bayes algorithm. The naïve classifier output helps to identify the application type and its QoS requirements. The QoS requirements of the application traffic are mapped with application profile to route based on the policy. For example, most of the users prefer the WLAN network to handle bulk and transactional application traffic and interactive application traffic through UMTS.

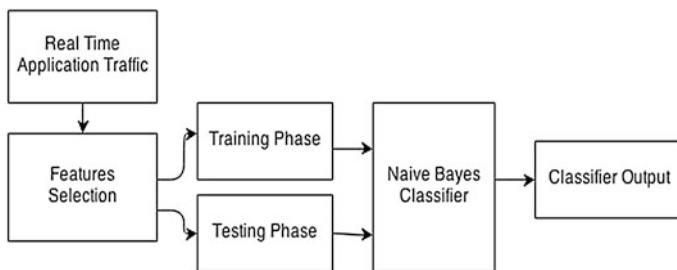


Fig. 2 Classification of application traffic

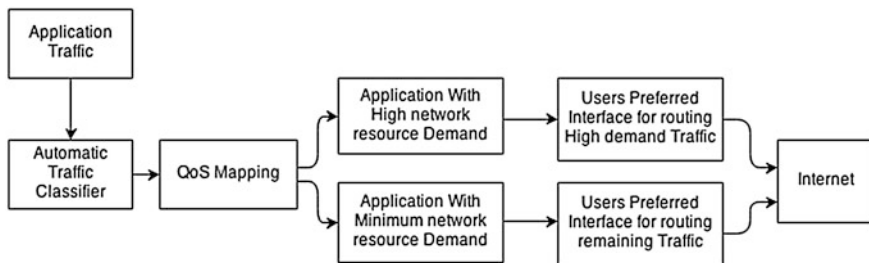


Fig. 3 Routing the traffic based on user policy

3 Results and Analysis

The proposed automatic traffic classification for enabling policy-based routing in UMTS–WLAN interworking is implemented using Wireshark 1.8.2 for real-time traffic data set collection and Weka 3.6.9 for naïve Bayes classifier. The time complexity of naïve Bayes classifier is $O(\text{features}) + O(\text{classes})$ [2].

3.1 Defining Application Profiles

Based on the QoS requirements, the application has to be mapped with application profiles. The total application traffic is classified into five major categories based on the Cisco recommendation (Table 1).

3.2 Features Considered

To classify the real-time application traffic, the naïve Bayes classifier has to be trained using the training data set. Every application will have some sort of pattern

Table 1 Defining application profile

Application type	Example application	Application properties	Message size
Interactive	Telnet, Oracle Thin-Clients, Yahoo! Instant Messenger, Conference	Highly interactive applications	Max message size <1 KB
Transactional	B2B, Application Server, Oracle 8i Database, Microsoft SQL	TCP and FTP sessions running simultaneously	Size varies from 1 KB to 50 MB
Bulk	Database syncs, network-based backups, video content distribution, large FTP file transfers	Long file transfers	Average message size 64 KB or greater
Best effort	All non-critical traffic, HTTP Web browsing	HTTP-based applications	Average packet size 557 bytes

to send data that can be used as features to the naïve Bayes classifier to classify the application traffic. The training data set includes the unique feature that identifies application traffic of associated class (protocol). The features used in this approach are packet level (packet length, byte counts, idle time), flow level (flow duration, data volume per flow, number of packets per flow), connection level (window sizes, throughput distribution, connection duration), and connection features (packet interarrival statistics). Along with these features, traffic and application profiles are included. This feature helps to assist in the identification and classification process of real-time application traffic. Table 2 shows the results of the proposed approach.

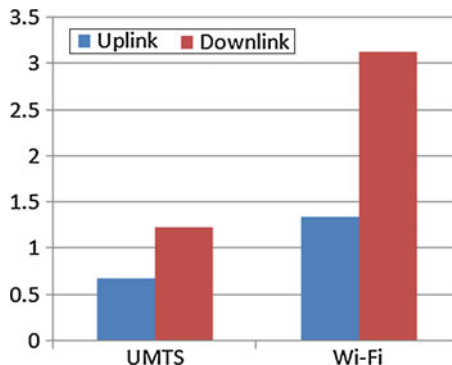
The naïve Bayes classifier has accuracy up to 88.28 % in classifying real-time application traffic with classification time of 0.08 s. But it is found that unsupervised clustering technique [2] has accuracy up to 91 %, but it takes more classification time (2,100 s) to classify 1,000 samples. But in real time, the traffic may be relatively higher than 1,000, and in such cases, the unsupervised learning algorithm will take long time to classify the traffic. But the traffic classification for policy-based routing in vertical handover decision-making problem does not require much accuracy and efficiency but demands minimum classification time. Hence, unsupervised approach is not suitable for policy-based routing in vertical handover problem.

Figure 4 shows the amount of traffic routed through UMTS and WLAN using the proposed approach. It clearly shows that high network-resource-demanding application traffics are routed through WLAN interface.

Table 2 Data set details and accuracy of the proposed algorithm

Naïve Bayes classifier	Number of data set samples	Time taken to build model (s)	Traffic class	Correctly classified instances	Incorrectly classified instances	Precision TP/ TP + FP	Recall TP/ TP + FN
Training phase	2,628	0.06	14	2,320 (88.28 %)	308 (11.72 %)	0.869	0.883
Testing phase	21,453	0.08	17	18,909 (88.14 %)	2,544 (11.86 %)	0.875	0.881

Fig. 4 Policy-based routing of application traffic in UMTS–WLAN interworking



4 Conclusion

The proposed approach uses fuzzy logic multiple-attribute decision-making to identify the optimal network to route high demand/bulk data. It maintains active connections in UMTS and WLAN interface simultaneously to offer best connectivity to the user. This algorithm is capable of forwarding data packets to appropriate attachment point to maximize battery lifetime and to maintain load balancing. The application traffic is classified using naïve Bayes classifier, and it shows that naïve Bayes has an accuracy of 88.28 % with the classification time of 0.08 s for data set with 21,453 samples. The proposed algorithm routes 71.99 % of application traffic through Wi-Fi network, and 28.008 % of application traffic is through UMTS network to reduce the service cost. This work can be extended to increase the accuracy of the traffic identification with suitable unsupervised learning algorithm which exhibits minimum classification time and reduced manual intervention in classification.

Acknowledgments We are highly indebted to the authorities of Mobile and Wireless Networks Research Laboratory of CSE Department of Amrita Vishwa Vidyapeetham for providing necessary hardware resources and test bed for carrying out this research work.

References

1. B. Hu, Y. Shen, Machine learning based network traffic classification: a survey. *J. Inf. Comput. Sci.* **9**, 3161–3170 (2012)
2. T.T.T. Nguyen, G. Armitage, A survey of techniques for internet traffic classification using machine learning. *Commun. Surv. Tutorials IEEE* **10**(4), 56–76 (2008)
3. A. Dainotti, A. Pescapé, K.C. Claffy, Issues and future directions in traffic classification. *Netw. IEEE* **26**(1), 35–40 (2012)
4. J. Erman, A. Mahanti, M. Arlitt, QRP05-4: internet traffic identification using machine learning, in *Global Telecommunications Conference* (2006) pp. 1–6
5. Y. Kirsal, E. Ever, G. Mapp, O. Gemikonakli, Enhancing the modeling of vertical handover in integrated cellular/WLAN environments, in *Advanced Information Networking and Applications* (2013) pp. 924–930
6. L. Ning, Z. Wang, Q. Guo, K. Jiang, Fuzzy clustering based group vertical handover decision for heterogeneous wireless networks, in *Wireless Communications and Networking Conference (WCNC)*, vol. 7(10) (IEEE, 2013) pp. 1231–1236
7. A.D. Grishaeva, V.Y. Voropayeva, Development of the vertical handover algorithm for heterogeneous wireless networks, in *Microwave and Telecommunication Technology. 23rd International Crimean Conference*, vol. 8(14) (2013) pp. 480–481
8. M. Kassar, B. Kervella, G. Pujolle, An overview of vertical handover decision strategies in heterogeneous wireless networks. *Comput. Commun.* **31**(10), 2607–2620 (2008)
9. A. Mehbodniya, F. Kaleem, K.K. Yen, F. Adachi, A fuzzy MADM ranking approach for vertical mobility in next generation hybrid networks, in *Ultra Modern Telecommunications and Control Systems and Workshop* (2012) pp. 262–267
10. Y. Wang, Y. Xiang, S.Z. Yu, Automatic application signature construction from unknown traffic, in *Advanced Information Networking and Applications IEEE* (IEEE, 2010) pp. 1115–1120

11. Y. Wang, Y. Xiang, S. Yu, Internet traffic classification using machine learning: a token-based approach, in *Computational Science and Engineering*, (IEEE, 2011) pp. 285–289
12. S. Zander, T. Nguyen, G. Armitage, Automated traffic classification and application identification using machine learning, in *Local Computer Networks*, (IEEE, 2005) pp. 250–257

Sequential Decision Making Using Q Learning Algorithm for Diabetic Patients

Pramod Patil, Parag Kulkarni and Rachana Shirsath

Abstract In sequential decision making, we program agent by reward and punishment. In this, agent learns to map situations to actions which results in maximizing rewards gained. This agent is also known as decision makers. It is difficult to take decision about giving specific kind and quantity of insulin dose to the diabetes patient in a critical system of insulin pump control. This paper implements the Q learning algorithm on diabetes data streams. This helps in classifying the data for diabetes dose and also helps in making decision about giving particular kind and quantity of insulin dose by generating various rules.

Keywords Decision making · Diabetes · Reinforcement learning · Q learning algorithms

1 Introduction

Learning is usually formulated as a search conducted in an abstractly defined space, and a large collection of understanding and designing procedures, or algorithms, for enabling a device or program to improve its performance over time. The behavior observed in classical conditioning experiments is far from computationally trivial; its strongest ties are to mathematical theories and computational procedures that are exceedingly useful in practice and surprisingly complex.

The state of system at a particular time is a description of the condition of the system at that time that it is sufficient to determine knowledge of the system's future input. Whatever happened to the system in the past that is relevant to its future

P. Patil (✉) · P. Kulkarni
College of Engineering Pune, Pune, India
e-mail: pdpatiljune@gmail.com

R. Shirsath
Dr. D.Y. Patil Institute of Engineering and Technology, Pimpri, Pune, India
e-mail: rachana.deshmane@gmail.com

behavior is summed up in its current state-future behavior does not depend on how the system arrived at its current state. In a decision-making system which is simply known as agent, the agent interacts with a system in such a way that the beginning of each of a series of discrete time periods it observes the system and is able to determine the systems state at that time. On the basis of observed state, the agent performs an action, thereby causing the system to deliver to the agent a ‘payoff,’ which we think of as a number whose value depends on the system state, on the agent’s action. The system then makes a transition to a new state determined by its current state, by the agent’s action. Upon observing the new state, the agent performs another action and receives another payoff, and the system changes state again. This cycle of state observation, action, payoff, and state change repeats for a sequence of time periods.

In case of critical system for controlling insulin pump, it is necessary to analyze the blood glucose level (BGL). After blood glucose analysis, the insulin requirement is computed, and then, insulin is given to the patient. In case of diabetic patient, it is necessary to take decision of insulin dose according to his historical data and symptoms. The proposed methodology helps in making decisions about insulin dose depending on the preconditions.

2 Literature Survey

Kaelbling et al. [1] have discussed issues of reinforcement learning (RL), which include trading off exploration and exploitation. It also provides the foundations of field using Markov decision theory.

Wiering and van Hasselt [2] have presented several ensemble methods used for combining multiple different RL algorithms using single agent. Learning speed and final performance get enhanced by combining the chosen actions or action probabilities of different RL algorithms.

Mill’an-Giraldo et al. [3] have discussed about some applications in which data arrive sequentially and they are not available in batch form, what makes difficult the use of traditional classification systems. In addition, some attributes may lack due to some real-world conditions. For this problem, a number of decisions have to be made regarding how to proceed with the incomplete and unlabeled incoming objects, how to guess its missing attributes values, how to classify it, whether to include it in the training set, or when to ask for the class label to an expert.

Zhiguo Shi et al. [15] have discussed about the Q learning to the swarm robot system. Q learning estimates the value function of the state and action. Q learning studies the mapping from environment status to action. The robot discovers which action should gain reward. So that selected action affects the current reward as well as next state.

3 Mathematical Framework

Sequential decision-making system has a finite set of states, X . At any time step $t = 0, 1, 2, \dots, n$, the system can be in a state $x \in X$. After observing the system state at time step t , the agent selects an action from a finite set of possible action A . Suppose that at time step t , the agent observes state X and selects action A . Independent of its history, the system makes a transition from state X to state Y .

In case, an agent begins interacting with a system at time step $t = 0$ and is able to continue for an infinite number of time steps. Using discount factor, the measure of the total amount of payoff that the agent will receive over this infinite time period is

$$r_1 + \gamma r_2 + \gamma^2 r_3 + \dots + \gamma^n r_{n+1} \tag{1}$$

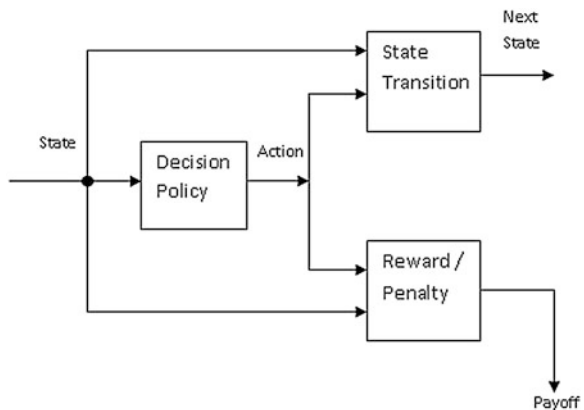
where $0 < \gamma < 1$ and the powers of weighting the payoffs form a decreasing sequence so that later payoffs are weighted less than earlier once, with payoffs in the far distant future contributing negligibly to the return. The discount factor adjusts the degree to which the long-term cost of actions must be accounted for in a decision takes and influences the rate at which learning occurs (Fig. 1).

The expected value of this sum over all possible decision tasks starting with initial state x when the agent uses policy π is where indicates that the expected value is taken under the assumption that policy π is used to select actions. We can think of this quantity as the value of a function, denoted V_π , assigning an expected return to each system state X :

$$v_\pi(x) = E_\pi \left[\sum_{i=1}^{\infty} \gamma^i r_i + 1 | x_0 = x \right]. \tag{2}$$

The function is the evaluation function for policy π . For each state x , $V_\pi(x)$ is the expected return over the infinite number of time steps beginning at $t = 0$ under the conditions that the system begins in state x and the agent uses policy π , where it is

Fig. 1 State of action and payoff change over time



understood that the discount factor has some specified value. We call $V\pi(x)$ the evaluation of state x . For the route finding example, the evaluation of location x for policy π depends on the expected number of time steps that agents take to reach the goal from location x using policy π . If π always brings the agent to the goal in say, n time steps from a location x , then.

Let π and π' be any two policies. Policy π' is an improvement over policy π if the expected return of π' is no smaller than that for π for any state and is larger for at least one state. More precisely, π' is an improvement over π if $V\pi'(x) \geq V\pi(x)$ for all states x , with strict inequality holding for at least one state. A policy is an optimal policy, which we denote π^* , if no policy is an improvement over it. Because the optimality of policies depends on the discount factor, technically we should refer to γ -optimal policies. As γ changes, different policies become optimal because a policy best for short-term versions of a task will generally not be best for long-term versions. Whatever policies are compared, they are compared according to expected returns defined for the same γ . For any optimal policy π^* ,

$$v\pi^*(X) \geq v\pi(X). \quad (3)$$

For all states x , an agent using an optimal policy will maximize the expected return from any system state. The object of a sequential decision task is to find one of the optimal policies.

4 Proposed Methodology

In case of critical system for controlling insulin pump, it is necessary to analyze the BGL. After blood glucose analysis, the insulin requirement is computed, and then, insulin is given to the patient. Depending on previous conditions of the patient, the decision about the next dose is taken (Fig. 2).

We have defined different states and action like.

Initially, we have to take training raw data, and then, $Q(s, a)$ is assigned arbitrary. After assigning Q value, the Q learning algorithm is applied. The patient blood glucose is analyzed, and it has various states as shown in Table 1.

After blood glucose analysis, we have to take action, i.e., if BGL is constant, then we give the constant insulin dose to patient, or else, we have to compute the requirement of insulin dose. There are different types of insulin doses that act as actions in system. These actions are as follows: normal, NPH, and UltraLente.

Our system helps in making the decision regarding the quantity and kind of insulin dose. Depending on previous conditions of the patient, the decision (i.e., kind of dose and quantity of dose) which has maximum rewards is chosen. If we achieve goal, i.e., for next measurement if BGL of patient is constant, then we assign reward to the decision made else penalty is given. Values for reward and punishment are as follows: Reward = +1 and Penalty = -1.

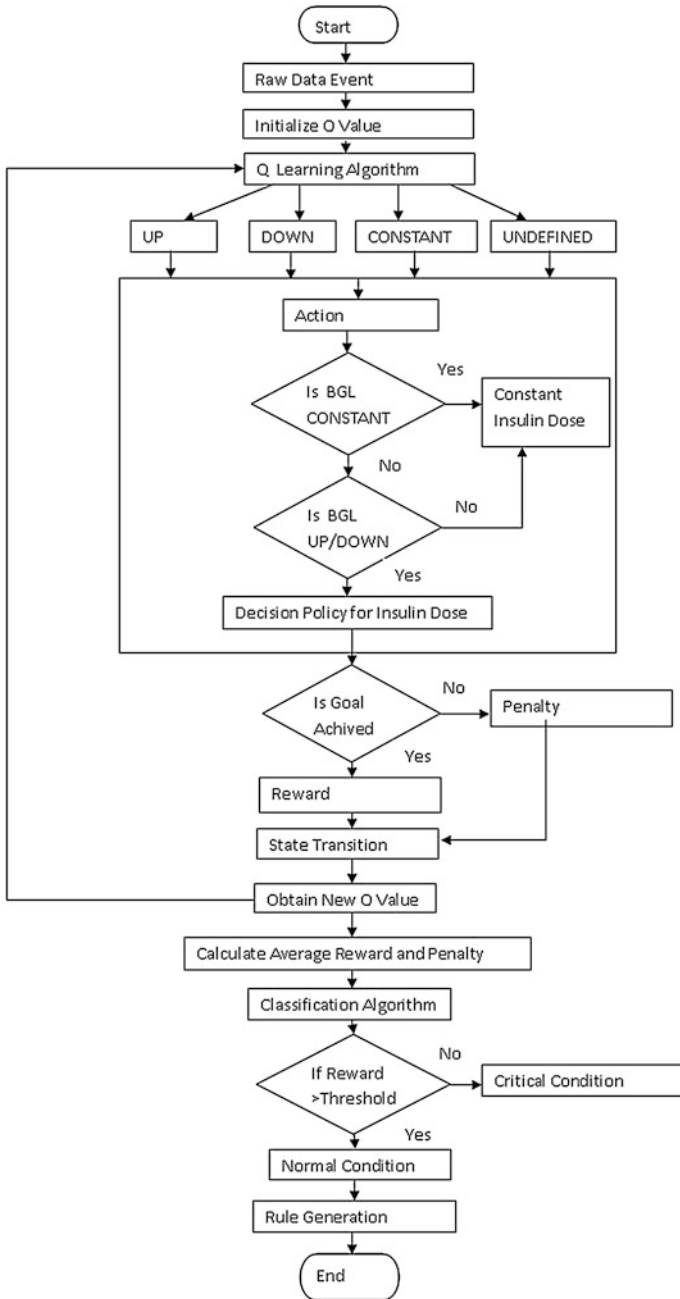


Fig. 2 Data flow architecture

Table 1 Various states of system

Before meal BGL	After meal BGL	State
70–100	80–140	Constant
<70	<80	Down
>100	>140	Up

Table 2 Variation in $Q(s, a)$ with respect to reward and penalty

	Reward/Penalty	$Q(s, a)$
1		6
2	1	6.6
3	1	6.84
4	1	8.3
5	-1	6.32
6	1	6.728
7	1	6.891

Every decision had given some reward and penalty. After this state transition is happened, Q value is computed by following formula.

$$Q(s, a) = (1 - \alpha)Q(s, a) + \alpha[r + \min(Q(s', a'))] \tag{4}$$

where

- α = learning rate where $0 < \alpha < 1$
- $r = 1$ if goal achieved
- $= -1$ if goal not achieved.

$Q(s, a)$ must be greater for the next state than the previous $Q(s, a)$. Value of $Q(s, a)$ is depend on the reward and penalty gain by the decision.

Initially, we assume $Q(s, a)$ arbitrarily, e.g., we consider $Q(s, a) = 6$ and $\alpha = 0.6$. Then, we observe variation in $Q(s, a)$ as follows (Table 2):

If system gets continuous reward (+1), then $Q(s, a)$ increases or else, for penalty (-1), the $Q(s, a)$ decreases.

5 Experimental Setup

In this critical system for insulin pump controller, blood glucose sensor checks the different parameters of the blood. Then, blood glucose analysis is done. For each patient, insulin requirement is computed, and insulin delivery controller controls quantity of insulin dose and gives commands to the pump controller. Insulin pump injects insulin to the patient (Fig. 3).

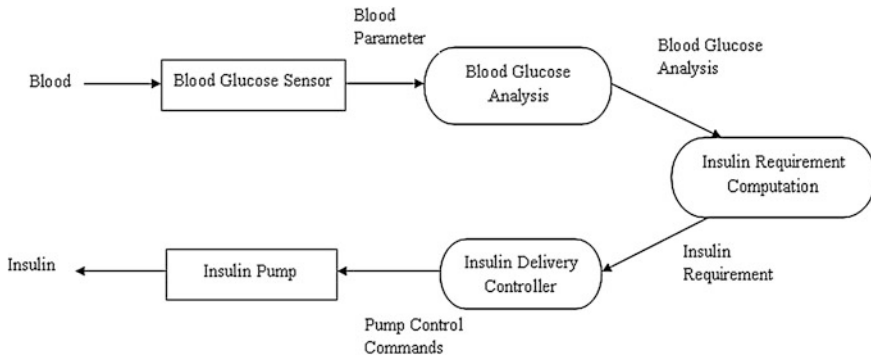


Fig. 3 Critical system for insulin pump controller

In this critical system, major problem is to compute insulin dose. Our system helps in computing insulin dose for each patient. We have to compute insulin dose so that during next measurement, BGL of the patient is in range of constant.

We use diabetes data set of 70 patients. For each patient, number of records are available. These diabetes record files for each patient consist of four fields per record. Each field is separated by a tab, and each record is separated by a new line. This data set is available on UC Irvine Machine Learning Repository (<http://archive.ics.uci.edu/ml/datasets/Diabetes>).

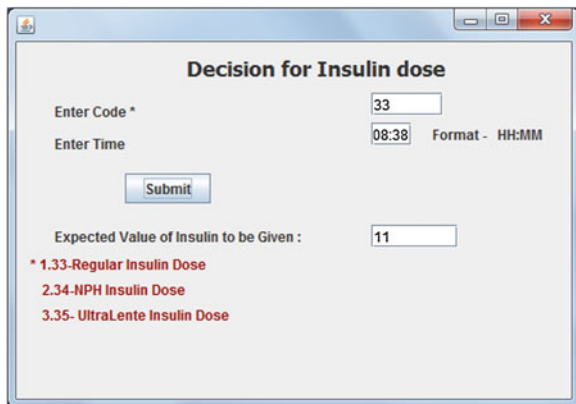
Sample data set is given in the following table.

Figure 4 shows the decision made by system. For example, for regular insulin dose, we have to give 11 unit of insulin to the patient.

BGL of the diabetes patient is always changing over time. Figure 5a shows the variation of BGL at different time intervals.

Figure 5b shows variation of BGL before meal, and next graph shows variation of BGL after meal. In Fig. 5c, we see that due to decision making, we are able to make BGL of patient constant except P3 where we see BGL is up.

Fig. 4 Decision for insulin dose



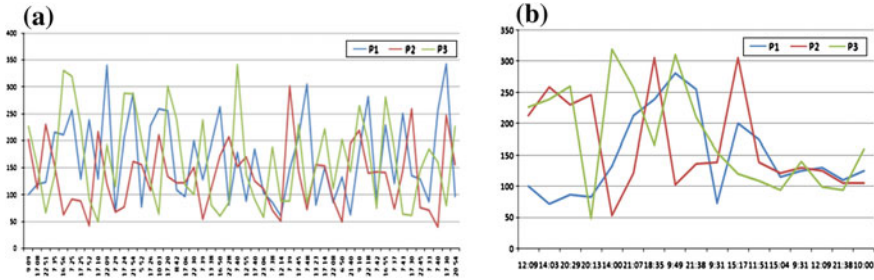


Fig. 5 Variation in blood glucose level: **a** blood glucose level variation before meal, **b** blood glucose level variation after meal

Table 3 Reward and penalty gain by patients

Day	Patient	Rewards gain	Penalty gain	Classification
22-Apr	P1	5	2	Normal
22-Jul	P2	4	2	Critical
11-Oct	P3	4	3	Normal

Average reward and penalty for each patient is calculated. Depending on average reward and penalty, patient is classified as normal and critical. If patient gets maximum reward, then we say that he is in normal condition or else, he is in critical condition (Table 3).

5.1 Decision Rules

At last, various rules are generated for each patient.

These rules contain user Id, code, value, and count of decision, i.e., counts of rewards gain so far. As shown in table for user 1 and code 33 (i.e., normal insulin dose), if we give 11 unit of insulin, then system gets 90 rewards (Table 4).

Table 4 Generated rules

User Id	Code	Value	Count of decision (rewards/penalty)
<i>Reward = +1</i>			
1	33	11	90
2	33	10	100
3	33	3	70
<i>Penalty = -1</i>			
1	33	10	35
2	33	14	55
3	33	1	4

6 Conclusion

Sequential decision-making agents interact with the system and take action a , and this selected action gives agent some reward and penalty, without considering how that task is achieved. This reward and penalty helps in making further decision. Using Q learning algorithm on diabetes, data help in making effective decision about giving specific quantity of insulin dose that should be given to the patient at particular time. Rewards and penalties are given to the decision which helps in classifying patients in normal or critical condition. Various rules are generated which helps in taking decision for insulin dose.

References

1. L.P. Kaelbling et al., in *Reinforcement Learning: A Survey*. AI Access Foundation and Morgan Kaufmann Publishers. All rights reserved (1996)
2. M.A. Wiering, H. van Hasselt, Ensemble algorithms in reinforcement learning. *IEEE Trans. Syst., Man, Cybern.—Part b: Cybern.* **38**(4)(2008)
3. M. Mill'an-Giraldo, et al., in *On-line Classification of Data Streams with Missing Values based on Reinforcement Learning*. Institute of New Imaging Technologies
4. M.M. Gaber, A. Zaslavsky, S. Krishnaswamy. Mining data streams: a review. *SIGMOD Rec.* **34**(2) (2005)
5. R. Garnett, S.J. Roberts, in *Learning from Data Streams with Concept Drift*. Technical Report PARG-08-01, Hilary Term (2008)
6. L. Busoniu et al., A comprehensive survey of multiagent reinforcement learning. *IEEE Trans. Syst., Man, Cybern.—Part c: Appl. Rev.* **38**(2) (2008)
7. C. Gaskett, D. Wettergreen, A. Zelinsky, Q-learning in continuous state and action spaces. *Int. J. Comput. Sci. Eng. (IJCSE)*
8. H. van Seijen, H. van Hasselt, S. Whiteson, M. Wiering. *A Theoretical and Empirical Analysis of Expected Sarsa*
9. M.A. Wiering, QV (λ)-learning: a new on-policy reinforcement learning algorithm, in *Proceedings of the 7th European Workshop on Reinforcement Learning* (2005) pp. 29–30
10. A.G. Barto et al. in *Learning and Sequential Decision Making*. pp. 530–599
11. G. Ditzler, R. Polikar, Incremental learning of concept drift from streaming imbalanced data. *IEEE Trans. Know. Data Eng.* **25**, 2283–2301 (2013)
12. C. Gaskett, D. Wettergreen, A. Zelinsky, Q-learning in continuous state and action spaces . *Int. J. Comput. Sci. Eng. (IJCSE)*
13. S. Harm van Seijen, H. van Hasselt, S. Whiteson, M. Wiering. A theoretical and empirical analysis of expected

Design of Quantum Cost and Delay-Optimized Reversible Wallace Tree Multiplier Using Compressors

A.N. Nagamani and Vinod Kumar Agrawal

Abstract Compressors play a specific role in realizing high-speed arithmetic circuits in particular multipliers. The increase in the demand of fast multiplication has attracted many researchers to design higher order compressors which enhance the speed of computation by reducing the critical path delay of the processing unit. In this paper, quantum cost and delay-optimized compressors are proposed. The compressors are designed using existing reversible gates such as Feynman, Fredkin, and Peres gates (PG). Using these optimized compressors, 8×8 Wallace multiplier is designed and the performance parameters are compared with the existing designs in the literature. It is evident from the results that this design exhibits better performance parameters and lesser delay and hence it is faster compared to existing designs in the literature. Thus, this design is suitable for high-speed arithmetic circuits such as FFTs, IFTs in modern DSP design.

Keywords Compressors · Fast arithmetic · Multipliers · Nanotechnology · Reversible logic

1 Introduction

Reversible computation is a bijection logic, where there exists a one-to-one mapping between the inputs and outputs. This means that the inputs of any reversible logic circuits can be uniquely determined by knowing the outputs which in conventional

A.N. Nagamani (✉)

Department of Electronic and Communication, PES Institute of Technology,
Bangalore, India
e-mail: nagamani@pes.edu

V.K. Agrawal

Department of Information Science and Engineering, PES Institute of Technology,
Bangalore, India
e-mail: vk.agrawal@pes.edu

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_36

323

logic is not possible [1]. It is determined that irreversible computing generates heat of $kT\ln 2$ joules for each bit of information loss. In [2], it is demonstrated that $kT\ln 2$ energy loss is directly proportional to the number of lost bits during the computation.

Multiplication is basic arithmetic operation in applications such as DSPs that depends on efficient implementation of ALU to execute dedicated operations such as multiplication, convolution, and filtering. The multiplication process consists of generation of partial products, partial product reduction, and final carry-propagating addition for computing the final product. The critical issue in multiplication is reduction of partial products, which contributes to overall delay, area, and power. Compressors are basic elements that are frequently used to reduce partial product matrix [3–6]. In this paper, reversible 3-2 and 4-2 compressors are designed using existing reversible logic gates and also present an improved design of Wallace tree multiplier proposed in [7].

Rest of the paper is organized as follows. In Sect. 2, we outline the motivation for reversible logic and fundamentals of reversible logic gates. In Sect. 3, design methodologies of various compressors are proposed along with the prior work on compressor designs. Section 4 presents the improvement of proposed compressor designs over existing designs with the verification of designs. Sections 5 and 6 present results and conclusions of the work and reference as follows.

2 Reversible Logic

2.1 Motivation for Reversible Logic

Reversible logic circuits are those circuits that do not lose information. Hence, reversible logic operations dissipate very less heat compared to conventional circuits [2]. Due to this reason, reversible logic is likely to be in demand for high-speed digital designs. Reversible logic circuits can be used in optical computing [8], quantum computing [9], and low power CMOS design [10]. One of the major applications of reversible logic is in designing quantum computers. The efficient reversible logic circuit must contain less number of performance parameters. Important performance parameters that are used for the evaluation of reversible circuit are as follows:

1. Gate count (GC): It is the number of reversible logic gates required for implementing the desired circuit.
2. Quantum cost (QC): It is the number of preliminary gates required to realize a reversible logic circuit.
3. Garbage outputs (GO): It is the number of unused outputs in a reversible logic circuit.
4. Constant inputs (CI): It is the number of constant inputs either '0' or '1' required to get desired function from the reversible logic gates.

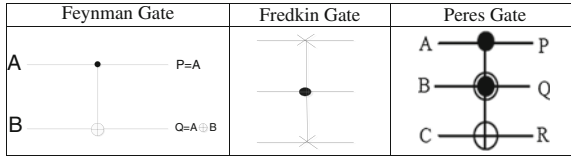


Fig. 1 Symbol of basic reversible gates

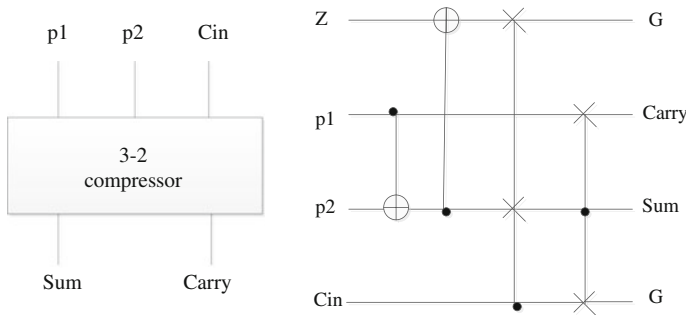


Fig. 2 Block diagram and quantum diagram of 3-2 compressor

5. Delay: Delay is an important parameter that is used to determine the efficiency of the reversible circuits. Here, delay represents the critical delay of the circuit. In delay calculations, we use the logical depth as the measure of the delay [11]. The delays of all 1×1 gate and 2×2 reversible gates are taken as unit delay, denoted as Δ . Any 3×3 reversible gate can be designed from 1×1 reversible gates and 2×2 reversible gates, such as the CNOT gate, the controlled-V and the controlled-V + gates. Thus, the delay of a 3×3 reversible gate can be computed by calculating its logical depth when it is designed from 1×1 and 2×2 reversible gates. Figures 1 and 2 show the logic depth in the quantum implementation of Feynman and Fredkin gate. It can be seen that the Fredkin gate (FRG) has the delay of 5Δ . Each 2×2 reversible gate in the logic depth contributes to 1Δ delay. Similarly, Feynman gate has delay of 1Δ .

2.2 Basic Reversible Logic Gates

2.2.1 The Feynman Gate (FG)

The Feynman gate or the controlled NOT gate (CNOT) has 2 inputs and 2 outputs reversible gate with inputs (A, B) which are mapped to outputs ($P = A$, $Q = A \oplus B$). It has a quantum cost of 1 [12]. Figure 1 shows symbol of the Feynman gate. The

Feynman gate is widely used for either copying the signal A when $B = 0$ or inverting A when $B = 1$.

2.2.2 The Fredkin Gate (FRG)

It has 3 inputs, 3 outputs reversible gate with inputs (A, B, C) and outputs $P = A$, $Q = A'B + AC$, $R = AB + A'C$. It has quantum cost five [13]. Figure 1b shows the symbol of Fredkin gate. Fredkin gate is widely used as multiplexer.

2.2.3 The Peres Gate (PG)

Figures 2a and 1c show a symbol of Peres gate (PG), [14]. It is a 3×3 reversible gate having inputs (A, B, C) and outputs $P = A$, $Q = A \oplus B$, $R = AB \oplus C$. The quantum cost of PG is 4 [14].

3 Prior Work

Many researchers have worked on various architectures of conventional compressors [15–20]. These compressor architectures are developed based on reducing the logic depth and are used for designing conventional multipliers and their performance is evaluated based on the speed of computing the product [3–6, 21–23]. 4-2 compressor is designed using new TSG gate [7] and Wallace tree multiplier is designed using this compressor, since Wallace tree structure is one of the fast multiplication schemes [24]. In [25], 4-2 compressors are designed using new HNG/RFA gate. This is proven to be the fastest among the compressor designs in conventional CMOS implementation in terms of delay, logic depth, and power [19]. The main focus in this work was to design compressors and 8×8 Wallace tree multiplier using existing reversible logic gates for those the synthesis algorithms and techniques have been defined in the reversible logic literature [26].

4 Proposed Designs

4.1 Reversible 3-2 Compressor

A 3-2 compressor has three inputs p_1 , p_2 , and C_{in} and generates outputs as sum and carry bit. Its functionality is same as a full adder. Here, C_{in} is the carry from the previous stage.

Table 1 Comparison of reversible 3-2 compressors

Performance parameters		GC	GO	CI	QC	Critical path delay in Δ
Designs		4	2	1	12	7Δ
Existing TSG [7]	3-2 compressors	1	2	1	12	12Δ
Percentage of improvement		-75 %	-	-	-	41 %
Proposed design	4-2 compressors	8	16	4	2	8Δ
Existing TSG [7]		2	24	4	2	24Δ

The basic 3-2 compressor equations are

$$p1 + p2 = \text{sum} + 2 * \text{carry} \quad (1)$$

$$\text{carry} = (p1 \text{ xor } p2) \cdot \text{cin} + (\overline{p1 \text{ xor } p2}) \cdot p1 \quad (2)$$

Thus, the equations can be modified for optimum delay as

$$\text{sum} = (p1 \text{ xor } p2)\overline{\text{cin}} + (\overline{p1 \text{ xor } p2})\text{cin} \quad (3)$$

$$\text{carry} = (\overline{p1 \text{ xor } p2})p1 + (p1 \text{ xor } p2)\text{cin} \quad (4)$$

Using these equations, the compressor is designed using Feynman gate and Fredkin gate, which has 2Δ -XOR + 1Δ -MUX in carry path. Here, z is the constant input required to generate intermediate terms or to store some intermediate values computed. The block diagram and quantum implementation of 3-2 compressor is shown in Fig. 2. Its performance parameters are listed in Table 1.

4.2 Reversible 4-2 Compressor

The block diagram and quantum implementation of 4-2 compressor is shown in Fig. 3 and its performance parameters are listed in Table 1. Design equations formulated in the proposed procedure are,

$$\text{sum} = (p1 \text{ xor } p2 \text{ xor } p3 \text{ xor } p4 \text{ xor } \text{cin}) \quad (5)$$

$$\text{cout} = (p1 \text{ xor } p2)p3 + (\overline{p1 \text{ xor } p2})p1 \quad (6)$$

$$\text{carry} = (p1 \text{ xor } p2 \text{ xor } p3 \text{ xor } p4)\text{cin} + (\overline{p1 \text{ xor } p2 \text{ xor } p3 \text{ xor } p4})p4 \quad (7)$$

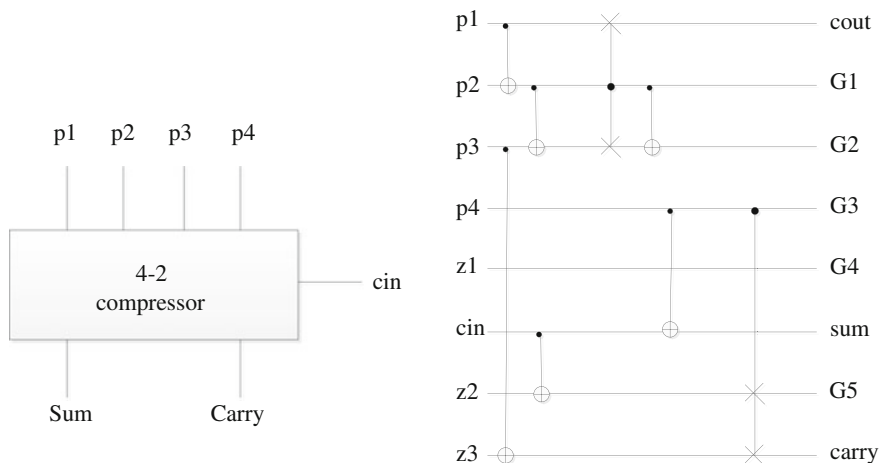


Fig. 3 Block diagram and quantum diagram of 4-2 compressor

Here, critical path delay of proposed implementation is $3\Delta\text{-XOR} + 1*\Delta\text{-MUX}$ (since XOR in terms of different inputs can be done in concurrency), i.e., 3Δ for XOR and 5Δ for multiplexer which evaluates to 8Δ . In this design, three constant inputs are required and generated garbage outputs of 5. The total quantum cost of the circuits is 16.

4.3 Design of Wallace Tree Multiplier Using 4-2 Compressor

Wallace tree is one of the fast multiplier architectures proposed by Wallace [24]. The first attempt to design this Wallace multiplier using reversible logic was proposed by Thapliyal [7] with a new gate called TSG gate. The design using TSG is optimized in terms of gate count and garbage outputs, but the quantum cost and the total delay are high. Also, the gate used is a 4×4 new gate. There are three stages involved in multiplication using Wallace multiplication algorithm. They are as follows: stage 1: partial product generation, stage 2: partial product addition, and stage 3: final addition stage. The most crucial stage is partial product addition, which takes longer time for this process. Hence, making this stage faster improves the overall performance of the multiplication reducing the delay. This partial product addition stage can be speeded up by the use of compressors [20]. Here, 4-2 compressors are used to speed up the addition process. The optimized compressor further enhances the multiplication performance. In the literature, 8×8 Wallace multiplier is designed using 4-2 compressor that is designed using TSG gate [7]. In the proposed work, an optimized 4-2 compressor is designed with less quantum cost and depth. Figure 4 shows the architecture of 8×8 Wallace tree multiplier. Initially,

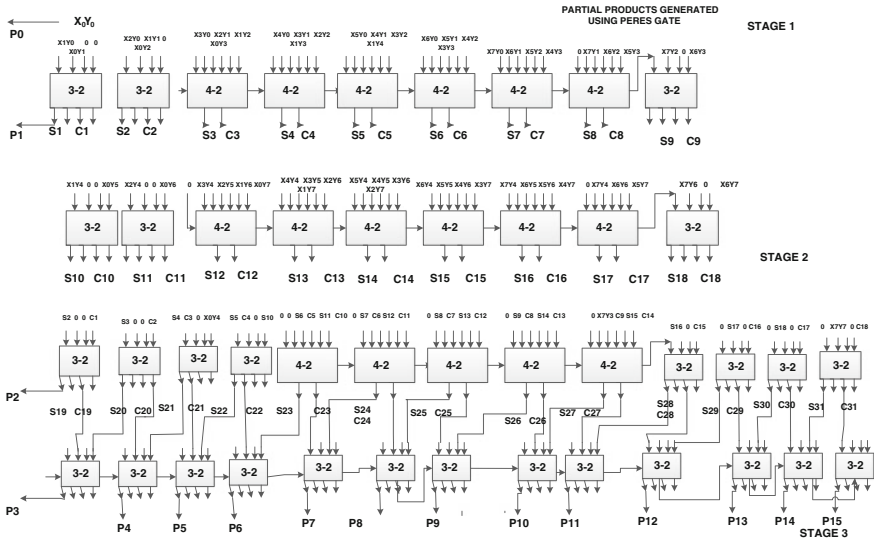


Fig. 4 Reversible 8×8 Wallace tree multiplier using 3-2 and 4-2 compressors

all the partial products are generated by Pere’s gate. Then, according to the weights of the partial products, they are added using 3-2 compressors or 4-2 compressors. Again, the sum and carry generated from stage 1 are added depending upon their weights using 4-2 compressors. In final stage, the results are obtained by using parallel adder structure. This architecture requires twenty-seven 3-2 compressors, seventeen 4-2 compressors, and sixteen partial product generators, which results in quantum cost of $27 \times 12 + 16 \times 17 + 16 \times 4$ which adds to 660. Since the initial partial products are AND ing of input bits, PG with control input $C = 0$ will give AND gate output. The performance parameter is evaluated for this multiplier and tabulated in Table 1.

5 Results and Discussions

The results tabulated in Tables 1 and 2 show that the proposed design is better in terms of quantum cost and hence the delay. The 8×8 Wallace multiplier performance parameters are tabulated in Table 2, which reveals that proposed design

Table 2 Quantum cost comparison of proposed and existing reversible 8×8 Wallace multiplier

Design	Existing design TSG [7]	Proposed	Percentage of improvement (%)
Quantum cost	796	660	18
Critical path delay in Δ	64	27	57

offers 18 % less quantum cost and 57 % reduction in critical path delay compared to the existing Wallace multiplier design.

6 Conclusions

This work presents a novel design of reversible compressors using existing reversible logic gates, and using these compressors, 8×8 Wallace multiplier is designed and it is evident from the results that the proposed designs have less quantum cost and delay compared to existing designs and thus suitable for designing high-speed arithmetic circuits. The parameters such as gate count, constant inputs, and garbage outputs are though more compared to existing design, the trade-off of higher speed is acceptable.

References

1. R. Landauer, Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **3**, 183–191 (1961)
2. C.H. Bennett, Logical reversibility of computation. *IBM J. Res. Dev.* (1973)525–532
3. V.G. Oklobdzija, D. Villeger, S.S Liu, A method for speed optimized partial product reduction and generation of fast parallel multipliers using an algorithmic approach. *IEEE Trans. Comput.* **45**(3) (1996)
4. P. Stelling, C. Martel, V.K. Oklobdzija, R. Ravi, Optimal circuits for parallel multipliers. *IEEE Trans. Comput.* **47**(3), 273–285 (1998)
5. V. Oklobdzija, High-speed VLSI arithmetic units: adders and multipliers, in *Design of High-Performance Microprocessor Circuits* (2000)
6. H.T. Bui, A.K. Al-Sheraidah, Wang, Y, Design and analysis of 10-transistor full adders using novel XORXNOR gates in *Proceedings International Conference Signal Processing 2000* (2000)
7. H. Thapliyal, M.B. Srinivas, Novel reversible ‘TSG’ gate and its application for designing components of primitive/reversible quantum ALU, in *Proceedings of 5th IEEE International Conference on Information, Communications and Signal Processing* (2005) 1425–1429
8. E. Knill, R. Laflamme, G.J. Milburn, A scheme for efficient quantum computation with linear optics. *Nature* 46–52 (2001)
9. M. Nielsen, I. Chuang, in *Quantum Computation and Quantum Information* (Cambridge University Press 2000)
10. G. Schrom, in *Ultra Low Power CMOS Technology*. Ph.D. Thesis, Technischen Universitat Wien (1998)
11. H. Thapliyal, H.V. Jayashree, A.N. Nagamani, H.R. Arabnia, Progress in reversible processor design: a novel methodology for reversible carry look-ahead adder. *Trans. Comput. Sci.* **XVII**, **7420**, 73–97 (2012)
12. R. Feynman, Quantum Mechanical Computers. *Optics News*. **11**, 11–20 (1985)
13. E. Fredkin, T. Toffoli, Conservative logic. *Int. J. Theory Phys.* **21** 219–253 (1982)
14. A. Peres, Reversible logic and quantum computers. *Phys. Rev. A, Gen. Phys.* **32**(6) 3266–3276 (1985)

15. P. Gopineedi, H. Thapliyal, M.B. Srinivas, H.R. Arabnia, in Novel and efficient 4:2 and 5:2 compressors with minimum number of transistors designed for low-power operations in *Proceedings of the 2006 International Conference on Embedded Systems and Applications (ESA'06)* 017(5)160–166 (2006)
16. K. Prasad, K.K. Parhi, Low-power 4-2 and 5-2 compressors, in *Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers* 129–133 (2001)
17. K. Ohsang, N. Kevin, E. Earl, Jr Swartzlander, A 16-Bit by 16-Bit MAC design using fast 5:3 compressor cells. *J. VLSI Sig. Proc.* **31** 77–89 (2002)
18. R. Mahnoush, K. Omid, P.M. Amir, J.J. Somaye, N. Keivan, A new design for 7:2 compressors, in *Proceedings. IEEE/ACS International Conference on communication, Networking and Broadcasting: computing and Processing (Hardware/Software)* (2007)
19. V. Sreehari, M.K. Kirthi, A. Lingamneni, R.P. Sreekanth, M.B. Srinivas, Novel architectures for high-speed and low-power 3-2, 4-2 and 5-2 compressors, in *20th IEEE International Conference on VLSI Design (VLSID'07)* (2007)
20. C.H. Chang, J.M. Gu, M. Zhang, Ultra low voltage low-power CMOS 4-2 and 5-2 compressors for fast arithmetic circuits. *IEEE Trans. Circ. Syst.-I: Reg. Papers* **51**(10), 1985–1997 (2004)
21. J. Yingtao, A novel multiplexer-based low-power full adder. *IEEE Trans. Circ. Syst. II. Exp. Briefs* **51**(7) (2004)
22. P.J. Song, G.D. Micheli, Circuit and architecture tradeoffs for high-speed multiplication. *IEEE J. Solid-State Circ.* **26**, 1184–1198 (1991)
23. V.J. Oklobdzija, D. Villegier, S.S. Liu, A method for speed optimized partial product reduction and generation of fast parallel multipliers using an algorithmic approach. *IEEE Trans. Comput.* **45**, 294–305 (1996)
24. C.S. Wallace, A suggestion for a fast multiplier. *IEEE Trans. Elec. Comput.* 14–17 (1964)
25. D. Krishnaveni, M. Geetha Priya, K. Baskaran, Design of an efficient reversible 8×8 wallace tree multiplier. *World Appl. Sci. J.* **20**(8), 1159–1165 (2012)
26. J. Donald, N.K. Jha, Reversible logic synthesis with Fredkin and Peres gates. *ACM J. Emerg. Technol. Comput. Syst.* **4** (2008)
27. P.W. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *IEEE Computer Society Press.* 124–134 (1994)
28. P.W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *quant-ph/9508027.* 2 (1997)

A Novel Authentication Framework for Hadoop

P.K. Rahul and T. GireeshKumar

Abstract Hadoop is an open-source file system used to process and carry Big data which is out of the reach of normal softwares to be work on. This paper analyses the data security issues in Hadoop and proposes a new authentication framework for clients. The framework uses cryptographic functions such as public key cryptography, private key cryptography, hashing functions, and random number generator. This framework will define a new key for each client and authenticate all clients and services using this key. The authentication agent offers user data protection, way for privilege separations, and basic security needs for data stored in Hadoop file system.

Keywords Hadoop distributed file system · Authentication protocols · Unix · Cryptographic functions

1 Introduction

The amount of data in our industry and world is exploding. According to statistics, limits on size of data sets that are feasible to process are in the order of exabyte of data. The capacity to store information has roughly doubled every 40 months since 1980s, as of 2013 everyday roughly 2.5 exabyte of data are created [13]. These data sets are so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications [1]. This gave birth to a new term called Big data. Big data require “massively parallel software running on tens or even thousands of servers.” Typical problems faced

P.K. Rahul (✉) · T. GireeshKumar

Tifac Core in Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, Tamilnadu, India
e-mail: pkraahul.pk870@gmail.com

T. GireeshKumar
e-mail: gireeshkumart@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_37

333

when dealing with Big data include capture, storage, dissemination, search, analytics, and security [2]. As a solution for all these problems, Apache foundation developed a java-based open-source framework named, Hadoop. Hadoop consists of the Hadoop common package, which provides OS-level abstraction, necessary Java Archives (JAR) files and scripts needed to start Hadoop, a MapReduce engine and the Hadoop distributed file system (HDFS) [3].

Hadoop allows for the distributed processing of large data sets across clusters of computers using simple programming models. Files on Hadoop are split into blocks typically of size 64 MB and stored in a redundant fashion across multiple machines to ensure their durability to failure and high availability to parallel applications [3]. Map reduce is a programming model for processing large data sets used in Hadoop. Hadoop is used as a shared multi-tenant service and is used to store sensitive data; as a result, strong authentication and authorization is necessary to protect data [4].

Initially, Hadoop was designed for private clusters behind organization's fire-wall, but the need for processing big file systems has grown. Hadoop has been employed over public clusters and then the concentration focused toward adding security in Hadoop clusters, but it was not an easy task due to many reasons. Remaining sections of paper are organized as follows: Sect. 2 discusses the problems associated with Hadoop data security, Sect. 3 proposes design of new authentication framework to make Hadoop stronger, and Sect. 4 concludes the paper and proposes future enhancements.

2 Related Works

A HDFS cluster has two types of nodes operating, NameNode and a number of DataNodes. They operate in a master-slave pattern [3, 4]. NameNode is the master of Hadoop file system and manages the file system namespace. It maintains the file system tree and the metadata for all the files and directories in the tree. It also knows the DataNodes on which all the blocks for a given file are located. DataNodes are the worker nodes of the file system. They store and retrieve blocks according to commands from clients or NameNode [5]. The NameNode distributes the blocks of data over different DataNodes. Periodically, DataNodes report back to the NameNode with lists of blocks that they are storing. Hadoop provides replication of data by distributing same data block on number of DataNodes. Usually, each block of data is replicated three times to provide proper availability of data in case of failure of any DataNode. Usually in a cluster, there will be a single NameNode and a number of DataNodes as shown below in Fig. 1.

Even though HDFS was designed to solve problems faced in data storage, availability, and processing, but still it has many loopholes, due to which organizations are not ready to accept this open-source framework. We will discuss some of the problems associated with HDFS.

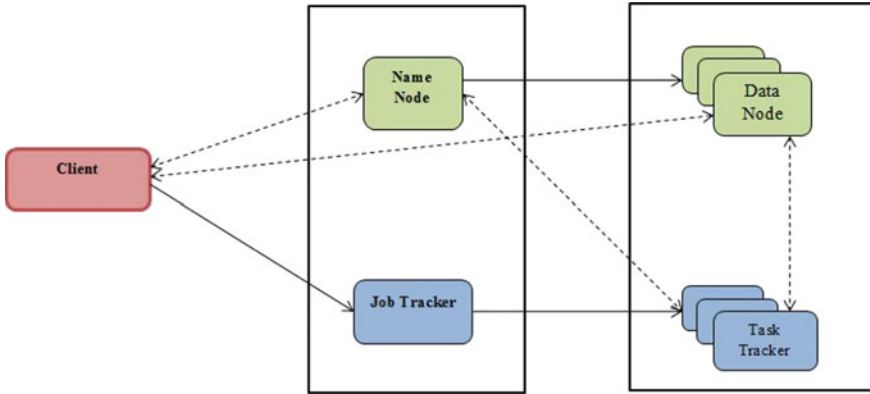


Fig. 1 Hadoop architecture

Hadoop didn't authenticate users or services, and also there was no data privacy. It was designed to execute codes over a distributed environment so anyone could submit code and it would be executed [7]. Although auditing and authorization controls were implemented, such controls were easily circumvented because user could impersonate as other user with a command line switch.

The access levels in Hadoop is designed in such a way that all users and programmers had the same level of access to all of the data in cluster, so any job given by any user could access any data in the cluster, and could read any data set. A submitted batch job is executed at later time on nodes different from the node on which the client authenticated and submitted job [3].

The Authentication protocols used username as key to authenticate users but if same username is used in same cluster then datasets could be read by other users as well. Hadoop uses UNIX "who am i" utility to authenticate user and gives access levels according to rules defined. Hadoop has a Superuser who has same username as NameNode name. Superuser has all the privileges to access any data set in that cluster [3].

As data is distributed over a large number of nodes it is very difficult to use any kind of encryption technique for data privacy. An unauthorized user may submit a job to a queue or delete or change priority of the queue. Even though Kerberos is suggested as mechanism for authentication of clients in Hadoop but still it is not sufficient due to size of datasets and clients [7].

Kerberos has to define authentication for clients and services separately and then these values has to be stored in Access Control Lists (ACL) associated with each files which will only increase overhead for proper data distribution over cluster [5]. Intermediate Map reduce function output is not stored in HDFS but on local machine so any user can use the host operating system interfaces to access those data.

DataNodes do not enforce any access control or authentication control on access to its data blocks. This makes it possible for an unauthorized client to read a data block as long as it can supply its block ID [7].

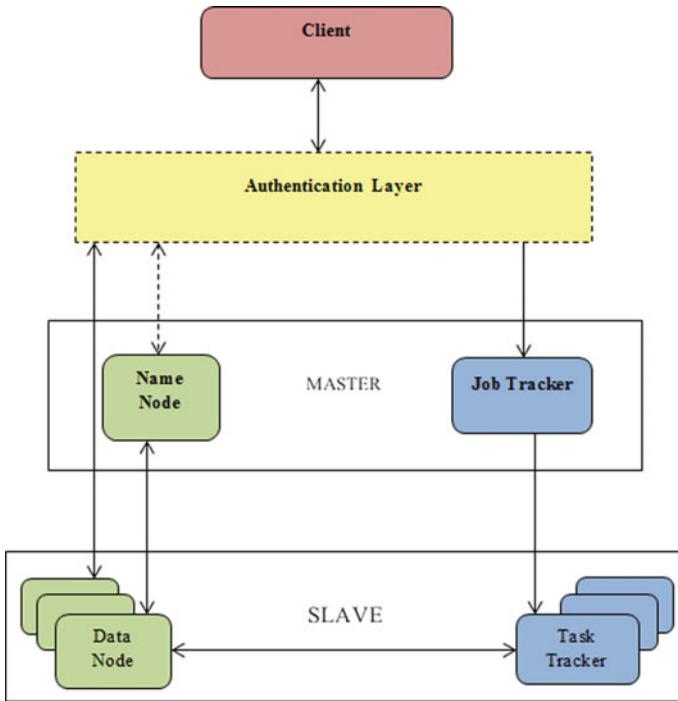


Fig. 2 Proposed Hadoop architecture

3 Framework

Hadoop is often provided as a shared distributed service and is used to store sensitive data; as a result adding security to Hadoop is very important and need of the time. Here, we create a novel authentication approach to boost up the security of Hadoop. Data will be stored along with the newly derived authentication key to differentiate between clients and their respective data sets. This key is used to provide data privacy and proper authentication for clients and services used by clients. The framework introduces two servers, User Server and Data Server. Architecture is shown in Fig. 2.

The above-proposed architecture adds an extra layer to the already existing Hadoop. All the clients are authenticated first and given a key, this key is used in all later communication. The new authentication protocol defines two new rules as mentioned below.

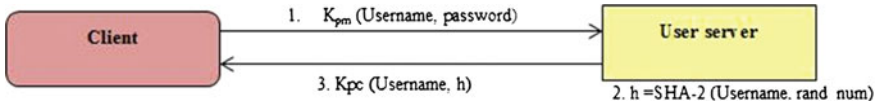


Fig. 3 Username registration

3.1 User Registration

Username was the most important authentication key to be used in all earlier versions of Hadoop authentication protocols, but as this username can easily be deceived, we introduce a new and more unique element for authentication key calculation, a random number which is being generated by server system. These random numbers must be purely random because pseudorandom numbers can produce confusion if they are used with same usernames. The combination of random number and username can solve our problem of using same username and compromising client data privacy. The protocol is defined in Fig. 3.

- (1) Client sends the username and password chosen for registration using the public key of user server.
- (2) Server computes a hash value; hashing algorithm SHA-2 is used, and components used are username and a random number (rand_num) generated by server, because if same username is used by different users, we can still differentiate them by this random number.
- (3) User server sends the username and hash value h back to client using his public key.

Here in this protocol, public key has been used to communicate. This protocol will run only once for a single user only during his registration to the cluster.

3.2 Data Control

This is the second phase of the authentication protocol defined. During this protocol, both the servers, User server and Data servers, are being used. Protocol uses combination of public key and symmetric key. Protocol type depends on the value given by client during first message sent to User server. Paper proposes a set of values {store, retrieve, cmd} as “action” to be used during the first message from client to User server. Further protocol will depend upon the value used by action from the given set. If client sends “store” in place of action, then protocol will work to store data in DataNodes, if client sends “retrieve” as action, then protocol is meant to retrieve data from DataNode corresponding to client; similarly, if action is defined to be a “cmd,” then given service will be launched for the client. Protocol is defined in Fig. 4.

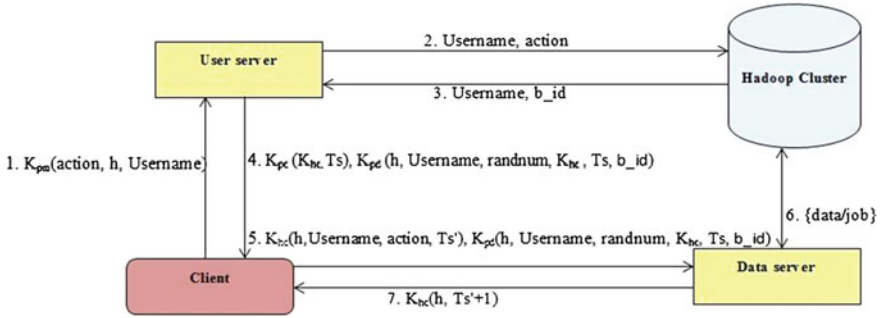


Fig. 4 Data control protocol

- (1) Client uses the public key of User server to send the action, hash value, and username combination; as mentioned earlier, action can be either of {store/retrieve/job}. The first element of this protocol is very much important because remaining activity of protocol depends on this, whether it is to store data, or retrieve data, or to do some job on data.
- (2) User server will authenticate the user and sends this username and intended action over the data to the Hadoop cluster. This cluster will contact the NameNode intended to be used for accessing data.
- (3) Hadoop cluster will find out the DataNodes where data are being stored and this data block ids will be sent to User server which was earlier sent to the client directly.
- (4) User server creates a symmetric key to be used between client and Data server for all remaining communications. User server also creates a time stamp to avoid key reuse.
- (5) Client sends the message intended for Data Server, and using the intended key, it will send the username, hash value, and new Timestamp. Hash values and Usernames are being used to confirm the user.
- (6) Data Server communicates with the Hadoop clusters submits job intended to user, stores, or retrieves data.
- (7) Sends back the result for the given job to client using symmetric key, thus maintains data privacy, authentication, integrity, and many other issues which were faced in earlier versions of Hadoop.

There will be a slight difference in the protocol if client wants to use some services of the Hadoop. In this case, the first message will contain “cmd” in place of action as shown in Fig. 5.

- (1) Client uses the public key of User server to send the cmd, hash value, and username combination.
- (2) User server will authenticate the user and sends this username and intended action over the data to the Hadoop cluster. This cluster will contact the NameNode intended to be used for accessing data.

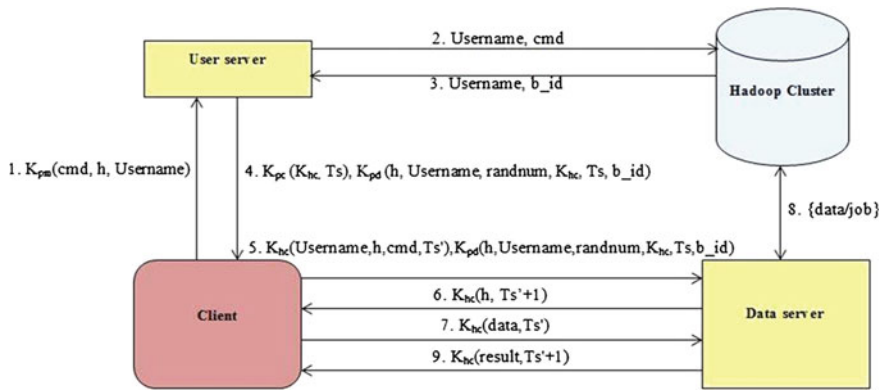


Fig. 5 Service request protocol

- (3) Hadoop cluster will find out the DataNodes where data are being stored and this data block ids will be sent to User server which was earlier sent to the client directly.
- (4) User server creates a symmetric key to be used between client and Data server for all remaining communications. User server also creates a time stamp to avoid key reuse.
- (5) Client sends the message intended for Data Server, and using the intended key, it will send the username, hash value, and new time stamp. Hash values and Usernames are being used to confirm the user.
- (6) Data Server recognizes that client is requesting for some services, so it will just acknowledge the symmetric key and request the client to send the service command to be executed.
- (7) Client sends the service command along with hash value and new time stamp.
- (8) Data Server checks whether the client is authorized to run that command and after that makes changes according to command in Hadoop file system.
- (9) Data Server sends back acknowledge or any data according to the demand.

4 Conclusions

This paper proposes a new authentication framework deployed over already existing Hadoop cluster architecture. The proposed authentication protocols enhance Hadoop security by adding a unique key to be used along with all clients. Each client data are stored along with this key which is a random number produced by authentication servers. Earlier the block id of DataNode used to be in plain text which was used by hacker by now it will be delivered in an encrypted form. All type of communication goes through the layer with is encrypted using symmetric key. The layer is proposed to provide strong data privacy along with data integrity,

authentication, and proper authorization of services in Hadoop. The generation of random number raises security level as they cannot be hacked easily. So the utilization of this protocol will make the Hadoop environment more secure.

References

1. T. White, *Hadoop: The Definitive Guide*, 3rd edn. (2012), pp. 2–14
2. M. Smith, C. Szongott, B. Henne, Big data privacy issues in public social media, in *IEEE International Conference on Digital Ecosystems Technologies* (2012)
3. D. Borthakur, in *The Hadoop Distributed File System: Architecture and Design*. Apache Software Foundation
4. P. Malik, Governing big data: principles and practices. IBM J. Res. Dev. (2013)
5. D. Das, O.O'. Malley, S. Radia, K. Zhang, Adding security to apache Hadoop. Horton Technical Report, IBM
6. A.B. Patel, M. Birla, U. Nair, Addressing big data problem using Hadoop and MapReduce, in *Nirma University International Conference on Engineering (NUICONE)* (2012)
7. O.O'. Malley, K. Zhang, C. Harnell, R. Marti, S. Radia, Hadoop security design. Yahoo Inc. (2009)
8. Zettaset, The big data security gap: protecting the Hadoop cluster. White Paper
9. H. Jing, L. Renfa, T. Zhuo, The research of the data security for Cloud disk based on the Hadoop framework, in *4th International Conference on Intelligent Control and Information Processing (ICIP)* (2013)

Fuzzy ART-Based User Behavior Trust in Cloud Computing

M. Jaiganesh, M. Aarthi and A. Vincent Antony Kumar

Abstract Nowadays, cloud has evolving rapidly for more scientific applications and communication across millions of people. It shares the resources which are available on-demand to the user as they needed. The virtual machines which resides the cloud resources can be shared by entire cloud environment and can easily be prone to attacks and threats. In this growing concern, ensuring the trust of virtual machines plays a major role in cloud computing environment. One of the most important evolving concerns is that the virtual client's perspective data or resources are hacked by the attackers easily. To overcome this behavior, we proposed a fuzzy logic technique called Fuzzy ART, where the consumption of resources is periodically scanned. Based on the traced-out behaviors, the virtual machine states are classified into categories from stable to attackers. The benefit of the proposed technique is an unsupervised learning.

Keywords Cloud computing · Fuzzy ART · Unsupervised learning · Random access memory

1 Introduction

Cloud computing has become an increasingly popular enterprise model where the resources are available on-demand to the user as they needed. It facilitates to access content across the Internet independently without reference to the underlying

M. Jaiganesh (✉) · M. Aarthi · A. Vincent Antony Kumar
Department of Information Technology, PSNA College of Engineering and Technology,
Dindigul 624 622, Tamilnadu, India
e-mail: jaidevlingam@gmail.com

M. Aarthi
e-mail: aarthimaris91@gmail.com

A. Vincent Antony Kumar
e-mail: vincypsna@rediffmail.com

hosting infrastructure. Cloud service provider provides services to users and manages the entire cloud. They are assigned to design the services based on scalable application. It establishes new data centers for hosting cloud computing applications in various locations around the world to provide redundancy and ensure reliability. The data center includes the resources where they are highly centralized and trusted. It provides more services and covers the maximum number of users. So, the cloud service providers must be prepared in better tolerance to manage and update the data centers. The main goal of data center is to provide access to a potentially vast amount of computing resources in an easy and user-centric way. The resources are shared and managed through virtualization technology employed in cloud computing environment [1]. It uses hypervisor within cluster environment that allows multiple virtual machines to share the resources allocated to them. The virtual clients must ensure that the data they receive, applications, and services are secure [2]. The clients have unauthorized access to the resources (memory, disk space, etc.) of their neighborhood and these vulnerabilities in turn make the platform more vulnerable to attacks. To overcome the above behaviors in the cloud, trustworthy of clients is considered to be major issue in cloud. Trust is always made only if sufficient services and expectation is attained. The challenges of trusting the cloud do not lie entirely in the technology; it also involves customer confidence that stems from loss of control over data assets. Though cloud computing is designed to provide better utilization of resources, the user behavior trust is essential. The user behavior trust includes the maintenance of software without malware, subjective of virtual client to hack the others, damage of infrastructure. No matter what causes the user to mistrust, but the cloud service provider must to monitor user behavior in order to ensure the credibility of the user's identity and behavior. To manage these uncertainty problems in cloud, a soft computing technique is proposed which an unsupervised learning technique called Fuzzy ART to classify the virtual clients based on their behavior.

This paper aims to provide behavior of different virtual clients in the cloud based on their resource consumption, and the trust rate evaluation is done. The rest of the paper is organized as follows: Sect. 2 discusses the problem formulation with their factors. Section 3 presents the Fuzzy ART technique to categorize the virtual clients based on behavior in their cloud environment. Section 4 shows the performance analysis of the proposed technique.

2 Problem Formulation

The virtual clients requesting for service suffer from the problem line sharing the resources available. It is solved virtually with the help of hypervisors available. The virtual clients request for service from the cloud (data center) by making hyper calls to the hypervisor. The sharing of resource is done by the hypervisor but the most imperative circumstance to be noted is the virtual hypervisor security. The major issue in hypervisor security is virtual isolation problem. The virtualization in the guest

operating system (OS) on a host is managed by the hypervisor. It usually controls the instruction flow that takes place between guest OSs and also physical hardware such as memory, disk space, and CPU [3]. Resources are partitioned by hypervisor either physically or logically. While physically partitioned hypervisor assigns separate physical resources to each guest OS such as disk space. While it is logically partitioned, it allows multiple guest resources to share same physical resources such as random access memory (RAM) and processors. The hypervisor function is to partition the system’s resources and to isolate the guest OSs, so that each can access only the resources allocated to them. But sometimes, they can share shared resources such as files on host OS. The hypervisor is capable to monitor the trust of each guest OS running within it. The most important issue to be considered is that the hypervisor should be carefully monitored for signs of compromise. By its functionality, it prevents unauthorized access to resources and also helps to prevent one OS from the other, i.e., the malware guest OS cannot inject the insecure files into another guest OSs memory. The main aim to isolate the guest OSs from one another and hypervisor is to be aware of side channel attacks. These attacks can exploit the physical properties of hardware to reveal information about the usage of amount of memory access, CPU, and other resources. Attackers try to break out of a guest OS so that they easily access the other guest OS and hypervisor too. Sometimes, the attacker compromises the hypervisor and gain the control over its entire guest OSs. So it is important to secure each hypervisor by some security policies. The hypervisor must be designed in some concern that they cannot be detected by the attackers.

2.1 Preliminaries

Theorem 1 *In Fuzzy ART architecture, all the templates are distinct. Assume that Fuzzy ART creates two templates W1 and W2 are equal [4]. W1 is created first and template W2 is created by the input pattern coded by template W2*

$$W2 = \beta(I \wedge W2') + (1 - \beta)W2' = W1. \tag{1}$$

Theorem 2 *In a Fuzzy ART architecture with sufficient number of nodes in the F2 layer, the size of a template is larger than $\alpha M / (\alpha + M)$*

$$|W| > \alpha M / (\alpha + M). \tag{2}$$

Theorem 3 *In Fuzzy ART architecture, if a node J in the F2 layer has perfectly learned an input pattern I, then when I is presented, it will directly access node J.*

$$T_j = |I > W_j| / (\alpha + W_j). \tag{3}$$

3 Proposed Method

The system is composed of several elements such as hypervisor, virtual clients, virtual service provider. In cloud computing environment, multiple instances of a variety of OSs may share the virtualized hardware resources. Hypervisors are programs that allow multiple OSs, known as guests, to run in virtual machines in an isolated fashion, and thus share a single physical machine, or host. It is responsible for sharing the data center resources between virtual clients. The virtual clients request for their service from the cloud (data center) by making hyper calls to the hypervisor and utilize the resources [5]. While the virtual client requests for their service, it suffers from the problem of sharing the resources available. It is solved virtually with the help of hypervisor available. The sharing of resources is done by the hypervisor, but the most imperative circumstance to be noted down is the virtual hypervisor security [6]. Since hypervisor-based cloud servers are always exposed to attacks, it can easily be exploited to take down the whole cloud along with its resources. In Fig. 1, each client in the cloud utilizes its resources allocated by the data center. When the targeted virtual client is attacked by the attacker, the resources will be hacked by the malware-injected virtual client and it utilizes the resources of other virtual clients. Even though hypervisor interaction is reduced, the malware-injected virtual client starts gaining its control over the hypervisor and it may also get attacked sometimes. To illustrate this problem, the proposed system uses Fuzzy ART [7]. This strengthens the hypervisor to find out the trust in the usage of resources by the virtual clients. The fuzzy adaptive resonance theory (ART) technique uses the input factors such as (memory, CPU and disk space) for each virtual client and follows an unsupervised learning method to train and test the virtual clients. With this fast learning algorithm, the virtual clients are classified into four categories such as secure, vulnerable, modified, and anomaly based on the vigilance threshold. Smaller the vigilance threshold, higher categories are obtained.

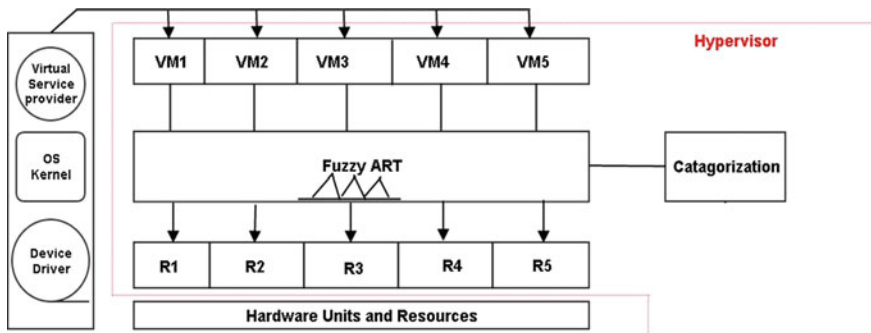


Fig. 1 Deployment of user behavior trust using Fuzzy ART categorization

3.1 Categorization of Attacks Using Fuzzy ART

In Fuzzy ART architecture, if an input pattern is different from the patterns (i.e., weights) that have stored in the network, then it will create new category and different input pattern will be associated with it. The ART1 was designed to process binary data only, whereas the proposed Fuzzy ART is designed to learn both the binary and analog patterns [8]. The Fuzzy ART uses unsupervised learning category [9]. This network consists of two layers, the input layer F1 (resources: memory, GFLOPS, disk space) for each virtual client in the network and the output layer F2 (categories: secure, vulnerable, modified, anomaly). The F1 neuron i and the F2 neuron j are interconnected by top-down weight and bottom-up weight W_{ij} where W_{ij} is chosen as discussed in Theorem 1. On the other hand, the orienting subsystem has a classification precision called vigilance parameter (ρ). The learning Fuzzy ART algorithm is as follows:

- Step 1: The input resources (memory, CPU, disk space) of each client is normalized as $I \in [0, 1]^n$ is given to input layer F1 satisfying the Eq. 2,

$$NI_{i,j} = (I_{i,j} - \text{Min}_{i,j}) / (\text{Max}_{i,j} - \text{Min}_{i,j}) . \tag{4}$$

where $i = \text{Virtualclient} (1, 2, \dots, m)$; $j = \text{Resources} (1, 2, 3)$

- Step 2: Initially, the weights are assigned as 1, $W_{i,j,s}(0) = 1$ with category set as 1 ($s = 1$).
- Step 3: Determine the dynamics of Fuzzy ART network where
 - $\alpha \in [0, 1]$ Choice parameter—It controls the choice of category whose weight vector W_{ij} is the largest coded subset of I
 - $\beta \in [0, 1]$ Learning Rate—It defines the degree to which weight vector W_{ij} is updated
 - $\rho \in [0, 1]$ Vigilance Parameter—It defines the required level of similarity of patterns within clusters

- Step 4: To categorize the input virtual clients with the specified usage of resources, the output node receives net input in the form of choice function for individual client as T_i , defined in Eq. 3, for any client in the cloud, the choice function value

$$T_i = |(NI_{i,j} - W_{ij})| / (\alpha + |W_{ij}|) . \tag{5}$$

where, $j \in \{1, 2, 3\}$ and $(a \wedge b)_i = \text{Min}(a_i, b_i)$; $|A| \equiv \sum^m |a_i|$

- Step 5: The winning category of individual client i , whose maximal T_i is found based on the resource usage limitation, $T_i = \max(T_i : i = 1, 2, \dots, m)$. If T_i is maximum for more than one, category i with smallest index is chosen.
- Step 6: To accept the virtual client, it should be nominated to a particular category the matching function; it should exceed the vigilance parameter (ρ),

$$M_{i,s} = |(NI_{ij} - W_{ij})|/|NI_{i,j}|. \quad (6)$$

If $M_{(i, s)} \geq \rho$, then pass that virtual client to existing category, else if $M_{(i, s)} < \rho$, then create a new category C_{s+1} .

- Step 7: The weight factor winning category is updated as follows

$$W_j^{\text{new}} = \beta(NI_{ij} - W_j^{\text{old}}) + (1 - \beta)(NI_{ij} - W_j^{\text{old}}). \quad (7)$$

- Step 8: This algorithm is repeated for each virtual client from step 4 to step 7; finally, the virtual clients are categorized into 4 categories as *secure*, *vulnerable*, *modified*, and *anomaly*.

4 Performance Analysis

Fuzzy ART analysis: We analyzed the proposed method to detect the abnormal handling of resources by the virtual clients in the cloud using the technique Fuzzy ART. This experiment is conducted in the MATLAB 7.1, and the results are interpreted. The Fuzzy ART is modeled with input factors such as memory, disk space, and GFLOPS. The Fuzzy ART configuration uses three predetermined values such as vigilance threshold, choice parameter, and learning rate. In Fuzzy ART, the lower the vigilance value struggles to precisely separate input patterns belonging to different input classes, whereas the higher the vigilance values favor good but result in multiple categories representing single input classes. The data interpreted in the Table 1 are trained by Fuzzy ART with above-predefined parameters. These are sample data in a single cloud service provider, but in a cloud, millions of clients are available and they are clearly categorized with this learning technique. To evaluate our system, sample statistics of virtual clients are categorized based on the limit in their usage of resources allocated to them. The

Table 1 Base data for training in Fuzzy ART

Resources	VC1	VC2	VC3	VC4	VC5
Memory	0.1818	1.0021	0.0013	0.2727	0.1515
Disk space	0.3784	0.9981	0.3538	0.0031	0.041
GFLOPS	0.0012	0.9998	0.1085	0.0182	0.0012

Table 2 Output classes of Fuzzy ART

Virtual clients	Categories
VC2	Secure
VC3, VC5	Vulnerable
VC2	Modified
VC4	Attacker

percentage of such categories in the network has a direct relationship with the network vigilance and performance. The greater accuracy in categorizing is evaluated using matching test function to finely categorize.

The output is categorized into 4 classes (secure, vulnerable, modified, and anomaly) shown in Table 2. The results clearly illustrate Fuzzy ART outperforms better than other learning techniques under tested conditions. Based on these results, it is clear that the Fuzzy ART configuration was able to achieve a training accuracy of 100 with only 4 categories. Thus, categorization used in Fuzzy Art is far better than other techniques with higher accuracy.

The resource for every virtual machine in the cloud depends on their usage which they demand from cloud service provider. The virtual machines may consume the resources from the data center, where the resources are centralized. The resources can be accessed by service which differs for every virtual machine in the cloud. They are bounded on certain limits for accessing the resources. The resource consumption rate for some virtual machines are rated based on their requirements

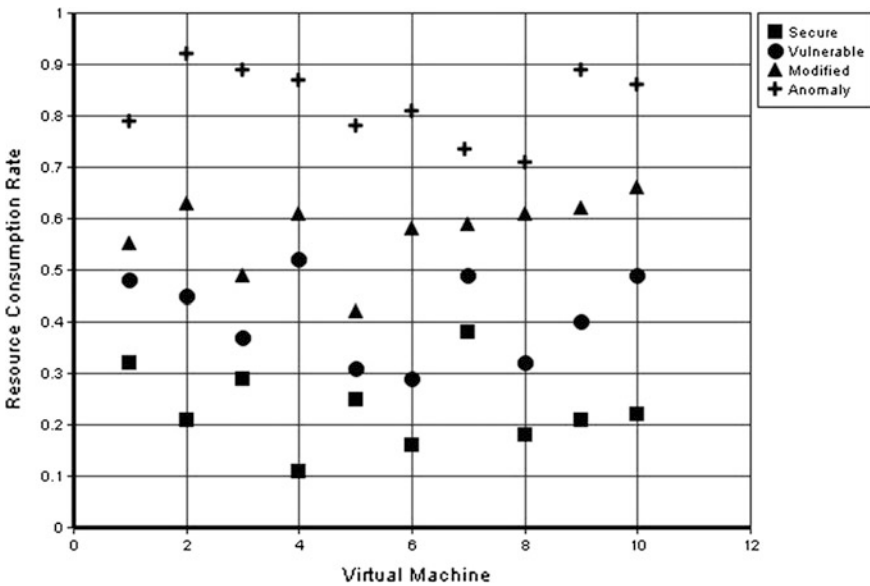


Fig. 2 Categorization of virtual machines into four categories based on their individual resource consumption rate

and their categorization classes are depicted with their usage of resources which is shown in Fig. 2. Based on these ratings, the virtual machines are classified by Fuzzy ART learning progress.

5 Conclusion

Thus, the contribution of this paper is to demonstrate that the classification of virtual machines is done in the cloud environment based on the behavior. The proposed model follows an unsupervised learning method to categorize the clients. Thus, the resources which are shared by the cloud are hacked and the priorities of clients are given. The clients which consume resources from others are traced out, and 4 categories of virtual clients are obtained. This allows it to learn a new input with minimum impact on its existing knowledge and accuracy.

References

1. K. Scarfone, M. Souppaya, P. Hoffman, *Guide to Security for Full Virtualization Technologies*. NIST, US Department of Commerce, Special Publication, pp. 125–800
2. M. Armbrust, A. Fox, R. Griffith, *Above the Clouds: A Berkeley View of Cloud Computing*
3. M. Jaiganesh, A.V. Antony Kumar, B3: fuzzy based data center load optimization in cloud computing. *Math. Probl. Eng.* **1**, 1–11 (2013)
4. J. Huang, M. Georgiopoulos, G.L. Heileman, Fuzzy ART properties. *Neural Netw.* **8**(2), 203–213 (1995)
5. M. Kuehnhausen, V.S. Frost Gary, J. Minden, Framework for assessing the trustworthiness of cloud resources, in *Proceedings of the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*, pp. 142–145 (2012)
6. M. Jaiganesh, A.V. Antony Kumar, B. Ramadoss, Hypervisor hardware fuzzy trust monitor in cloud computing, in *Elsevier Proceedings of the Eighth International Conference on Communication Networks*, pp. 11–19 (2013)
7. G.A. Carpenter, S. Grossberg, D.B. Rosen, Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Netw.* **4**(6), 759–771 (1991)
8. G. Aydn Keskin, S. Ilhan, C. Zkan, The fuzzy ART algorithm: a categorization method for supplier evaluation and selection. *Expert Syst. Appl.* **37**(2), 1235–1240 (2010)
9. C.-J. Lin, C.-T. Lin, An ART-based fuzzy adaptive learning control network. *IEEE Trans. Fuzzy Syst.* **5**(4), 477–496 (1997)
10. J. Huang, D.M. Nicol, Trust mechanisms for cloud computing. *J. Cloud Comput. Adv. Syst. Appl.* **2**(9), 1–14 (2013)
11. M. Nelson, C. Charles, Fernando, F. Marcos, A. Tereza, M. Pourzandi, A quantitative analysis of current security concerns and solutions for cloud computing. *J. Cloud Comput. Adv. Syst. Appl.* **1**(11), 1–21 (2012)
12. M. Jaiganesh, A.V. Antony Kumar, JNLP based secure software as a service in cloud computing, in *Proceedings of the International Conference on Mathematical Modeling and Scientific Computations (ICMMSC'12)—Communications in Computer and Information Science*, Springer-Verlag, Berlin, pp. 495–504 (2012)

Semi-supervised Learning Algorithm for Online Electricity Data Streams

Pramod Patil, Yogita Fatangare and Parag Kulkarni

Abstract Recent developments in electricity market deregulation, the prices are not fixed. In such application, class labels are not available directly and potentially valuable information is lost. A learning model of electricity demand and prices needs to be adaptive for dynamic changes in massive data streams. This paper presents adaptive building of learning model for electricity demand supply and prices by detecting and adapting changes in trends and values. A proposed framework is build with four main challenges such as future assumptions, data stream summarization, change in data stream trend clusters, and learner adaptivity and model. A proposed online algorithm for not only considering data values by avoiding trends of the streams. A correlation-based similarity method is used to produce concept clusters to handle unlabeled data and trend analysis, change detection type in terms of variation between past concept clusters and current ones, and predict future assumptions. An adaptive classify algorithm for the predictive ability evaluation on the test set. Results of experiments using electricity data confirm applicability of methodology with more than 80–85 % unlabeled data.

Keywords Multiple data streams · Data summarization · On-line update · Synopsis · Single scan · Clustering

1 Introduction

Electricity load depends on predictable load factors such as weather conditions, temporal factors, and customer characteristics. The planning of electricity production and prices also depends on daily peak load. It is essential to get the information of the local system demand in next minutes, hours, and days so that the generators

P. Patil (✉) · Y. Fatangare · P. Kulkarni
Department of Computer Engineering, College of Engineering, Pune, India
e-mail: pdpatiljune@gmail.com

Y. Fatangare
e-mail: yogita.fatangare@gmail.com

with various start-up times and start-up cost can be changed as per the requirement and knowledge gained from the historical data collected. In this paper, our proposed methodology is projected for several industry/organization to optimize energy usage. Electricity demand and pricing analysis are challenges in electricity planning and management.

A fundamental challenge in electricity data stream mining is to develop an online technique to summarize multiple data streams. A general-purpose summary can be utilized by a large number of data mining and management tasks over multiple streams, such as clustering, classification, change detection, statistical monitoring, selectivity estimation, query optimization, and query processing. While reducing the size of the data, the resulting summaries are typically designed to have certain qualities needed for the applications, such as preserving the original pairwise distances as much as possible. Also, the proposed system is not only considering data values by avoiding trends of the streams.

In this paper, we proposed multiple data stream clustering algorithm using correlation analysis. Correlation analysis on multiple data streams with the single-scan requirement is technically challenging since we cannot store the raw data as raw data are huge. We are using this technique that compresses the incoming data online and stores only the compressed measures, called the history concept, in the online system. We proposed a new theory that computes the correlation coefficients based on the history concept. The correlation coefficients are used to specify distances between different streams. These coefficients use dynamic online algorithm to generate the clustering results. A weighted history concept coefficient is also generated to understand the evolving behaviors of data streams and adjust the clusters according to it dynamically. However, due to the large stream number and the huge data volume in electricity data stream management, reclustering streams are very costly. Furthermore, periodical clustering is not able to cope with the electricity data streams with different evolving speeds. If the values of data streams are relatively steady, most of the clustering tasks are unnecessary since the resulting clusters are likely to remain the same. On the other hand, if the values of data streams are relatively fluctuant, we may lose some cluster information when the fixed time period is too long. Concluding from the above issues, we need a solution that is able to perform clustering whenever it is necessary.

This paper is organized as follows: Section 2 represents various clustering algorithms for data streams. In Sect. 3, framework design and various modules of proposed system are described. Mathematical model is presented in Sect. 3, and Sect. 4 represents methodology and algorithms for the proposed system. Finally, in Sect. 5, results are discussed.

2 Related Clustering Algorithm

The k-means algorithm is defined by Hartigan¹. This algorithm divides the data into k clusters, where k is given by the user as an input. The feature that differentiates the incremental version of k-means algorithm from standard k-means is an

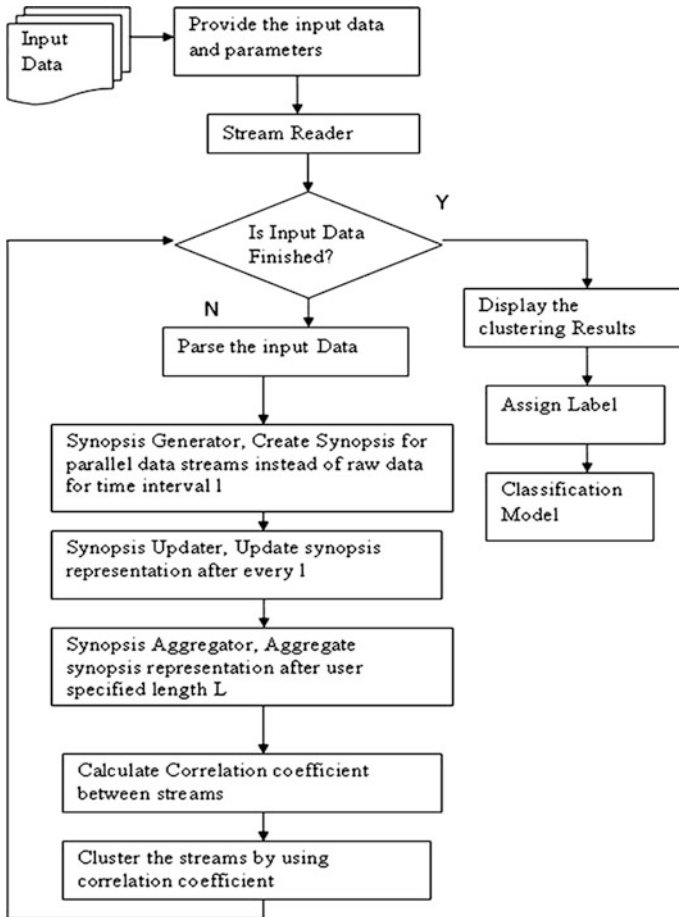
incremental use of the cluster number K . This adaption is very important for the applications where the clustering structure can change time to time. Deciding the right number of K is important in standard k-means also. Yang in [2] measures the distance between two streams using weighted aggregation of snapshot deviations. This technique observes the similarity of data values but ignores the trends of streams. Beringer and Hüllermeier in [3] proposed a technique that uses a discrete Fourier transform (DFT) approximation of the original data. This technique ignores important trends in the information contained in the data streams. Reason behind it is the data streams with similar trends may be not closed to each other in the Euclidean distance. O’Challaghan et al. [5] have proposed STREAM and LOCALSEARCH algorithms for high-quality data stream clustering. The STREAM algorithm starts by determining the size of the sample and then applies the LOCALSEARCH algorithm if the sample size is larger than a prespecified equation result. This process is repeated for each data chunk. Finally, the LOCALSEARCH algorithm is applied to the cluster centers generated in the previous iterations. Aggarwal et al. [6] have proposed a framework for clustering data streams called CluStream algorithm. They [7] have recently proposed HPStream: a projected clustering for high-dimensional data streams. HPStream has outperformed CluStream in recent results. Keogh et al. [8] have proved empirically that most highly cited clustering of time series data stream algorithms proposed so far in the literature come out with meaningless results in subsequence clustering. They have proposed a solution approach using k-motif to choose the subsequences that the algorithm can work on to produce meaningful results.

The COD framework [9] has two main features that are single pass for online statistics collection and compact multiresolution approximations. These two features are designed to address, respectively, the time and the space constraints in a data stream environment. With the use of multiresolution approximations of data streams, flexible clustering demands can be supported.

3 A Framework for Semi-supervised Learner for Online Data Streams

3.1 Block Diagram

The following figure explains design and its procedure to implement semi-supervised learning algorithm for online data streams.



4 Methodology and Algorithm

The proposed framework has two phases, i.e., online phase and offline phase. In online phase, electricity data streams are processed, and summarized format of same is created. In offline phase, clustering is performed.

4.1 Basic Concepts

Attenuation Coefficient: To give high priority to new data records than old data records, we are using attenuation coefficient λ [0 1], which slowly reduces the importance of each data record over time. If t is current time and x_i is received at time i , then we have formula to replace the original value of x_i ,

$$Xi(t) = \lambda^{t-i} xi \quad (1)$$

4.2 Online Phase

Data streams are divided into time segments: For efficiency, we are dividing the electricity data streams of length L into m time segments of equal length l . When new segment of length l completely arrived, recalculate clustering results.

Summary generation over multiple data streams: The formula for finding correlation coefficient between two data streams $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ is as follows:

$$\rho_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{j=1}^n (y_j - \bar{y})^2}}, \quad (2)$$

where

$$\bar{x} = \left(\sum_{i=1}^n x_i \right) / n \text{ and } \bar{y} = \left(\sum_{i=1}^n y_i \right) / n.$$

If the value of $|\rho_{XY}| = 1$, there is strong correlation between streams X and Y and $\rho_{XY} = 0$ indicates that X and Y are uncorrelated. In data stream world due to very large data, it is not possible to store all the data in the stream. So, there is need to summarize the data and only store the summary of each time segments each data stream. To calculate the correlation coefficient ρ_{XY} between two streams, we need to save $\sum_i x_i$, $\sum_i x_i^2$, $\sum_i y_i$, $\sum_i y_i^2$ information for each stream and $\sum_i x_i y_i$ between any two streams. We need to simplify Eq. 2. If current time is t and time segment length is l , then we need to store only the following values in

$$\begin{aligned} S_i &= \sum_{k=t-l+1}^t x_{ik}(t), \quad i = 1, \dots, n \\ Q_i &= \sum_{k=t-l+1}^t x_{ik}^2(t), \quad i = 1, \dots, n \\ C_{ij} &= \sum_{k=t-l+1}^t x_{ik}(t)x_{jk}(t), \quad i, j = 1, \dots, n, \quad i < j \end{aligned} \quad (3)$$

For two streams X_i and X_j , $i, j = 1 \dots n$, their correlation coefficients can be computed as follows:

$$\rho_{x_i x_j} = \frac{c_{ij} - \frac{1}{n} S_i S_j}{\sqrt{(Q_i - \frac{1}{n} S_i^2) (Q_j - \frac{1}{n} S_j^2)}} \tag{4}$$

Updating Summary Representation: Once the summarized correlation coefficient of summary structure at current time t_c is calculated, our next task is to update the saved information for time $t > t_c$ as follows:

$$\vec{S}'_i = \lambda^{A_t} \vec{S}_i, \vec{Q}'_i = \lambda^{2A_t} \vec{Q}_i, C'_{ij} = \lambda^{2A_t} C_{ij} \tag{5}$$

Update of summary structure will take place at every 1 steps.

Aggregation of Summary Structure: For a user-specified length of clustering L , we need to calculate the streams within the time $[t - L + 1, t]$. We have to calculate the summarized correlation structure (SCS) for the time segment with length l in the above sections. Now, we need to combine them to calculate the summarized correlation structure $_L$, i.e., for time window $[t - L + 1, t]$,

$$\text{Summarized correlation structure}_L = \text{SCS}_L = (\vec{S}_L, \vec{Q}_L, C_L) \tag{6}$$

We have, at time t

$$\begin{aligned} \vec{S}_{Li} &= \sum_{k=t-L+1}^t x_{ik}(t) = \sum_{v=1}^m \left(\sum_{k=t-vl+1}^{t-(v-1)l} x_{ik}(t) \right) \\ &= \sum_{v=1}^m \vec{S}_i(v), \quad \forall_i = 1, \dots, n, \end{aligned} \tag{7}$$

Similarly, we have

$$\vec{Q}_L = \sum_{v=1}^m \vec{Q}(v), \text{ and } C_L = \sum_{v=1}^m C(v). \tag{8}$$

Use the above equations to calculate summarized correlation structure $_L$ after receiving the first m time segments. For further updates, there is no need to recalculate the summations. We can incrementally update summarized correlation structure $_L$ as follows:

$$\vec{S}'_i = \vec{S}_L + \vec{S}_{(\text{new})} - \vec{S}_{(1)}, \vec{Q}'_i = \vec{Q}_L + \vec{Q}_{(\text{new})} - \vec{Q}_{(1)}, C'_i = C_L + C_{(\text{new})} - C_{(1)} \tag{9}$$

4.3 Offline Phase

We are using k-means clustering algorithm for generating clusters. In this algorithm, distance between two streams X and Y is measured as follows:

$$d(X, Y) = 1/\rho_{XY} \quad (10)$$

and clustering quality is measured using the objective function:

$$G = \sum_{i=1}^k \sum_{j=1}^n \left(1/\rho_{X_j} C_i\right). \quad (11)$$

Data streams generate very huge amount of data at a very high speed, and it changes over time so there are chances of creation of new clusters. Problem here is that user needs to specify the number of clusters.

Change of Cluster Number: To solve the above problem, we need to use the technique that continuously updates the number of clusters. Consider that k is the number of clusters given by previous clustering algorithm. Currently, we have $k + 1$ or $k - 1$ clusters. Then, we choose k' as the current number of clusters. We choose k' as one that produces smallest objective function G .

$$G_{k'} = \min\{G_{k-1}, G_k, G_{k+1}\} \quad (12)$$

4.4 Algorithm

Algorithm 1

It shows overall procedure of online clustering of multiple data streams.

Input: New data records from various streams.

Output: Clustering of data streams.

1. Initially $t = 0$;
2. Read new data $X_k(t)$; $k = 1 \dots n$, one from each of n data stream;
3. if $(t \bmod l == 0)$, then
4. Calculate the SCS of time segment $[t \ l + 1; t]$;
5. Update other $m - 1$ SCS, where $m = Ll$;
6. if $t == L$, then compute initial SCS_L, for $[t \ L + 1; t]$;
7. else incrementally update SCS_L;
8. Use online clustering algorithm();
9. Use update cluster algorithm to update the number of clusters;
10. Show clustering result;

5 Experimental Setup and Evaluation

5.1 Data set

In this section, data set used for testing is electricity data set. The data set that we use contains daily electricity demands from various cities. Each city is considered as one stream contains 2,000 points.

Results of Electricity Data Set: Our proposed algorithm generates clusters of most recent data elements of streams of fixed time length. The summary of data stream over user-specified fixed time length is computed in single pass only. Following are the results obtained from running clustering algorithm over multiple data streams. We ran clustering algorithm on electricity data set to cluster cities based on the recorded daily electricity demand. We set $L = 160$ and $l = 30$. The input of the algorithm is the daily electricity demand of cities. Our proposed system gave two clusters, each of which contains cities mostly in the same continent and belonging to the same electricity demand. The correct rate is around 80–85 %. The results are shown in Fig. 1, where each graph shows one cluster.

5.2 Evaluation Parameters

Speed: Processing speed of proposed algorithm that uses correlation distance to find trend similarity is higher than that of DFT cluster that uses Euclidean distance. The average processing time per segment for proposed cluster is 0.928 s, whereas average processing time for DFT cluster is 1.2 s. The results are shown in Figs. 2 and 3.

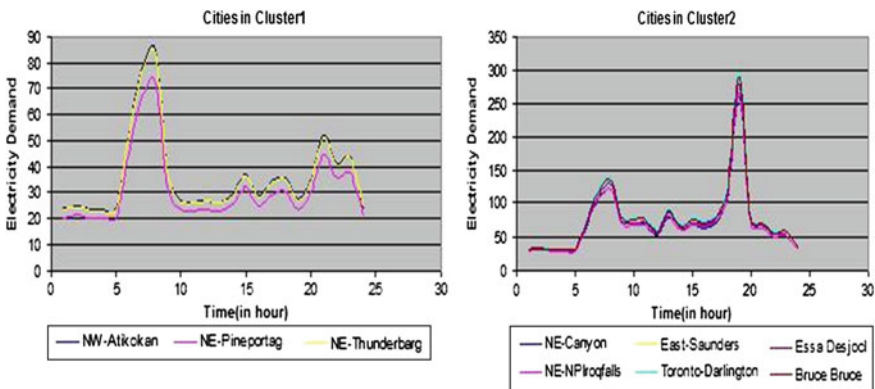


Fig. 1 Cities in cluster 1 and cluster 2

Fig. 2 Clustering speed of correlation-based and DFT cluster

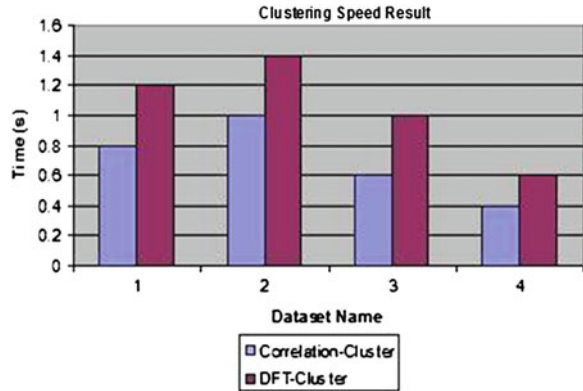
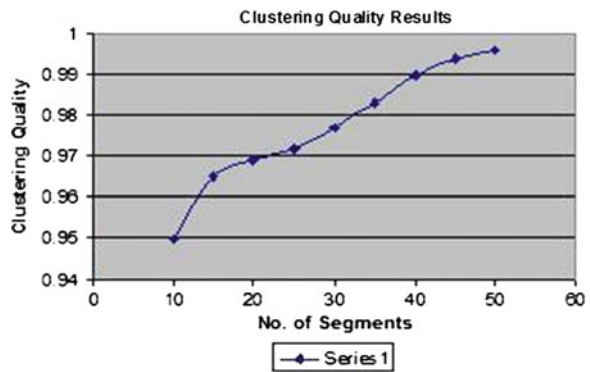


Fig. 3 Clustering quality varies as the number of segments varies



6 Conclusion

We proposed semi-supervised learning algorithm for electricity data streams, which uses correlation-based distance to find out similarity between data streams instead of Euclidean distance. It is used for online summarization of multiple data streams of fixed length L . Due to the use of compression technique, it creates synopsis of most recent data elements of data streams in single pass. So there is no need to store all the data elements in data stream. This algorithm can be used for determining the number of clusters dynamically to adjust with evolving nature of data streams. These cluster results can be used in other data mining purposes.

References

1. M.R. Anderberg, *Cluster Analysis for Applications* (Academic Press Inc, New York, NY, 1973)
2. J. Yang, Dynamic clustering of evolving streams with a single pass, in *Proceedings The 19th International Conference on Data Engineering* (2003), pp. 695–697
3. J. Beringer, E. Hüllermeier, Online-clustering of parallel data streams. *Data Knowl. Eng.* **58** (2), 180–204 (2006)
4. L. O’Callaghan, N. Mishra, A. Meyerson, S. Guha, R. Motwani, Streaming-data algorithms for high quality clustering, in *Proceedings of IEEE International Conference on Data Engineering* (2002)
5. C. Aggarwal, J. Han, J. Wang, P.S. Yu, A framework for clustering evolving data streams, in *Proceedings 2003 International Conference on Very Large Data Bases* (2003)
6. C. Aggarwal, J. Han, J. Wang, P.S. Yu, A framework for projected clustering of high dimensional data streams, in *Proceedings of 2004 International Conference on Very Large Data Bases* (2004)
7. E. Keogh, J. Lin, W. Truppel, Clustering of time series subsequences is meaningless: implications for past and future research, in *Proceedings of the 3rd IEEE International Conference on Data Mining* (2003)
8. B.R. Dai, J.W. Huang, M.Y. Yeh, M.S. Chen, Adaptive clustering for multiple evolving streams. *IEEE Trans. Knowl. Data Eng.* **18**(9) (2006)
9. M.-Y. Yeh, B.-R. Dai, M.-S. Chen, Clustering over multiple evolving streams by events and correlations. *IEEE Trans. Knowl. Data Eng.* **19**(10) (2007)
10. L. Kaufmann, P. Rousseeuw, Finding groups in data: an introduction
11. R. Ng, J. Hahn, Efficient and effective clustering methods for spatial data mining (1994)
12. T. Zhang, R. Ramakrishnan, M. Livny, BIRCH: A new data clustering algorithm and its applications. *Data Min. Knowl. Discov.* **1**, 141–182 (1997)
13. S. Guha, A. Meyerson, N. Mishra, R. Motwani, Clustering data streams: theory and practice. *IEEE Trans. Knowl. Data Eng.* **15**(3), 515–528 (2003)
14. C.C. Aggarwal, J. Han, J. Wang, P.S. Yu, A framework for clustering evolving data streams, in *Proceedings of conference of very large databases* (2003), pp. 81–92
15. A. Franzblau, *A Primer of Statistics for Non-Statisticians* (Harcourt, Brace, and World, California, 1958)
16. S. Guha, N. Mishra, R. Motwani, L. O’Callaghan, Clustering data streams, in *Proceedings of Annual Symposium Foundations of Computer Science* (2000)
17. C.C. Aggarwal, J. Han, J. Wang, P.S. Yu, A Framework for clustering evolving data streams, in *Proceedings of Very Large Data Bases Conference* (2003)

An Efficient Continuous Speech Recognition System for Dravidian Languages Using Support Vector Machine

J. Sangeetha and S. Jothilakshmi

Abstract This paper mainly focuses on developing a novel speech recognition system for Dravidian languages such as Tamil, Malayalam, Telugu, and Kannada. This research work targets to afford a well-organized way for human to interconnect with computers absolutely for people with disabilities who façade variety of stumbling blocks while using computers. This work would be very helpful to the native speakers in various applications. The proposed CSR system comprises of three steps namely preprocessing, feature extraction, and classification. In the preprocessing step, the input signal is preprocessed through the steps such as pre-emphasis filter, framing, windowing, and band stop filtering in order to remove the background noise and to enrich the signal. The best-filtered and the enriched signal from the preprocessing step is taken as the input for the further process of CSR system. The speech features being the most essential segment in speech recognition system. The most powerful and widely used short-term energy (STE) and zero-crossing rate (ZCR) are used for continuous speech segmentation, and Mel-frequency cepstral coefficients (MFCC) and shifted delta cepstrum (SDC) are used for recognition task. Feature vectors are given as the input to the classifier such as support vector machine (SVM) for classifying and recognizing Dravidian language speech. Experiments are carried out with real-time Dravidian speech signals, and the results reveal that the proposed method competes with the existing methods reported in literature.

Keywords Dravidian languages · CSR system · Support vector machine · Automatic speech recognition · Large-vocabulary speech recognition

J. Sangeetha (✉) · S. Jothilakshmi
Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar,
Chidambaram, India
e-mail: sangita.sudhakar@gmail.com

S. Jothilakshmi
e-mail: jothi.sekar@gmail.com

1 Introduction

Speech recognition is the capability of a computer to identify a universal, naturally flowing utterance from a wide variety of users. In recent years, with the new generation of computing technology, speech technology becomes the next major innovation in man-machine interaction. Automatic speech recognition (ASR) system takes a human speech utterance as an input and returns a string of words as output. Current research works on automatic continuous speech recognition (CSR) has led to a variety of applications such as hands-free and eyes-free applications, voice user interfaces, simple data entry, forensic applications, voice authentication, biometrics, robotics, air traffic controllers, preparation of medical reports, learning tools for handicapped, and reading tools for blind people. Even though research in speech recognition in English language attained certain maturity, speech interfaces in Indian languages are still in the start-up level.

In spite of the improvements made in this area, machines cannot tie the performance of human beings in terms of accuracy and speed particularly in the case of speaker-independent speech recognition systems. Since speech is the most important means of communication between people, research in automatic speech recognition and speech synthesis by machine have attracted a great deal of devotion over the past five decades [1, 2]. A host of methodologies for effective speech recognition have been articulated and evaluated using SPINE corpus. A grapheme-based automatic speech recognition system that jointly models phoneme and grapheme information using Kullback-leibler divergence-based HMM has been presented and investigated for English language using DARPA resource management (RM) corpus [3].

A novel context-dependent (CD) model for large-vocabulary speech recognition (LVSR) has been proposed that leverages recent advances in using deep belief networks for phone recognition [4]. There have been certain noticeable advances in discriminative training such as maximum mutual information (MMI) estimation [5], minimum classification error (MCE) training, and minimum phone error (MPE) training [6]. In large-margin approaches (such as large-margin estimation [7]), large-margin MCE [8], and boosted MMI, as well as in novel acoustic models (such as conditional random fields (CRFs) [9] has been proposed. Spoken word recognition strategy for Tamil Language based on HMM and AANN has been proposed [10].

Many research and progresses have been taken place in different Indian languages during the current years. However, Dravidian language speech recognition is still in its embryonic stage and very less work has been reported in Dravidian languages. This research work is predominantly applicable for native speakers where the people do not know any other languages other than native languages and it is applied in many real-time backgrounds such as railway, ATM, and weather forecasting.

The rest of the paper is organized as follows: A brief description about the method of extracting the features from the speech signal for automatic speech

segmentation and speech recognition is described in Sect. 2. Section 3 briefly reviews the support vector machine (SVM). The proposed algorithm for speech segmentation and the speech recognition task is presented in Sect. 4. Section 5 provides the experimental results for the proposed CSR system. Section 6 gives the conclusions and describes the future work.

2 Feature Extraction

Speech signals need to be parameterized prior to the identification process. Parameterization consists of the extraction of a set of features from the speech waveform, which may present two main characteristics: They must provide a reasonable and compact representation of the speech signal, and they must have adequate discrimination capabilities for discriminating between sounds. Three important features are used in this paper namely and MFCC feature of the speech recognition process.

2.1 Mel-Frequency Cepstral Coefficients

MFCC have proven to be one of the most successful feature representations in speech-related recognition tasks. The mel-cepstrum exploits auditory principles as well as the decorrelating property of the cepstrum. Figure 1 illustrates the computation of MFCC features for a segment of speech signal, which is described as follows [11]:

- The speech waveform is first windowed with analysis window and the discrete short-time Fourier transform (STFT) is computed.
- The magnitude is then weighted by a series of filter frequency responses whose center frequencies and bandwidths roughly match those of the auditory critical band filters. These filters follow the mel scale whereby band edges and center frequencies of the filters are linear for low frequency and logarithmically increase with increasing frequency as shown in Fig. 3. We call these filters as mel-scale filters and collectively a mel-scale filter bank. This filter bank, with 24 triangularly shaped frequency responses, is a rough approximation to actual auditory critical band filters covering a 4,000 Hz range.
- The log energy in the STFT weighted by each mel-scale filter frequency response is computed.
- Finally, discrete cosine transform (DCT) is applied to the filter bank output to produce the cepstral coefficients.

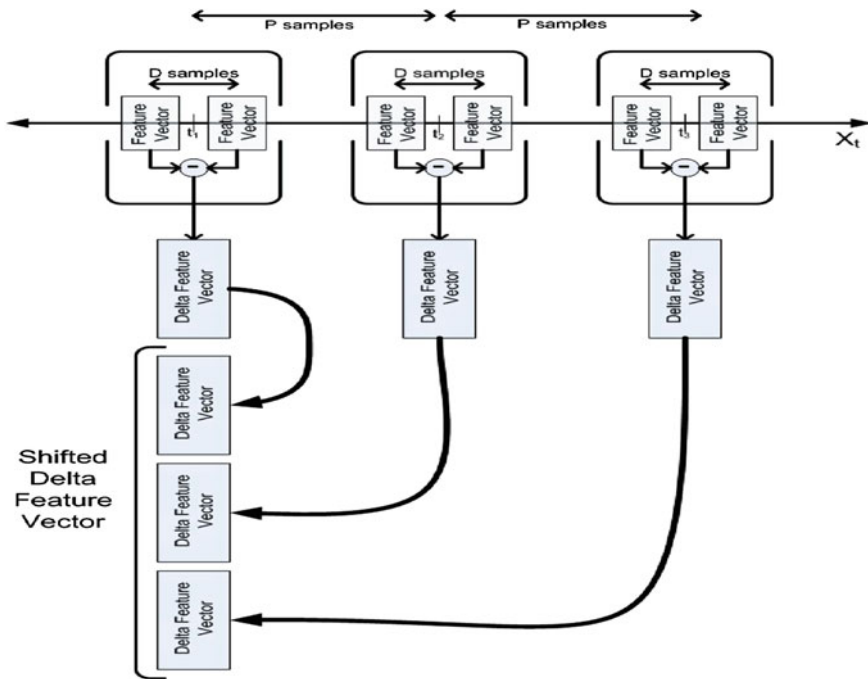


Fig. 1 Calculation of the shifted delta feature vectors

2.2 Shifted Delta Cepstrum (SDC)

The shifted delta cepstral features have been introduced to improve the recognition performance with respect to the classical cepstral and delta cepstral features [12]. The SDC coefficients are computed, for a cepstral frame at time t , according to:

$$\Delta c_n(t, i) = c_n(t + iP + D) - c_n(t + iP - D), \tag{1}$$

$$n = 0, \dots, N - 1, i = 0, \dots, k - 1$$

where n is the n th cepstral coefficient, D are the lag of the deltas, P is the distance between successive delta computations, and i is the SDC block number. The final feature vector is obtained by concatenation of k blocks of N parameters. The computation of the shifted delta feature vectors is a relatively simple procedure. The process is as follows:

The MFCC feature vectors are first computed as described above. Then, the acoustic feature vectors spaced D sample frames apart are first differences. Then, k differences feature vector frames, spaced P frames apart, are then stacked to form a new feature vector. Figure 1 gives a graphical depiction of this process.

3 Support Vector Machine (SVM)

Support vector machine (SVM) is based on the principle of structural risk minimization (SRM). Like RBFNN, support vector machines can be used for pattern classification and nonlinear regression. SVM constructs a linear model to estimate the decision function using nonlinear class boundaries based on support vectors. If the data are linearly separated, SVM trains linear machines for an optimal hyperplane that separates the data without error and into the maximum distance between the hyperplane and the closest training points. The training points that are closest to the optimal separating hyperplane are called support vectors. SVM maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen a priori. A linear decision surface is then constructed in this high-dimensional feature space. Thus, SVM is a linear classifier in the parameter space, but it becomes a nonlinear classifier as a result of the nonlinear mapping of the space of the input patterns into the high-dimensional feature space.

The support vector machine is a valued machine-learning technique that has been effectively applied in the pattern-recognition tasks [12, 13]. If the data are linearly indivisible but nonlinearly separable, the nonlinear support vector classifier will be applied. The fundamental idea is to make over input vectors into a high-dimensional feature space by means of nonlinear transformation and then to do a linear separation in feature space.

The SVM algorithm can build a variety of learning machines by use of different kernel functions. Four kinds of kernel functions such as linear kernel, Polynomial kernel, Gaussian radial basis function (RBF), and Sigmoidal kernel are usually used.

4 Proposed Continuous Speech Recognition

There is variety of automatic speech recognition methodologies presented such as neural networks, hidden Markov models, Bayesian networks, and dynamic time warping [14]. Among these methods, neural networks (NNs) have confirmed to be a powerful tool for resolving problems of forecasting, classification, and pattern-recognition issues [15].

4.1 Signal Preprocessing

It is very essential to preprocess the speech signal in the applications where silence or background noise is completely objectionable. In the preprocessing step, the input signal is preprocessed through the steps such as pre-emphasis filter, framing and windowing, and band stop filtering has been in order to remove the background

noise and to enrich the signal. The best-filtered and the enriched signal from the preprocessing step is taken as the input for the further process of CSR system

4.2 Speech Segmentation

Automatic speech segmentation is a necessary step that used in speech recognition and synthesis systems. Speech segmentation is breaking continuous streams of sound into some basic units such as words, phonemes, or syllables that can be recognized [13]. Following procedure has been used for automatically marking the boundaries in the sound file.

- Short-term energy and zero-crossing rates are computed for the preprocessed frames.
- Some threshold value that is dynamically generated has been taken and signals having a value less than this threshold value has been changed to zero as signal having syllable will have a data value more than the threshold value.
- Then, signal has been checked for value not equal to zero and greater than some particular value and that point will be marked as starting location of the boundary.
- After getting the starting location, the zero values of signal have been checked and if there are suitable numbers of continuous zeros, then it has been defined as the end of the boundary. Once an endpoint has been detected, we can precede analyzing signal from the endpoint of the first one looking for the starting position of next one.

4.3 Classification

The segmented speech frames are classified using SVM network to identify the spoken utterances of specific words in the input speech. The vocabulary includes the words in the speech corpus. Figure 2 shows the overview of the proposed CSR system.

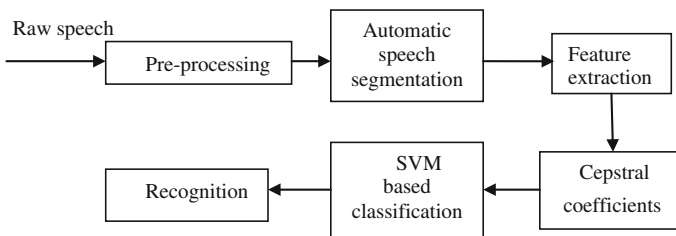


Fig. 2 System overview of continuous speech recognition system based on AANN

5 Experimental Results

5.1 Databases

In this research work, real-time speech database is considered for the banking sector and ATM centers applications. For the experimental simulation, around 125 people's voice samples were collected for various Dravidian languages such as Tamil, Malayalam, Telugu, and Kannada and evaluated. These samples were collected over a period of time and are used in this experimental simulation. The dataset is divided into the development corpus and evaluation corpus. The development corpus is used for training the system and tuning the parameters, which is composed of six 20 min speech of each language. The evaluation corpus is composed of six 10 min speech of each language, which is for validation. The performance of speech recognition systems is usually specified in terms of accuracy, error rate, and speed. The word recognition rate is used as the performance measure.

- Word Error Rate (WER).

Word error rate is a common metric for measuring the performance of a speech recognition system. Word error rate can be computed as

$$\text{WER} = \frac{S + D + I}{N} \quad (2)$$

where S is the number of substitutions, D is the number of the deletions, I is the number of the insertions, and N is the number of words in the reference. When reporting the performance of a speech recognition system, the word recognition rate (WRR) is used mostly.

$$\text{WRR} = 1 - \text{WER} \quad (3)$$

The recognition accuracy for testing set with the features MFCC and SDC for the proposed speech recognition system is presented in Table 1. Based on this, it is clearly shown that the SDC feature provides the better performance compared to MFCC and also provides the better performance.

Table 1 Performance of the proposed system based on MFCC and SDC features

Language	Total number of sentences	Total number of words present	Number of words tested	WRR using MFCC (%)	WRR using SDC (%)
Tamil	250	1,378	100	93	96
Malayalam	250	1,257	100	91	94
Telugu	250	1,550	100	89	95
Kannada	250	1,345	100	91	96

6 Conclusion

In this paper, such a significant effort has been carried out for recognizing Dravidian Languages using acoustic features. To achieve this job, the desired feature extraction is done after performing required preprocessing techniques. The most extensively used MFCC and SDC are used to extract the substantial feature vectors from the enriched speech signal, and they are given as the input to the SVM classifier. The adopted SVM classifier is trained with these input and target vectors. Experiments were carried out to contrast the performance of the system with MFCC with delta and acceleration coefficients and SDC individually. The system with SDC features performs better than the system with MFCC features. The results with the specified parameters were found to be agreeable considering the less number of training data. The more number of continuous speeches has to be trained and tested with this network in future. As the Dravidian languages are alike in characteristics, designing a lesser amount of intricate system with the best performance is a challenging task. This work is the principal step in this track.

References

1. V. Radha, Efficient speaker independent isolated speech recognition for tamil language using wavelet denoising and hidden markov model. in *Proceedings of the Fourth International Conference on Signal and Image Processing* (2012)
2. M. Magimai Doss, R. Rasipuram, G. Aradilla, H. Bourlard, Grapheme-based automatic speech recognition using KL-HMM, in *Proceedings of Inter Speech* (2011)
3. B. liu, Research and implementation of the speech recognition technology based on DSP, in *International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)* (2011)
4. G.E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **20**(1) (2012)
5. E. McDermott, T. Hazen., J.L. Roux, A. Nakamura, S. Katagiri, Discriminative training for large vocabulary speech recognition using minimum classification error. *IEEE Trans. Speech Audio Process.* **15**(1), 203–223 (2007)
6. D. Povey, P. Woodland, Minimum phone error and smoothing for improved discriminative training, in *Proceedings of ICASSP* (2002), pp. 105–108
7. H. Jiang, X. Li, Incorporating training errors for large margin HMMs under semi-definite programming framework, in *Proceedings of ICASSP*, vol. 4 (2007), pp. 629–632
8. D. Povey, D. Kanevsky, B. Kingsbury, B. Ramabhadran, G. Saon, K. Visweswariah, Boosted MMI for model and feature space discriminative training, in *Proceedings of ICASSP* (2008), pp. 4057–4060
9. G. Heigold, A log-linear discriminative modeling framework for speech recognition. Ph.D. dissertation, Aachen University, Aachen, Germany, 2010
10. A.N Sigappi, Spoken word recognition strategy for Tamil Language. *Int. J. Comput. Sci. Issues* **9**(3) (2012)
11. HTK book (2002)
12. A. Geetha, V. Ramalingam, B. Palaniappan, S. Palanivel, Facial expression recognition—a real time approach. *Expert Syst. Appl.* **36**(1), 303–308 (2009)

13. Saheli A.A., Abdali G.A., A.A. Suratgar, Speech recognition from PSD using neural network, in *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, vol. I, Hong Kong, 18–20 March 2009
14. M. Antal, Speaker independent phoneme classification in continuous speech. *Studia Univ. Babeş-Bolyai Informatica* **49**(2) (2004)
15. B. Yegnanarayana, S.P. Kishore, AANN: an alternative to GMM for pattern recognition. *Neural Networks* **15**, 459–469 (2002)

Video Surveillance Based Tracking System

R. Venkatesan, P. Dinesh Anton Raja and A. Balaji Ganesh

Abstract The paper presents video surveillance-based tracking system for the outdoor environment. The video processing has been done using both LabVIEW and MATLAB, and the comparisons are illustrated. In LabVIEW, self-learning and real-time moving object tracking are implemented for the surveillance of environment. In a two-dimensional color pattern matching, a self-learned template has to be located on real-time video, regardless of the template's position, color, and shape. It is done by organizing a set of feature vectors that encompass all the variations in the self-learned template. Matching is then done by determining the best similarity between the feature vectors extracted from the video image and the self-learned template set. Template learning time and elapsed time are taken as parameters for the comparison. In MATLAB, tracking is achieved using Camshift algorithm. The feature vectors include position, orientation, and size of the object to be tracked.

Keywords Camshift algorithm · Self-autonomous video surveillance · Tracking system · Region of interest · LabVIEW and MATLAB

1 Introduction

Self-autonomous video surveillance is most required to minimize the crimes especially in the outdoor environment. Video surveillance involves series of video tracking processes [1]. Based on the performance and the requirement of manual

R. Venkatesan · P.D.A. Raja · A.B. Ganesh (✉)
Department of TIFAC CORE, Velammal Engineering College, Chennai, India
e-mail: abganesh@velammal.edu.in

R. Venkatesan
e-mail: venky88an@gmail.com

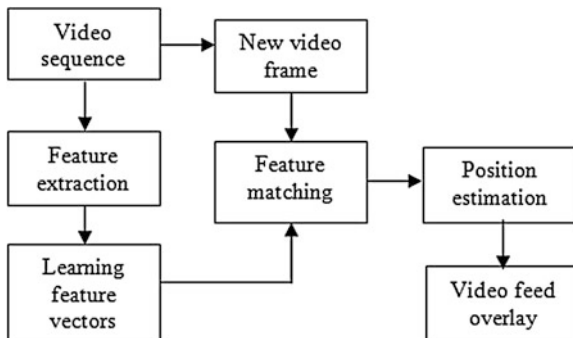
P.D.A. Raja
e-mail: pdaraja@gmail.com

intervention in tracking a particular object, the video surveillance has been grouped into three types, namely manual, semi-autonomous, and self-autonomous. Manual video surveillance involves analysis of the video content by a human. Such systems are currently widely used. Semi-autonomous video surveillance involves some form of video processing, but with significant human intervention. Typical examples are systems that perform simple motion detection. Only in the presence of significant motion, the video is recorded and sent for analysis by the human expert. By a fully autonomous system, only input is the video sequence taken at the scene where surveillance is performed. In such a system, there is no human intervention, and the system does both the low-level tasks, such as motion detection and tracking, and also high-level decision-making tasks, such as abnormal event detection and gesture recognition.

2 Tracking System

Figure 1 shows the block diagram of tracking. It illustrates the overall methodology of tracking system. Video sequence involves the continuous sequence of frames (images) which are used for the process of tracking. Feature extraction illustrates the parameters, such as color, shape, and orientation that are to be extracted in the captured image. These features are extracted by specifying the region of interest (ROI). The template is an idealized representation of a feature in the image. When the ROI is drawn on the target object, the feature is extracted from that object. So then, the template generated depends upon the features. The learned template will be used for matching performance. Position estimation finds the specific location of the object in the current frame.

Fig. 1 Block diagram of video processing-based tracking system



3 Object Tracking Mechanism

The algorithm of object tracking has been implemented using both LabVIEW and MATLAB.

LabVIEW is a system design platform and development environment for virtual programming language. The vision development module is the tool box which has functions of image processing. The program will be constructed in the form of graphical interface functions. The functions that used for the implementations are as follows: IMAQ setup learn color pattern, the learn mode, and feature mode. The color score weight (between 0 and 1,000) determines the percent contribution of the color score to the final color pattern matching score. The software uses the color score weight for the final match ranking. If the color score weight is specified as 1,000, the algorithm finds each match using both color and shape information and then ranks the matches based on their color scores. If the weight is 0, the matches are ranked based on their shape scores. The default is 500, indicating that the match score uses an equal combination of the color and shape scores. Match mode specifies the invariance mode to use when looking for the color template pattern in the inspection image [2].

The algorithm is illustrated as flowchart model and shown in Fig. 2. Each operation is carried out in LabVIEW.

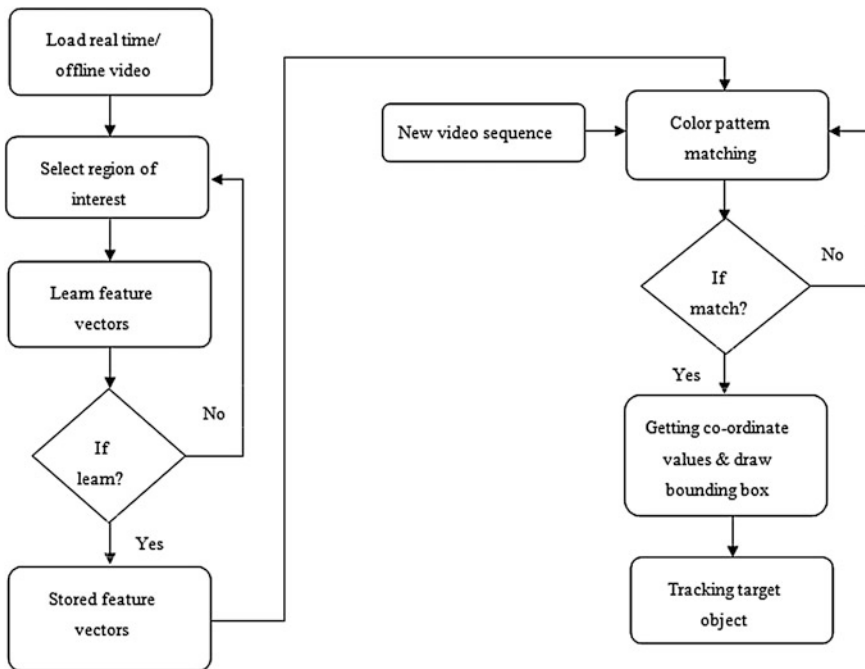


Fig. 2 Flowchart model for the implementation of tracking

The LabVIEW image processing sequence receives the captured video file directly from the camera interface.

$$\text{Match score}_{\text{out}} = \frac{\text{Matched pixels}}{\text{Total pixels in ROI}} \times 1000. \quad (1)$$

The Eq. (1) represents the match score calculation criteria. A factor of 1,000 has been taken to avoid decimal values and provide a higher precision to the checking of the image. The pixel matching technique is being used here [2]. The machine vision application then searches for instances of the template in each acquired image, calculating a score for each match. This score relates how closely the template resembles the located matches. Then, the pattern matching finds template matches regardless of lighting variation, shifting, and rotation [3]. The matching algorithm used depends on whether the user has specified shift-invariant matching (finding the template at any location in the search image) or rotation-invariant matching (finding the template at any location and rotation in the search image). Shift-invariant matching is based on the normalized cross-correlation. The following is the basic concept of correlation,

$$C(i \cdot j) = \sum_{x=0}^{L-1} \sum_{y=0}^{k-1} w(x, y) f(x + i, y + j). \quad (2)$$

Consider a sub-image $w(x, y)$ of size $K \times L$ within an image $f(x, y)$ of size $M \times N$, where $K \leq M$ and $L \leq N$. The correlation between $w(x, y)$ and $f(x, y)$ at a point (i, j) is given. Where $i = 0, 1, \dots, M - 1$, $j = 0, 1, \dots, N - 1$, and the summation is taken over the region in the image where w and f overlap. If the matching is done, the target object is overlaid by the bounding box. This is achieved by getting the coordinate values of the object to be tracked. It results in the object tracked with bounding box.

3.1 Functionality and Implementation in MATLAB

The mean-shift algorithm is a nonparametric method [4]. It provides accurate localization and efficient matching without expensive exhaustive search method. In this, the size of the searching window is fixed [5]. It is an iterative process, first compute the mean-shift value for the current point position, then move the point to its mean-shift value as the new position, and then compute the mean-shift until it fulfills certain conditions. For a frame, it used distribution of levels of gray which gives the description of the shape and to converge on the center of mass of the object calculated by means of moments. The inputs of this module are the initial position, width, and length of the initial search window of the object. Let $I(x, y)$ the image of "2D", (x, y) is the position of the object in the image. The zero-order

moment is given in Eq. (3), and first order moment spatial and the center of mass are given in Eqs. (4) and (5). These parameters are the output of mean-shift algorithm module and the input for the Camshift algorithm module. The moment of order zero (M_{00}) represents the area occupied by the shape of the frame [1].

$$M_{00} = \sum_x \sum_y I(x, y). \quad (3)$$

The moments of order one (M_{10} , M_{01}) are calculated by Eq. (4).

$$M_{10} = \sum_x \sum_y x \times I(x, y); M_{01} = \sum_x \sum_y y \times I(x, y). \quad (4)$$

The center of mass for an object can be calculated by means of the moments of zero order, and one (x_c , y_c) is expressed by the Eq. (5).

$$x_c = \left(\frac{M_{10}}{M_{00}} \right); y_c = \left(\frac{M_{01}}{M_{00}} \right). \quad (5)$$

Camshift algorithm. The principle of the Camshift algorithm is based on the principles of the algorithm mean-shift. Camshift is able to handle the dynamic distribution by adjusting the size of the search window for the next frame based on the moment zero of the current distribution of images [6–8]. In contrast to the algorithm mean-shift who is conceived for the static distributions, Camshift is conceived for a dynamic evolution of the distributions. It adjusts the size of searching window by invariant moments. This allows the algorithm to anticipate the movement of objects to quickly track the objects in the next frame. Even during the fast movements of an object, Camshift is still capable of tracking well. It occurs when objects in video sequences are tracked, and the object moves such that the size and location of the change in probability distribution change in time. The initial search window was determined by a detection algorithm or software dedicated to video processing. The initial search window was determined by a detection algorithm or software dedicated to video processing. The Camshift algorithm calls upon the mean-shift one to calculate the target center in the probability distribution image, but also the orientation of the principal axis and dimensions of the probability distribution. Defining the first and second moments is defined by Eqs. (6–8).

$$M_{20} = \sum_x \sum_y x^2 \times I(x, y). \quad (6)$$

$$M_{02} = \sum_x \sum_y y^2 \times I(x, y). \quad (7)$$

$$M_{11} = \sum_x \sum_y x \times y \times I(x, y). \quad (8)$$

The orientation of the major axis and the scale of the distribution are determined by finding an equivalent rectangle that has the same moments as those measured from the 2D probability distribution image. The orientation is defined by the Eq. (9).

$$2\theta = \arctan \left(\frac{2 \times \left(\frac{M_{11}}{M_{00}} \right) - (x_c \times y_c)}{\left(\left(\frac{M_{20}}{M_{00}} \right) - x_c^2 \right) - \left(\left(\frac{M_{02}}{M_{00}} \right) - y_c^2 \right)} \right). \quad (9)$$

The first two Eigen values (the length and width of the probability distribution) are calculated in closed form as follows in the Eqs. (10–12).

$$a = \frac{M_{20}}{M_{00}} - x_c^2. \quad (10)$$

$$b = 2 \times \left[\frac{M_{11}}{M_{00}} - x_c \times y_c \right]. \quad (11)$$

$$c = \frac{M_{02}}{M_{00}} - y_c^2. \quad (12)$$

4 Results and Discussions

4.1 Results for LabVIEW

As mentioned above, color pattern matching has been implemented which includes shift-invariant and rotated-invariant algorithm to track exact moving object in real-time/off-line video as effective as possible. The paper is proposed to include learning of template based on color and shape of moving objects in traffic environment which provide high-accuracy and less computational complexity. Color score weight chosen for exact learned object also plays an important role in tracking of moving object in real-time/off-line environment. The following results also include learning and elapsed time of moving objects in traffic environment for both shift- and rotation-invariant algorithm.

Figure 3 shows the front panel of LabVIEW which obtain the corresponding input video off-line/real time meant for select particular moving object, learn template and choose exact color score weight to track accurate object for a traffic environment. Comparisons of accuracy levels of tracking objects with respect to color score weight in traffic environment are illustrated.

Figures 4 and 5 show the exact tracking of moving objects in traffic environment which is based on color score weight. The different color score weights are compared and illustrated here. Also, the learning time and elapsed time are illustrated.



Fig. 3 Moving object tracking







video 1	Iteration no	Color score weight Range (0-1000)	Whether tracked the template	Reason	Learning time (secs)	Elapsed time (secs)	video 1	Iteration no	Color score weight Range (0-1000)	Whether tracked the template	Reason	Learning time (secs)	Elapsed time (secs)
	1	400	Yes	It tracks the target object	7.07	0.2		1	300	Yes	It tracks the target object	3.93	0.25
	2	500	Yes	It tracks the target object									
	3	600	No	It tracks no objects with same color & different shape									
	4	1000	No	It tracks many objects with same color									
Learned Template 1	Color score weight	Tracking target object					Learned Template 2	Color score weight	Tracking target object				
	400							300					

Fig. 4 Comparisons for template 1 and template 2

Figure 6 shows tracking of object which has been rotated 180° using rotation-invariant algorithm.

4.2 Results for MATLAB

Figure 7 shows the tracked object at different position by updating the search window that includes both color and orientation of moving object in the consecutive video frames using Camshift algorithm.




video 1	Iteration no	Color score weight Range (0-1000)	Whether tracked the template	Reason	Leaming time (secs)	Elapsed time (secs)	
Leamed Template 3 	1	300	Yes	It tracks the target object	11.09	0.15	
	2	400	Yes	It tracks the target object			
	3	600	No	It tracks two objects with same color & different shape			
	4	700	No	It tracks many objects with same color			
Leamed Template 3	Color score weight	Tracking target object					
	300						

Fig. 5 Comparisons for template 3

Leamed Template for rotation invariant	Color score weight	Target tracking
	800	

Fig. 6 Tracking for rotation invariant



Object to be tracked	Tracked object at different video frames	Elapsed time (secs)
		<p>0.12</p>

Fig. 7 Tracking using Camshift algorithm

5 Conclusions

The paper describes the procedures for the implementation of semi-autonomous video surveillance system using both LabVIEW and MATLAB. The performance comparisons are done between LabVIEW and MATLAB. The evaluation parameters considered are size, orientation, and color of the pattern to be tracked. Also, tracking and elapsed time are calculated. In MATLAB, tracking is achieved using Camshift algorithm. Both techniques are found suitable for the surveillance of vehicles. However, the functions are in LabVIEW very limited when it has been compared with MATLAB.

Acknowledgments The authors wish to thank Department of Science and Technology for awarding a project under Cognitive Science Initiative Programme (DST File No.: SR/CSI/09/2011) through which the work has been implemented.

References

1. G. Sharma, S. Sood, G.S. Gaba, N. Gupta, Image recognition system using geometric matching and contour detection. *Proc. Int. J. Comput. Appl.* **51**(17), 48–53 (2012)
2. P. Armen, Vision system for disabled people using pattern matching algorithm. In *Proceedings of the Seventh International Conference on Computer Science and Information Technologies*. (2009), pp. 343–346

3. A. Salhi, A.Y. Jammoussi, Object tracking system using Camshift, meanshift and kalman filter. World Academy of Science Engineering and Technology. (2012)
4. S. Avidan, Support vector tracking. in *IEEE Conference on Computer Vision and Pattern Recognition*. **8**, 184–191 (2001)
5. G. Bradski, T. Ogiuchi, M. Higashikubo, Visual tracking algorithm using pixel-pair feature. Int. Conf. Pattern Recogn. **4**, 1808–1811 (2010)
6. Y. Ruiguo, Z. Xinrong, The design and implementation of face tracking in real time multimedia recording system. *IEEE Trans.* **3**, 1–3 (2009)
7. E. David, B. Erich, K. Daniel, S. Anselm, Fast and robust Camshift tracking. *IEEE Trans.* **8**, 1–8 (1982)
8. B. Wang, Q. Yang, C. liu, M. Cui, An efficient method for face feature extraction based on contourlet transform and fast independent component analysis. 4th *International Symposium on Computational Intelligence and Design*. (2011), pp. 344–347

Stroke Detection in Brain Using CT Images

S. Neethu and D. Venkataraman

Abstract Computed tomographic (CT) images are widely used in the diagnosis of stroke. The objective is to find the stroke area from a CT brain image and also improve the visual quality. The proposed algorithm helps to detect the stroke part in the absence of radiologist or doctors. Seed region growing (SRG) technique is the most popular method for segmentation of medical images because of high-level knowledge of anatomical structures in seed selection process. The proposed method consists of three steps: preprocessing, feature extraction, and segmentation. Feature extraction is done based on texture using the Gabor filter, and segmentation is done using SRG algorithm.

Keywords Wiener filter · Histogram equalization · Gabor filter · Seed region growing algorithm

1 Introduction

Stroke is the rapid loss of brain function due to disturbance in the blood supply to the brain. Stroke is a disease which affects vessels that supply blood to the brain. It is defined in medical term as sudden death of brain cells due to lack of oxygen. Stroke can be divided into two types: ischemic stroke and hemorrhagic stroke. About 85 % of all strokes are of ischemic type. It occurs as a result of an occlusion of arteries due to thrombus. Hemorrhagic stroke occurs due to rupturing of a weakened blood vessel. There is a possibility for co-occurrence of both ischemic stroke and hemorrhagic stroke.

S. Neethu (✉) · D. Venkataraman
Computer Vision and Image Processing, Department of Computer Science,
Amrita Vishwa Vidyapeetham University, Coimbatore, India
e-mail: neethussneethu.s@gmail.com

D. Venkataraman
e-mail: dvenkataramanindia@gmail.com

Stroke results in serious long-term disability or death. The symptoms of stroke are weakness, clumsiness, altered feeling on one side of the body, speech disturbance, loss of vision, or dizziness. The following are the risk factors for stroke: old age, high blood pressure, previous stroke or transient ischemic attack (TIA), diabetes, high cholesterol, tobacco smoking, and atrial fibrillation. High blood pressure is the most important modifiable risk factor of stroke. Stroke is diagnosed through several techniques. These techniques are neurological examination (such as the NIHSS), Computed tomographic (CT) scans (most often without contrast enhancements) or MRI scans, Doppler ultrasound, and arteriography strokes. This can be due to ischemia (lack of blood flow) caused by blockage (thrombosis, arterial embolism), or a hemorrhage, i.e., due to blockage/blood clot or inside bleeding [1]. In CT images, a hemorrhage appears as a bright region and ischemic stroke appears as a dark region with the contrast, relative to its surrounds. The Digital Imaging and Communications in Medicine (DICOM) standard allows to share information and resources among various medical imaging techniques [2].

2 Survey on Previous Works

In this literature survey, we found that many stroke detection and classification techniques from a series of CT scan brain images have been proposed by different authors.

Li et al. [3] proposed a system for the detection of subarachnoid hemorrhage (SAH) and segmentation of subarachnoid space. In their method, they have applied morphological operation and fuzzy c means clustering to extract the brain part. After that feature extraction based on distance to found the five land mark of brain using distance map [4]. Classification based on Bayesian decision theory was used to two categories for each pixel, being either within or outside SAS.

Chawla et al. [5] proposed an automated method to detect and classify an abnormality into acute infarct, chronic infarct, and hemorrhage at the slice level of non-contrast CT images. The knowledge about the anatomical structure of the skull and the brain is used to detect abnormalities. A two-level classification to identify abnormal and normal slices was performed by the detection algorithm. Histogram features were used in the first level, while in the second level, wavelet-based features were used [6]. The advantage was the ability to detect all types of strokes (acute, chronic infarcts, and hemorrhages) even if different types are present in the same slice. Their system failed only if same type of stroke occurred symmetrically in both hemispheres.

Kyaw [7] proposed an easy method for preprocessing and presegmentation, which helped to detect the abnormal regions in the brain image. He proposed a method in which the given image was divided into four quadrants and the histogram was calculated. Then, the statistical features, i.e., mean and standard deviation, were used to distinguish the normal and abnormal regions. Jeena and Kumar [8] proposed a new method using which they compared the stroke disease in CT and MRI

modality. In their method, texture-based segmentation was performed by using the Gabor filter. After that, they performed the seed region growing (SRG) algorithm. They concluded that the MRI images gave better result.

Rai and Nair [9] introduced a method in CT angiography images. Region growing algorithm depends on the initial seed selection and criteria used to terminate the recursive region growing process [10]. This algorithm works based on the following criteria: region homogeneity, object contrast with respect to background, strength of the region boundary, size, and conformity to desired texture features like texture, shape, and color. Mostly used is region homogeneity. Thakur et al. [4] introduced a method for disease detection based on the texture property. For the texture-based segmentation, they have used Gabor filter [10]. It was found that the Gabor filter improved the data transfer rates, provided efficient noise reduction, less power consumption, and required less memory storage. This method was found to be helpful in detecting early stages of disease.

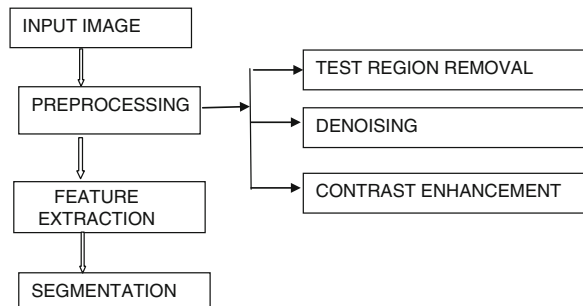
3 Proposed Method

The proposed system consists of three modules, namely preprocessing, feature extraction, and segmentation. The system design is described in Fig. 1. Each module will be detailed in the upcoming sections.

3.1 Data Set Details

The data set is collected from “Malankara Orthodox Syrian Church Medical College Hospital” (MOSC MCH) Kolenchery. Eighty CT images are collected with the format of DICOM. Thickness of the image is 5 mm and interval 10 mm. Size is 512×512 . The standard window size is 40 WL and 80 WW. Total images are divided into three categories as 30 block images, 22 hemorrhage images, and 24 normal images. Some images are downloaded form www.radiographers.org.

Fig. 1 System design



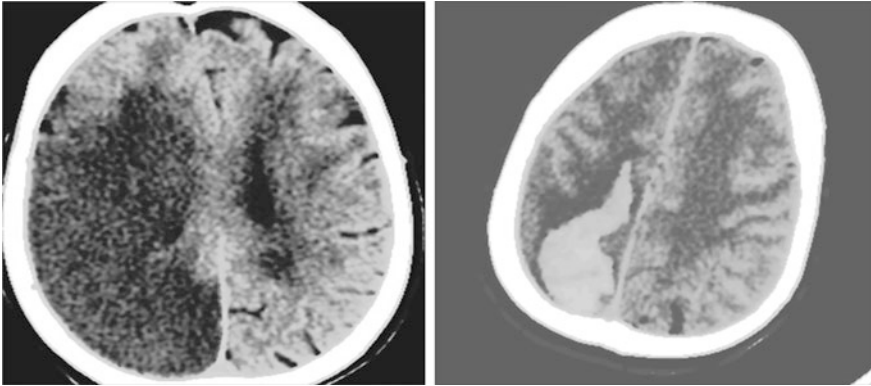


Fig. 2 Preprocessing result of stroke image and hemorrhage image

3.2 Preprocessing

In the preprocessing stage, we are performing three steps. In the first step, the text region in the images is removed using morphological operations. In the second step, the noise is removed. The noise removal is performed using median filter, Weiner filter, and wavelet decomposition, after which the results are compared using their PSNR values and the best method is chosen. In our experiment, we found that using Weiner filter gave better results for noise removal. Finally, we apply histogram equalization on the image obtained after noise removal for contrast enhancement. This helped in better visualization of the stroke area in the brain CT images (Fig. 2).

3.3 Find Whether the Stroke Is Present or Not

After preprocessing, we have to check the presence of stroke. In this paper, we are using the method that the given input image is divided into four regions. And then find the statistical measures like mean and standard deviation for each region. From that, we can find the region of the stroke. And the remaining feature extraction and segmentation is performed on that particular region (Fig. 3).

3.4 Feature Extraction

Texture-based feature extraction is important role in segmenting the identification of uniform regions within a given image. Gabor filtering method is to extract the edge, and contours of the image anatomical structure changes in object location, scale, and orientation can be detected in the Gabor feature space. Frequency and

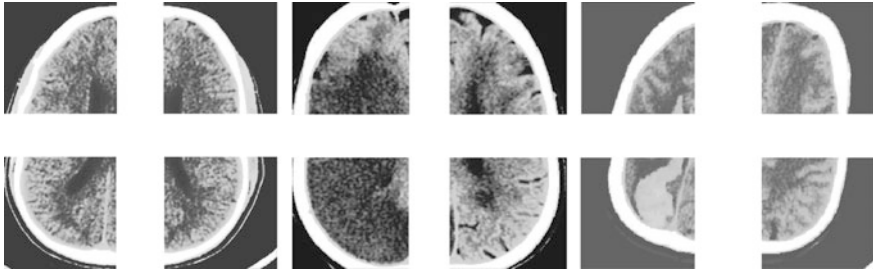


Fig. 3 Normal brain image, stroke image, and hemorrhage image divided into four regions

orientation representations of Gabor filters are similar to those of the human visual system. The Gabor function is a sinusoidal modulated Gaussian in the spatial frequency domain. The complex Gabor function in space domain is given

$$g(x, y) = S(x, y) * w_r(x, y) \tag{1}$$

where $S(x, y)$ is a complex sinusoid, known as the carrier, and $w_r(x, y)$ is a 2-D Gaussian-shaped function, known as the envelope (Fig. 4).

The complex sinusoid is defined as follows

$$S(x, y) = \exp(j(2\Pi(u_0x + v_0y) + P)) \tag{2}$$

where (u_0, v_0) and P define the spatial frequency and the phase of the sinusoid, respectively.

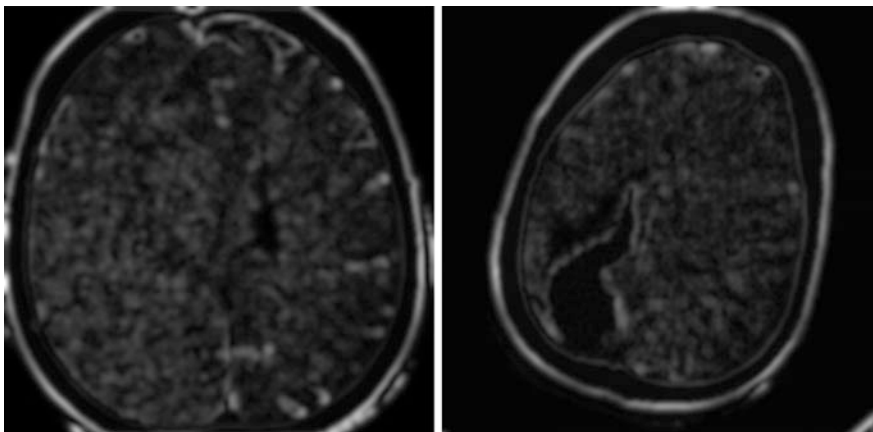


Fig. 4 Gabor filter output of stroke image, and hemorrhage image

The envelope is a Gaussian function:

$$W = K * \exp(-\Pi(a^2(x - x_0)r^2 + b^2(y - y_0)r^2)) \quad (3)$$

where (x_0, y_0) is the peak of the function, a and b are scaling parameters of the Gaussian, and r subscript stands for a rotation operation such that

$$(x - x_0)r = (x - x_0)\cos \Theta + (y - y_0)\sin \Theta \quad (4)$$

$$(y - y_0)r = (x - x_0)\sin \Theta + (y - y_0)\cos \Theta \quad (5)$$

Here the feature vector we are using mean and standard deviation and also the filter size used is 7×7 of ones. Spatial frequencies and their orientations are important characteristics of textures in images. We use the orientation θ at 0, 45, 90, and 135, and the phase shifted to 0–180 because of symmetry. The complex Gabor function has five different parameters: K the magnitude of the Gaussian envelope, (a, b) the two axes of the Gaussian envelope, θ rotation angle of the Gaussian envelope, (x_0, y_0) location of the peak of the Gaussian envelope, and (u_0, v_0) spatial frequencies of the sinusoid carrier in Cartesian coordinates.

3.5 Segmentation

Region-based segmentation is one of the fast scanning algorithms, which includes the seeded and unseeded region growing. Region growing is a procedure that groups pixels or subregions into larger regions based on the criteria for growth. The basic idea behind is that to start with a set of “seed” points and grow regions by adding to each seed those neighboring pixels that have predefined properties similar to the seed. Region growing is a simple region-based image segmentation method.

The first step in region growing is to select a set of seed points. Seed point selection is based on some user criterion (for example, pixels in a certain gray-level range, pixels evenly spaced on a grid, etc.). The initial region begins as the exact location of these seeds. The regions are then grown from these seed points to adjacent points depending on a region criterion. The criterion could be pixel intensity, gray-level texture, or color.

There is a very simple example followed below. Here, we use 4 connected neighborhoods to grow from the seed points. We can also choose 8 connected neighborhoods for our pixels’ adjacent relationship. And the criteria we make here is the pixel have same mean and variance value. That is, we keep examining the adjacent pixels of seed points. If they have the same intensity value with the seed points, we classify them into the seed points. Region is growing based on mean value and variance.

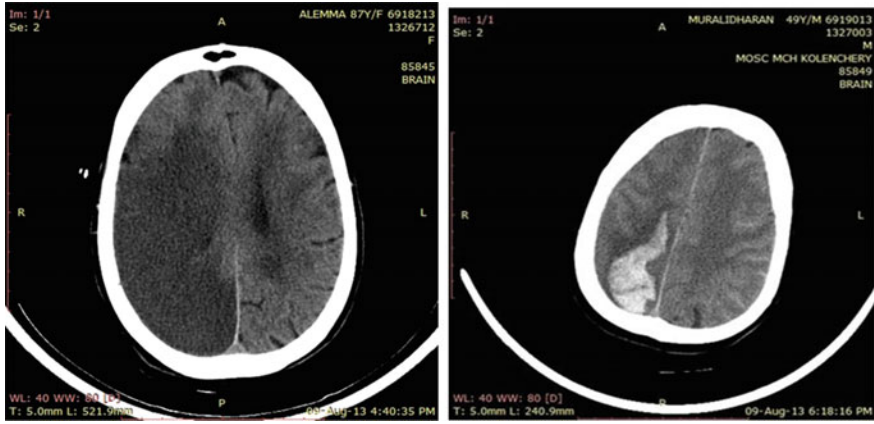


Fig. 5 Input stroke image and input hemorrhage image

4 Result and Analysis

In preprocessing, denoising is performed with three different methods: Wiener filter, median filter, and discrete wavelet decomposition. The analysis is done with the PSNR value. The PSNR for Wiener filter, median filter, and discrete wavelet decomposition are 59.4391, 59.3675, and 52.3255. Then, we conclude that best result occurs in Wiener filter. So the proposed system we use is the Wiener filter for denoising.

To find the presence of stroke using statistical analysis method of four region. From the result, we conclude that for normal brain image the mean value ranges from 0.5100 to 0.5900. In the case of Ischemic stroke image, in the stroke region, mean and standard deviation are minimum, and in the case of Hemorrhage stroke region, mean and standard deviation are maximum (Fig. 5).

5 Conclusion

We presented a method for detecting stroke in CT images. In this proposed work, for better visualization use histogram equalization. Even though the stroke part is enhanced in the first step, the segmentation is done to clearly separate the stroke and non-stroke area. Statistical analysis helps to find presence of stroke. We conclude that mean and standard deviation are minimum for ischemic stroke region and maximum in the case of hemorrhage. The texture-based segmentation gives better result than intensity-based segmentation. Here, we are using texture-based segmentation is done using the Gabor filter. By using the given data set, Gabor filter gives better result in hemorrhage images compared to ischemic stroke images. In ischemic stroke images, the edges are not clearly separated. The input image is

divided into four regions and corresponding histograms are plotted for each region. Based on the histogram matching technique, find whether the given input image is affected by stoke or not. The proposed method gives significantly better visual quality and also reduces the cost of work. In future, we plan to do the classification of different types of strokes.

References

1. A. Padma, R. Dr Sukanesh, A wavelet based automatic segmentation of brain tumor in ct images using optimal statistical texture features. *Int J Image Process (IJIP)*. **5**(5), 552 (2011)
2. T.-L. Tan, K.-S. Sim, A.-K. Chong, Contrast enhancement of ct brain images for detection of ischemic stroke. *International Conference on Biomedical Engineering (ICoBE)*, (2012) pp. 27–28
3. Y.-H. Li, L. Zhang, Q.-M. Hu, H.-W. Li, F.-C. Jia, J.-H. Wu, Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans in Springer. (2012) **7**, 507–516
4. R.R. Thakur, S.R. Dixit, A.Y. Deshmukh, VHDL design for image segmentation using gabor filter for disease detection. *Int. J. VLSI Design Commun. Syst. (VLSICS)*. **3**(2) (2012)
5. M. Chawla, S. Sharma, J. Sivaswamy, L.T. Kishore, A method for automatic detection and classification of stroke from brain CT images Engineering in medicine and biology society. *Annual International Conference of The IEEE*. (2009), pp. 3581–3584
6. A. Padma, R. Sukanesh, Automatic classification and segmentation of brain tumor in CT images using optimal dominant gray level run length texture features . *Int. J. Adv. Comput. Sci. Appl.* **2**(10) (2011)
7. M.M. Kyaw, Pre-segmentation for the computer aided diagnosis system. *Int. J. Comput. Sci. Inf. Technol.* **5**(1) (2013)
8. R.S. Jeena, S. Kumar, A Comparative analysis of MRI and CT brain images for stroke diagnosis. *Int. Conf. Microelectron. Commun. Renew. Energy*. (2013), **6**,1–5
9. G.H. Rai, T.R. Nair, Gradient based seeded region grow method for CT angiographic image segmentation. *Int. JRI Comput. Sci. Netw.* **1**(1) (2009)
10. I. Wu, S. Poehlman, M.D. Noseworthy, M.V. Kamath, Texture feature based automated seeded region growing in abdominal MRI segmentation. *J. Biomed. Sci. Eng.* **2**, 1–8 (2009)
11. B. Sharma, K. Venugopalan, Automatic segmentation of brain CT scan image to identify hemorrhages. *Int. J. Comput. Appl.* **40**(10),0975–8887 (2012)
12. S. Ibrahim, N.E.A. Khalid, M. Manaf, Computer aided system for brain abnormalities segmentation. *Fac Comput Math. Sci.* (2010)

Preemptive Appliances Scheduling in Smart Home Using Genetic Algorithm

A. Ranjini and B.S.E. Zoraida

Abstract Smart grid is the newer version of electricity network that incorporates the information and communication technology in addition with the renewable energy resources. Smart grid offers electricity for the users at different rates at different time. Smart grid employs smart home where there is a set of electrical appliances that need to be scheduled in such a way to reduce their electricity charges. Many bio-inspired algorithms are of greater interest for the recent years and are being able to solve several complex problems in the real-world situations; the smart home appliances scheduling problem is solved using the genetic algorithm. Two new operators namely appliance-based one-point crossover operator and appliance-based rotation mutation operator are used for solving the problem.

Keywords Smart grid · Genetic algorithm · Scheduling

1 Introduction

The main goal of the smart grid is to provide reliable and secured electricity network and guarantee to match the power demand and supply. The smart grid is empowered with various features such as advanced metering infrastructure, distributed storage management, and use of renewable energy resources for power generation. Among the various subsystems of smart grid, the energy subsystem is of major concern. The demand side management system tries to reduce the power demand as an alternative to increase the power generation [1]. It induces the customers either to shift or reduce their power demand by their response to the price information provided. A smart home can be defined as a home that is equipped with

A. Ranjini (✉) · B.S.E. Zoraida
Bharathidasan University, Tiruchirappalli, India
e-mail: ranjini.anbu@gmail.com

B.S.E. Zoraida
e-mail: b.s.e.zoraida@gmail.com

smart devices. A home network makes it possible to transfer information between devices, while the residential gateway connects a smart home to the Ethernet or Internet for downloading new services [2]. The energy management controller is built into the home gateway which provides signals to control the smart appliances in a smart home. The energy management system allows the two-way flow of information such as the price information and power consumption profile. Smart grid offers different types of pricing model such as time of use pricing, critical peak pricing, real-time pricing, etc. To avoid the high peak-to-average (PAR) of power, the real-time pricing model and inclining block rate model can be combined [3]. In this paper, this combined price model is utilized. In real-time pricing, the price of electricity changes often for a particular period of time usually, hourly basis. In the inclining block rate model, the price changes when the total load reaches a particular threshold value. Due to the varying price of electricity, the consumers in the smart grid can reduce their electricity charges by scheduling the appliances in a smart home in an optimal manner. Energy scheduling algorithm is proposed such that the monetary expense of a customer is minimized while considering the uncertainties in appliance operation time and intermittent renewable generation [4]. It is also possible to store electricity from the external grid in a battery by the residential customers. In order to minimize the cost of the energy drawn from the external grid, a scheduling algorithm is proposed while considering the usage of appliances which is subjected to individual constraints [5]. Game theory is used to solve the problem with the objective of minimizing the cost of energy and PAR ratio. Game theory is used to formulate an energy consumption scheduling game, where the players are the users and their strategies are the daily schedules of their household appliances and loads [6]. They try to minimize the cost of energy and also to balance the total load. The electricity provider's cost minimization problem is solved while considering the consumers' device-specific scheduling flexibility [7]. Game theory is now becoming an important tool for solving the various issues in smart grid [8].

2 Problem Statement

A smart home can consists of a set of electrical appliances connected through a network, which may be either non-preemptive or preemptive operation-type appliances. In this paper, the set of preemptive appliances are considered for scheduling where the price information is given. The objective of the problem is to schedule the preemptive appliances for a specified amount of time in order to minimize the electricity charge for a consumer.

In this paper, H denotes the specified amount of time for which the appliances are needed to be scheduled, which is usually some hours. This is because the inclining block pricing model is used in this paper which changes the price of electricity every one hour. The price of electricity at any hour h can be given by the following price function as:

$$pr_h(tp_h) = \begin{cases} a_h & \text{if } 0 \leq tp_h \leq c \\ b_h & \text{if } tp_h > c \end{cases} \tag{1}$$

where $1 \leq h \leq H$.

Here, tp_h denotes the total power consumption of all appliances in the home during the h th hour, a_h and b_h denote the price of electricity at h th hour, and c denote the threshold value above which if the total power consumption exceeds the price of electricity is a_h and it is b_h if not exceeds the value. Before to schedule the appliances, it is necessary to divide the 1 h into time units called *slots*, for which the appliances need to be scheduled. The time slot can have any value (within the interval of 1–60) which is suitable for customer needs and appliances operation. Usually, this value is the minimum amount of time need by an appliance for its operation and it must be a factor of 60. This is because the price of electricity changes for every one hour as stated earlier. For example, if the shortest operation of any appliance is 10 min, then time unit is set to 10 and the number of slots in 1 h is 6. So, the appliance with shortest operation needs one time slot, whereas other appliances need different number of slots according to its operation time. So, the total number of slots for each appliance must be either equal to greater than the actual operation time needed. It is assumed that the power consumption value for each time slot for each appliance is fixed.

Let n be the number of appliances to be scheduled, t is the total number of time slots in H , the matrix S denotes the schedule where each element of S is s_{ij} denote the either the on or off state of the appliance i on the j th time unit. Hence, the value for s_{ij} can be either 0 or 1 which implies the on or off state of each appliance, respectively. The values of each row stand for the schedule of each appliance, whereas the values of each column stand for the schedule for each slot. P denotes the matrix where each element p_i represent amount of power needed for each appliance i . ST denotes the matrix where each element st_i represent number of slots needed for each appliance i , and it must be equal to the number of ones in each column of matrix S . The matrix SP is given by the product of the two matrices namely S and P . Each element in matrix SP represents the total consumption in h .

$$tp_h = \sum_{i=1}^m (SP)_i \text{ where } l = (((h - 1) * 60) / t) + 1, m = ((h * 60) / t) \text{ and } 1 \leq h \leq H \tag{2}$$

Now, the power consumption scheduling problem can be stated as follows:

$$\text{Minimize } \sum_{h=1}^H (pr_h(tp_h) \times tp_h) \tag{3}$$

s.t (1) and (2).

3 Methodology

Genetic algorithms are bio-inspired search algorithms based on the evolutionary ideas of natural selection and genetics that are used successfully to solve problems in many different disciplines. Genetic algorithm uses the technique that resembles natural selection in the biological process. Genetic algorithm was developed by John Holland, his colleagues, and his students at the University of Michigan in the 1960s and the 1970s. Due to the robustness of genetic algorithms on problems of high complexity, it has an increasing number of applications in the fields of artificial intelligence, numeric and combinatorial optimization, business, management, medicine, computer science, engineering, etc.

3.1 Binary Encoded Chromosome Representation

The term chromosome in genetic algorithm represents the solution in the search space where it is a schedule for this problem. Different types of chromosome encoding are available and now the binary encoding is adopted in this paper. Each gene in the chromosome represents the on or off state of an appliance at a particular time slot. Example for the chromosome with 6 time slots and 5 appliances is shown in Fig. 1.

3.2 Initial Population

The initial population in the genetic algorithm affects the speed of convergence and so the better initial solutions might provide better results. The diversity among the individuals is also important for avoiding premature convergence. In this paper, it is created by using random permutation of appliances' schedule.

3.3 Fitness Calculation

Since it is a minimization problem, the fitness function for the scheduling problem can be defined as $\text{fitness} = (1/\text{objective function})$.

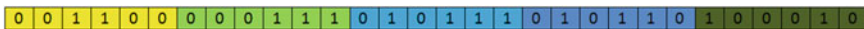


Fig. 1 Example for the chromosome with 6 time slots and 5 appliances

3.4 Selection

Different selection strategies have different methods of calculating the selection probability of an individual. In this paper, Roulette wheel selection is applied which is a fitness appropriate selection method.

3.5 Crossover

The crossover operator takes more than chromosomes and combines them in such a way to produce new chromosomes to inherit some of the information contained their parents. In this paper, the new crossover operator namely appliance-based one-point crossover is utilized, and its operation is in the following Fig. 2. This crossover alters the schedule of the appliances only, not the number of slots required for each appliance.

3.6 Mutation

The appliance-based rotation mutation rotates the genes within each appliance and so avoids the production of infeasible solutions. The example for this mutation operation is shown in Fig. 3.

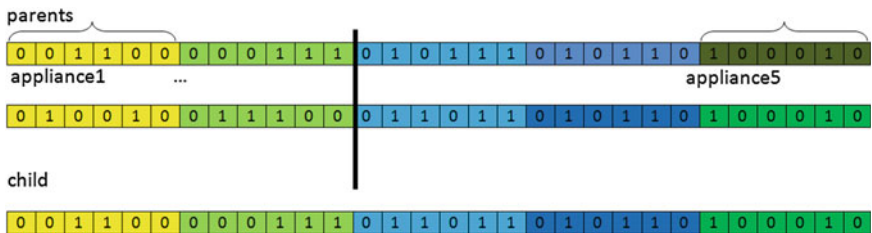


Fig. 2 Appliance-based one-point crossover

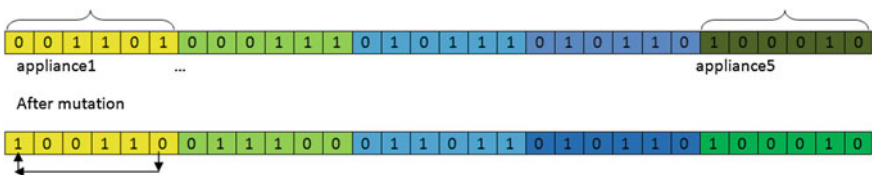


Fig. 3 Appliance-based mutation

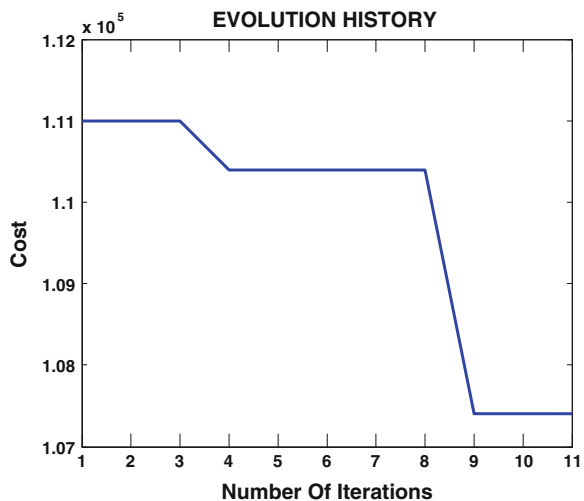
3.7 Case Study

The implementation of genetic algorithm for solving the scheduling problem in a smart home is analyzed using a sample consumer power demand with the given price information. In this example, the 5 appliances are namely space heater, clothes dryer, dishwasher, electrical car, and oven which 1, 1, 2, 3, and 0.5 (h), respectively. The following list of parameters was used to evaluate the performance of the proposed system.

- Total number of appliances = 5
- Total number of time slots (H) = 6
- Time for one slot (h) = 1 h
- P (in Watt) = [1,000 5,000 1,200 3,500 5,000]
- S = [2 1 2 1 1]
- ST (price in cents) = [4 4 4 6 6 8; 6 6 6 9 9 12]
- C = 3,500
- Population size = 20

The optimum schedule for this scenario produces the schedule with the price 107,400 cents. In order to compare the result of genetic algorithm with the optimum schedule, all the feasible schedules are evaluated and the optimum schedule is caught. The total number of feasible schedules for this example problem is 1,350,000. But, using genetic algorithm, only part of the search space is traversed and the optimum schedule is caught. In this example, the genetic algorithm searches through only 180 schedules and caught the optimum schedule which is shown in the Fig. 4.

Fig. 4 Evolution history of the cost of the schedules generated using genetic algorithm



4 Conclusion

In this paper, a smart home having five preemptive appliances has taken into consideration for scheduling with the objective of minimizing the electricity expense of a single consumer and the genetic algorithm was successfully applied for solving this scheduling problem. In order to evaluate the performance of the genetic algorithm, all the possible schedules for this scheduling problem were generated and the optimum schedule is caught. However, within a few iterations, the optimum schedule was generated by the genetic algorithm. Through the use of problem-specific crossover and mutation operator, the algorithm produces only feasible solutions which remarkably improve speed of the algorithm. Since the genetic algorithm is suitable for combinatorial optimization problems and the execution time is usually few minutes, it is successfully applied for the scheduling in smart homes. The work can be extended to schedule both preemptive and non-preemptive appliances in addition with the use of storage devices.

References

1. X. Fang, S. Misra, G. Xue, D. Yang, Smart grid—the new and improved power grid: a survey. *IEEE Commun. Surv. tutorials*. **14**(13) (2012)
2. Y.-C. Huang, C.-M. Huang, K.-Y. Huang, C.-Y. Liu, Energy optimization approaches for smart home applications. *Internal Conference on International Conference on Artificial Intelligence and Soft computing*, Lecture Notes in Information Technology. vol. 12 (2012)
3. Z. Zhao, W.C. Lee, Y. Shin, K.-B. Song, An optimal power scheduling method applied in home energy management system based on demand response. *J. Electron. Telecommun. Res. Inst. (ETRI)* **35**, 677–686 (2013)
4. X. Chen, T. Wei, S. Hu, Uncertainty-aware household appliance scheduling considering dynamic electricity pricing in smart home. *IEEE Trans Smart Grid*. (2013)
5. S. Chen, P. Sinha, N.B. Shroff, Heterogeneous delay tolerant task scheduling and energy management in the smart grid with renewable energy. *IEEE J. Sel. Areas Commu.* **31** (2013)
6. A.-H. Mohsenian-Rad, V.W.S. Wong, J. Jatskevich, R. Schober, A. Leon-Garcia, Autonomous demand side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Trans Smart Grid*. **1** (2010)
7. C. Joe, S. Sen, S. Ha, M. Chiang, Optimized day-ahead pricing for smart grids with device-specific scheduling flexibility. *IEEE J Sel Areas Commu.* **30** (2012)
8. W. Saad, Z. Han, H.V. Poor, T. Basar, Game—theoretic methods for the smart grid. *IEEE Signal Process. Mag.* **29**(5) (2012)

Credible Secure Data Aggregation in Wireless Sensor Networks

M.P. Anuradha and Gopinath Ganapathy

Abstract Wireless sensor networks contain large number of low-cost sensor devices called nodes that have severely limited in sensing, computation, and communication abilities. In order to improve the sensor network lifetime, the amount of data transmission minimization is a vital issue. Data aggregation is the technique which is used to minimize the rate of data transmission among the sensor networks. Data aggregation is the process of merging sensor data in order to moderate the quantity of data transmission in the network. The results of data aggregation are typically used to make serious decisions; hence, the correctness of final aggregation results is very important. In this paper, credible-based secure data aggregation (CDA) is proposed to ensure the quality of data aggregation. The main idea of this scheme is that secured, accurate data aggregation is achieved based on the credible value of the sensor nodes. The simulation results show that protocol CDA considerably expands the system reliability through secured data aggregation in the occurrence of compromised nodes.

Keywords Credibility · Data aggregation · Wireless sensor networks

1 Introduction

Wireless sensor networks (WSN) contain large sensor nodes [1] are used to measure environmental conditions such as temperature, sound, pressure, etc., and to supportively transmit data through the intermediate sensor nodes in the network to a sink. The WSN is built of hundreds or thousands “nodes,” where each node is connected with different sensors. Sensor node coverage, fault tolerance, reliability,

M.P. Anuradha (✉) · G. Ganapathy
School of Computer Science, Engineering and Applications, Bharathidasan University,
Tiruchirapalli 620023, TamilNadu, India
e-mail: anushivam@yahoo.com

G. Ganapathy
e-mail: gganapathy@gmail.com

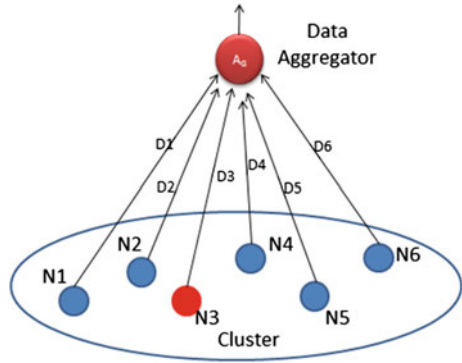
node to node communication, group communication, communication speed, computation rate, and security are the important properties in WSN [2]. The typical aspect of sensor node is that adjacent sensor nodes frequently have overlapping sensing ranges, and therefore, the data made by adjacent sensor nodes are redundant. Furthermore, the amount of data processed by the base station is typically huge. To manage the base station and to preserve the energy and bandwidth of resource controlled WSN, redundancy in sensor data is minimized and eliminated at in-between nodes by executing data aggregation. Data aggregation is the process of uniting data from multiple sensor nodes from the predefined location to eliminate redundant data before the transmission and provide significant information to the base station [1–5]. The sensor nodes are used to perform some critical tasks [4] also deployed in unattended environments; hence, the security of WSN must be sensibly considered. Due to the lack of security in WSN, the impostors can easily add the false node to the network, to subvert the network operation. The identification of compromised node and minimizing the effects of compromised nodes in the network are the challenging tasks of the protocol designer.

This paper proposes a credible secured data aggregation called CDA that considers the credibility value of each sensor nodes during data aggregation [6] and transmission to increase the reliability of aggregated data. In every cluster, there is non-supporting node (compromised node) which is recognized by a credibility-based data aggregation process. This calculation typically identifies the non-supporting node which deviates high in value in comparison with the other node values. In an instant of time, the credibility value is calculated as the product of aging factor, old credibility value summed with products of a number of supporting nodes with increment assurance factor and product of a number of non-supporting nodes with decrement assurance factor. The rest of the paper is organized as follows. In Sect. 2, the state of the art in secure data aggregation is presented and also explains the system model of CDA protocol. Performance evaluation is presented in Sect. 3, and concluding remarks are made in Sect. 4.

2 Credible Data Aggregation (CDA) in WSN

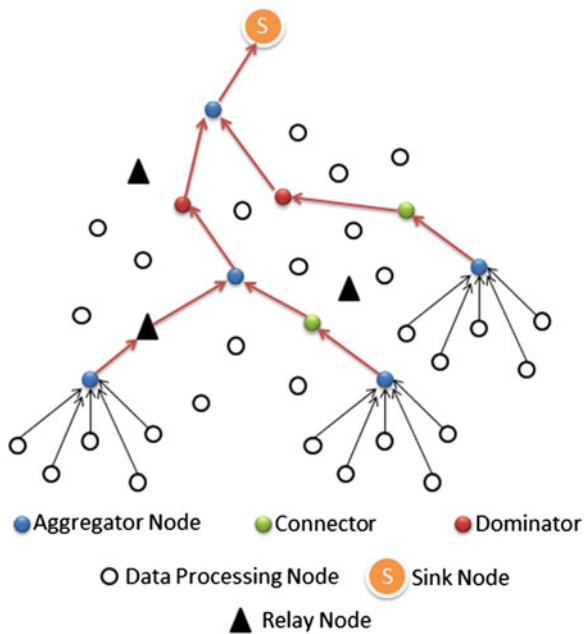
Due to the condensed deployment, sensor nodes have overlapping sensing regions, and same trials are perceived by many sensor nodes. Hence, data aggregation [7] is preferred to decrease the number of data transmission. Some sensor nodes are energetically nominated as data aggregators to aggregate the data from their adjacent nodes. The main intention of WSN is to distribute the information about the external atmosphere and to distribute information properly with smaller amount energy consumption. Present data aggregation techniques of sensor network adopt that the nodes are preprogrammed and the sensed data are communicated to a central sink for interrogating and investigation. It faces two main disadvantages. First, the system behaviors cannot be altered during runtime. Then, the energy consumption due to high transmission will decrease the complete network lifetime.

Fig. 1 Data aggregation



The general assembly of the data aggregation [8] is described in Fig. 1. A sample data aggregation process is described in Figs. 2 and 3, where a collection of sensor nodes are used to gather the data from a predefined section. When the base station wants to know some information from the particular set of sensor nodes, it broadcasts the queries to the network, instead of sending the data from each sensor node to base station, and any one sensor nodes will perform the predefined data aggregation functions [9], named data aggregator node [7], gathers the data from its adjacent nodes, combined them, and communicate the combined data to the base station. For example, $N1$'s aggregated data are equal to the aggregate of $(D1||D2||D3||D4)$ and A_j 's aggregated data are equal to the aggregate of $(D1||D2||D3||D4||D5||D6)$. Due

Fig. 2 Cluster-based data aggregation



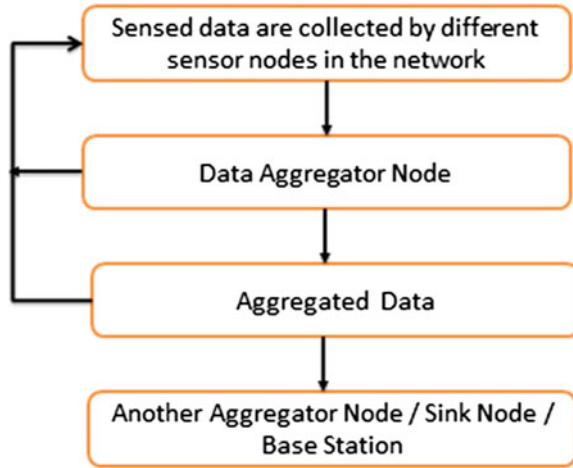


Fig. 3 Data aggregation process

to the reduction of the number of transmission among the nodes, the energy utilization of the sensor nodes and network bandwidth are improved. The effectiveness of data aggregation is described in Fig. 4.

Some cases compromised sensor nodes [10], and an unauthorized sensor node communicates false-sensing data to the aggregator node and it deludes the consistency of aggregated data in WSN [11]. Typically, it is very difficult for data aggregators to verify the accuracy of the received sensor data. However, due to the solid placement of sensor nodes in the networks, near-end sensor nodes in the network frequently having common sensing ranges and data detected by adjacent sensor nodes are interrelated [1]. So identification of false data at the aggregate is a

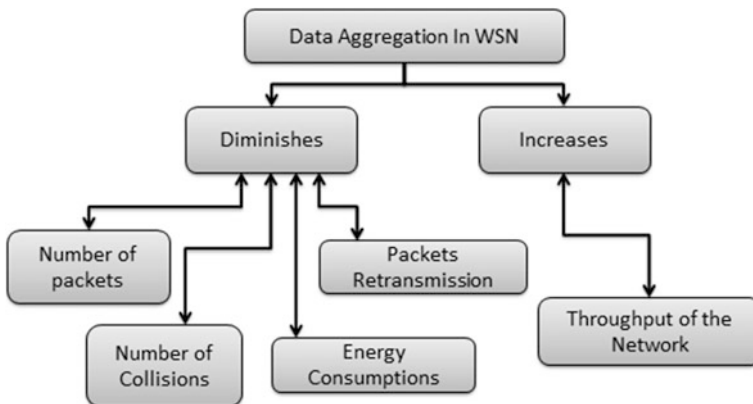
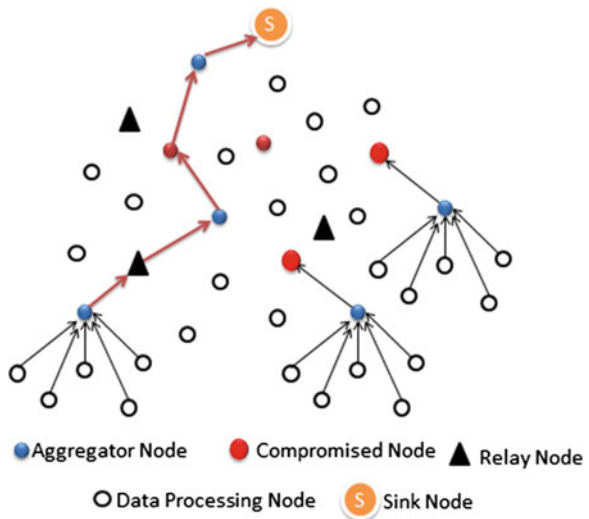


Fig. 4 Effectiveness of data aggregation

Fig. 5 Impact of compromised nodes in WSN



key issue in WSN. Due to the sensing misbehaviors, the integrity of the data aggregation process [9, 12, 13] will get affected.

Even if a compromised node is nominated as aggregator node, it can insert wrong data into the network. It will affect the integrity of the system and point in the wrong direction of aggregated data at the base station. Hence, the main intention of the proposed work is to diminish the effects of false data and related occurrences. Due to compromised node in the network, the aggregated data are not able to reach the sink node appropriately is described in Fig. 5. Hence, in data aggregation, the identification of reliable node [9] is an essential one.

The existing aggregation technique [8, 13–18] has many key aspects, but does not consider the credibility value. In every cluster, there is non-supporting node, which is recognized by a credibility-based data aggregation process. Such a calculation typically identifies the non-supporting node which deviates high in value in comparison with the other node values. In an instant of time, the credibility value is calculated as the product of aging factor, old credibility value summed with product of a number of supporting nodes with increment assurance factor and product of a number of non-supporting nodes with decrement assurance factor. After the first instant of time, the precomputed credibility value is taken as new credibility value for the next time instant. This process is described in Figs. 6 and 7, for example, N1, N2, N3, and N4 send the sensed data to the data aggregator. The data of N1, N2, and N4 are relatively same, but the data of N3 are different from other data. Hence, the data aggregator will conclude that N3 is compromised node. In same case, another data aggregator concludes that the N7 is compromised node.

Fig. 6 Identification of compromised node

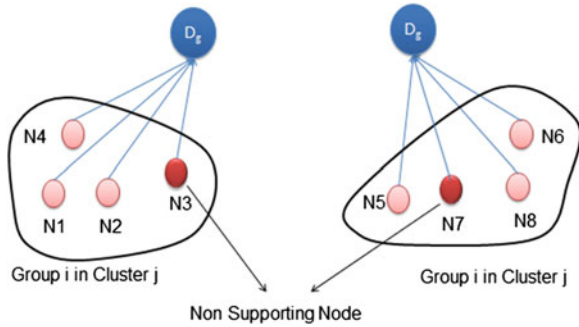
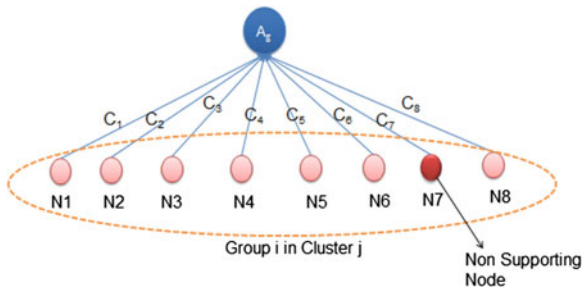


Fig. 7 Credibility calculations



Algorithm: Credible Data Aggregation

INPUTS: Data Aggregator DA, Readings D_i from nodes $N_i \in$ Cluster N

OUTPUT: Aggregated data D

Step 1: Aggregator DA requests nodes $N_i \in$ Cluster N to send their sensor readings

Step 2: Nodes transmit their data D_i to the aggregator DA

Step 3: DA compares the sensor reading of each node N_i with every other node $N_j \in N$

(Where $i \neq j$) to determine the number of supporting and non-Supporting nodes.

Step 4: Credibility of each node is calculated as

$C_i = (\text{Old credibility} \times \text{Aging factor}) + (\text{Number of supporting nodes} \times \text{Factor of Increment}) + (\text{Number of non-supporting nodes} \times \text{Factor of decrement})$

Step 5: If for each node, the number of non-supporting nodes exceeds half the total number of nodes in the cluster, then the node N_i is flagged as False.

Step 6: The aggregated Data D is calculated as

$$D = \frac{\sum (\text{Flag} * D_i * C_i)}{\sum (\text{Flag} * C_i)} \tag{1}$$

For secure transmission [18] and data aggregation, the less overhead public key excess N crypto system [18] is proposed in this work. In order to reduce the high

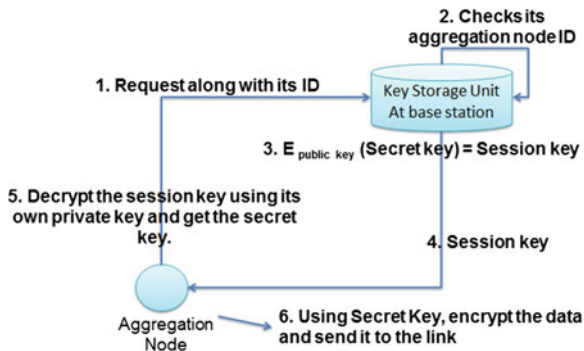


Fig. 8 Public key excess N crypto system

computation in implementation of security in WSN, the public key excess N crypto system is introduced in CDA. Using this technique, the secret key is attained by a data aggregator in a secure way from the key storage unit at the base station.

The steps involved in public key excess N crypto system are as follows:

1. Data aggregator sends its request to the base station. It will send its aggregator node ID.
2. The base station verifies its aggregator ID for authentication purpose, and once it confirmed means it encrypts the secret key using its own public key and generates the session key.

$$E_{\text{public key}}(\text{Secret Key}) = \text{Session key} \tag{2}$$

3. The session key is communicated from the base station to the aggregator node.
4. The data aggregators decrypt the session key using its own private key and obtain the secret key.
5. Using a secret key, it will encrypt the aggregated data. The encrypted aggregated data through the dominators will reach the sink node (Fig. 8).

3 Performance Evaluation

Protocol CDA is simulated using NS2 simulator [12]. Cluster-based sensor network is considered to monitor the temperature of a terrain. The different numbers of sensor nodes are randomly deployed in the predefined ranges. This analysis evaluates the performance of the proposed CDA approach [8, 9] with different numbers of sensor nodes. In this analysis, the system accuracy, system efficiency, and

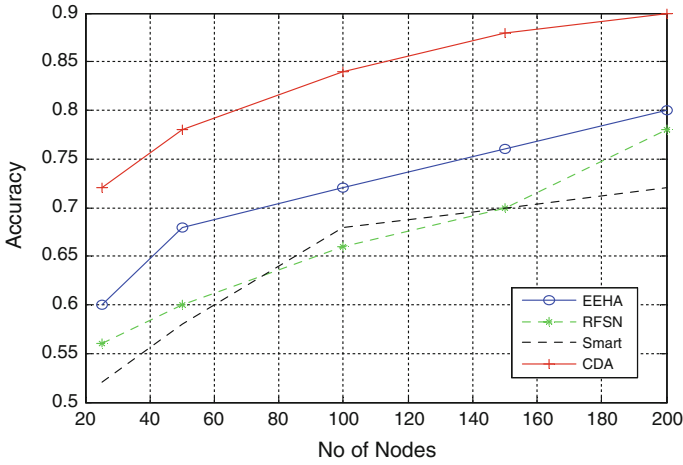


Fig. 9 Accuracy versus number of nodes

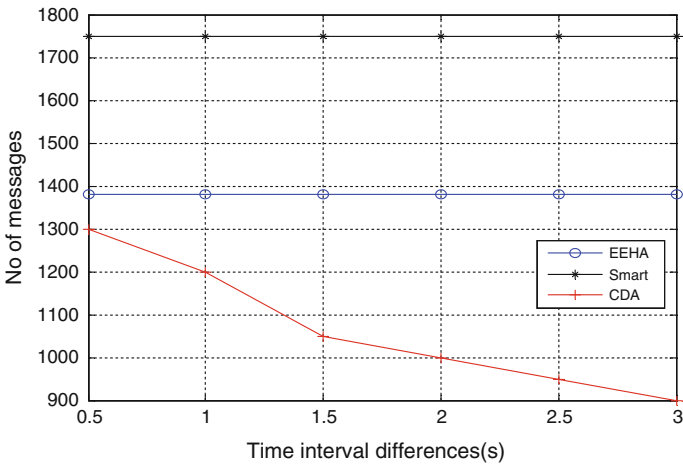


Fig. 10 No. of messages versus time

number of messages involved for the data transmission factors are measured. EEHA, RFSN, and SMART [8, 9] methods are implemented with CDA approach. The simulated results for Credible Data Aggregation approach compared with an existing algorithm are described in Figs. 9, 10 and 11 respectively. From the simulated results, the communication overhead of CDA is not substantial, thereby making the implementation of CDA feasible. The real outcome of the proposed work is tabulated in Table 1.

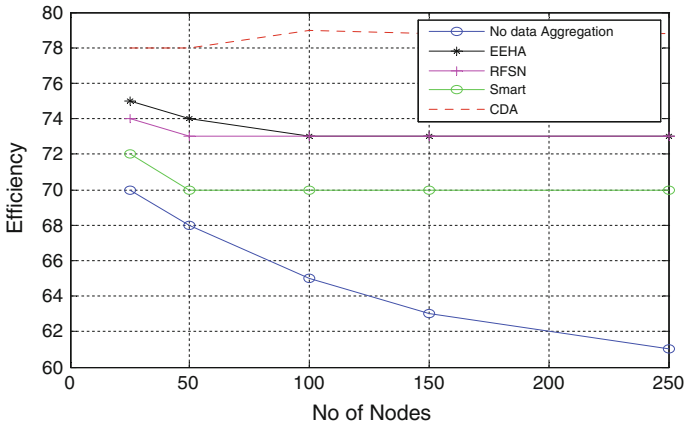


Fig. 11 Efficiency versus number of nodes

Table 1 Outcome of the proposed work

Criteria	Existing aggregation	Proposed aggregation technique (CDA)
Identification of false data	Not possible	Possible
Data aggregation	At cluster head only	At multiple level
No. of aggregation point	Minimum	Maximum
Total number of message	Maximum	Reasonable
Packet loss	Maximum	Minimum
Aggregation overhead	Normal	Reasonable

4 Conclusion

In wireless sensor networks, compromised sensor nodes can mislead the reliability of data by directing false data reports, inserting false data in data aggregation, and disturbing the data transmission of aggregated data. Subsequently, cryptographic determinations are not adequate to avoid these problems. This paper has presented novel reliable data aggregation techniques called CDA; it enhances the reliability of the data aggregation process with the help of credibility calculations. CDA improves the reliability of aggregated data by assessing sensor nodes and data aggregators via credibility values of each sensor nodes. The performance evaluation indicates that CDA ensures the reliability of data aggregation and transmission in the presence of compromised nodes. In addition, the simulation results show the efficiency, communication overhead, and system accuracy of CDA.

References

1. I.F. Akyildiz, I.H. Kasimoglu, Wireless sensor and actor networks: research challenges. *Ad Hoc Netw.* (2004) pp. 351–367
2. G. Pottie, W. Kaiser, Wireless integrated network sensors. *Commun. ACM* **43**(5), 51–58 (2000)
3. V. Mhatre et al., Design guidelines for wireless sensor networks: communication, clustering and aggregation. *Ad Hoc Netw. J* **2**(1), 45–63 (2004). Elsevier Science
4. E. Stavrou, A. Pitsillides, A survey on secure multipath routing protocols in WSN's. *Comput. Network.* **54**, 2215–2238 (2010)
5. O. Cheikhrouhouet al., LNT: a Logical neighbor tree for secure group management in wireless sensor networks. *Procedia Comput. Sci.* (2011) pp. 1877–0509
6. C. Intanagon Wiwat, D. Estrin, R. Govindan, J. Heidemann, Impact of network density on data aggregation in wireless sensor networks, in *Proceedings International Conference on Distributed Computing Systems*. (2002) pp. 457–458
7. M. Ding, X. Cheng, G. Xue, Aggregation tree construction in sensor networks. *IEEE Xplore* (2003) 7803-7954-3/03
8. Hongjuan Li, Kai Lin, K. Li, Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks. *Comput. Commun.* **34**, 591–597 (2011)
9. S. Ozdemir, Functional reputation based reliable data aggregation and transmission for wireless sensor networks. *Comput. Commun.* **31**, 3941–3953 (2008)
10. K. Lu, L. Huang, Y. Wan, H. Xu, Energy-efficient data gathering in large wireless sensor networks, in *Second International Conference on Embedded Software and Systems*. (2005) pp. 5–10
11. S. Ozdemir, Y. Xiao, Secure data aggregation in wireless sensor networks: A comprehensive overview. *Comput. Netw.* **53**, 2022–2037 (2009)
12. R. Roman, C. Alcaraz, J. Lopez, A survey of cryptographic primitives and implementations for hardware-constrained sensor network nodes, *Mob. Netw.* (2007) pp. 231–244
13. H. Sethi, R.B. Patel, EIRDA: an energy efficient interest based reliable data aggregation protocol for wireless sensor networks. *Int. J. Comput. Appl.* **22**(7), 0975–8887 (2011)
14. E. Fasolo, M. Rossi, J. Widmer, M. Zorzi, In-network aggregation techniques for wireless sensor networks: a survey. *IEEE Wirel. Comm.* **14**(2), 70–87 (2007)
15. L.A. Villas, A. Boukerche, R.B. Araujo, A.A. Loureiro, A Reliable and Data Aggregation Aware Routing Protocol for Wireless Sensor Networks, in *Proceedings 12th ACM Int'l Conference Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*. pp. 245–252
16. P. SudipMisra, D. Thomasinos, A simple, least-time, and energy-efficient routing protocol with one-level data Aggregation for wireless sensor networks. *J. Syst. Softw.* **83**, 852–860 (2010)
17. L. Villas, A. Boukerche, R.B. de Araujo, A.A.F. Loureiro, Highly dynamic routing protocol for data aggregation in sensor networks, in *Proceedings IEEE Symposium Computers and Communication (ISCC)*. (2010) pp. 496–502
18. L.A. Villas, A. Boukerche, H.S. Ramos, A.B.F. de Oliveira, R.B. de Araujo, A.A.F. Loureiro. DRINA: A lightweight and reliable routing approach for in-network aggregation in wireless sensor networks. *IEEE Trans Comput.* **62**(4) (2013)

Zumkeller Labeling Algorithms for Complete Bipartite Graphs and Wheel Graphs

B.J. Balamurugan, K. Thirusangu and D.G. Thomas

Abstract Let $G = (V, E)$ be a graph. An injective function $f: V \rightarrow N$ is said to be a Zumkeller labeling of the graph G , if the induced function $f^*: E \rightarrow N$ defined as $f^*(xy) = f(x)f(y)$ is a Zumkeller number for all $xy \in E, x, y \in V$. A graph $G = (V, E)$ that admits a Zumkeller labeling is called a Zumkeller graph. In this paper, we provide polynomial time algorithms for Zumkeller labeling of complete bipartite graphs and wheel graphs.

Keywords Polynomial time algorithms · Zumkeller labeling · Zumkeller numbers · Complete bipartite graph · Wheel graphs

1 Introduction

A positive integer n is called a perfect number [1] if n equals the sum of its proper positive factors. Generalizing this concept in 2003, Zumkeller [2] published in Sloane's sequences of integers A083207 a sequence of integers n with the property that the positive factors of n can be partitioned into two disjoint parts so that the sums of the two parts are equal.

The study of graph labeling has focused on finding classes of graphs, which admits a particular type of labeling. Many practical problems in real-life situations

B.J. Balamurugan (✉)

Department of Mathematics, Agni College of Technology Thalambur, Chennai 600130, India
e-mail: balamuruganbj@yahoo.com

K. Thirusangu

Department of Mathematics, SIVET College Gowrivakkam, Chennai 600073, India
e-mail: kthirusangu@gmail.com

D.G. Thomas

Department of Mathematics, Madras Christian College Tambaram, Chennai 600059, India
e-mail: dgthomasccc@yahoo.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_45

405

have motivated the study of labelings of the graph $G = (V, E)$ subject to certain conditions. A systematic presentation of diversion applications of graph labelings is presented in [3]. The problems arising from the effort to study various labeling schemes of the elements of a graph are a potential area of challenge. Most of the labeling techniques found their origin with graceful labeling introduced by Rosa (1967) [4]. There is an enormous literature dealing with several kinds of labelings of graphs over the past three decades, and for a survey of various graph labeling problems, we refer to Gallian [5]. For all notations and terminology in graph theory, we follow Harary [6].

Graph labeling is a strong communication between number theory and structure of graphs. The study of Zumkeller numbers [2] forms part of the branch of mathematics called number theory. Some parts of number theory have found a use in code-making and breaking. Labeled graphs are becoming an increasingly useful family of mathematical models for a broad range of applications. Since Zumkeller numbers are a recent developments of number theory and graph labeling has often been motivated by practical problems, we are interested in Zumkeller labeling of graphs. The concept of strongly multiplicative Zumkeller labeling of graphs has been introduced and investigated in the literature [7]. It is proved that some well-known graphs, namely path, cycle, wheel, star, bipartite graphs, etc. are strongly multiplicative Zumkeller graphs.

In this paper, we provide algorithms to obtain Zumkeller labeling for complete bipartite graph [8] and wheel graphs. We refer a programming language slightly adapted to C, given by Frank Buss [9] to identify the Zumkeller numbers and Zumkeller partitions.

2 Zumkeller Numbers and Their Properties

In this section, we recall the notion of Zumkeller numbers and few properties of Zumkeller numbers [2].

Definition 1 A positive integer n is said to be a Zumkeller number if all the positive factors of n can be partitioned into two disjoint parts so that the sum of the two parts is equal.

We shall call such partition as Zumkeller partition.

Example 1 6, 12, 20 are Zumkeller numbers.

For 12, the factors of 12 are 1, 2, 3, 4, 6, 12.

We can partition these factors into 2 disjoint sets, such as,

$A = \{2, 12\}$, $B = \{1, 3, 4, 6\}$.

The sum of the elements of A is $2 + 12 = 14$.

The sum of the elements of B is $1 + 3 + 4 + 6 = 14$.

Therefore, 12 is a Zumkeller number.

Properties of Zumkeller Numbers

1. Let the prime factorization of an even Zumkeller number n be $2^k p_1^{k_1} p_2^{k_2} \dots p_m^{k_m}$. Then, at least one of k_i must be an odd number.
2. If n is a Zumkeller number and p is a prime with $(n, p) = 1$, then np^ℓ is a Zumkeller number for any positive integer ℓ .
3. Let n be a Zumkeller number and $p_1^{k_1} p_2^{k_2} \dots p_m^{k_m}$ be the prime factorization of n . Then, for any positive integers $\ell_1, \ell_2, \dots, \ell_m$, $p_1^{k_1 + \ell_1(k_1 + 1)} p_2^{k_2 + \ell_2(k_2 + 1)} \dots p_m^{k_m + \ell_m(k_m + 1)}$ is a Zumkeller number.
4. For any prime $p \neq 2$ and a positive integer k with $p \leq 2^{k+1} - 1$, $2^k p$ is a Zumkeller number.

3 Main Results

In this section, we introduce new labeling of graphs using Zumkeller numbers and prove that complete bipartite graphs [8] and wheel-related graphs admit Zumkeller labeling.

Definition 2 Let $G = (V, E)$ be a graph. An injective function $f: V \rightarrow N$ is said to be a Zumkeller labeling of the graph G , if the induced function $f^*: E \rightarrow N$ defined as $f^*(xy) = f(x)f(y)$ is a Zumkeller number for all $xy \in E, x, y \in V$.

Definition 3 A graph $G = (V, E)$ that admits a Zumkeller labeling is called a Zumkeller graph.

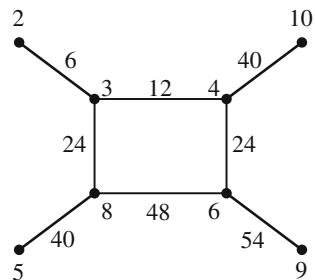
Example 2 The Zumkeller labeling of a graph is given in Fig. 1.

Here, the integers 6, 12, 40, 24, 54, 48, 40, 24 on the edges are Zumkeller numbers.

Proposition 1 A non-totally disconnected subgraph of a Zumkeller graph is a Zumkeller graph.

Proof Let $G = (V, E)$ be a Zumkeller graph. Let $G' = (V', E')$ be a subgraph of G obtained by removing some vertices or some edges or both such that G' is not a

Fig. 1 A Zumkeller



totally disconnected. Since the Zumkeller numbers on the edges of G' are obtained by multiplying the labels of the end vertices of the edges in G , obviously, the non-totally disconnected subgraph G' is a Zumkeller graph. \square

Definition 4 The complete bipartite graph on m and n vertices, denoted $K_{m,n}$, is the simple graph whose vertex set is partitioned into sets V_1 with m vertices and V_2 with n vertices in which there is an edge between each pair of vertices v_1 and v_2 where v_1 is in V_1 and v_2 is in V_2 .

Example 3 The complete bipartite graph $K_{2,4}$ is given in Fig. 2.

Now, we present an algorithm for the Zumkeller labeling of a complete bipartite graph.

Algorithm 1 (Zumkeller Labeling of a Complete Bipartite Graph)

This algorithm labels the edges of a complete bipartite graph with Zumkeller numbers.

Input : A complete bipartite graph $K_{m,n}$ having $m+n$ vertices and mn edges.

Output : Zumkeller complete bipartite graph.

Procedure : Zum_lab_complete bipartite graph.

$U = \{u_i \mid 1 \leq i \leq m\}$ and $V = \{v_j \mid 1 \leq j \leq n\}$ be the two disjoint vertex sets of the graph $K_{m,n}$.

$E = \{e_{ij} = u_i v_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ be the edge set of the graph $K_{m,n}$.

$p_1 :=$ a prime number $\neq 2 < 10$

$p_2 :=$ a prime number $\neq 2 < 10$

$p_1 \neq p_2$

for $i := 1$ **to** m **do**

$f(u_i) = p_1 2^i$

for $j := 1$ **to** n **do**

$f(v_j) = p_2 2^j$ // f is an injective function on the vertex set of $K_{m,n}$.

if $i \neq j$ **then**

begin

for $i := 1$ **to** m **do**

begin

for $j := 1$ **to** n **do**

$f^*(u_i v_j) = f(u_i) f(v_j)$ // f^* is an induced function

defined on the edge set of $K_{m,n}$.

end

end

else

$f^*(u_i v_i) = f(u_i) f(v_i)$

end Zum_lab_complete bipartite graph.

The time complexity of this algorithm is tightly bounded and it runs in polynomial time. The time complexity is $\theta(mn)$.

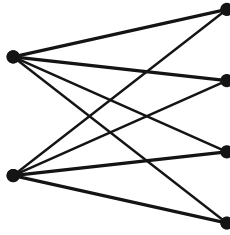


Fig. 2 Complete bipartite graph $K_{2,4}$

Theorem 1 *The complete bipartite graph $K_{m,n}$ is a Zumkeller graph.*

Proof Let the vertex set W of $K_{m,n}$ be partitioned into two disjoint sets $U = \{u_1, u_2, \dots, u_m\}$ and $V = \{v_1, v_2, \dots, v_n\}$ and let $E = \{e_{ij} = u_i v_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ be the edge set of the graph $K_{m,n}$.

Define a function $f: W \rightarrow N$ such that

$$f(u_i) = p_1 2^i \quad \forall u_i \in U, \quad 1 \leq i \leq m$$

$$f(v_j) = p_2 2^j \quad \forall v_j \in V, \quad 1 \leq j \leq n$$

where p_1, p_2 are distinct prime numbers which are greater than 2 but less than 10 and an induced function $f^*: E \rightarrow N$ such that

$$f^*(e_{ij}) = f^*(u_i v_j), \quad 1 \leq i \leq m, \quad 1 \leq j \leq n$$

$$= f(u_i) f(v_j).$$

We have two cases. □

Case 1: When $i = j$,

$$f^*(e_{ii}) = f^*(u_i v_i) = f(u_i) f(v_i) = p_1 2^i p_2 2^i = p_1 p_2 2^{2i}$$

which is a Zumkeller number.

Case 2: When $i \neq j$,

$$f^*(e_{ij}) = f^*(u_i v_j) = f(u_i) f(v_j) = p_1 2^i p_2 2^j = p_1 p_2 2^{i+j}$$

which is a Zumkeller number.

Hence, $K_{m,n}$ is a Zumkeller graph.

Corollary 1 *Every bipartite graph admits a Zumkeller labeling.*

Proof Since every bipartite graph is a non-totally disconnected subgraph of a complete bipartite graph and complete bipartite graphs are Zumkeller graphs, the result follows. □

Example 4 The complete bipartite graph $K_{5,4}$ is a Zumkeller graph for $p_1 = 3$, $p_2 = 5$, which is given in Fig. 3.

Algorithm 2 (Zumkeller Labeling of Wheel Graph $W_n = K_1 + C_n$)

This algorithm computes the integers to the vertices of the wheel graph $W_n = K_1 + C_n$ to label the edges with Zumkeller numbers.

Input : A wheel graph $W_n = K_1 + C_n$

Output : Zumkeller wheel graph.

Procedure : Zum_lab_wheel graph $W_n = K_1 + C_n$

Let v_0 be the central vertex of the wheel W_n

// v_1, v_2, \dots, v_n be the vertices of C_n in W_n .

// $E = \{e_i = v_i v_{i+1} \mid 1 \leq i \leq n-1\} \cup \{e_n = v_n v_1\} \cup \{e'_i = v_0 v_i \mid 1 \leq i \leq n\}$ be the edge set of the wheel graph W_n .

$p_1 :=$ a prime number $\neq 2 < 10$

$p_2 :=$ a prime number $\neq 2 < 10$

$p_3 :=$ a prime number $\neq 2 < 10$

$p_1 \neq p_2, p_2 \neq p_3, p_1 \neq p_3$

$f(v_0) = 2p_1$ // f is an injective function on the vertex set of W_n .

if $n \equiv 1 \pmod{2}$ **then**

begin

for $i = 1, 3, 5, \dots, n-2$ **do**

begin

$$f(v_i) = 2^{\frac{i+1}{2}}$$

$$f(v_{i+1}) = p_2 2^{\frac{i+1}{2}}$$

end

if $i = n$ **then**

$$f(v_n) = 2p_3$$

end

else

begin

for $i = 1, 3, 5, \dots, n-1$ **do**

begin

$$f(v_i) = 2^{\frac{i+1}{2}}$$

$$f(v_{i+1}) = p_2 2^{\frac{i+1}{2}}$$

end

end

end Zum_lab_wheel graph $W_n = K_1 + C_n$.

The time complexity of this algorithm is tightly bounded and it runs in polynomial time. The time complexity is $\theta(n)$.

Theorem 2 *The wheel graph $W_n = K_1 + C_n$ is a Zumkeller graph.*

Proof Let v_0 be the central vertex and v_1, v_2, \dots, v_n be the rim vertices of W_n .

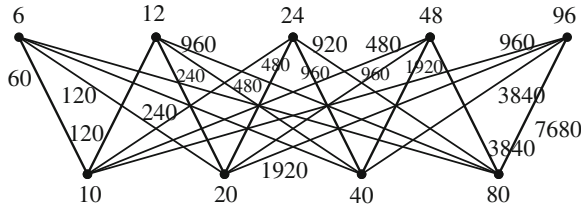


Fig. 3 Zumkeller labeling of $K_{5,4}$

Let $E = \{e_i = v_i v_{i+1} | 1 \leq i \leq n - 1\} \cup \{e_n = v_n v_1\} \cup \{e'_i = v_0 v_i | 1 \leq i \leq n\}$ be the edge set of the wheel graph W_n . We have the following cases. □

Case (i) $n \equiv 1 \pmod{2}$

Define an injective function $f: V \rightarrow N$ such that

$$f(v_0) = 2p_1$$

$$f(v_i) = 2^{\frac{i+1}{2}}, f(v_{i+1}) = p_2 2^{\frac{i+1}{2}}, i = 1, 3, 5, \dots, n - 2$$

$f(v_n) = 2p_3$ where p_1, p_2, p_3 are distinct prime numbers which are greater than 2 but less than 10 and an induced function $f^* : E \rightarrow N$ such that

$$f^*(e_i) = f^*(v_i v_{i+1}) = f(v_i) f(v_{i+1}), 1 \leq i \leq n - 2$$

$$f^*(e_{n-1}) = f^*(v_{n-1} v_n) = f(v_{n-1}) f(v_n)$$

$$f^*(e_n) = f^*(v_n v_1) = f(v_n) f(v_1) \text{ and}$$

$$f^*(e'_i) = f^*(v_0 v_i) = f(v_0) f(v_i), 1 \leq i \leq n$$

Now, we have to prove that the numbers on the edges are Zumkeller numbers.

- (i) $f^*(e_i) = f^*(v_i v_{i+1}) = f(v_i) f(v_{i+1}) = 2^{\frac{i+1}{2}} p_2 2^{\frac{i+1}{2}} = p_2 2^{i+1}, 1 \leq i \leq n-2$
which is a Zumkeller number by a property of Zumkeller numbers.
- (ii) $f^*(e_{n-1}) = f^*(v_{n-1} v_n) = f(v_{n-1}) f(v_n) = p_2 2^{\frac{n-1}{2}} 2p_3 = p_2 p_3 2^{\frac{n+1}{2}}$
which is a Zumkeller number and
- (iii) $f^*(e_n) = f^*(v_n v_1) = f(v_n) f(v_1) = 2 p_3 2 = 4p_3$
which is also a Zumkeller number
- (iv) If i is an odd number then
 $f^*(e'_i) = f(v_0) f(v_i) = 2p_1 2^{\frac{i+1}{2}} = p_1 2^{\frac{i+3}{2}}$
which is a Zumkeller number.
And $f^*(e'_{i+1}) = f(v_0) f(v_{i+1}) = 2p_1 p_2 2^{\frac{i+1}{2}} = p_1 p_2 2^{\frac{i+3}{2}}$ which is also a Zumkeller number.
- (v) $f^*(e'_n) = f(v_0) f(v_n) = 2p_1 2p_3 = p_1 p_3 2^2$
which is also a Zumkeller number.

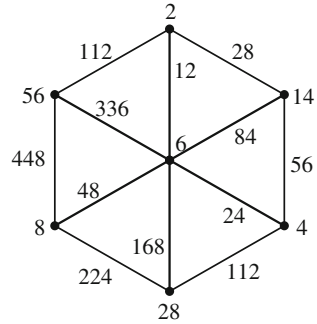
Case (ii) $n \equiv 0 \pmod{2}$

Define an injective function $f: V \rightarrow N$ such that

$$f(v_0) = 2p_1$$

$$f(v_i) = 2^{\frac{i+1}{2}}, f(v_{i+1}) = p_2 2^{\frac{i+1}{2}}, i = 1, 3, 5, \dots, n-1$$

Fig. 4 Zumkeller labeling of W_6



where p_1, p_2 are distinct prime numbers which are greater than 2 but less than 10 and an induced function $f^* : E \rightarrow N$ such that

$$\begin{aligned}
 f^*(e_i) &= f^*(v_i v_{i+1}) = f(v_i) f(v_{i+1}), 1 \leq i \leq n - 1 \\
 f^*(e_n) &= f^*(v_n v_1) = f(v_n) f(v_1) \text{ and} \\
 f^*(e'_i) &= f^*(v_0 v_i) = f(v_0) f(v_i), 1 \leq i \leq n
 \end{aligned}$$

From case (i), it is true that $f^*(e_i)$ is a Zumkeller number for $1 \leq i \leq n - 1$ and $f^*(e'_i)$ is also a Zumkeller number for $1 \leq i \leq n$. □

Now, $f^*(e_n) = f(v_n) f(v_1) = p_2 2^{\frac{n}{2}} 2 = p_2 2^{\frac{n+1}{2}}$ is also a Zumkeller number. Hence the wheel graph $W_n = C_n + K_1$ is a Zumkeller graph.

Corollary 2 A cycle C_n admits a Zumkeller labeling.

Proof The cycle C_n is a non-totally disconnected subgraph of W_n , obtained by removing the central vertex of W_n , and hence, C_n admits a Zumkeller labeling. □

Corollary 3 The path P_n with n vertices is a Zumkeller graph.

Example 5 For $n = 6$, the wheel graph W_6 is a Zumkeller graph, its Zumkeller labeling with $p_1 = 3, p_2 = 7$ is given in Fig. 4.

4 Conclusion

In this paper, we have proved the existence of Zumkeller labeling for complete bipartite graph and wheel graphs by presenting algorithms.

References

1. A.K. Srinivasan, Practical numbers. *Curr. Sci.* **17**, 179–180 (1948)
2. Y. Peng, B. Rao, K.P.S. Rao, On Zumkeller numbers. *J. Number Theor.* **133**(4):1135–1155

3. G.S. Bloom, S.W. Golomb, Applications of numbered undirected graphs. *IEEE* **165**(4), 526–570 (1977)
4. Rosa, A., On certain valuations of the vertices of a graph, in ed. by N.B. Gordan, Dunad, *Theory of Graphs. International Symposium.* (1966) pp. 349–359
5. J.A. Gallian, A dynamic survey of graph labeling. *Electron J Comb.* 16(DS6) (2013)
6. F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1972)
7. Balamurugan, B.J., Thirusangu, K., Thomas, D.G.: Strongly multiplicative Zumkeller labeling of graphs, in *International Conference on Information and Mathematical Sciences* (Elsevier, 2013), pp. 349–354
8. R. Johnsonbaugh, *Discrete Mathematics.* Pearson Education. Asia (2001)
9. Frank Buss.: *Zumkeller numbers and partitions*

An Unidentified Location-Based Efficient Routing Protocol in VANET

**P. Dharani, S. Sibi Chakkaravarthy, M. Ganesan,
Ethala Kamalanaban, P. Visu, Pravin R. Patil
and C. Mahesh**

Abstract Vehicular ad hoc networks (VANETs) use unidentified routing protocols that hide node uniqueness and which protects the node from outside observers to provide privacy protection. However, existing unidentified routing protocols depend on either hop-by-hop encryption or unnecessary traffic whichever makes high cost or cannot provide full privacy protection to data, sender, receiver, and routes. The high cost intensifies the essential resource constraint problem in VANET. To offer high privacy protection at a low cost, we propose an unidentified location-based efficient routing protocol. Here, using digital signatures and public key infrastructure (PKI) and mix zone-based security to protect message integrity is sufficient taking into account multilateral security. Main goal is to develop security architecture for VANETs that balances security requirements of all participants and also tries to identify to develop feasible mechanisms that fit in this architecture. This balances security requirements of all participants while keeping in mind the real time.

P. Dharani (✉) · M. Ganesan · C. Mahesh
Department of Information Technology, Vel Tech University, Chennai, Tamil Nadu, India
e-mail: dharaniarun007@gmail.com

M. Ganesan
e-mail: mganesh558@gmail.com

C. Mahesh
e-mail: chimahesh@gmail.com

S.S. Chakkaravarthy · E. Kamalanaban · P. Visu
Department of Computer Science and Engineering, Vel Tech University, Chennai
Tamil Nadu, India
e-mail: sb.sibi@gmail.com

E. Kamalanaban
e-mail: kamalanaban2009@gmail.com

P. Visu
e-mail: pandu.visu@gmail.com

P.R. Patil
Department of Computer Science and Engineering, Pune Institute of Computer Technology,
Pune, Maharashtra, India
e-mail: Prpatil@pict.edu

Keywords Vehicular ad hoc networks • Public key infrastructure • RSU • Public key cryptographic infrastructure • ADHOC mode

1 Introduction

VANET means vehicular ad hoc network it's the comes under the wireless networking and communication it's the part of mobile communication network here also used for the sensor network concepts because of sensor network is basically derived and used for the wireless networking. VANET is the subclass of mobile ad hoc network (MANET), and it is a similar work done in a process model, but it is dissimilar only in the mobile node (wireless telephone nodes for moving node) and vehicle node (moving vehicles connecting with wireless protocols). The vehicles can be have some history like first cars were pure mechanical then today's cars with almost fully controlled software devices and future smart cars fully controlled by software's (concept model of Audi A9 will repairs itself). Initially the people can be having the vehicles anyone and vehicle used for the rare cases but now any one can be have the vehicles and some people can be have the multiple vehicles. Because of the VANET, security providing is important one and it can develop the good motivation that is avoiding theft, and we can motivate the rules' followers. The VANET can be important one in the future world because today, many people died in vehicle accidents. Because of the human being safety and the vehicles crashes (damages) can be waste the most money therefore the VANET security is important one in the human life. The VANET explains the human life must have the safety plus secure policies. Initially, the people avoid the mobile ad hoc network, but it can be very useful for the current situations such as the VANET is also useful one for the emerging world for support the new kind of technologies. The VANET is used for the similar components of mobile ad hoc network using, and it can develop some future enhancements of networking, algorithms, routing protocols, scenarios, identifications, reporting, etc.

The VANET can have the working procedure model, and it can be deployed by the manufacturer- or government-authorized trusted third party (TTP). First upon, the VANET manufacturing can be fixed the unique vehicle identification number for the wireless networking. And it can derive all cities covered by the road side unit (RSU). The RSU can have the communication between the vehicles and certification authority (CA), and they store its details in the database storage in online, and it checks whether the vehicle is authorized or not. The vehicles can have the communication range with the coverage area with the help of the road side unit. Generally, the VANET is having the transmission and receiving the data in the format of the halo packets with the help of the beacon signals. The initial stage of the vehicle can be having the beacon communication of the vehicle to vehicle communication (i.e. V2V propagation) and the developing the technology can be have for the vehicle to infrastructure communication (i.e. V2I propagation). The

both types of communication are the important one in the vehicle communication technology, and these are developed with the help of the short-range communication and the distance-range communications. The RSU communicates within the coverage area (wireless broadcast area) vehicles tamper proof devices (TPD). The tamper proof devices can communicate with the vehicle wheels, and it can calculate the like speed and accelerations of the vehicles.

The vehicle tamper proof devices can have the digital signatures (DS) and public key cryptographic infrastructure (PKI) with itself. The vehicular networks can be emulated and increase the technology development, and it supports data transmission, and receiving without packet loss is the main challenge. The vehicular communication can have some basic components such as road side unit, certified authority, Internet, tamper proof devices, digital signature, on-board unit, and the public key infrastructure. The wireless access in vehicular environment (WAVE): IEEE 1609.2 standard also called as DSRC (distance short-range communication) 802.11p, and it supports the coverage range 1,000 m. The vehicle ad hoc network can be used for the pre-crash warning: warning messages from the authorized vehicles, jamming, and denial of services (DoS). And the traffic condition warning is as follows: tracking the theft vehicles and rules over tracker. The VANET need various privacy need is following: anonymity, unlink ability, restricted credential usages, perfect forward privacy. The security requirements need for VANET is authentication, availability, accountability, non-repudiation, verification of data consistency, and credential revocation. Advantages of the vehicular network are public safety, traffic management, traveller information support, traffic coordination and assistance, comfort, air pollution emission measurement, and finding the level for help the its reduction. The vanet can be having the drawbacks are data packet losses, routing optimization, flooding in the route discovery initial phases: wasted band width, delay, increasing networking congestion, external sources for destination location and it have the bad performances for the long distance between the source and destinations. Nowadays many researchers can do well and interested in this vehicular ad hoc network area it can be interesting and it can motivate the social engineering knowledge and develop the people safety.

2 Related Work

Existing vehicle network can be used for the two kinds of communication that are as follows:

1. Hop-by-hop communication
2. Zone-based communication

The hop-by-hop communication is basically using method, and it is the initial method. Here, the vehicle can have the unique identification and it can communicate only the nearest vehicle with the help of the beacon signals (beaconing). The node-to-node communication can be followed the small transmission, and it

transmits only the halo packets. This method can communicate only the vehicle-to-vehicle communication. The node-to-node communication can be used for the simple node and geographic routing protocol algorithms. It takes more cost and more communication and key distribution.

The zone-based communication can be the next level of the vehicular ad hoc networking, and it follows some additional futures. They can create the zone, and it can broadcast with the help of the road side unit (RSU). The RSU can cover the many zones, and it participates in the communication between the vehicles and certification authority (CA). The zone based communication used for the zone routing protocol (ZRP) and follow the zone based algorithm used in different kind of usages. The zone routing communication can follow the vehicle-to-vehicle communication (V2V) and the vehicle-to-infrastructure communication (V2I). The zone-based communication can have only the public key, so it is not able fully protection.

MANET is used for the unidentified location-based communication, and it can follow the same rules of the VANET. They used the dynamic pseudonym and location service, zone partitions, relay node, and random forwarding. The MANET provides the security to the source and destination, and it can hide the node with the help of anonymity routing protocol.

3 Working Methodology

Now, the VANET can have only the node-based security, and it can be the chances having attacks. We provide the mix zone for the VANET and hide the node identification for the vehicles and also hide the source, destination, and routes (data transfer). They used some requirements. They are as follows: digital signatures, public key infrastructure, mix zone, and CMIX encryption. The every vehicle can have the unique identification, providing the trusted third party (TTP); they can communicate with the help of the tamper proof devices (TPD). The vehicle has the on-board unit, vehicle sensors, and tamper proof devices; these things can communicate with the road side unit for data transferring.

The road side unit can have the coverage area, but we create the mix zones to avoid the key establishment, key forwarding, and key updating. The mix zone can be communicated if the vehicle can travel more road side unit means it will help. Now the normally the vehicle v1 and v2 communicated means the vehicle v1 can communicate it on board unit (OBU) it calculate the vehicle conditions (speed, acceleration and etc..) with help of vehicle sensors.

The on-board unit transfers the data to the tamper proof devices; they can have the digital signature and PKI itself. The vehicle communicates with tamper proof device; it can provide the digital signature (i.e., electronic signature) and can communicate with the public key infrastructure (PKI), the PKI through the data to road side unit; the PKI can provide the public key for the secure communication; and the private key (cryptographic key) provides for the certification authority. The

data can convey to the RSU, and it is fully in charge of all operations; it is like a relay node.

The road side unit can get the data from the vehicle v1 (hide the vehicle id using cryptographic method using cmix key) tamper proof device. And it can check whether the vehicle is authorized or not with the help of the certification authority (CA). The CA can have for all vehicles' information in its database, and the database is cloud as well as data center maintenance. The certification authority is the trusted third party; they get a license form the government for maintenance. The CA will check if the data source should be the authorized means transfer into the road side unit otherwise they aborted the connection and inform to the police.

Now, the road side unit can transfer the data to the vehicles if the vehicle is present within the coverage area; otherwise, it can communicate with the other road side unit and transfer the data in RSU to vehicle v2. The way of transmitting and receiving the data with the help of that algorithm hides the node id, and the attacker cannot find out the data sources, destinations, and data routes. The data can transfer safely with the help of CMIX algorithm, and it can follow the mix zone routing protocols.

The unidentified location-based efficient routing protocol in VANET is following these kinds of technologies in the real-time simulation in the mobility simulator. The VANET can be used for hiding node identification, and data packets are used to save the public people and its vehicles from the bad people. The bad people can be the attacker, and he cannot see the vehicle id and the data routes; also, it can do nothing. The VANET is implemented the road maps, and they can follow such rules and safety materials also.

3.1 Algorithm

```

Begin
Function VANET
    Input:MAP(.xml)
    Output:Vcap/frames
    Step:Pre-processing
RSu, CA, PSK;
PKI(horizontal,MAP);
PKI(vertical,MAP);
Plot(octagon);
Plot(rectangle);
Dilate(MAP);
    for i=1:m
        for j=1:n
            Findnode(min(n),max(m),min(m),max(n));
        End
    End
routing(Findnode);
End

```


4 Results and Implementation

See Fig. 1.

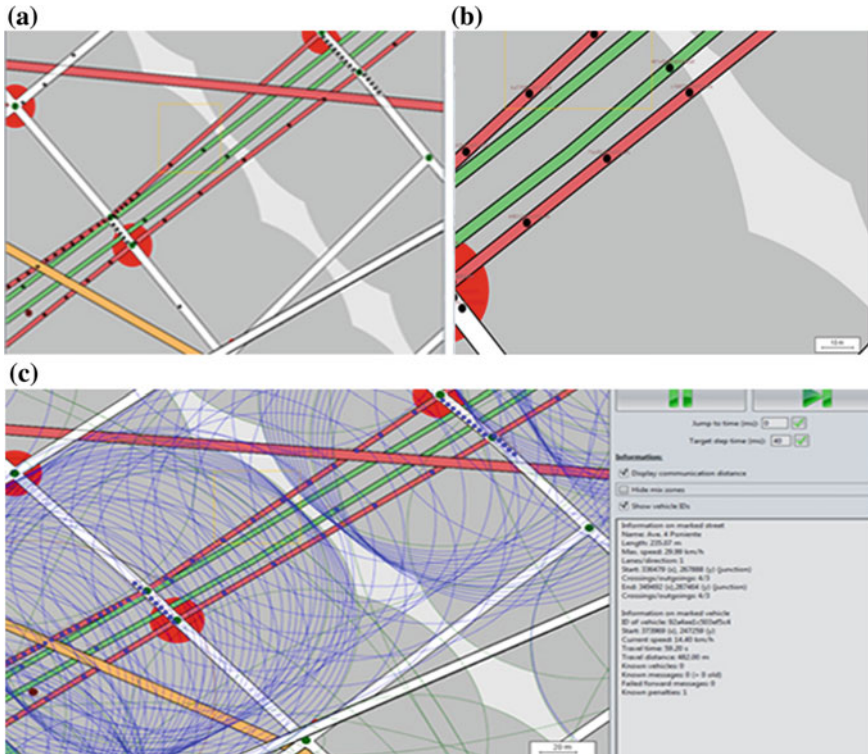


Fig. 1 Hide vehicle id, data route, and communication, **a** denotes vehicle identity for every node, **b** vehicle identity with tamper proof device, and **c** communication channel associated with RSU unit

5 Conclusion and Future Enhancements

In real-world scenario, identifying the vehicle and its position by tuning various GPS monitoring system is quite tedious in nature. Here, we proposed a hybrid algorithm called “unidentified routing algorithm” for monitoring the active vehicles in the ad hoc mode with highness of security. Since routing is clearly done through the art of the characteristic “invisbleness” and integrity. Here, mixed zone-level security is provided with public key infrastructure, due to its standalone privacy policy, and it utilizes the node integrity in ad hoc mode. The VANET can provide

only security now, but in the future, it can implement its routing optimization. The VANET RSU can have some limit for the vehicle traveling, and it can travel more RSU, and they can have a chance for the data packet loss.

References

1. H. Shen, L. Zhao, Alert: an anonymous location-based efficient routing protocol in MANETs. *IEEE Trans. Mob. Comput.* **12**(6) (2013)
2. P. Sermpezis, G. Koltsidas, F.-N. Pavlidou, Investigating a junction-based multipath source routing algorithm for VANETs. **17**(3) (2013)
3. J.-H. Song, V.W.S. Wong, V.C.M., Leung, Wireless location privacy protection in vehicular ad-hoc networks
4. J. Freudiger, M. Raya, M. Félegyházi, P. Papadimitratos, J.-P. HubauxEPFL, Mix-zones for location privacy in vehicular networks
5. K.I. Mershad, H. Artail, M. Gerla, We can deliver messages to far vehicles. *IEEE Trans. Intell. Transp. Syst.* **13**(3), 1099 (2012)

A Secure and Efficient Binding Update Scheme with Decentralized Design for Next Generation IP Mobility

Senthil Kumar Mathi, M.L. Valarmathi and Srilakshmy

Abstract Mobile Internet Protocol version 6 (MIPv6) maintains convergence for communication of mobile technologies with the ability to change their IP address. In MIPv6, a mobile device transmits location update messages which informs the current care-of address to its home agent indirectly, or to correspondent node directly, is termed as binding update. Security issues in MIPv6 are of vital significance, and the binding update messages must be secured against any malicious attacks that might attempt to get unauthorized access from the participating entities. This paper improves binding update scheme using optimal asymmetric encryption, which enhances security without relying on a third party. The proposed scheme achieves mutual authentication between the communicants and significant reduction in computational cost. The scheme is simulated and validated using the security tool AVISPA. Also, mitigation for various malicious attacks, such as replay attack, man-in-the-middle attack, and false binding update attack, have been addressed.

Keywords Binding update · Route optimization · Cryptographically generated address · Mutual authentication · Optimal asymmetric encryption

S.K. Mathi (✉) · Srilakshmy

Department of Computer Science and Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham University, Coimbatore, Tamil Nadu, India
e-mail: msenthil_cse@yahoo.co.in

Srilakshmy

e-mail: srilakshmykrish@gmail.com

M.L. Valarmathi

Department of Computer Science and Engineering, Government College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: ml_valarmathi@rediffmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_47

1 Prologue

With the growth of telecommunication sector, there is a gradual hike in the cardinality of mobile devices used till date which implies precisely that the mobile and other IP applications will be drifted with their sudden growth in near future. The security requirements of mobile IP [1] have been an active research area for recent years mainly focusing on authentication of binding update procedure. Internet protocol (IP) is a set of predefined technical rules that tells how hosts can communicate over a network. There are two versions used for IP mobility, namely, IP version 4 (IPv4) and version 6/next generation (IPv6/IPng). The former version IPv4 has limited addresses. But in the near future, it may be replaced by the latter with greater address space of 128 bits. The MIPv6 environment as shown in Fig. 1 includes a mobile node (MN), home agent (HA), and correspondent node (CN), unlike IPv4 which makes use of another entity called foreign agent (FA). In MIPv6, the MN is allowed to move from one subnetwork to another maintaining a reachable communication. On the other hand, in home network, MN uses a static IP address called home of address (HoA) to communicate with its CN, while MN's present address called as the care-of address (CoA) [2] changes dynamically with its mobility. However, the CoA of MN will be informed to HA at regular intervals to get the current attachment point.

A lack of protection of binding update (BU) and binding acknowledgement (BA) messages and would cause security attacks such as man-in-the-middle attack (MIMA), denial-of-service or false BU attacks, etc. Besides, concealing of CoA must be addressed through an authenticated channel. Hence, this paper focuses on a BU scheme by incorporating optimal asymmetric encryption which makes use of two hash functions based on random oracle model to enhance signaling security of BU and BA messages in IPv6 mobility. This paper is organized into five sections.

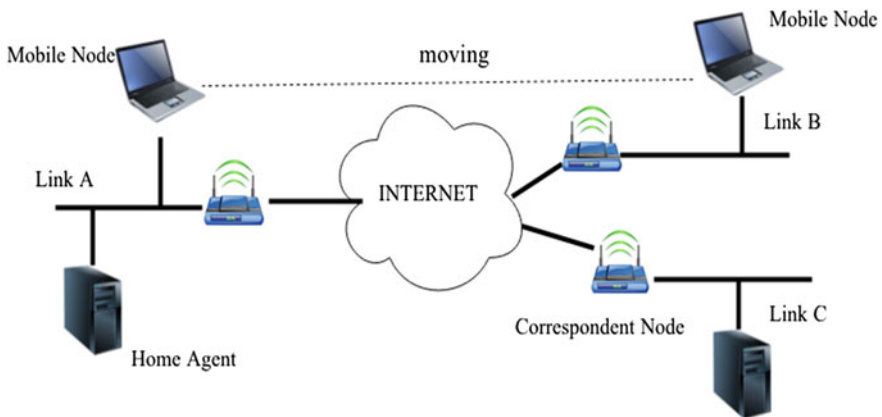


Fig. 1 Mobile IPv6 model

Section 2 discusses the related works of existing BU schemes. The intricacies of the proposed schemes are explained in Sect. 3. Results and analysis are discussed in Sect. 4. Section 5 describes conclusion and scope for future work.

2 Related Works

Numerous studies are investigated in [3, 4] for location update schemes in IPv6 mobility. RR-IBE [5] is return routability scheme integrated with ID-based encryption which makes use of a third party, known as private key generator (PKG) to issue private key during communication. Before the communication lays its foundation, each participating principal must know the identity of other communicants. The main feature of this procedure is that it neither requires an additional database for key recovery, nor a public key infrastructure using certificate authority. However, it depends on PKG for retrieving the private key. Subsequently, the research work [6] suggested a scheme (PKBU) to protect BU messages between MN and CN unlike return routability protocol where both messages from MN and CN are relayed by HA. It protects signal messages by integrating a private key-based model between MN and CN. The security is ensured only when a valid CoA is mapped to its HoA; however, there are chances of MN being malicious and CoA being fake/faulty. Next, the investigation [7] verifies a batch of binding update messages (BBU) using elliptic curve discrete logarithmic problem. Here, to ensure the address ownership, multi-key cryptographically generated address is used. This investigation also adopts traditional signatures to solve the security problems for route optimization which often leads to heavy computational costs when a large number of signatures are needed to be verified.

3 Proposed Binding Update Schemes

In this paper, we propose two binding update schemes with decentralized design using optimal asymmetric encryption for CoA generation.

3.1 Preliminaries

Optimal asymmetric encryption [8] involves message padding procedure, which uses a randomized function and two hash functions. It is based on key exchange attributes and with the restriction that the public key must be preshared. Unlike, discrete logarithm problem which depends on heavy computation, this method makes use of less cryptographic operations. This encryption is a four step procedure where initially two random oracle numbers are encrypted and appended to the text

Table 1 Notations used in the proposed BU schemes

Notation	Meaning
N_i	Nonce of an entity i
MN_{imp}	Temporary identity of MN
H	Hash function
MN_{sig}	Signature of MN
$E_{sym-mn}[M]$	Encryption of M using symmetric key of MN
K_{mn-ha}	Shared key between MN and HA

message. The receiving entity decrypts the message using the predefined public key and retrieves the contents. Also, it selects its own random numbers and computes a nonce resulting from the hash function. This step indicates the one-way authentication of the encryption scheme, and it is repeated with random values to complete mutual authentication mechanism. The proposed scheme uses the notations listed in Table 1.

3.2 CoA Generation Using Cryptographically Generated Address

The CoA (IPv6) is generated by MN by using cryptographically generated address (CGA) [9, 10]. The 128-bit IPv6 address consists of subnet prefix (leftmost 64 bits) and interface identifier (rightmost 64 bits). Here, the interface identifier is calculated using one-way hash function performed by SHA-1 algorithm. The input parameters defined for the CoA generation includes a randomized 128-bit modifier value, a 64-bit subnet prefix, a HA's public key, random nonces N_{mn} and N_{ha} , random values r_1 , and r_2 , and a security parameter "sec" predefined which has a value between 0 and 7 binary values (000 to 111), a collision count initially set to "0." Two SHA-1 passing is used for the address generation in which the first passing uses the 128-bit modifier, 9 octets of zero that is, 18 hexadecimal 0's, public key, extension fields. The leftmost 16 bits of the resulting hash1 (112 bits) is multiplied with sec and checked whether the value is zero or not if so, the modifier value is incremented by one bit. The loop proceeds until the first leftmost $16 * sec$ values is zero, then the second passing of SHA-1 follows with the new arguments of modifier, subnet prefix, public key, and the extension fields. Here, the first 64 bits of this hash2 is set as the interface identifier with the 6th and 7th bit set to zero. Now, the new IPv6 address is generated by concatenating earlier defined subnet prefix and the newly obtained interface identifier.

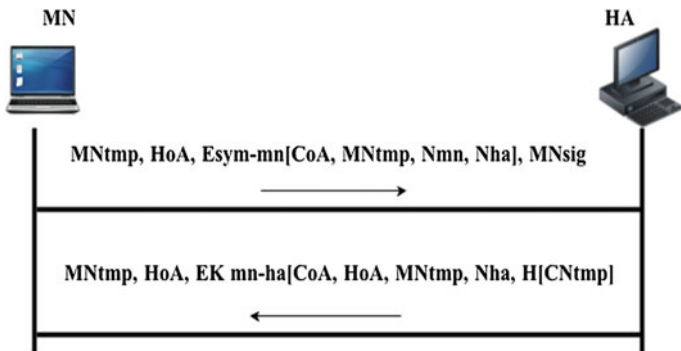


Fig. 2 Binding update with HA

3.3 The Proposed Binding Update with HA

The proposed BU procedure with HA consists of the following steps (Fig. 2): In step 1, MN sends a BU request $\{MN_{tmp}, HoA, E_{sym-mn}[CoA, MN_{tmp}, N_{mn}, N_{ha}], MN_{sig}\}$ to HA. Here, the intended CoA along with the random nonces of MN, HA and temporary identity of MN are encrypted by the preshared symmetric key of MN. In step 2, HA sends a BA $\{MN_{tmp}, HoA, E_{mn-ha}[CoA, HoA, MN_{tmp}, N_{ha}, H[CN_{tmp}]]\}$ to MN. Here, HA generates hash value using temporary identity of CN and MN decrypts and recalculates the hash to obtain CN_{tmp} , thus verifying the request.

3.4 The Proposed Binding Update with CN

The proposed BU procedure with CN consists of the following steps (Fig. 3): In step 1 MN sends a BU request $\{MN_{tmp}, HoA, E_{sym-mn}[CoA, MN_{tmp}, N_{mn}, N_{cn}], M_{sig}, H[MN_{tmp}]\}$,

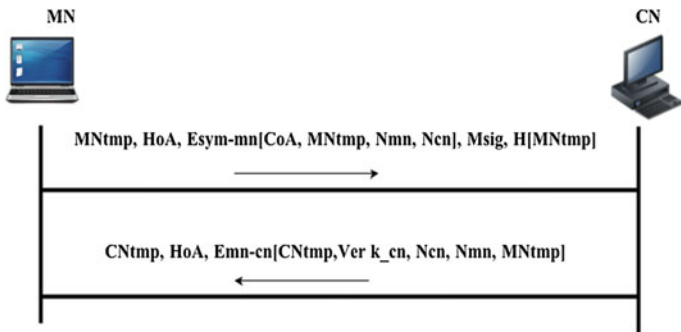


Fig. 3 Binding update with CN

$M_{\text{sig}}, H[MN_{\text{tmp}}]$ to CN where the temporary identity of MN is hashed and encrypted using symmetric key MN. In step 2, CN sends a BA message $\{CN_{\text{tmp}}, \text{HoA}, E_{\text{mncn}}[CN_{\text{tmp}}, \text{Ver}_{k_{\text{cn}}}, N_{\text{cn}}, N_{\text{mn}}, MN_{\text{tmp}}]\}$ to MN with encrypted message.

4 Security Analysis and Formal Analysis Using AVISPA Tool

This section discusses the security analysis of the proposed scheme. In the proposed scheme, MN signs each of the messages; each communicant is mutually authenticated. The location update messages to HA and CN are encrypted using encryption through a preshared key which is known only to either of the parties to provide secrecy in the BU and BA messages. The integrity can be achieved by providing a hashed value in each of the message sequence. The property with which either of the communicating parties accepts their intended message without denying is termed as non-repudiation. Here, an entity cannot repudiate because it uses the signature.

The CGA allows no duplication of a particular node, i.e., binding request comes from a valid and expected node, and there is a no chance of it being fake/spoofed. Since the random oracle parameters are shared only between the pairs MN-HA and MN-CN. As each message is concealed using the preshared key and also MN generates the CoA based on random oracle parameters the proposal prevents from man-in-the-middle attack. Also, every message in our scheme has been adopted in the specific time interval with the provision of randomized nonce either communicating parties; they are protected from replay attacks [11]. Table 2 shows the security considerations. The proposed scheme is verified using automated validation of Internet security protocols and applications (AVISPA) tool [12]. AVISPA has its own modular and role-based language for specifying security properties.

Table 2 Security considerations

	RRP	RRP-IBE	PKBU	NEMO	Proposed
Third party involvement	FA	HA	HA	HA	Directly to CN
Authentication	□	□	□	□	□
Confidentiality	□	□	□	□	□
Non-repudiation	∅	□	□	□	□
Integrity	□	□	đ	□	□
Man-in-middle attack	#	□	□	□	□
Replay attack	#	□	□	□	□
False binding update	#	đ	□	□	□
CN's memory saturation	#	đ	đ	đ	□

□ Provided, đ not provided, ∅ partial, # not discussed, T_{exp} time (expression), T_h time (hash)


```

role Mobile_Node(MN, HA: agent, sym-mn: public_key, MNmp, HoA, CoA, Nha, Nma: text, SND, RCV: channel (dy))
played_by MN def=
local
    State : nat
init
    State := 0
transition
0. State = 0 ∧ RCV(start) => State' = 1 ∧ MNmp' := new() ∧ HoA' := new() ∧ CoA' := new() ∧ Nha' := new() ∧ Nma' :=
new() ∧ SND(MNmp'. HoA'. { CoA', MNmp', Nma', Nha' }_Ksym-mn)
2. State = 1 ∧ RCV(MNmp'. HoA'. { CoA', MNmp', Nma', Nha' }_Ksym-mn) => State' = 2
end role
    
```

Fig. 4 Time complexity versus number of BU messages

Table 3 Security considerations

	Base RRP	RRP-IBE	PKBU	BBU	Proposed
Payload	3Ph + 4P	6T _{exp} + 1T _h	5T _{exp} + 2T _h	n*(3T _h + 4T _{exp})	n*(2T _{exp} + 2T _{hash})
Security parameters	1H	1H	2H	3H	2-way H
Number of messages	4	7	5	4	4

Ph Phases; P pairing; T_{exp} time (exponentiation); T_h time (hash); H hash function

Table 4 Execution time for cryptographic calculations

Number of BU messages	T _{sig}	T _{ver}	T _{hash}	T _{mul}	T _{exp}	T _{rsa}
20	4.32	0.025	0.43	0.55	0.027	0.527
40	4.35	0.057	0.432	0.551	0.027	0.530
60	4.363	0.107	0.436	0.56	0.029	0.567
80	4.37	0.115	0.440	0.568	0.03	0.6
100	4.40	0.127	0.441	0.575	0.324	0.67

The simulated role definition of our proposed scheme in AVISPA for MN is shown in Fig. 4. Let T_{oper} be the execution time for cryptographic operations such as hash, exponentiation. It can be observed from Table 2 that our proposed schemes ensure integrity as well as protect from replay attack and false BU attack. Table 3 discusses about the computation payload and security parameters where in on an average it takes two-way hash function to provide the required level of security. The total time calculations for proposed schemes have been estimated and shown in Table 4 and its subsequent calculation on the number of BU messages (Table 5).

Table 5 Computation cost for BU schemes

Number of BU messages	T_{RRP}	$T_{RRP-IBE}$	T_{PKBU}	T_{BBU}	$T_{Proposed}$
20	2.108	2.038	2.004	1.878	1.474
40	2.120	2.057	2.023	1.921	1.506
60	2.127	2.06	2.026	1.945	1.518
80	2.154	2.073	2.051	1.964	1.543
100	2.170	2.075	2.101	1.989	1.551

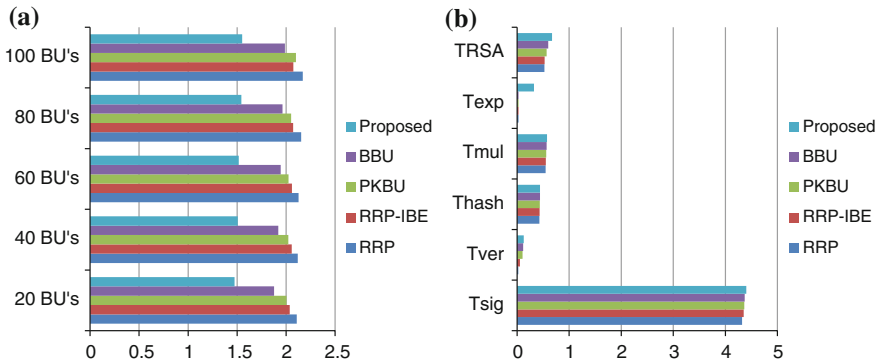


Fig. 5 Time complexity versus number of BU messages. **a** Binding update latency in ms. **b** Cryptographic execution time in ms

Figure 5 depicts various schemes on the basis of time complexity (average) and their approximations when the number of BU messages increases.

5 Conclusion

The proposed BU schemes that carry the current CoA of MN can be protected to a great extent so that possible attacks can be avoided and the CoA can be concealed. The proposed scheme incorporates a scheme of symmetric encryption, which ensures mutual authentication in the CoA generation. That is, the newly generated CoA address which is embedded in the communication can be protected by the use of optimal asymmetric encryption in which both CN and MN can be mutually authenticated. Both schemes are validated using the security tool AVISPA. From the security considerations, it can be concluded that the proposed schemes provide all security requirements with minimal computational delay and protection from all possible attacks such as MIMA, replay attack, DoS attack, false BU attack.

References

1. K. Elgoarany, M. Eltoweissy, Security in mobile IPv6.: a survey. *Inf. Secur. Tech. Rep.* **12**(1), 32–43 (2007)
2. J. Guan, I. You, C. Xu, H. Zhou, H. Zhang, Survey on route optimization schemes for proxy mobile IPv6, in *6th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IEEE* (2012), pp. 541–546
3. H. Modares, A. Moravejosharieh, J. Lloret, R. Salleh, A survey of secure protocols in mobile IPv6. *J. Netw. Comput. Appl.* (2013)
4. S.K. Mathi, M.L. Valarmathi, A secure and decentralized registration scheme for IPv6 network-based mobility. *Int. J. Eng. Technol.* **5**(5) (2013)
5. W.A.A. Alsalihi, M.S.S. Alsayfi, Integrating identity-based encryption in the return routability protocol to enhance signal security in mobile IPv6. *Wirel. Pers. Commun.* **68**(3), 655–669 (2013)
6. H. Modares, R. Salleh, A. Moravejosharieh, H. Malakootikhah, Securing binding update in mobile IPv6 using private key base binding update protocol (2012)
7. L.Y. Yeh, C.C. Yang, J.G. Chang, Y.L. Tsai, A secure and efficient batch binding update scheme for route optimization of nested network mobility in VANETs. *J. Netw. Comput. Appl.* **36**(1), 284–292 (2013)
8. J. Liu, J. Li, A novel key exchange protocol based on RSA-OAEP, in *10th International Conference on Advanced Communication Technology, ICACT 2008* (2008), p. 3
9. S. Qadir, M.U. Siddiqi, Cryptographically generated addresses (CGAs): a survey and an analysis of performance for use in mobile environment. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **11**(2), 24–31 (2011)
10. C. Castellucia, G. Montenegro, Securing group management in IPv6 with cryptographically generated addresses, in *8th IEEE International Symposium Computers and Communication* (2003), pp. 588–593
11. S. Sehwa, C. Hyoung-Kee, K. Jung-Yoon, A secure and lightweight approach for routing optimization in mobile IPv6. *EURASIP J. Wirel. Commun. Netw.* (2009)
12. Armando, A., Basin, D., Boichut, Y., Chevalier, Y., Compagna, L., Cuéllar, J., Vigneron, L.: The AVISPA tool for the automated validation of internet security protocols and applications. In *Computer Aided Verification*. (2005) 281-285

Stego with Color Cue

**N.R. Raajan, G. Balasubraminayan, B. Barath Krishna,
S. Ramya, M. Malligaraj and K. Karthikeyan**

Abstract The ontogenesis of technology has paved way for an upsurge in cyber-crime too. In fact, the rate at which this happens overhauls even the Moore's law. Steganography is a tool which supports us in elucidating the above facts. Here, we have proposed four methodologies in image steganography. First method deals with gray image, and the subsequent methods encounter color image. Pixel indicator technique is used for selecting the plane of embedding for color images. Embedding is done in block segmentation fashion. We have devised a novel mod4 methodology which tells us about the order of embedding, number of bits to be embedded in each pixel as well as about indicator channel in PIT and also tremendously escalates the randomization. OPAP technique reduces error for an enhanced image quality. MSE, PSNR, and the brute-force attack values of the above proposed methodologies have been simulated and displayed.

Keywords LSB substitution · OPAP · Space-filling curve (SFC) · Pixel indicator

N.R. Raajan (✉) · G. Balasubraminayan · B. Barath Krishna · S. Ramya · M. Malligaraj · K. Karthikeyan

School of Electrical and Electronics Engineering, SASTRA University, Tanjore
Tamil Nadu, India
e-mail: nrraajan@ece.sastra.edu

G. Balasubraminayan
e-mail: balu_eie@eie.sastra.edu

B. Barath Krishna
e-mail: barath.oct@gmail.com

S. Ramya
e-mail: ramya.selvan@gmail.com

M. Malligaraj
e-mail: m.malligaraj@gmail.com

K. Karthikeyan
e-mail: krishkarthikeyan1991@gmail.com

1 Introduction

The very existence of mankind dictated a need for communication which then translated to a need for secure communication with the mankind growing greedy and needy. With the advent of the industrial revolution, technology found itself in a new dimension opening up the fancied portfolio ‘cryptography.’ This was short-lived when the number of hackers rocketed up exponentially. More efficient methods of hacking came in leading to the idea of ‘cryptography’ being outdated. These shortcomings were overcome by the simple but a deeper concept of ‘steganography.’ This efficient method had no marks of flaw in contrast to its predecessor, and its ‘cover image’ served as a perfect blockage for secret data giving the user a feeling of void in it. Researchers started to introduce many randomization algorithms and procedures to change the data, but keep the outlook simple and clean. Steganography’s fascinating part was that the data were intact even after many randomizations attracting many toward this field of study. The opacity of the ‘cover image’ left the hackers awestruck. Apart from images, steganography has the special property of operating on any hardware (here sources) such as audio files and videos. This versatile field of ‘steganography’ is inevitable in all walks of life and shall prove excellent for constructive purposes.

2 Proposed Methodology

In this work, four new-fangled algorithms are proposed for Randomized Image Steganography, one is for gray images and the others are for color images. In gray scale algorithm, image is segmented into 8×8 blocks, and the order of embedding is shown in Table 1.

The matrix exhibited above is composed in a fashion such that in first row, first four numbers are filled by deserting one cell in clockwise and final four numbers in anticlockwise. In the next row, consecutive values from first are written in anticlockwise and from last in clockwise direction, and this setup is adhered throughout.

Then, three methods for color images which increases the complexity in step by step are proposed. In PI technique, indicator channel for the first method is chosen as RED. Image will be block segmented into a 4×4 block, and this algorithm is designed to process on this 4×4 alone. Each individual block will be taken, processed, and then will be stored as stego image. In the next method, indicator channel will be decided based on the user’s wish during the run time, and in the last method, indicator channel selection is also done by mod4-based method, thus increases the complexity of the algorithm.

The so-called space-filling curve which has been dealt here will go around a 4×4 block. That particular 4×4 random matrix for our corresponding SFC is shown in Table 2.

Table 1 8×8 random matrix

1	64	2	63	3	62	4	61
57	8	58	7	59	6	60	5
9	56	10	55	11	54	12	53
49	16	50	15	51	14	52	13
17	48	18	47	19	46	20	45
41	24	42	23	43	22	44	21
25	40	26	39	27	38	28	37
33	32	34	31	35	30	36	29

Table 2 4×4 random matrix

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

Table 3 mod4 matrix

4	11	7	3
8	13	15	10
12	16	14	6
1	5	9	2

This matrix, firstly, converted to its mirror image and secondly that mirror imaged matrix is transposed. With this transposed matrix, the position of the pixels is determined. Let that transposed matrix be ‘random.’ And number of bits to be embedded in every pixel will be decided by taking mod4 of the particular transposed matrix, ‘random,’ in each position. So it is just enough to have the SFC that means the random matrix alone in both the sides to determine the ‘k’ value using mod4 method. Let that matrix be ‘mod4.’ So mod4 will be as in Table 3.

2.1 Method I

2.1.1 Embedding Algorithm

Inputs: cover image (Gray) (C), secret data (d), and number of bits per pixel (k).

Output: stego image (S) with secret data embedded in it.

1. Get the binary stream from the secret data.
2. Get 'random' matrix from SFC methodology.
3. Let $x = 1$, a variable which decides the position of embedding.
4. Go to ' x ' in the random matrix.
5. Embed ' k ' number of bits of secret data in the LSB of the ' x ' pixel.
6. Now, $x = x + 1$ and go to step 4 until the ' x ' becomes 64 in the specified matrix sized 8×8 .
7. If $x = 64$, then go to the next 8×8 block in the cover image in the next cycle and similarly repeat the steps 3–7.
8. Perform the steps 3–7 until all the data are embedded into the image and then stop the embedding procedure.

2.1.2 Recovery Algorithm

Inputs: stego image (S), number of bits per pixel (k). *Output:* secret data (D).

1. Split the cover image and get 'random' matrix from SFC methodology.
2. Let $x = 1$, a variable which decides the position of embedding.
3. Go to ' x ' in random matrix.
4. Assume index of ' x ' to be (i, j) in random matrix.
5. Read k bits of secret data in the LSB of the current pixel and concatenate to D.
6. Now, $x = x + 1$ and go to step 4 until the x becomes 64 in the random matrix sized 8×8 .
7. If $x = 64$, then go to the next 8×8 block in the cover image in the next cycle and similarly repeat the steps 2–6.
8. Store the resulting recovered secret data D.
9. Do the same until the entire data are recovered.

2.2 Method II

2.2.1 Embedding Algorithm

Inputs: cover image (C) and secret data (d). *Output:* stego image (S) with secret data embedded in it.

1. Get the binary stream from the secret data
2. Split the RGB color planes of the cover image and get 'mod4,' 'random' matrix from SFC by mod4 methodology.
3. Segment entirely the R, G, and B indicators into 4×4 sized blocks, namely RB, GB, and BB and do the following procedure for each of the pixel:
4. Let $x = 1$, a variable which decides the position of embedding.
5. Go to the assumed ' x ' in the random matrix.

6. Assume index of 'x' to be (i, j) in random matrix.
7. Let k = value in indicated position (i, j) of the mod4 matrix.
8. For each pixel in the cover image in the order of our random matrix, ref = the last two LSBs of current pixel in RB should be taken.
9. If ref = 00, then move to the next pixel in the order of position.
 Else if ref = 01, then embed k bits of secret data in the current pixel of B.
 Else if ref = 10, then embed k bit of secret data in the current pixel of G.
 Else, embed k bits of secret data in the current pixel's G as well as B color plane.
10. Now, $x = x + 1$ and go to step 5 until the x becomes 16 in the specified matrix sized 4×4 .
11. If $x = 16$, then go to the next 4×4 block in the cover image in the next cycle and similarly repeat the steps 4–10.
12. Perform the steps 4–11 until all the data are embedded into the image and then stop the embedding procedure.

2.2.2 Recovery Algorithm

Inputs: stego image (S). *Output:* secret data (D).

1. Split the RGB color planes of the cover image and get 'mod4,' 'random' matrix from SFC by mod4 methodology.
2. Segment entirely the R, G, and B indicators into 4×4 sized blocks, namely RB, GB, and BB and do the following procedure for each of the pixel:
3. Let $x = 1$, a variable which decides the position of embedding.
4. Go to the assumed 'x' in random matrix.
5. Assume index of 'x' to be (i, j) in random matrix.
6. Let k = value in indicated position (i, j) of the mod4 matrix.
7. For each pixel in the cover image in the order of our random matrix, ref = the last two LSBs of current pixel in RB should be taken.
8. If ref = 00, then move to the next pixel in the order of position.
 Else if ref = 01, then read k bits of secret data in the current pixel of B and concatenate to D.
 Else if ref = 10, then read k bit of secret data in the current pixel of G and concatenate to D.
 Else, read k bits of secret data in the current pixel's G as well as B color plane and concatenate to D.
9. Now, $x = x + 1$ and go to step 4 until the x becomes 16 in the random matrix sized 4×4 .
10. If $x = 16$, then go to the next 4×4 block in the cover image in the next cycle and similarly repeat the steps 3–9.
11. Store the resulting recovered secret data D.

2.3 Method III

2.3.1 Embedding Algorithm

Inputs: secret data (D), cover image (C), and Indicator Plane Index (I).

Output: stego image (S) with secret data embedded in it.

1. Convert the secret data into binary stream.
2. Split the RGB color planes of the cover image and get 'mod4,' 'random' matrix from SFC by mod4 methodology.
3. Segment entirely the R, G, and B indicators into 4×4 sized blocks, namely RB, GB, and BB and do the following procedure for each of the pixel.
4. Let I be the index of the plane to be chosen as indicator channel which abides by the definition from user.

If $I = 1$, then $P [1] = R$, $P [2] = G$, $P [3] = B$

Else if $I = 2$, then $P [1] = G$, $P [2] = B$, $P [3] = R$

Else if $I = 3$, then $P [1] = B$, $P [2] = R$, $P [3] = G$

5. Let $x = 1$, a variable which decides the position of embedding.
6. Go to the assumed 'x' in random matrix.
7. Assume index of 'x' to be (i, j) in random matrix.
8. Let $k =$ value in indicated position (i, j) of the mod4 matrix.
9. For each pixel in the cover image of our random matrix, ref = the last two LSBs of current pixel in color plane block specified by $P [1]$ should be taken.
10. If ref = 00, then move to the next pixel in the order of position.
 Else if ref = 01, then embed k bits of secret data in the current pixel in color plane block specified by $P [3]$.
 Else if ref = 10, then embed k bit of secret data in the current pixel in color plane block specified by $P [2]$.
 Else, embed k bits of secret data in the current pixel's $P [2]$ as well as $P [3]$ color plane.
11. Now, $x = x + 1$ and go to step 6 until the x becomes 16 in the specified matrix sized 4×4 .
12. If $x = 16$, then go to the next 4×4 block in the cover image in the next cycle and similarly repeat the steps 5–10.
13. Perform the steps 5–12 until all the data are embedded into the image and then stop the embedding procedure.

2.3.2 Recovery Algorithm

Inputs: stego image (S) and Indicator Plane Index (I). *Output:* secret data (D).

1. Split the RGB color planes of the cover image and get 'mod4,' 'random' matrix from SFC by mod4 methodology

2. Segment entirely the R, G, and B indicators into 4×4 sized blocks, namely RB, GB, and BB and do the following procedure for each of the pixel:
3. Let I be the index of the plane to be chosen as indicator channel which abides by the definition from user.
 If $I = 1$, then $P [1] = R$, $P [2] = G$, $P [3] = B$
 Else if $I = 2$, then $P [1] = G$, $P [2] = B$, $P [3] = R$
 Else if $I = 3$, then $P [1] = B$, $P [2] = R$, $P [3] = G$
4. Let $x = 1$ and move to the assumed 'x' random matrix.
5. Assume index of 'x' to be (i, j) in random matrix.
6. Let $k =$ value in indicated position (i, j) of the mod4.
7. For each pixel in the cover image of our random matrix, ref = the last two LSBs of current pixel in color plane block specified by $P [1]$ should be taken.
8. If ref = 00, then move to the next pixel in the order of position.
 Else if ref = 01, then read k bits of secret data in the current pixel in color plane block specified by $P [3]$ and concatenate to D.
 Else if ref = 10, then read k bit of secret data in the current pixel in color plane block specified by $P [2]$ and concatenate to D.
 Else, read k bits of secret data in the current pixel in color plane block specified by $P [2]$ as well as $P [3]$ color plane and concatenate to D.
9. Now, $x = x + 1$ and go to step 5 until the x becomes 16 in the specified matrix sized 4×4 .
10. If $x = 16$, then go to the next 4×4 block in the cover image in the next cycle and similarly repeat the steps 4–9.
11. Store the resulting recovered secret data D.

2.4 Method IV

2.4.1 Embedding Algorithm

Inputs: secret data (D), cover image (C), and SFC. *Output:* stego image (S) with secret data embedded in it.

1. Convert the secret data into binary stream.
2. Split the RGB color planes of the cover image and get 'mod4,' 'random' matrix from SFC by mod4 methodology.
3. Segment entirely the R, G, and B indicators into 4×4 sized blocks, namely RB, GB, and BB and do the following procedure for each of the pixel:
4. Let $x = 1$, a variable which decides the position of embedding.
5. Indicator plane will be decided based on the following condition:
 - 1) If mod $(x, 4)$ is 1, then $P [1] = R$, $P [2] = G$, $P [3] = B$
 - 2) Else if mod $(x, 4)$ is 2, then $P [1] = G$, $P [2] = B$, $P [3] = R$
 Else if mod $(x, 4)$ is 3, then $P [1] = B$, $P [2] = R$, $P [3] = G$
 Else, no embedding in that pixel.

6. Go to the assumed 'x' in random matrix.
7. Assume index of 'x' to be (i, j) in random matrix.
8. Let k = value in indicated position (i, j) of the mod4 matrix.
9. For each pixel in the cover image of our random matrix, ref = the last two LSBs of current pixel in color plane block specified by P [1] should be taken.
10. If ref = 00, then move to the next pixel in the order of position.
Else if ref = 01, then embed k bits of secret data in the current pixel in color plane block specified by P [3].
Else if ref = 10, then embed k bit of secret data in the current pixel in color plane block specified by P [2].
Else, embed k bits of secret data in the current pixel's P [2] as well as P [3] color plane.
11. Now, $x = x + 1$ and go to step 5 until the x becomes 16 in the specified matrix sized 4×4 .
12. If $x = 16$, then go to the next 4×4 block in the cover image in the next cycle and similarly repeat the steps 4–10.
13. Perform the steps 4–12 until all the data are embedded into the image and then stop the embedding procedure.

2.4.2 Recovery Algorithm

Inputs: stego image (S). *Output:* secret data (D).

1. Split the RGB color planes of cover image and get 'mod4,' 'random' matrix from SFC by mod4 methodology.
2. Segment entirely the R, G, and B indicators into 4×4 sized blocks, namely RB, GB, and BB and do the following procedure for each of the pixel:
3. Let $x = 1$, a variable which decides the position of embedding.
4. Indicator plane will be decided based on the following condition:
If mod $(x, 4)$ is 1, then P [1] = R, P [2] = G, P [3] = B
Else if mod $(x, 4)$ is 2, then P [1] = G, P [2] = B, P [3] = R
Else if mod $(x, 4)$ is 3, then P [1] = B, P [2] = R, P [3] = G
Else, no embedding in that pixel.
5. Go to the assumed 'x' in random matrix.
6. Assume index of 'x' to be (i, j) in random matrix.
7. Let k = value in indicated position (i, j) of the mod4 matrix.
8. If ref = 00, then move to the next pixel in the order of position.

Else if ref = 01, then read k bits of secret data in the current pixel in color plane block specified by P [3] and concatenate to D.
Else if ref = 10, then read k bit of secret data in the current pixel in color plane block specified by P [2] and concatenate to D.
Else, read k bits of secret data in the current pixel in color plane block specified by P [2] as well as P [3] color plane and concatenate to D.

9. Now, $x = x + 1$ and go to step 4 until the x becomes 16 in the specified matrix sized 4×4 .
10. If $x = 16$, then go to the next 4×4 block in the cover image in the next cycle and similarly repeat the steps 3–9.
11. Store the resulting recovered secret data D.

3 Error Metrics

The two main parameters used to estimate the quality of stego image are mean square error and peak signal to noise ratio. The MSE is calculated using the equation,

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (X_{ij} - Y_{ij})^2$$

Where, $X_{i,j}$ is stego value and $Y_{i,j}$ is the cover object.

The PSNR is calculated using the equation,

$$\text{PSNR} = 10 \log_{10} \left(\frac{I_{\max}^2}{\text{MSE}} \right) \text{dB}$$

Where I_{\max} is the intensity value of each pixel which is equal to 255 for 8 bit gray scale images.

Higher the value of PSNR, better the image quality. As we all know, we should have a low MSE and high PSNR values for a good stego system. Distortion analysis of stego image using the software-based secret sharing algorithm with 100 % embedding of data gave the following results. Here, in implementation Tree, Football, and Lena [256 × 256], color digital images have been taken as cover images and tested for full embedding capacity. The effectiveness of the stego process proposed has been studied by calculating MSE and PSNR for all the four digital images in RGB planes (Figs. 1, 2).



Fig. 1 Cover images

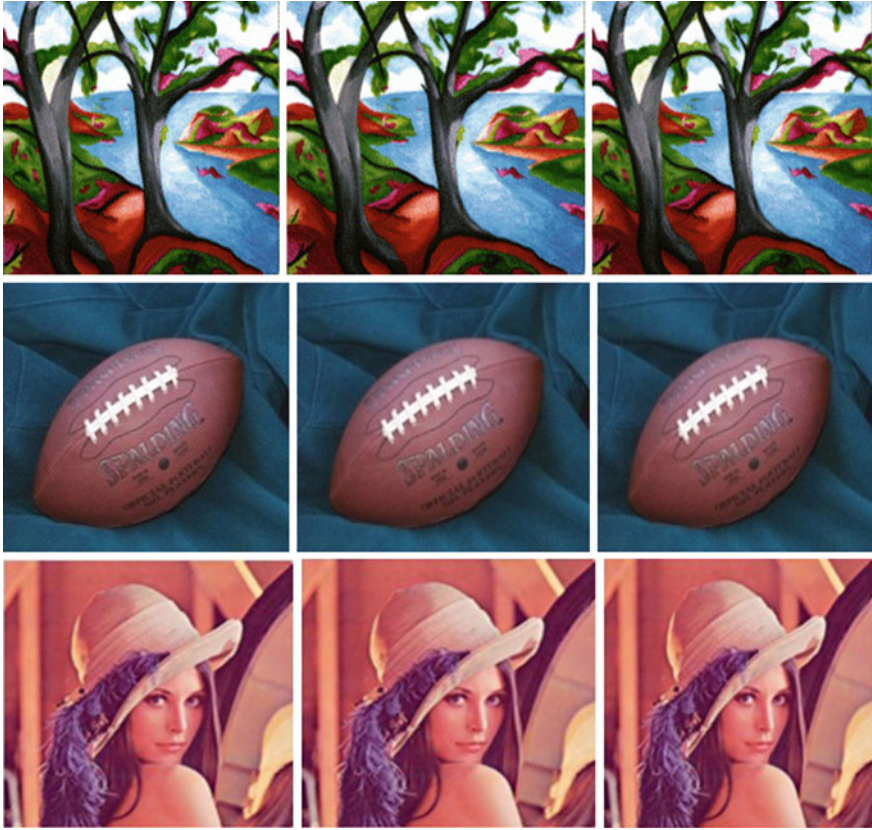


Fig. 2 Stego images obtained in method II, III, and IV, respectively (from top to bottom)

4 Tabulation and Results

See Figs. 3, 4, 5, 6.

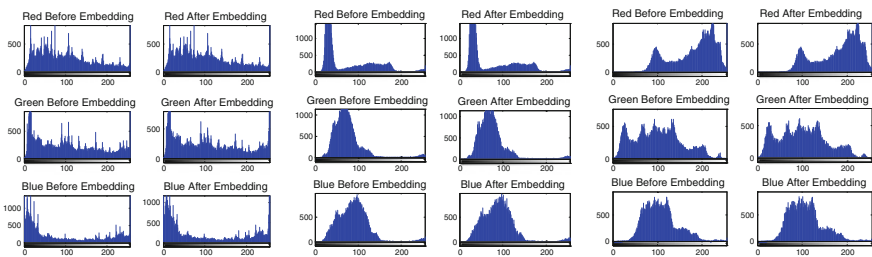


Fig. 3 Tree, football, Lena (red indicator)

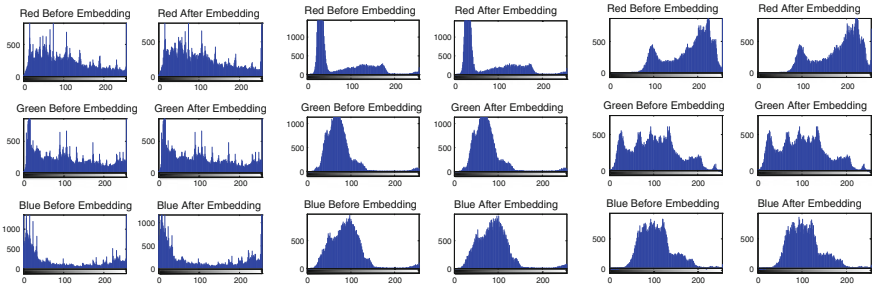


Fig. 4 Tree, football, Lena (*green* indicator)

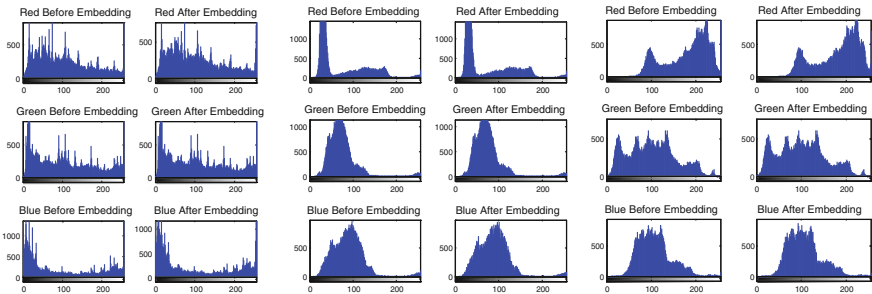


Fig. 5 Tree, football, Lena (*blue* indicator)

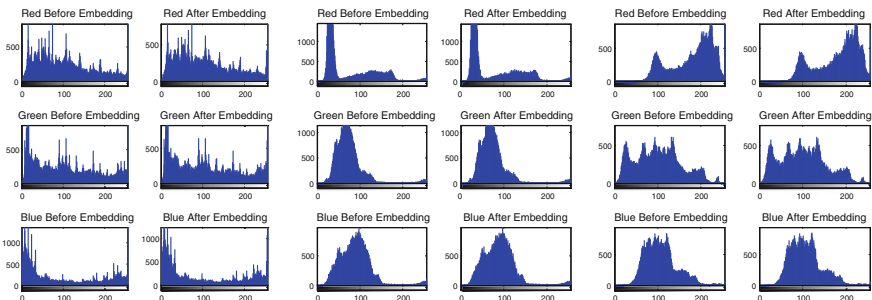


Fig. 6 Tree, football, Lena (Variable indicator)

Table 4 Method II tabulation

Cover image	Indicator channel	Channel 1 (red)		Channel 2 (green)		Channel 3 (blue)		Bits per pixel
		MSE	PSNR	MSE	PSNR	MSE	PSNR	
Tree	Red	0	∞	0.3185	53.0998	0.3132	53.1728	1.5523
Football	Red	0	∞	0.3140	53.1610	0.3160	53.1337	1.5031
Lena	Red	0	∞	0.3169	53.1221	0.3211	53.0649	1.5176

4.1 Method II

See Table 4.

4.2 Method III

See Table 5.

4.3 Method IV

See Table 6.

Table 5 Method III tabulation

Cover image	Channel 1 (red)		Channel 2 (green)		Channel 3 (blue)		Bits per pixel
	MSE	PSNR	MSE	PSNR	MSE	PSNR	
Tree	0.2917	53.4814	0.2474	54.1962	0.0862	58.7751	1.5393
Football	0.2879	53.5385	0.2484	54.1797	0.0836	58.9099	1.4945
Lena	0.2888	53.5250	0.2523	54.1124	0.0823	58.9781	1.5024

Table 6 Method IV tabulation

Cover image	Indicator	Channel 1 (red)		Channel 2 (green)		Channel 3 (blue)		Bits per pixel
		MSE	PSNR	MSE	PSNR	MSE	PSNR	
Tree	Red	0	∞	0.3185	53.0998	0.3132	53.1728	1.5523
	Green	0.3164	53.1291	0	∞	0.3159	53.1350	1.5372
	Blue	0.3168	53.1233	0.3123	53.1853	0	∞	1.5352
Football	Red	0	∞	0.3140	53.1610	0.3160	53.1337	1.5031
	Green	0.3128	53.1784	0	∞	0.3104	53.2121	1.5029
	Blue	0.3100	53.2175	0.3119	53.1907	0	∞	1.5007
Lena	Red	0	∞	0.3169	53.1221	0.3211	53.0649	1.5176
	Green	0.3171	53.1191	0	∞	0.3066	53.2652	1.4994
	Blue	0.3080	53.2457	0.3102	53.2150	0	∞	1.5025

5 Brute-Force Attack

In AES algorithm, there are three keys, 128 bit key, 192 bit key, and 256 bit key. To crack this key, we need 10.79×1025 years, 19.9×1045 years, and 36.71×1065 years, respectively. In the 4×4 matrix, the probability of embedding is only $12/16$. Among the nonzero pixels of 4×4 matrix, on selecting any one pixel will have three color channels. Among the three channels, we select one as indicator to decide the embedding capacity of the other two channels. The probability of choosing the indicator plane and embedding plane is 3×2 . The probability of embedding 1, 2, and 3 bit as well as not embedding is $4/16$. So the final complexity in embedding at least one bit is $2128 \times (3 \times 2) \times 1/((3/4) \times (4/16) \times (4/16) \times (4/16) \times (4/16))$.

6 Conclusion

In this work, we have proposed four pristine randomized methods and have made a comparative analysis of last three color image stego methods which are successfully simulated and results are presented. The MSE and PSNR of all the methods are also compared, and also, we have calculated the complexity of the brute-force attack and that has been displayed above. This shows how strong this work will be, and it will not be that easy for any sort of blind steganalysis attack. And of course, three major constrains of steganography viz. capacity, invisibility, and robustness are high and thus achieved.

References

1. A. Gutub, Pixel indicator technique for RGB image steganography. *J. Emerg. Technol. Web Intell.* **2**(1), 56–64 (2010)
2. B. Schneier, *Applied Cryptography Protocols, Algorithm and Source Code in C* (wiley, New Jersey, 2007)
3. C.K. Chan, L.M. Chen, Hiding data in images by simple LSB substitution. *Pattern Recogn.* **37**(3), 469–474 (2004)
4. S. Katzenbeisser, F.A.P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking* (Artech House, London, 2000)
5. L.M. Marvel, C.G. Bonchelet Jr, C.T. Retter, Spread spectrum image steganography. *IEEE Trans. Image Process.* **8**(8), 1075–1083 (1999)
6. J. Mielikainen, LSB Matching Revisited. *IEEE Sig. Process. Lett.* **13**(5), 285–287 (2006)
7. F.A.P. Petitcolas, R.J. Anderson, M.G. Kuhn, Information hiding-A survey. *Proc. IEEE* **87**(7), 1062–1078 (1999)
8. N. Provos, P. Honeyman, Hide and seek: an introduction to steganography. *IEEE Secur. Priv. Mag.* **1**(3), 32–44 (2003)
9. R. Gonzalez, R. Woods, *Digital image processing* (2002)
10. C.C. Thien, J.C. Lin, A simple and high-hiding capacity method for hiding digit-by-digit data in images based on modulus function. *Pattern Recogn.* **36**(11), 2875–2881 (2003)
11. T.-C. Lu, S.-R. Liao, P.-L. Chen, C.-C. Chang, Z.-H. Wang, Information hiding technology based on block-segmentation strategy, in *ISECS International Colloquium on Computing, Communication, Control, and Management* (2009), pp. 500–506

Secure Seed-Based Sturdy OTP via Convenient Carry-on Device

Ashok Kumar Mohan and T. Gireesh Kumar

Abstract The Internet users for the purpose of easy memorizing select a weak password and reuse it along many Web sites vulnerable to password stealing and reuse due to rapid growth of cloud computing. Sturdy one-time password (S-OTP) provides with easy remembrance and prevention of password reuse using personal Android mobile phone without the necessity of sending an SMS and can be integrated into any original user authentication system without contradicting the overall security. Also, in the worst cases, if the mobile phones are stolen, it is made unfeasible by means of two-dimensional SHA3 and MD5 forward hashing with unique and secure hard-coded seed information from mobile device.

Keywords Sturdy one-time password · Two-dimensional SHA3 and MD5 · Dynamic keypad lock · OTP · TOTP

1 Introduction

Cloud computing allows people to use computer on the Internet for all day-to-day activities, where all communication starts with user authentication. Existing solutions for traditional OTPs have been exposed to security vulnerabilities such as active eavesdropping and replay attacks. Traditional OTP suffers from a faked login Webpage that impersonates the latter to ask the service for the provider. Also, to enhance security, the users are forced to practice unfamiliar security procedures like dynamic keypad lock or RSA secure id token [1] to work.

A.K. Mohan (✉) · T. Gireesh Kumar
TIFAC CORE in Cyber Security, Amrita Vishwa Vidyapeetham University, Coimbatore,
Tamil Nadu, India
e-mail: ashokforgalaxy@gmail.com

T. Gireesh Kumar
e-mail: gireeshkumart@gmail.com

1.1 Evolution of OTP

The watchword S/Key familiarized in 1980s is one-time password valid only once and then thrown out of existence. Late in 1980, there evolved Kerberos in MIT university which maintains a set of secret keys granted through tickets generated by authentication server, one for every entity to be authenticated within a specific domain is deployed but its security completely have faith in server which has the whole set of passwords. SMS-based two-factor authentication [2] was introduced after 2000 due to the rapid increase in the usage of mobile phones. Then, to increase the randomness, time-synchronized OTP is usually related to a piece of hardware called a security token progressed. Open authentication (OATH) is a collaboration of HMAC-based scheme called as HOTP introduced in 2006 to enhance strong time-based authentication, and it released another version named as TOTP for providing custom-made enterprise solutions in 2011. Google Authenticator is a software-centered two-step authentication token technologically advanced by Google, specifically for Android mobile phones with HMAC-SHA1, which produces 80-bit secret key and yields six-digit OTP. In the next section, we will discuss about the general architecture of the system. Section 2 describes the growth of OTP, and the most recently occurring risk factors are associated with contracting security over traditional OTP. Section 3 reveals the truth that existing A-OTP via SMS has some vulnerability. Section 4 concludes with the idea to enhance this in the near future to organize S-OTP for safe and sound computing.

2 Traditional OTP

2.1 Related Works

For nearly a decade, risk has been applied to the problem of OTP for banking sectors, mail servers, and corporate login. High profile compromise and password stealing have proven that the traditional approaches such as SUAN [3], APPT [4], AOTP [5], and LumaCert [6] to OTP security are simply not sufficient, even with the durable perimeter and appropriately configured dynamic password mechanisms such as QR code [3], hash chains [7, 8], two-factor authentication [2], biometric verification [9, 10], and day-to-day sprouting hack-resistant protocols.

2.2 Cumulative Weakening of OTP

Information like OTP that should be protected is very often publicly available and revealed by careless or ignorant users. One such vulnerability in Android mobile phones exploited in the middle of July 2013, which exposes that the OTP

Fig. 1 OTP exposed in locked android phones



originating from SMS is vulnerable to phone phishing [4], shoulder surfing, and masquerading attacks shown via screenshots as shown in Fig. 1.

Certainly, when comes the mobility, many of the Gmail users have configured their mobile phone number for their account password recovery options mostly via Android-based smartphones [2], providing these services anytime and anywhere. Consumer can reset the account password by merely requesting Google to send a verification code on the preregistered mobile number. Now, consider that the phone screen is locked using pattern lock, PIN lock, or password lock highlighted as (1) in Fig. 1. But, the flaw, we are discoursing here, permits SMS content (OTP, in our case) to be displayed on the one-line notification panel at the top of the mobile display which can be read by anyone tinted as (2) and (3) in Fig. 1, even if the mobile phone is in security lock mode. The issue has been recognized and conveyed by the RnD Lab at Varutra Consulting, an information security consulting and training services company in Pune, and later, in July 2013, the same was alerted by ‘thehackernews’ Web site.

3 Sturdy OTP

As the name implies ‘Sturdy’ means robust, durable, strong, and secure factors that influence the user to lower their personal risk level using S-TOP compared to earlier flavors of OTP. The proposed OTP scheme is released as two versions namely S-OTP version 1.0 to send the SMS in the reverse direction to that of existing traditional OTP systems as in AOTP [5] with the addition of forward hashing functions [7] as described in this project as S-OTP version 2.0.

3.1 Framework Description

We propose a connectionless authentication mechanism and the ultimate intention of S-OTP to communicate without the need of an SMS with secure hash functions.

Registration Phase: Seed information is a combination of IMSI, IMEI, and registration time stamp [1]. IMEI and IMSI are extracted from the mobile device by the Android application, while the registration time stamp is the time in which the firmware is loaded into the mobile. So uncertainty if the mobile is stolen and if the same is used to authenticate S-OTP, the connection is refused as in Fig. 2.

IMEI, IMSI, and Time stamp (TS): IMEI and IMSI are general information, while time stamp is the time in which the Android firmware was installed by the legitimate user or manufacturer.

Login Phase: The combination of SHA3 (A) and MD5 (B) in a sequential order [7] over a two-dimensional field is shown in Fig. 3. Hash function is used to generate S-OTP; likewise, signature is created. Two hash functions are used as forward hash function in two different directions resulting in ${}^hB_y({}^hA_x(S_t^{OTP}))$ to provide complete anonymity over backtracking.

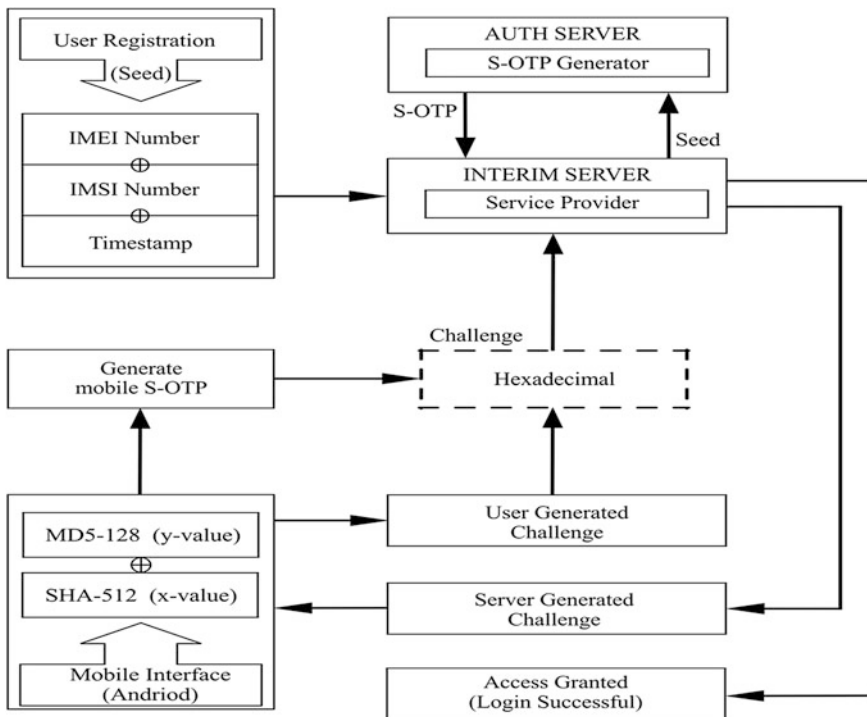
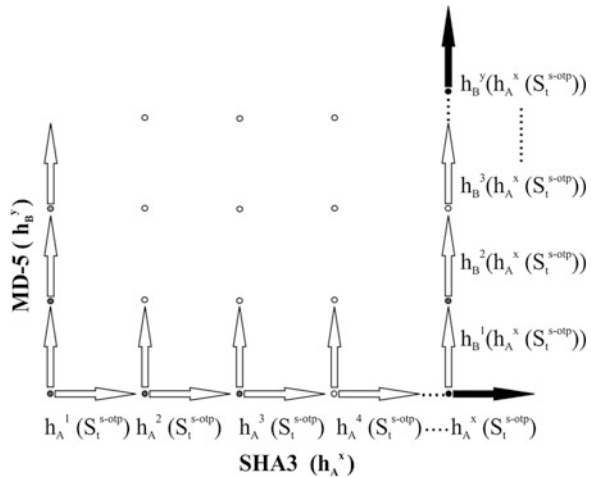


Fig. 2 S-OTP architecture

Fig. 3 Forward nested hash chains



3.2 Project Outgrowth (S-OTP Version 2.0)

To ensure that S-OTP remains assertive, it is important to follow the sequential flow of steps prescribed below and to maintain the state of security by providing the challenge response from server to mobile and vice versa, only as and when required to authenticate each other.

Note: The notations used below (1a, 1b... 4d) are as displayed in Fig. 4.

Registration Phase

- Step 1a: The mobile S-OTP (.apk) file extracts the IMEI, IMSI, and TS value (time stamp of the firmware installation, here it is Android version 4.0.3) and displays to the user.
- Step 1b: Then, the user visits the Banking Server with the help of the bank employee and registers their corresponding seed value (IMEI, IMSI, TS) to acquire/update their account number.
- Step 1c: All the above-mentioned details are stored in the Auth Server.
- Step 2: With the registered account number, the user browses the NEW USER page in bank Web site and creates the static username and password.

Login Phase

- Step 3a: The user mobile has USER PASSWORD button to create the initial OTP1 and corresponding X1 and Y1 pairs.
- Step 3b: Now, the user is redirected to the client authentication page where the user enters the generated OTP1, X1, and Y1.

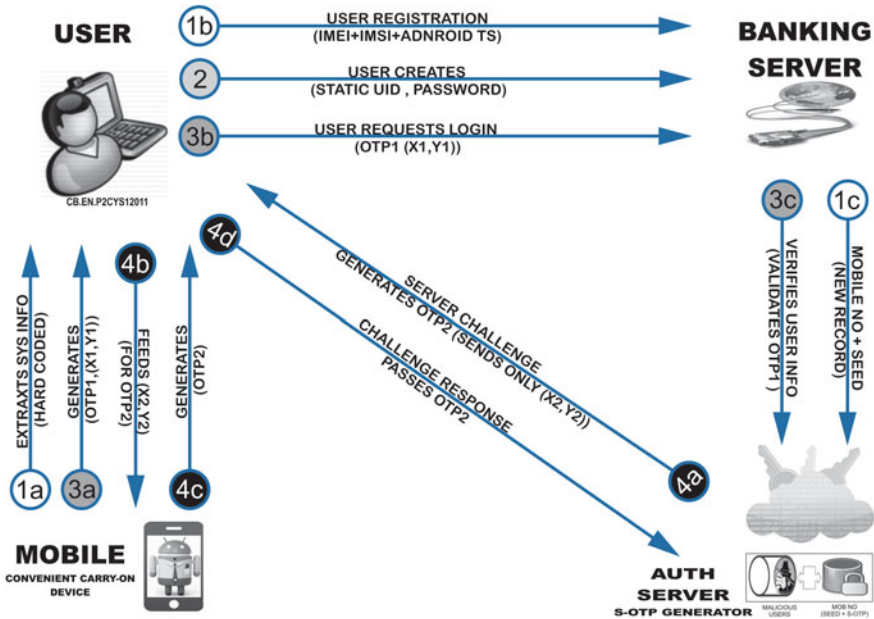


Fig. 4 Operational flow of S-OTP

Step 3c: These are validated with the user mobile number, and then, the corresponding OTP1 is generated in Auth Server and confirmed with OTP1.

Server Challenge Response

- Step 4a: The Auth Server generates OTP2 with respect to a new random X2, Y2 pair and sends only X2, Y2 to mobile device as challenge.
- Step 4b: The user clicks SERVER CHALLENGES button and enters X2, Y2 pair.
- Step 4c: As a response, the mobile generates the exact OTP2 and displays it.
- Step 4d: Finally, the user types the OTP2 in the bank Webpage which is examined by the Auth Server and login is provided if the OTP2 matches with previously stored OTP as shown in step (4a) (Fig. 5).

3.3 Putting into Practice

Result Analysis:

The comparison of S-OTP with the other flavors of OTP and proven to be the effective, cost-effective and easily deployable as shown in Table 1.

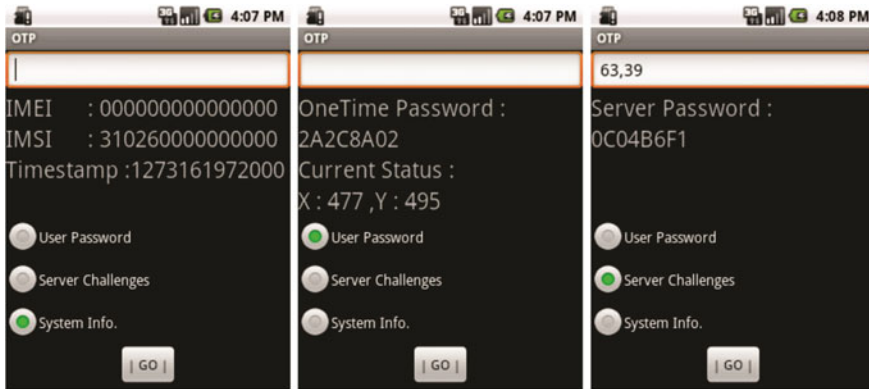


Fig. 5 Output of S-OTP

Table 1 All-time OTP comparisons

Type	Sturdy OTP v2.0	Sturdy OTP v1.0	Active OTP	Secure ID	SUAN via (QR code)	S/KEY
SMS/MMS	No (a)	Yes (b)	Yes (b)	No (a)	Yes (c)	No (a)
Additional hardware	No (a)	No (a)	No (a)	Yes (d)	Yes (e)	No (a)
Time restriction	No (a)	No (a)	Yes (f)	Yes (f)	Yes (f)	Yes (f)
Multiple encryption	Yes (g)	Yes (g)	No (h)	No (h)	No (h)	No (h)
Encryption algorithm	Yes (i)	Yes (i)	No (i)	Yes (k)	Yes (k)	No (j)
Attack during user authentication	No (l)	No (l)	Yes (m)	Yes (n)	Yes (o)	Yes (p)
Repudiation	No (q)	No (q)	Yes (r)	Yes (r)	Yes (r)	Yes (r)

Note The notations used below (a, b, c... r) are as represented in Table 1

It is not needed/not applicable for the given factor (a).

It needs to send OTP via SMS (b).

It needs to send QR code through MMS (c).

It requires hardware token like Best Buy's BestToken and RSA's Secure ID (d).

Here, camera is mandatory to capture user image (e).

It is most probably limited from 10 to 15 s (f).

It is achieved via two-dimensional forward hashing in two different directions (g).

It is not achieved (h).

Two hash functions SHA3 and MD5 are used one after another as parallel connections (i).

It is not applicable (j).

Only, simple hashing and RSA are employed (k).

No attacks are possible as both mobile and server are mutually authenticated in a unique approach (l).

All social engineering attacks such as shoulder surfing are proven to attack AOTP (m).

Mostly damage, loss, failure, and tampering of the device are possible (n).

It can be spoofed by prerecorded user image samples (o).

Session hijacking, eavesdropper, and replay attacks are most common in S/Key (p).

Both the mobile and server are authenticated individually; hence, user cannot repudiate their transaction (q).

It is possible for end user to claim not sending it or reporting absence of user device (r).

4 Conclusion and Future Works

Gone are the days when only banking sector and mailing agents needed to worry about securing their OTP. In today's highly competitive business environment, companies from all different sectors and size need to use OTP to authenticate their client user-friendly with most commonly available resources with them. The most needed in the current debate over OTP is a move away from fear-based doomsday thinking and a move toward more level-headed threat valuations that take into account the strategic agenda toward the secured practice of S-OTP. The biggest things that resulted from the comments and feedback received had to do with replacing the initial IMEI, IMSI, and TS with some simple keywords to be more user-friendly. After final testing, the S-OTP will be deployed in Google Play Store to get some responses and feedback to improve its security and make it effective to be proposed to some OTP security divisions in banking sector.

References

1. S. Indu, T.N. Sathya, V. Saravana Kumar, A stand-alone and SMS-based approach for authentication using mobile phone: *IEEE Trans.* (2013)
2. J.-Y. Hu, C.-C. Sueng et al., Android-based mobile payment service protected by 3-factor authentication and virtual private ad hoc networking, pp. 111–116 (2012)
3. Y.-W. Kao, G.-H. Luo, H.-T. Lin et al., Physical access control based on QR code, in *IEEE. 2011 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (2011), pp. 285–288
4. A.A. Khan, Preventing phishing attacks using one time password and user machine identification. *Int. J. Comput. Appl.* (2013)

5. C.-I. Fan, C.-N. Wu et al., *Active One-Time Password Mechanism for User Authentication* (Springer, Berlin, 2013), pp. 464–471
6. Artan Luma, Betim Prevalla et al., LumaCert: conception and creation of new digital certificate for online user authentication in e-Banking systems. *World Acad. Sci. Eng. Technol.* **78**, 1333–1391 (2013)
7. M.H. Eldefrawy, M.K. Khan, K. Alghathbar, *One-Time Password System with Infinite Nested Hash Chains*, vol. 122, CCIS (Springer, Berlin, 2010), pp. 161–170
8. V.P. Thakur, K.N. Hande, Hash based dynamic password authentication mechanism for kerberos environment. *IJERT*, www.ijert.org, pp. 2278–0181 (2013)
9. W. Jang, S. Cho et al., *User-Oriented Pseudo Biometric Image Based One-Time Password Mechanism on Smart Phone*, vol. 199 (Springer, Berlin, 2011), pp. 49–58
10. X. Liu, Y. Shen et al., A fingerprint-based user authentication protocol with one-time password for wireless sensor networks (2013)

A (t, n) Secure Sum Multiparty Computation Protocol Using Multivariate Polynomial Secret Sharing Scheme

K. Praveen and Nithin Sasi

Abstract A (t, n) threshold scheme is a method for sharing a secret among n shareholders so that the collaboration of at least t shareholders is required in order to reconstruct the shared secret. In this paper, we propose a (t, n) secure sum multiparty computation protocol using multivariate polynomial secret sharing scheme. In this scheme, any t or more shareholders acting in collusion can reconstruct the secret, but a particular shareholder's information is not revealed to other shareholders. This scheme can be applied for authenticating a selected single group of t participants securely without revealing their shares.

Keywords Multiparty computation · Secret sharing · Multivariate linear polynomial · Group authentication

1 Introduction

Secret sharing is a cryptographic method for distributing a secret among a group of n participants. The participants have to combine together their shares to reconstruct the original secret. Secret sharing was introduced by Shamir [1] and Blakley [2] independently in 1979. They introduced a (t, n) secret sharing scheme whereby at least t of them is required to cooperate to reconstruct the secret. Beimel [3] presented a survey of secret sharing schemes. He has described a number of secret sharing schemes based on various properties and application requirement. Shen et al. [4] describe a linear secret sharing scheme using multivariate polynomial for generating a group key. But because the reconstruction of group keys is not made

K. Praveen · N. Sasi (✉)
TIFAC-CORE in Cyber Security, Amrita Vishwa Vidyapeetham Coimbatore,
Coimbatore, Tamil Nadu, India
e-mail: nithinsasi@gmail.com

K. Praveen
e-mail: praveen.cys@gmail.com

private, the group key generated in the scheme [4] cannot be reused for multiple times. The goal of multiparty computation scheme is to create methods that enable parties to jointly compute a function over their inputs, while at the same time keeping these inputs private. Oded Goldreich has provided a working draft about secure multiparty computation (SMC) [5] in which he has described extensively about SMC. A distributed secure sum protocol for SMC was proposed by Sheikh et al. [6] in 2010.

In this paper, we propose a (t, n) secure sum multiparty computation protocol using multivariate polynomial secret sharing scheme. In this scheme, any t or more shareholders acting in collusion can reconstruct the secret, but a particular shareholder’s information is not revealed to other shareholders. In 2013, Lein Harn proposed a group authentication scheme [7]. Our scheme can also be used for group authentication.

2 Preliminaries

2.1 A Multivariate Linear Polynomial Secret Sharing Scheme

In this section, we describe a multivariate linear polynomial secret sharing scheme [12]. We give the different phases of work as follows:

1. Select parameters
 - (a) The group selects the group public parameter m , where m is a larger integer.
 - (b) The group selects a $n \times t$ matrix.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1t} \\ \dots & \dots & \dots & \dots \\ x_{t1} & x_{t2} & \dots & x_{tt} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nt} \end{bmatrix}, x_{ij} \in \mathbb{Z}_m, 1 \leq i \leq n, 1 \leq j \leq t.$$

For any $t \times t$ submatrix X_t of the matrix X , it satisfies $\gcd(\det X_t, m) = 1$.

2. Determine a threshold function and a secret key
 - (a) The group selects a function y such that,

$$y = f(x_1, x_2, \dots, x_t) = a_1x_1 + a_2x_2 + \dots + a_tx_t$$

$$y = f(x_1, x_2, \dots, x_t) \in \mathbb{Z}_m[x_1, x_2, \dots, x_t]$$

We know it is a threshold function. Suppose the secret key is k ($k \in \mathbb{Z}_m$); we can write

$$k = f(1, 1, 1, \dots, 1) = \sum_{i=1}^t a_i$$

3. Determine the secret shares

- (a) The group computes secret shares y_i for each user $i(1 \leq i \leq n)$ as

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{it}) = a_1x_{i1} + a_2x_{i2} + \dots + a_t x_{it} \pmod m, (1 \leq i \leq n)$$

- (b) The group sends $(y_i; X_i)(1 \leq i \leq n)$ to each user $i(1 \leq i \leq n)$ in a secret manner, where $X_i = (X_{i1}, x_{i2}, \dots, x_{it})$

4. Note of selecting the matrix X

From Sect. 2 (1), (b), any $t \times t$ submatrix X_t of a matrix X must satisfy $\gcd(\det X_t, m) = 1$, so we should compute any $t \times t$ submatrix X_t of the matrix X when we select a matrix X in general. However, we should do more computations about determinants of the submatrix for larger integer's t and m . Therefore, while selecting a matrix X , we should notice the calculation of these determinants. For example, we can select the following $n \times t(n > t, t > 2)$ matrix X as

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{t-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{t-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n-1} & x_{n-1}^2 & \dots & x_{n-1}^{t-1} \\ 1 & x_n & x_n^2 & \dots & x_n^{t-1} \end{bmatrix} \tag{1}$$

For any $t \times t$ submatrix of the above matrix, the value of its determinant can be calculated easily. In fact, let matrix X_i be $t \times t$ submatrix of the matrix X ,

$$X_i = \begin{bmatrix} 1 & x_{i_1} & x_{i_1}^2 & \dots & x_{i_1}^{t-1} \\ 1 & x_{i_2} & x_{i_2}^2 & \dots & x_{i_2}^{t-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{i_{t-1}} & x_{i_{t-1}}^2 & \dots & x_{i_{t-1}}^{t-1} \\ 1 & x_{i_t} & x_{i_t}^2 & \dots & x_{i_t}^{t-1} \end{bmatrix}$$

where $i_1 < i_2 \dots < i_t$, and $i_1, i_2, \dots, i_t \in \{1, 2, \dots, n\}$. By Vander monde determinant [14], we know

$$\det X_i = \begin{bmatrix} 1 & x_{i_1} & x_{i_1}^2 & \dots & x_{i_1}^{t-1} \\ 1 & x_{i_2} & x_{i_2}^2 & \dots & x_{i_2}^{t-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{i_{t-1}} & x_{i_{t-1}}^2 & \dots & x_{i_{t-1}}^{t-1} \\ 1 & x_{i_t} & x_{i_t}^2 & \dots & x_{i_t}^{t-1} \end{bmatrix} = \prod_{1 \leq k < j \leq t} (x_{ij} - x_{ik}) \quad (2)$$

So, for any $t \times t$ submatrix X_t of the matrix X , by Eq. (2), if $\gcd((x_{ij} - x_{ik}), m) = 1, 1 < k < j < t$, then $\prod_{1 \leq k < j \leq t} \gcd((x_{ij} - x_{ik}), m) = 1$. So $\gcd(\det X_t, m) = 1$. Therefore, when we select a matrix of the form (1), we only notice the second row elements of the matrix satisfied $\gcd((x_i - x_j), m) = 1, 1 < j < i < n$.

(5) Generate a group key

In this (t, n) threshold scheme, any subgroup of t or more shareholders of the designated group can generate a valid group key. If any t shareholders say 1, 2, ..., t , act in collusion, every one of them contributes his/her share $(y_i; X_i)(1 \leq i \leq t)$. Then, through the congruence

$$\begin{bmatrix} x_{11} & x_{21} & \dots & x_{1t} \\ x_{21} & x_{22} & \dots & x_{2t} \\ \dots & \dots & \dots & \dots \\ x_{t1} & x_{t2} & \dots & x_{tt} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_t \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} \pmod m \quad (3)$$

They can compute $(a_1, a_2, \dots, a_t) \pmod m$. By the Cramer’s rule [8], the solution $(a_1, a_2, \dots, a_t) \pmod m$ is unique. So they can generate a valid group key $k = f(1, 1, 1, \dots, 1) = \sum_{i=1}^t a_i$ as

$$y = f(x_1, x_2, \dots, x_t) = a_1x_1 + a_2x_2 + \dots + a_t x_t$$

But any subgroup $t_1(t_1 < t)$ or less shareholders of the designated group cannot generate a valid group key. From reference [9], the solution $(a_1, a_2, \dots, a_t) \pmod m$ of the congruence (3) is not unique. So they cannot generate a valid group key.

3 Proposed System

3.1 (t, n) Threshold Secure Sum Secret Sharing Scheme Using Multivariate Polynomial and Multiparty Computation

The proposed scheme in this paper is carried out between a dealer D and a set of participants holding shares $P = \{P_1, P_2, P_3, \dots, P_n\}$. We assume that the communication channel between D and P provides security and data integrity for the retrieval of information.

Our basic idea to realize this scheme can be summarized as follows:

- (1) Initialization by dealer
 - (a) Dealer D initiates the protocol by randomly choosing a secret S . Dealer then divides S into t parts $K_1, K_2, K_3, \dots, K_t$ where $S = \sum_{i=1}^t K_i$ and t is the threshold.
 - (b) Dealer applies the multivariate linear polynomial secret sharing scheme in Sect. 2.1 for each of $K_1, K_2, K_3, \dots, K_t$.
 - (c) Shares of K_1 are $P_{1i}(1 \leq i \leq n)$ and those of K_2 are $P_{2i}(1 \leq i \leq n), \dots, K_t$ and $P_{ti}(1 \leq i \leq n)$.
- (2) Sending of generated shares to the participants
 - (a) Dealer sends $(P_{1i}, P_{2i}, \dots, P_{ti}; X_i)(1 \leq i \leq n)$ to each user $P_i(1 \leq i \leq n)$ in a secret manner, where $X_i = (x_{i1}, x_{i2}, \dots, x_{it})$.
- (3) Generating keys
 - (a) All P_i will send P_{ji} to $P_j(1 \leq j \leq t; j \neq i)(1 \leq i \leq t)$.
 - (b) Using the reconstruction procedure given in Sect. 2.1, if any t participants combine, P_i reconstructs K_i .
- (4) Finding S

Two methods can be followed:

Method 1:

- (a) P_1 adds a random number r_1 to K_1 and sends it to P_2 . P_2 adds a random number r_2 and K_2 to the received value from P_1 and sends it to P_3 . Process goes on till P_t adds a random number r_t and K_t to received value from P_{t-1} and sends it to P_1 . P_1 subtracts r_1 from the received value from P_t and sends back the result to P_2 . P_2 subtracts r_2 and sends back the result to P_3 . Process goes on till P_t subtracts r_t . Finally, P_t gets S . P_t can now transmit S to P_1, P_2, \dots, P_{t-1} .

Method 2:

- (a) Each party P_i breaks its data K_i into t segments $d_{i1}, d_{i2}, \dots, d_{it}$ such that $\sum d_{ij} = K_i$ for $j = 1$ to t .
- (b) Each party keeps any one segment with it and distributes $t - 1$ segments to other parties.
- (c) Each party reshuffles the received segments randomly.
- (d) Assume $c = t$ and $S_{ij} = 0$, where S_{ij} is the partial sum and c is a counter.
- (e) while $c \neq 0$
 - begin
 - for $j = 0$ to $t - 1$
 - for $i = 0$ to $t - 1$
 - P_i sends $S_{ij} = S_{ij} + d_{ij}$ to $P_{(i+1) \bmod t}$
 - $c = c + 1$

end

- (f) The protocol initiator party broadcasts sum S_{ij} to all the parties.
- (g) Here, $S = S_{ij}$

We can see that in this scheme the shares are not revealed even after the reconstruction of the secret. The disadvantage of this scheme is that the shares are not verifiable and the scheme cannot detect cheating parties.

4 Application

Finally, we present an application of our general-purpose construction, namely for group authentication. The group authentication can be used to determine whether all users belong to the same group or not. We have used a many-to-many type of authentication. Consider we have t users, $P_i, i = 1, 2, \dots, t$, participating in a group-oriented application. These users want to make sure whether they all belong to the same group of n group members at the beginning of the application. Initially, the group manager (*GM*) registers all group members. During registration, the *GM* uses our secret sharing scheme to issue a private token to each group member. Later, all users participate in the group authentication work to authenticate each other. There are two possible outcomes of the group authentication: Either all users are member of the same group or they are non-members of the group. The group authentication is sufficient to check whether all users are members of a group. However, if there are non-members, it can be used as a preprocess before applying conventional user authentication to identify non-members.

5 Conclusion

In this paper, we have presented a secret sharing based on multivariate linear polynomials and SMC. The scheme introduced is a (t, n) secret sharing scheme where secrets are shared among n shareholders whereby at least t of them are required to cooperate before the secret can be reproduced. Until t individuals act in collusion, they will get no information about the secret. Our proposed scheme is efficient since it is based on (t, n) multivariate linear polynomials secret share scheme and SMC and also because the computations involve only polynomial operations. We have applied our scheme to group authentication, which can authenticate multiple users at once. We have shown how our scheme could be used in (t, n) group authentication scenario. The application of this scheme could also be extended to signature verification, some of the electronic voting protocols, e-auctions, etc.

References

1. A. Shamir, How to share a secret. *Comm. ACM.* **22**(11), 612–613 (1979)
2. G.R. Blakley.: Safeguarding cryptographic keys, in *Proceedings of the AFIPS National Computer Conference*, vol. 48 (1979), pp. 313–317
3. A. Beimel, Secret-Sharing Schemes: A Survey. *Coding Cryptology* **6639**, 11–46 (2011)
4. Z. Shen, X. Yu.: A multivariate linear polynomial secret share scheme and its applications. *J. Computat. Inf. Syst.* **7**(3), 904–915 (2011) (Springer, Heidelberg)
5. O. Goldreich, Secure Multi-Party Computation (Working Draft). (1998)
6. R. Sheikh, B. Kumar, D.K. Mishra, A distributed k -secure sum protocol for secure multi-party computations. *J. Comput.* **2**(3), 2151–9617 (2010)
7. L. Harn, Group authentication. *IEEE Trans. Comput.* **62**(9) (2013)
8. W. Bin, P.H. Dong, L.J. Hua, A secure (t, n) threshold signature scheme. *J. Shanghai Jiaotong Univ.* **36**(9), 1333–1336 (2002)
9. A. Shamir, A polynomial time algorithm for breaking the basic Merkle-Hellman Cryptosystem, in *Proceeding of the 23 IEEE Symposium Found on Computer Science* (1982), pp. 142–152
10. B. Zhao, Secret sharing in the encrypted domain with secure comparison. Global Telecommunications Conference (GLOBECOM 2011) (2011)

Small-World Particle Swarm Optimizer for Real-World Optimization Problems

Megha Vora and T.T. Mirnalinee

Abstract Many real-world problems from different domains, viz. engineering, data mining, biology, can be formulated as the optimization of a continuous function. These problems require the estimation of a set of model parameters or state variables that provide the best possible solution to a predefined cost or objective function, or a set of optimal trade-off values in the case of two or more conflicting objectives. Locating global optimal solutions becomes challenging especially in the presence of high dimensionality, nonlinear parameter interaction, insensitivity, and multi-modality of the objective function. These conditions make it very difficult for any search algorithm to find high-quality solutions quickly without getting stuck in local optima. Unfortunately, these difficulties are frequently encountered in real-world optimization problems when traversing the search space en route to the global optimum. Small-world PSO has been proven to be effective in solving global function optimization problems. After all, every optimization algorithm has to be applied to some real-world problems. This paper evaluates the performance of small-world PSO algorithm on two real-world function optimization problems. Comparative study with state of the art demonstrates the effectiveness of small-world PSO.

Keywords Small-world PSO · Frequency modulation · Genetic algorithm with a new multi-parent crossover · SLPSO · SWPSO-I algorithm

M. Vora (✉) · T.T. Mirnalinee
Department of Computer Science and Engineering, SSN College of Engineering, Anna University, Chennai, India
e-mail: meghavora25@gmail.com

T.T. Mirnalinee
e-mail: mirnalineeett@ssn.edu.in

1 Introduction

Small-world concept was originally proposed by Milgram [1]. It was stated that social network exhibits small-world phenomenon in which any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances. To demonstrate the universality of this phenomenon in network arising in nature and technology, Milgram et al. in 1960s [1, 2] performed a series of striking experiments and reported their results. This was later modeled by Watts and Strogatz [3]. However, the model was insufficient to explain the striking algorithmic component of Milgram's original findings that how the individuals using local information are collectively effective at actually constructing a short path between two points in a social network. Kleinberg [4] further generalized Watts–Strogatz model and showed that there is a decentralized algorithm capable of finding short paths with high probability. In [5], small-world particle swarm optimization (SW-PSO) algorithm was proposed where small-world concept given by Jon Kleinberg was incorporated in Von Neumann architecture and on the newly obtained network topology particle swarm optimization algorithm was applied. The algorithm was applied on the four well-known benchmark functions; comparative study was performed with other PSO variants and was proven better. Later, in [6] SWPSO-I algorithm was proposed which took forward the concept of SW-PSO. Different from the work in [5] where inertia weight parameter was set constant, in [6], linearly decreasing inertia weight had been used. This led to a more balanced exploration–exploitation trade-off. Moreover, particle's velocity and position was updated at individual level rather than at population level to make it adaptive. SWPSO-I was applied on 12 benchmark functions, and comparative study with the other PSO variants was performed and was proven better.

In this paper, effectiveness of SWPSO-I [6] for solving real-world parameter optimization is evaluated. Many real-world problems from different domains, viz. engineering, data mining, biology, can be formulated as the optimization of a continuous function. These problems require the estimation of a set of model parameters or state variables that provide the best possible solution to a predefined cost or objective function, or a set of optimal trade-off values in the case of two or more conflicting objectives. Locating global optimal solutions becomes challenging especially in the presence of high dimensionality, nonlinear parameter interaction, insensitivity, and multi-modality of the objective function. These conditions make it very difficult for any search algorithm to find high-quality solutions quickly without getting stuck in local optima. Unfortunately, these difficulties are frequently encountered in real-world optimization problems when traversing the search space en route to the global optimum.

2 Small-World PSO for Real-World Function Optimization

As mentioned in the introduction, small-world network has a distinctive combination of highly clustered and small characteristic path length network. This leads to a balanced trade-off between exploration and exploitation. Here, small-world network topology on Von Neumann's network model is considered. This is achieved by adding few random particles (in this case two) and retaining four immediate Von Neumann neighbors of a current particle (see Figs. 1 and 2). These random particles are treated as additional neighbors of current particle.

The intuitive advantage of these additional random neighbors is the way small world is formed within the current swarm. Each particle \vec{x}_k in the swarm represents the candidate solution to the considered problem. Particles travel in the solution space and attempt to move toward a better solution by changing its direction and speed based on its own past experience P_{best} and the experience of the particles in its small-world neighbor SWN_{best} . Velocity v_k and position \vec{X}_k of the particle are updated using Eqs. 1 and 2, respectively.

Fig. 1 A swarm of particles 6×6

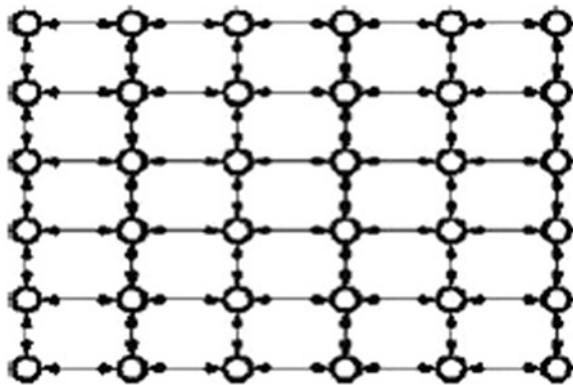


Fig. 2 Small world with 2 random particles

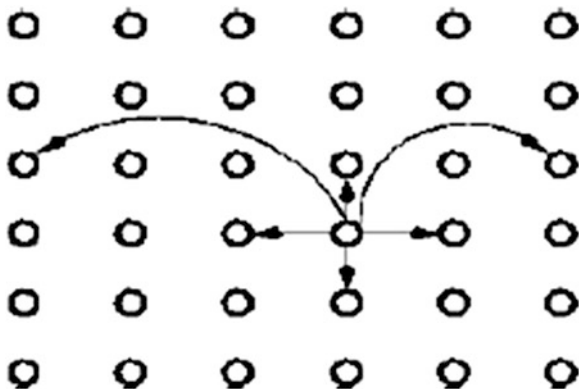
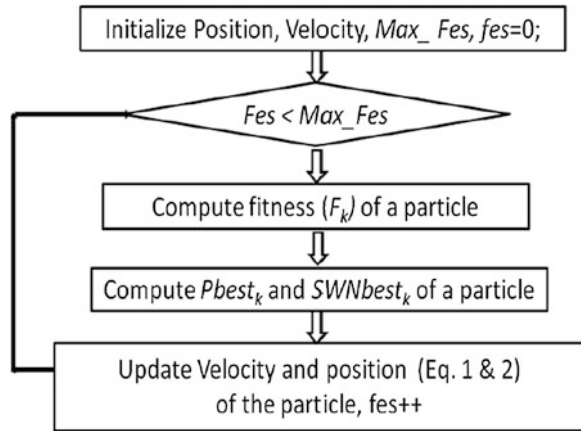


Fig. 3 Flowchart of SWPSO-I



$$\vec{v}'_k = \omega \cdot \vec{v}_k + C_1 R_1 (X_{pbest_k} - \vec{X}_k) + C_2 R_2 (X_{gbest_k} - \vec{X}_k) \quad (1)$$

$$\vec{X}'_k = \vec{X}_k + v'_k \quad (2)$$

where ω is inertia weight; C_1 and C_2 are acceleration constants; R_1 and R_2 are random numbers uniformly generated from the interval $[0, 1]$.

Figure 3 depicts the flowchart of the proposed algorithm. Max_Fes is the maximum fitness evaluation.

3 Experimental Results

3.1 Description of Data Sets

In order to test the effectiveness of SWPSO-I on real-world application, two real-life problems are chosen: design of a gear train [7] and parameter estimation for frequency modulation (FM) sound waves [8]. The first problem involves optimization of the gear ratio for a compound gear train that contains four gears. The gear ratio for gear train is defined as the ratio of the angular velocity of the output shaft to that of the input shaft. It is to be designed that the gear ratio is as close as possible to $1/6.931$. For each gear, the number of teeth must be between 12 and 60. The mathematical model of this problem can be described as follows:

$$f(x) = \left(\frac{1}{6.931} - \frac{x_1 * x_2}{x_3 * x_4} \right)^2 \quad (3)$$

where $x_i \in [12, 60]$, $i = 1, 2, 3, 4$.

The second problem is to estimate the parameters of an FM synthesizer [8]. It has important role in several modern music systems. It is a six-dimensional optimization problem where the vector to be optimize is $X = [a_1, \omega_1, a_2, \omega_2, a_3, \omega_3]$. The expression of the estimated sound wave is given by the following equation:

$$y(t) = a_1 * \sin(\omega_1 * t * \theta + a_2 * \sin(\omega_2 * t * \theta + a_3 * \sin(\omega_3 * t * \theta))) \quad (4)$$

and the expression of the target sound waves is given by

$$y_0(t) = 1 * \sin(5 * t * \theta + 1.5 * \sin(4.8 * t * \theta + 2 * \sin(4.9 * t * \theta))) \quad (5)$$

where $\theta = 2 * \pi/100$ and the parameters are constrained in $[-6.4, 6.35]$. The problem is to generate the estimated sound similar to the target sound. Thus, fitness function is the summation of square errors between the estimated wave and the target wave. It is defined as follows:

$$f(\vec{x}) = \sum_{t=0}^{100} (y(t) - y_0(t))^2 \quad (6)$$

3.2 Results

Table 1 shows the mean, standard deviation, and best (minimum) and worst (maximum) values obtained over 30 independent runs. Each run of the algorithm is terminated when the number of function evaluations (*Max Fes*) exceeds $3e+04$. Simulations were done considering population size of 20. Acceleration coefficient parameters C_1 and C_2 are set to 1.494. These values were chosen to ensure good convergence [12]. Inertia weight parameter is varied linearly from 0.9 to 0.4 according to the Eq. 7 for both SWPSO-I and SLPSO, while it is set to 0.72 for SWPSO.

$$\omega(fes) = ub - (lb + 0.1 * fes) / T_Fes \quad (7)$$

The larger step size at the beginning offers a better exploration and the smaller step size toward the end offers better exploitation, thus decreasing chances of missing a global optima. For SWPSO and SWPSO-I, two small-world randomized particles were chosen. Experiments with different numbers (1, 2, and 3) of randomized particles showed that using two randomized particles, convergence to optimal solution was faster.

The results obtained by the proposed SWPSO-I algorithm are compared with those obtained by previously proposed SWPSO [5] algorithm. Beside this, it is also compared with the latest version of standard particle swarm optimization algorithm, namely SPSO-11 [9] and self-learning particle swarm optimizer (SLPSO) [10] where each particle has a set of four strategies to cope with different situations. A particle can choose an appropriate strategy at any instance according to the property

Table 1 Comparison on two real-world problems

Algorithm	Train gear				FM parameter estimation			
	Mean	Std. dev	Best	Worst	Mean	Std. dev	Best	Worst
SWPSO-I	0.0000	0.0000	0.0000	0.0000	8.74	7.172	0.0000	18.66
SWPSO [5]	0.0000	0.0000	0.0000	0.0000	13.02	6.759	0.0000	21.1
SPSO-11 [9]	6.06e-11	2.47e-10	2.70e-12	1.36e-9	17.6	5.55	5.77e-19	23.7
SLPSO [10]	1.62e-9	8.69e-9	2.31e-11	6.51e-9	6.96	31.28	7.25e-27	15.14
GAMPC [11]	2.43e-4	5.63e-4	3.18e-21	1.96e-3	16.56	1.19e-14	16.56	16.56

of its local search space, and finally, it is also compared with a variant of genetic algorithm with a new multi-parent crossover (GAMPC) for solving a variety of optimization problems. This algorithm uses both a randomized operator as mutation and maintains an archive of good solutions [11]. These results are depicted in Table 1. Results in the bold indicate the best one.

It is seen from Table 1 that for train gear problem, both SWPSO and SWPSO-I outperform other state-of-the-art algorithms, while for parameter estimation for frequency modulation (FM) sound waves, it performs second best in terms of mean result after SLPSO [10]. However, it should be noted that SLPSO fails to achieve the optimal result in any of the iterations, whereas SWPSO-I does. Also, the standard deviation of SLPSO is very high compared to SWPSO-I. Thus, the overall performance of SWPSO-I seems to be better than the others.

4 Conclusions

In this paper, performance of SWPSO-I algorithm for solving real-world optimization problem is evaluated. Experiments were carried out on two real-world problems. Mean, standard deviation, and best and worst results obtained over 30 runs are reported. Comparative study of the proposed algorithm with other state-of-the-art techniques justifies the efficiency of the proposed SWPSO-I algorithm for solving real-world problem.

References

1. S. Milgram, The small world problem. *Psychol. Today* **2**, 60–67 (1967)
2. J. Travers, S. Milgram, An experimental study of the small world problem. *Sociometry* **32**, 425 (1969)
3. D. Watts, S. Strogatz, Collective dynamics of small-world networks. *Nature* **393**, 440–442 (1998)
4. J. Kleinberg, The small-world phenomenon: an algorithmic perspective. Technical report, Cornell University Ithaca (1999)
5. A.K. Saxena, M. Vora, Novel approach for the use of small world theory in particle swarm optimization, in *16th International Conference on Advanced Computing and Communications*, IEEE (2008), pp. 363–366
6. M. Vora, T.T. Mirmalinee, Small world particle swarm optimization for global function optimization, in *Pattern Recognition and Machine Intelligence*, vol. 8251, ed. by P. Maji, A. Ghosh, N.M. Murty, K. Ghosh, S.K. Pal (Springer, Heidelberg, 2013), pp. 575–580
7. E. Sandgen, Nonlinear integer and discrete programming in mechanical design optimization. *J. Mech. Des. (ASME)* **112**, 223–229 (1990)
8. S. Das, P.N. Suganthan, Problem definitions and evaluation criteria for CEC 2011 competition on testing evolutionary algorithms on real world optimization problems. Technical report (2010)

9. M. Zambrano-Bigiarini, R. Rojas, M. Clerc, Standard particle swarm optimisation 2011 at CEC-2013: a baseline for future PSO improvements, in *Congress on Evolutionary Computation (CEC)*, IEEE (2013)
10. C. Li, S. Yang, T. Nguyen, A self-learning particle swarm optimizer for global optimization problems. *IEEE Trans. Syst. Man Cybern. Part B* **42**(3), 627–646 (2012)
11. S. Elsayed, R. Sarker, D. Essam, GA with a new multi-parent crossover for solving IEEE-CEC2011 competition problems, in *Congress on Evolutionary Computation (CEC)* (2011), pp. 1034–1040

A Comparative Study of Feature Ranking Methods in Recognition of Handwritten Numerals

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

Abstract Feature selection is an important task during classification of any pattern. In this paper, we compute and compare the strengths of five most widely used feature ranking techniques in identifying the optimal subset of features for best classification results. The feature ranking measurements that are used here are information gain (IG), gain ratio (GR), correlation, symmetrical uncertainty (SU), and chi-square (CS). For evaluation purpose, recognition of handwritten numeral samples from five popular Indic scripts—Bangla, Hindi, English, Telugu, and Arabic—are used. These ranking methods are applied over quadtree-based longest run feature set. Experimental results are drawn and compared using support vector machine (SVM)-based classifier.

Keywords Feature selection · Handwritten numeral recognition · Filters

A. Roy (✉) · N. Das · A. Saha · R. Sarkar · S. Basu · M. Kundu · M. Nasipuri
Department of Computer Science and Engineering, Jadavpur University,
Kolkata 700032, India
e-mail: abhinabaroy1990@gmail.com

N. Das
e-mail: nibaran@gmail.com

A. Saha
e-mail: iamitcse@gmail.com

R. Sarkar
e-mail: ramsarkar@gmail.com

S. Basu
e-mail: bsubhadip@gmail.com

M. Kundu
e-mail: mahantapas@gmail.com

M. Nasipuri
e-mail: mitanasipuri@gmail.com

1 Introduction

Feature selection has been an active research field for a while among researchers in different topics such as pattern recognition, machine learning, and data mining. It has found enormous utility in different areas, especially in forecasting, document classification, bioinformatics, and object recognition or in modeling of complex technological processes. Feature selection reduces the dimensionality of feature space and removes redundant, irrelevant, or noisy features. Thus, speeds up the learning time, improves the data quality and therefore the performance of classification process, and also increases the comprehensibility of the results.

Traditionally, feature selection techniques are divided into two broad categories: filters and wrappers [1, 2]. Filter methods judge the merit of features by using heuristics based on general features of the data, whereas wrappers evaluate the merit of features using the learning algorithm that is to be applied to the dataset. Although wrappers have been used for feature selection and region selection [3], filters [4–9] are more preferred. In this empirical study, we investigated five different standard filter-based feature ranking techniques (rankers), chi-square (CS), information gain (IG), gain ratio (GR), symmetrical uncertainty (SU), and correlation. In order to evaluate the effectiveness of these methods, we used classification model built with support vector machine (SVM)-based classifier on the smaller subsets of selected attributes. The created classification model is assessed on the basis of its success rate (recall value).

The empirical validation of the models has been implemented through a case study of five datasets of handwritten numerals. Each dataset holds the same numbers of classes (ten) but different numbers of training and testing samples. We have used one of the most popularly used feature-longest run features. The results demonstrate that different feature ranking techniques impact the evaluation outcome differently for different datasets. For instance, one ranker may perform better than another ranker for one dataset but gives worse result for another dataset in comparison with the same ranking technique.

The main contribution of this work is to provide an assessment and comparison of five filter-based feature ranking techniques for the task of handwritten numeral recognition.

The rest of the paper is organized as follows. Section 2 provides detailed information about the techniques used in the study. A brief description of datasets and the feature sets used are given in Sect. 3. Section 4 presents the experimental results and analysis. Finally, the conclusion is summarized in Sect. 5.

2 Methodology

2.1 Filter-based Feature Ranking Techniques

Filter-based feature ranking techniques rank features independent of any learning algorithm. Feature ranking consists of scoring each feature according to a particular trait or characteristic of the dataset involved in training, then selecting features based on the scores of that characteristic. This work employs some commonly used filter-based feature ranking techniques including CS, IG, GR, SU, and correlation.

The CS [10] test is used to check whether there is “no association” between two attributes, i.e., whether the two variables are independent. IG, GR, and SU all come from the concept of entropy, which comes from information theory. IG [11] is the information provided about the target class feature Y; given the value of independent feature X. IG measures the decrease of the weighted average impurity of the partitions, compared with the impurity of the complete set of data. A drawback of IG is that it tends to prefer attributes with a larger number of possible values. One strategy to counter this problem is to use the GR, which penalizes multi-valued attributes. SU [12] is another way to overcome the problem of IG’s bias toward attributes with more values, doing so by dividing IG by the sum of the entropies of X and Y. The correlation feature selection (CFS) [5] measure tries to find out the subsets of features that are highly correlated with the target classification class, yet uncorrelated to each other.

3 Datasets and Feature Sets

3.1 Datasets

In this work, handwritten numeral datasets of five scripts Bangla [13], Hindi [14], English, Telugu [15], and Arabic [16] were used. The Bangla, Hindi, Telugu, and Arabic datasets are freely available at www.code.google.com/cmaterdb. The English numeral database is created by randomly selecting 600 data samples from each of the 10 classes of numerals from MNIST dataset.

3.2 Feature Sets

We used one of the most widely used feature set—longest run features for experimental purposes. Longest run features [17] are computed in four directions: row wise, column wise, and along the directions of two major diagonals. The row wise longest run feature is computed by considering the sum of the lengths of the longest bars that fit consecutive black pixels along each of all the rows of the region. Details regarding the features used and extraction of features can be found in [17, 18].

4 Experiments

4.1 Classifier

All related experiments in this work have been carried out using the standard WEKA [19] machine learning tool. SVM [20], which has been used by majority of the research community in recent times, is used as classifier. We used the LibSVM tool [21] for all experimental purposes.

4.2 Performance Metric

For any classification outcome, there can be four possibilities: true positive (TP) (i.e., correctly classified positive instances), false positive (FP) (i.e., negative instance classified as positive), true negative (TN) (i.e., correctly classified as negative instance), and false negative (FN) (i.e., positive instance classified as negative). The number of cases from these four sets form the basis for all possible classifier evaluation. We have used the recall value, which is most commonly used for classifier evaluation.

- Recall: For a class i (ρ_i) is defined as $\left(\frac{TP_i}{TP_i+FN_i}\right)$. The overall recall measure is obtained by summing over the individual recall measures of each class. So the net recall value is,

$$\rho = \frac{\sum TP_i}{\sum TP_i + FN_i} \quad (1)$$

4.3 Experimental Results

Out of the total 84 numbers of longest run features, best possible feature subsets have been identified using the filter-based ranking techniques. Once a ranking is drawn, the exact numbers of features that give best possible success rate (recall) have been determined empirically. From the results shown in Table 1, it is evident that for all the feature selection techniques, best performance is achieved with selected features in the range of 70–80. It is to be noted that majority of them are 75. Table 2 shows the recall values, i.e., success rate attained through the individual feature selection techniques. Table 2 shows the comparative performance of ranking methods for the individual datasets. It is visible that SU performs better than all other ranking techniques for all the datasets and CS perform worst among all the ranking techniques (Table 3).

Table 1 Number of features selected

Dataset	Feature selection techniques				
	Chi-square	Correlation	Gain ratio	Information gain	Symmetrical uncertainty
Arabic	75	75	75	80	70
Bangla	75	75	75	75	75
Hindi	70	75	75	75	75
English	75	80	75	70	80
Telugu	80	80	80	80	75

Table 2 Recall values

Dataset	Feature selection techniques					
	Without feature selection	Chi-square	Correlation	Gain ratio	Information gain	Symmetrical uncertainty
Arabic	94.9	95.2	95.3	95.4	95.9	96.9
Bangla	95.9	96.1	96.1	96.5	96.4	97.3
Hindi	95.7	95.7	95.8	95.9	96.1	96.5
English	96.9	97.1	97.2	97.0	96.9	97.2
Telugu	98.2	98.4	98.3	98.5	98.7	98.7

Table 3 Rank of rankers

Dataset	Ranker (best → worst)				
Arabic	SU ¹ >	IG ² >	GR ³ >	COR ⁴ >	CSQ ⁵
Bangla	SU>	GR ^c >	IG ^b >	COR ^d =	CSQ ^e
Hindi	SU>	IG ^b >	GR ^c >	COR ^d >	CSQ ^e
English	SU>	GR ^c >	IG ^b >	COR ^d >	CSQ ^e
Telugu	SU=	IG ^b >	GR ^c >	COR ^d >	CSQ ^e

¹ SU = Symmetrical Uncertainty

² IG = Information Gain

³ GR = Gain Ratio

⁴ COR = Correlation

⁵ CSQ = Chi-square

5 Conclusion

In this paper, we have evaluated the effectiveness of five filter-based feature ranking techniques with a standard classifier. From Table 2, it is quite evident that SU proves to be superior over the four other filter-based techniques used and CS filter performs the worst. Our work serves as a preamble to our work in [9]. The

effectiveness of AFSFS theory [9] is compared with the results found in this work, and its superiority is proved. It should be noted that the work carried out in this paper was carried out before that of in [9].

The current work is a comprehensive study of the application of five most widely used filter-based feature selection techniques in handwritten numeral recognition. This work can be used as a reference in future for evaluating the performance of other newly developed and more sophisticated feature selection techniques. Also, as a scope of future work, the performance of wrapper-based feature selection techniques can be compared with that of the filter-based techniques presented here.

References

1. I. Tsamardinos, C.F. Aliferis, 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, FL, USA (2003)
2. M.A. Hall, L.A. Smith, 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* (2003), p. 239
3. A. Roy, N. Das, S. Basu, R. Sarkar, M. Kundu, M. Nasipuri, Region selection in handwritten character recognition using artificial bee colony optimization, in *EAIT* (2012), pp. 189–192
4. P. Mitra, C. Murthy, S.K. Pal, Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 301–312 (2002)
5. M.A. Hall, Correlation-based feature selection for machine learning (The University of Waikato, Hamilton, 1999)
6. H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, in *Pattern Analysis and Machine Intelligence* (2005), pp. 1226–1238
7. H.-M. Lee, C.-M. Chen, J.-M. Chen, Y.-L. Jou, An efficient fuzzy classifier with feature selection based on fuzzy entropy. *Trans. Sys. Man Cyber. Part B.* **31**, 426–432 (2001)
8. P. Luukka, Feature selection using fuzzy entropy measures with similarity classifier. *Expert Syst. Appl.* **38**, 4600–4607 (2011)
9. A. Roy, N. Das, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri, An axiomatic fuzzy set theory based feature selection methodology for handwritten numeral recognition, in *ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India*, ed. by S.C. Satapathy, P.S. Avadhani, S.K. Udgata, S. Lakshminarayana, vol. 248 (2014), pp. 133–140
10. R.L. Plackett, Karl Pearson and the chi-squared test. *Int. Stat. Rev./Rev. Int. de Stat.* **11**, 59–72 (1983)
11. J.R. Quinlan, Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986)
12. M.A. Hall, G. Holmes, Benchmarking attribute selection techniques for discrete class data mining (Knowledge and Data Engineering 2003), pp. 1437–1447
13. S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri, D.K. Basu, A novel framework for automatic sorting of postal documents with multi-script address blocks. *Pattern Recogn.* **43**, 3507–3521 (2010)
14. A. Roy, N. Mazumder, N. Das, R. Sarkar, S. Basu, M. Nasipuri, A new quad tree based feature set for recognition of handwritten bangla numerals. *Eng. Edu. Innovative Pract. Future Trends (AICERA)* **2012**, 1–6 (2012)
15. XXX
16. YYY

17. XXXX
18. YYYY
19. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update. *ACM SIGKDD Explor. Newslett.* **11**, 10–18 (2009)
20. V.N. Vapnik, An overview of statistical learning theory. *Neural Networks* **10**, 988–999 (1999)
21. C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011)

Performance Evaluation of PET Image Reconstruction Using Radial Basis Function Networks

T. Arunprasath, M. Pallikonda Rajasekaran, S. Kannan
and Shaeba Mariam George

Abstract In this paper, for the reconstruction of the positron emission tomography (PET) images, Artificial Neural Network (ANN) method and Artificial Neural Network-Radial Basis Function (ANN-RBF) method are pursued. ANN is a dominant tool for demonstrating, exclusively when the essential data relationship is unfamiliar. ANN imitates the learning process of the human brain and can process problems involving nonlinear and complex data even if the data are imprecise and noisy. But, ANN calls for high processing time and its architecture needs to be emulated. So, ANN-RBF method is implemented which is a two-layer feed-forward network in which the hidden nodes implement a set of radial basis functions. Thus, the learning process is very fast. By the image quality parameter of peak signal-to-noise ratio (PSNR) value, the ANN method and the ANN-RBF method are compared and it was clinched that better results are obtained from ANN with RBF method.

Keywords ANN · ANN-RBF · PSNR value · Positron emission tomography · Radial basis function

T. Arunprasath (✉) · M.P. Rajasekaran · S. Kannan · S.M. George
Kalasalingam University, Krishnankoil, Virudhunagar 626126, Tamil Nadu, India
e-mail: arun.aklu@gmail.com

M.P. Rajasekaran
e-mail: mpraja80@gmail.com

S. Kannan
e-mail: kannaneeps@gmail.com

S.M. George
e-mail: shaebamgeorge@gmail.com

1 Introduction

In past years, the conventional practice in positron emission tomography (PET) has been to reconstruct images while ignoring the effects of anatomical motion in the patient. When motion is ignored, a PET scanner acts something like a conventional optical camera with the shutter open—one obtains the superposition of images of the object as it appears in various stages of motion [1]. Thus, the image parades a grade of blur that is correlated with the motion scale. The magnitude of anatomical motion was not always significant enough, in evaluation to the scanner resolution, for motion rectification to promise any Benet. However, with the enhancement of scanner resolution over time, motion-related blur is appealing a restrictive factor in the resolution possible in PET reconstruction [2, 3]. This problem has been compounded by changing trends in the application of PET.

In early years, PET was applied mainly in the research of the brain, where the comparatively small movements of the head made motion effects still easier to dismiss. Conversely, recent clinical PET practice has emphasized cancer treatment, and therefore, thorax scans have become more rampant [4]. As a result, the scanner resolution is increased, but awareness has got rid of over the years to the scan of anatomy where motion is much larger. As a result of these factors, the recent tomographic imaging literature has seen much interest in motion-correction techniques [5]. In this dissertation, our work is motivated by the problem of reconstructing motion-corrected images of the thorax from respiratory gated (histogram mode) data and its probable Benets to liver cancer treatment. In attaining respiratory gated records, one essentially makes a separate scan of the object for every different position of the liver. In PET scans of the thorax, image intensity in the region of a liver lesion can be used as an indicator of malignancy and also, if it is malignant, of how well it is retorting to cure [6, 7]. Enumerating this, concentration precisely is therefore desirable. However, blur associated with the motion of the liver can degrade quantization accuracy, unless effective motion-correction measures are active.

The liver is the largest organ in the body. It gets oxygen-rich blood from the heart through the main artery leading into the liver (hepatic artery). Another source is oxygen-poor blood containing nutrients, poisons, and other things that come from the intestines. The liver filters this blood, and then sends it on to the heart through the hepatic vein. The liver's job is to run over 500 bodily functions. It plays a role in the dispensation of food, sugar, and fat. It also plays a vital role in the body's defense system—the immune system. It processes almost everything a person eats, breathes, or absorbs over and done with the skin. Around 90 % of the body's nutrients pass through the liver from the small and large intestines. The liver changes food into energy, stores this energy, and sorts blood proteins. The liver also get rid of bacteria and poisons from the blood. Liver cells make bile, a greenish-yellow fluid that helps with the breakdown of fats and in absorbing nutrients. Waste made by the liver in the breakdown of food is carried in the bile and removed from the body. Liver cells also change heme (a portion of hemoglobin that is released when red blood cells are broken down) into bilirubin.

2 Methodology

2.1 ANN Method

In an artificial neural network (ANN) model, a neuron is an information-processing unit that is fundamental to the operation of a neural network [8]. In mathematical terms, we may describe a neuron by writing the following equation:

$$o_k = f\left(\sum_{j=1}^m w_{kj}i_j + b_k\right) \tag{1}$$

where k is the neuron number, i_1, i_2, \dots, i_m are the input parameters; $w_{k1}, w_{k2}, \dots, w_{kj}$ are the weights of neuron; b_k is the bias; O_k is the output parameter; f is the effect of applying an affine transformation to the output [9, 10]. Typically, a network as shown in Fig. 1 consists of a set of sensory units that constitute the input stratum, one or more hidden stratum, and an output stratum [9].

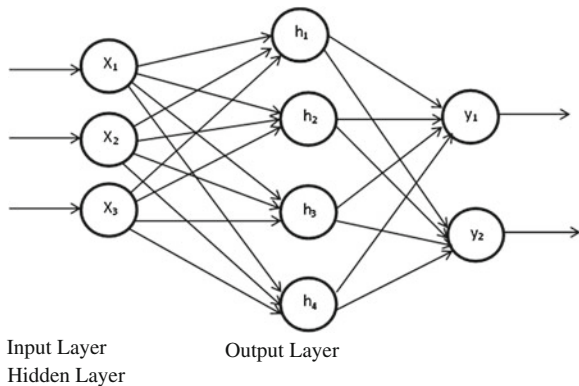
The training process is described by the following equations to update these weights, bias values. The output of the j th neuron in the hidden layer is calculated by following steps:

$$\text{net}_j = f\left(\sum_{i=1}^k w_{2,i}i_i + b_{2,j}\right) \tag{2}$$

$$a_j = f_{\text{hidden}}(n_j) \tag{3}$$

where net_j is the activation value, a_j is the output of the hidden layer, and f_{hidden} is called the activation function which is usually a linear, sigmoid, or tanisg function [11].

Fig. 1 The ANN model based on BP training algorithm: neural network framework for prediction



The sigmoid used in this model is described as:

$$f_{\text{hidden}}(x) = \frac{1}{1 + \exp(-x)} \quad (4)$$

The numerous steps intricate in the progress of ANN-based image reconstruction model are presented below.

A. Training data generation

The cohort of the proper training data is a vital step in the development of ANN models. For the ANN to precisely envisage the output, the training data should epitomize the ample range of effective state of affairs of the system under contemplation. For model expansion, a large quantity of training data is spawned through off-line power system replication. Assortment of suitable input for the ANN is very imperative [12].

B. Data normalization

The objective to train the network is to amend the weights so that application of a set of inputs yields the anticipated set of outputs. Formerly opening the training process, totally the weights must be modified to small haphazard information. This ought to guarantee that the network is not saturated by large values of weights [13]. This affects the network training to a great extent. To avoid this, the raw data is normalized before the actual application to the neural network. One way to standardize the data x is by means of the expression:

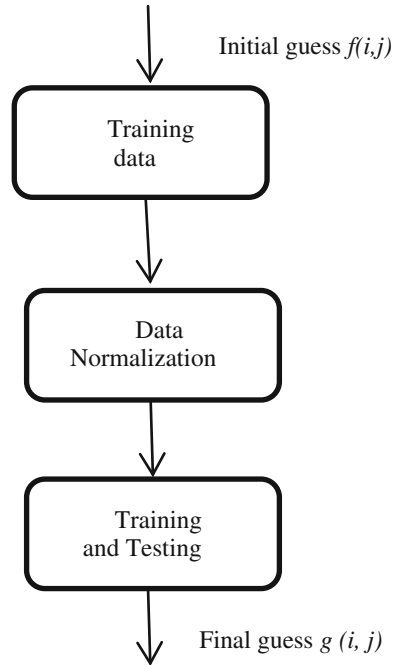
$$x_n = \frac{1(x - x_{\min})}{x_{\max} - x_{\min}} \quad (5)$$

where x_n is the normalized value and are the minimum and maximum values of the variable.

C. Training and testing of neural network

Training data prerequisite to train the neural network can be reserved from any image haphazardly or in a chronological pixel way. Patterns so far well thought out for training the network are 100. This can be reckoned as the most important trouble to select the training pattern, and hence, appropriate matching of training data with the network carefully chosen is quite a typical work based on trial and error based [13]. If the training pattern pixels have a very less intensity disparities, then this will result to diminish the quality of reconstructed image (Fig. 2).

Fig. 2 Flow chart shows the ANN-based image reconstruction steps in which the data to be trained is generated and then normalized, and finally, those data are trained and tested



2.2 ANN-RBF Method

The radial basis functions (RBFs) in the hidden layer produce a significant non-zero response only when the input falls within a small localized region of the input space. Each hidden unit has its own receptive field in input space. An input vector x_i which lies in the receptive field for center c_j would activate c_j and by proper choice of weights the target output is obtained [14].

It has a single hidden layer. The basic neuron model as well as the function of the hidden layer is different from that of the output layer [15]. The hidden layer is nonlinear but the output layer is linear. The activation function of the hidden unit computes the Euclidean distance between the input vector and the center of that unit. It establishes local mapping, hence capable of fast learning. It is a two-fold learning. Both the centers (position and spread) and weights have to be learned.

A RBF is a real-valued function whose value depends only on the distance from the origin. If a function 'h' satisfies the property $h(x) = h(|x|)$, then it is a radial function. Their distinctive feature is that their response decreases (or increases) monotonically with distance from a midpoint. The midpoint, the distance scale, and the specific shape of the radial function are strictures of the model, all stable if it is in lines [16].

Radial functions are merely a class of utilities. In principle, they could be engaged in any sort of model (linear or nonlinear) and any sort of network (single-layer or multilayer). RBF networks have conventionally been concomitant with radial functions in a single-layer network [17]. The distance between these values and midpoint values are set up and summed to form linear combination formerly the neurons of the hidden layer. These neurons are said to enclose the RBF with exponential formula. The RBF activation function output is supplementary sort out rendering to specific requirements.

3 Results and Discussion

Table 1 shows image quality parameters for two kinds of PET images using ANN method and ANN with RBF method. It was observed that the peak signal-to-noise ratio (PSNR) value is highest for Image II using ANN with RBF, and normalized absolute error (NAE) and root mean square errors (RMSE) values are lowest for the same. Thus, it was observed that ANN with RBF method is better when compared with ANN method.

$$PSNR = 20 \log \frac{\text{Max } N}{\sqrt{MSE}} \dots \tag{6}$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [f(i,j) - g(i,j)]^2 \dots \tag{7}$$

$$NAE = \frac{1}{K} \sum_{k=1}^K \frac{\sum_{i,j=0}^{N-1} f(i,j) * g(i,j)}{\sum_{i,j=0}^{N-1} f(i,j)^2} \dots \tag{8}$$

PSNR and signal-to-noise ratio (SNR) is a mathematical measure of image quality based on the pixel difference between two images. The SNR measure is an estimate of quality of reconstructed image compared with original image. PSNR is defined as in (6) where N is $m*n$ for a 8-bit image. Here, Max is maximum pixel value of image when pixel is represented by using 8 bits per sample. Mean square

Table 1 Image quality parameter

Methods	PET images	Image quality parameters		
		PSNR	NAE	RMSE
ANN	Image I	24.2349	0.2514	15.6602
	Image II	26.6472	0.0794	11.8626
ANN + RBF	Image I	24.2600	0.2718	15.6150
	Image II	27.2312	0.0757	11.0912

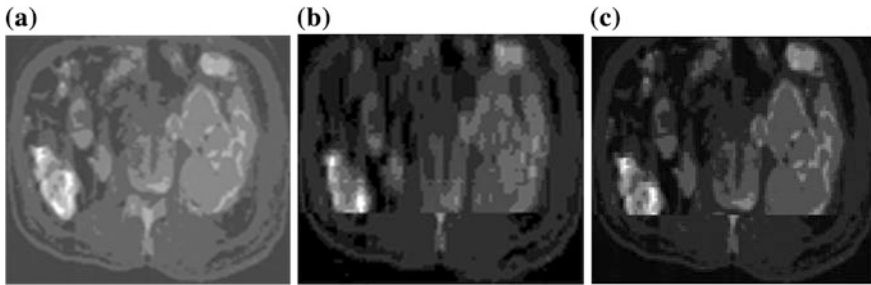


Fig. 3 a Image I. b ANN reconstructed image. c ANN with RBF reconstructed image

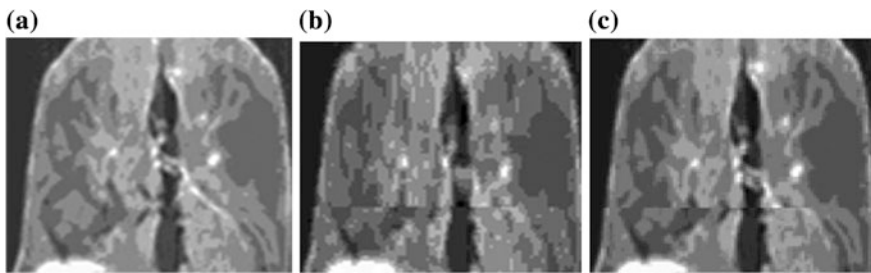


Fig. 4 a Image II. b ANN reconstructed image. c ANN with RBF reconstructed image

error (MSE) is computed by averaging the squared intensity of the original (input) image and the trained (output) image pixels as in (8) where f is the original image, g is the trained image, both of size $m \times n$. NAE is a measure of how far is the reconstructed image from the original image with the value of zero being the perfect fit. RMSE is given by the square root of MSE [18] (Figs. 3 and 4).

4 Conclusion

In this paper, image reconstruction is done using ANN with RBF. The image reconstruction result provides that the ANN with RBF produce the reconstructed image with better PSNR, thus proves the image quality is improved compared to ANN. Normalized and RMSE are also reduced significantly with reduced memory size for storage. The resulting framework sightseer optimally spatial dependencies between images content en route for nonlinear image reconstruction. The neural network-based RBF of image reconstruction illustrates the auspicious results. So,

ANN becomes more cognitive. Thus, it was concluded that high-quality reconstructed image is provided since the PSNR ratio of image is improved. The table of results describes that the performance of the ANN with RBF method was better than ANN.

Acknowledgment We thank Anderson Diagnostics and Lab, Chennai for providing PET images for our research and Department of Instrumentation and Control Engineering of Kalasalingam University (Kalasalingam Academy of Research and Education), Tamil Nadu, India for permitting to use the computational facilities available in biomedical Laboratory which was setup with the support of the Department of Science and Technology (DST), New Delhi under FIST Program.

References

1. G. Tarantola, F. Zito, P. Gerundini, PET instrumentation and reconstruction algorithms in whole-body applications. *J. Nucl. Med.* **44**, 756–769 (2003)
2. G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography* (Academic Press, New York, 1980)
3. V.Y. Panin, F. Kehren, C. Michel, M. Casey, Fully 3-D PET reconstruction with system matrix derived from point source measurements. *IEEE Trans. Med. Imag.* **25**(7), 907–921 (2006)
4. A.J. Rockmore, A. Macovski, A maximum likelihood approach to emission image reconstruction from projection. *IEEE Trans. Nucl. Sci.* **23**(4), 1428–1432 (1976)
5. S. Vandenberghe, S. Staelens, C.L. Byrne, E.J. Soares, I. Lemahieu, S.J. Glick, Reconstruction of 2D PET data with Monte Carlo generated system matrix for generalized natural pixels. *Phys. Med. Biol.* **51**(12), 3105–3125 (2006)
6. M. Rafecas, B. Mosler, M. Dietz, M. Pögl, A. Stamatakis, D.P. McElroy, S.I. Ziegler, Use of a Monte Carlo based probability matrix for 3D iterative reconstruction of MADPETII data. *IEEE Trans. Nucl. Sci.* **51**(5), 2597–2605 (2004)
7. R.D.L. Prieta, An accurate and parallelizable geometric projector/backprojector for 3D PET image reconstruction. *Lect. Notes Comput. Sci.* **3337**, 27–38 (2004)
8. P. Chawla, R. Mittal, K. Grewal, Hybrid filtering technique for image denoising using artificial neural network. *Int. J. Eng. Adv. Techno. (IJEAT)* **1**(3), 36–40 (2012)
9. D.B. Mantri, S.N. Kulkarni, S.S. Katre, Image compression using ebp with neural network. *J. Inf. Knowl. Res. Electron. Commun. Eng.* **1** (2011)
10. Z. Shi1, L. He, Application of neural networks in medical image processing, in *Proceedings of the Second International Symposium on Networking and Network Security (ISNNS'10)* (2010), pp. 2–4
11. J. Jiang, P. Trundle, J. Ren, Medical image analysis with artificial neural networks. *Comput. Med. Imaging Graph.* **34**, 617–631 (2010) (Elsevier)
12. D.C. Durairaj, M.C. Krishna, R. Murugesan, A neural network approach for image reconstruction in electron magnetic resonance tomography. *Comput. Methods Programs Biomed.* **37**, 1492–1501(2007) (Elsevier)
13. K.H. Su, L.C. Wu, J.S. Lee, R.S. Liu, J.C. Chen, A novel method to improve image quality for 2-D small animal PET reconstruction by correcting a Monte Carlo-simulated system matrix using an artificial neural network. *IEEE Trans. Nucl. Sci.* **56**(3), 704–714 (2009)
14. M.T. Munley, C.E. Floyd, J.E. Bowsher, R.E. Coleman, A spatially-variant SPECT reconstruction scheme using artificial neural networks, in *Proceedings of Conference on Record of IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. 2 (1992), pp. 1279–1281

15. M. Niranjan, F. Fallside, Neural network and radial basis function in classifying static speech patterns. *Comput. Speech Lang.* **4**, 275–289 (1990)
16. W. Hong, W. Chen, R. Zhang, The application of neural network in the technology of image processing, in *Proceedings of International Multi Conference of Engineers and Computer Scientists*, vol. 1 (2009), pp. 18–20
17. E. Floyd, An artificial neural network for SPECT image reconstruction. *IEEE Trans. Med. Imag.* **10**(3), 485–487 (1991)
18. T. ArunPrasath, M.P. Rajasekaran, S. Kannan, V.A. Kalasalingam, Reconstruction of PET brain image using conjugate gradient algorithm, in *Second World Congress on Information and Communication Technologies (WICT 2012)*, Trivandrum, India (2012)

Clustering for Knowledgeable Web Mining

**B.S. Charulatha, Paul Rodrigues, T. Chitralekha
and Arun Rajaraman**

Abstract Web pages nowadays have different forms and types of content. When the Web content is considered, they are in the form of pictures, videos, audio files, and text files in different languages. The content can be multilingual, heterogeneous, and unstructured. The mining should be independent of the language and software. Statistical features of the images are extracted from the pixel map of the image. The extracted features are presented to the fuzzy clustering algorithm (FCM) and Gath–Geva algorithm. The similarity metric being Euclidean distance and Gaussian distance, respectively. The accuracy is compared and presented.

Keywords Fuzzy clustering algorithms · Gath–Geva fuzzy clustering · Content mining

1 Introduction

The Web is ocean of data. The indexed WWW has more than 13 billion pages, more than 630,000,000 sites (Netcraft, Feb 2013), and 2 million scholarly articles published each year. 4 % growth rate each year more than 50,000,000 scholarly articles so far (Jinha 2010) [7]. But the drawback is that the Web pages are only in English language. Moreover, these data are in the form of databases or flat files or other repositories, structured or unstructured. The types of data found may be

B.S. Charulatha (✉)
JNTUK, Kakinada, India
e-mail: charu2303@yahoo.co.in

P. Rodrigues
Velammal Engineering College, Chennai, India

T. Chitralekha
Central University, Puduchery, India

A. Rajaraman
IIT Madras, Chennai, India

relational or object-oriented or object-relational or multimedia. But, however, these databases must be mined to extract useful information from the repository. Such mining is called Web mining.

Web mining tasks can be classified into three categories:

- Web content mining,
- Web structure mining,
- Web usage mining.

Next advancement in mining is the streaming, both audio and video. Hence, media mining is gaining popularity. The usage of Internet is for sharing of knowledge or entertainment. In the case of entertainment, reading news through Web page is much popular but is multilingual. The usage of the Internet has spread across the nook and corner of the world. In a country like India which has many official languages, the need for multilingual mining arises. The language need not be only in English but also in regional languages. The mining industry gives importance to English. Hence, the need arises for multilingual search or mining [6].

A multilingual Web page consists of text more than one language along with images, streaming videos, and scrolling news, which makes the user to see a lot of information on a single Web page. The focus of the present work is to extract the content from the multilingual Web pages and understand what actually the Web page considered tells about method should preferably be not in translation as in text mining. Method should preferably be computer understandable and not software dependent. Pixel-based processing to assess the overall content is the focus of the study. The objective of the present study is to extract the features dealing with Web documents either in English or in a regional language like Sanskrit [1]. The remainder of the paper is organized as follows: Sect. 2 discusses about the content representation of the images. Section 3 discusses about the statistical feature extraction. Section 4 discusses the fuzzy clustering. Section 5 discusses about Gath–Geva clustering. Section 6 contains the test images, the test of accuracy is presented, and Sect. 7 concludes the paper.

2 Content Representation

Here, the content can be in the form of the following flavors which are considered as initial setup:

- a) Pictorial representation,
- b) English equivalent of the image,
- c) Sanskrit equivalent of the image represented in Sanskrit,
- d) Sanskrit equivalent transliterated in English,
- e) Tamil equivalent in Tamil,
- f) Tamil equivalent in English.

Initially, five animals, namely cat, dog, horse, tiger, and lion, are taken. Pencil sketch of the images of above-said animals under various sizes is taken. In case of alphabetical representation, the words were taken under different font sizes and styles. When cat is taken, the various representations chosen are as follows:



cAt, बिडालः, BIDAL, புனை, s Poonai

The representation goes in the order mentioned above (The exact representation is scaled down for representation here). The content is taken, and the pixel map equivalent of the content is determined. Once the pixel map of the content in any of the above-said form is found, the same pixel map is fed as input to the feature extraction module [2].

Likewise, five animals, viz. lion, cat, tiger, dog, and horse in the specified forms, are taken for the training. As sample, only cat varieties are shown in the paper.

3 Statistical Feature Extraction

The pixel map of the image is of higher dimension. Hence, dimension reduction is done to the pixel map of the image using the nonzero concept. The reduced matrix size is 2×2 and 3×3 . The features extracted are the statistical properties of the images. To the pixel map of the image, the various attributes such as mean, standard deviation, and eigenvalue are considered for processing.

4 Fuzzy Clustering

Fuzzy C-means clustering (FCM) relies on the basic idea of hard C-means clustering (HCM), with the difference that in FCM, each data point belongs to a cluster to a degree of membership grade, while in HCM, every data point either belongs to a certain cluster or does not. So FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1. However, FCM still uses a cost function that is to be minimized while trying to partition the data set.

The membership matrix U is allowed to have elements with values between 0 and 1. However, the summation of degree of belongingness of a data point to all clusters is always equal to unity:

$$\sum_{i=1}^c u_{ij} = 1, \quad \forall j = 1, 2, 3, \dots, n \tag{1}$$

The cost function for FCM is as follows:

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 \tag{2}$$

where u_{ij} is between 0 and 1, c_i is the cluster center of fuzzy group i , and d_{ij} is the Euclidean distance between the i th cluster center and j th data point.

$M \in [1, \infty]$ is a weighting exponent.

The necessary condition for (2) to reach its minimum is (3) and (4)

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \tag{3}$$

$$u_{ij} = \frac{1}{\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{4}$$

The algorithm works iteratively through the preceding two conditions until no more improvement is noticed. In a batch mode, FCM determines the cluster centers c_i and the membership matrix U using the following steps:

- Step 1: Initialize the membership matrix U with random values between 0 and 1 such that the constraints in Eq. (1) are satisfied.
- Step 2: Calculate c fuzzy cluster centers c_i , $1, c$ using (3).
- Step 3: Compute the cost function according to (2). Stop either if it is below a certain tolerance value or if its improvement over previous iteration is below a certain threshold.

The performance of FCM depends on the initial membership matrix values; thereby, it is advisable to run the algorithm for several times, each starting with different values of membership grades of data points.

Euclidean Distance

The Euclidean distance or Euclidean metric is the ordinary distance between two points that one would measure with a ruler. It is the straight-line distance between two points. In a plane with p_1 at (x_1, y_1) and p_2 at (x_2, y_2) , it is $\sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$. The distance is calculated using the formula [4, 5]:

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{5}$$

5 Gath–Geva Fuzzy Clustering

FCM are usually based on minimization of the following objective function:

$$J = \sum_{i=1}^c \sum_{j=1}^q u_{ij}^m d_{ij} \tag{6}$$

where C and Q are the number of clusters and data points, respectively. U_{ij} is the membership degree of the j th data point to the i th cluster, and m is a positive integer greater than one, which defines the fuzzification degree of clusters. Selection of small values close to 1 for m leads the algorithm to crisp clustering, while values greater than 3 result in spiky clusters. In the GG method, d_{ij} , distance of the j th data point from the i th cluster, is defined as follows:

$$D_{ij} = \frac{P_i}{\sqrt{\det \Sigma_i}} * \exp^A \tag{7}$$

$$A = \frac{1}{2} (x_j - \mu_i)^T \sum_i^{-1} (x_j - \mu_i)$$

where the parameters of each cluster, μ_i and Σ_i , are, respectively, center and covariance of the i th cluster. P_i is also the coefficient designed for eliminating the sensitivity of the algorithm to the number of data points in different clusters, which is computed by the following formula:

$$P_i = \frac{\sum_{j=1}^Q u_{ij}^m}{\sum_{i=1}^c \sum_{j=1}^Q u_{ij}^m} \tag{8}$$

Minimization of the objective function (6) with respect to membership degree by considering the fact that the sum of membership values of a data point to all clusters becomes one, leads (Hoppner et al. 1999).

$$u_{ij} = \frac{\left(\frac{1}{d_{ij}}\right)^{\frac{1}{(m-1)}}}{\sum_{i=1}^c \left(\frac{1}{d_{ij}}\right)^{\frac{1}{(m-1)}}} \tag{9}$$

In the GG algorithm, center and covariance matrix of clusters and membership degree of data points are estimated in the following iterative process (Gath and Geva 1989).

1. Choose the number of clusters.
2. Find the centers using initial values of membership matrix and data set using

$$\mu_i = \frac{\sum_{j=1}^Q u_{ij}^m x_j}{\sum_{j=1}^Q u_{ij}^m} \quad (10)$$

and covariance matrix for each cluster using

$$\sum_i = \frac{\sum_{j=1}^Q u_{ij}^m (x_j - \mu_i)(x_j - \mu_i)^T}{\sum_{j=1}^Q u_{ij}^m} \quad (11)$$

3. Calculate the distances between data points of all clusters using (7).
4. Compute the degree of membership for all data points using (9).
5. Estimate the center μ and covariance matrix Σ for each cluster using

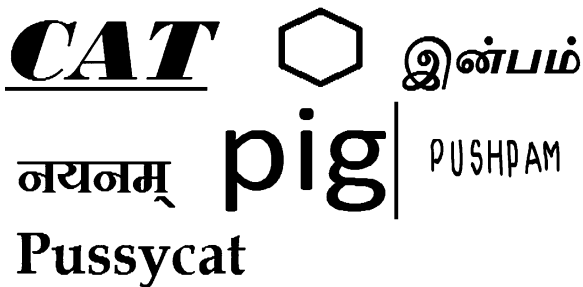
$$\mu_i = \frac{\sum_{j=1}^Q u_{ij}^m x_j}{\sum_{j=1}^Q u_{ij}^m} \quad (10)$$

$$\sum_i = \frac{\sum_{j=1}^Q u_{ij}^m (x_j - \mu_i)(x_j - \mu_i)^T}{\sum_{j=1}^Q u_{ij}^m} \quad (11)$$

6. Go to step 2 until the termination criterion is satisfied based upon minimization (6) [8].

6 Result

For testing, the following images are taken and the result analysis is provided below:



The results are provided in Tables 1, 2, 3, and 4.

All the test data belong to cluster animal.

Inference:

When accuracy is taken as measure, the results are as follows:

True positive = 42.85 %

False positive = 57.14 %

Table 1 Result using FCM and subtractive clustering

FCM	Centroid 1	0.330	0.0072	0.0118	0.1955
	Centroid 2	0.0460	0.0047	0.0175	0.9205
Gath–Geva	Centroid 1	0.0346	0.0060	0.0123	0.2125
	Centroid 2	0.0521	0.0083	0.0135	0.2052

Table 2 Euclidean distance with the two centroids

	FCM		Gath–Geva	
	Distance with Centroid 1	Distance with Centroid 2	Distance with Centroid 1	Distance with Centroid 2
I1	0.0072	0.4203	3.4058	17.7048
I2	0.0087	0.6642	1.2981	4.9868
I3	0.0030	0.6072	1.2719	5.1174
I4	0.0014	0.5132	1.7825	7.8562
I5	0.0041	0.4388	2.7685	13.5633
I6	0.0038	0.6057	1.3610	5.3170
I7	0.0040	0.4387	2.7418	13.5786

I1–I7 represent the seven test images

Table 3 Membership of the test data with the clusters using FCM

Animal	0.9997	0.9998	1.0000	1.0000	0.9999	1.0000	0.9999
Not animal	0.0003	0.0002	0.0000	0.0000	0.0001	0.0000	0.0001

Table 4 Membership of the test data with the clusters using Gath–Geva

Animal	0.8387	0.7935	0.8009	0.8151	0.8305	0.7962	0.8320
Not Animal	0.1613	0.2065	0.1991	0.1849	0.1695	0.2038	0.1680

7 Conclusion

The study is limited to the black-and-white images, the languages are confined to English, Tamil, and Sanskrit, and the word set is restricted to animals only. The number of clusters is restricted to the class of animal or not. The algorithms used to draw conclusion are restricted to two FCMs. This can be extended to the specific class of animal or the specific representation along with the algorithm types. The application will be in the field of content mining.

References

1. B.S. Charulatha, P. Rodrigues, T. Chitralekha, A. Rajaraman, Heterogeneous clustering
2. B.S. Charulatha, P. Rodrigues, T. Chitralekha, A. Rajaraman, Heterogeneous clustering on images, in *READ IT 2013 by IGCAR* (2013)
3. B.S. Charulatha, P. Rodrigues, T. Chitralekha, A. Rajaraman, A Comparative study on subtractive and FCM clustering algorithms, in *IEEE International Conference on Intelligent Interactive Systems and Assistive Technologies* (2013)
4. B.S. Charulatha, P. Rodrigues, T. Chitralekha, Fuzzy clustering algorithms—different methodologies and parameters—a survey, in *International Conference on Advances in Electrical and Electronics, Information Communication and Bio Informatics* (2012)
5. M. Gupta, J. Verma, V. Tomar, S. Roy, Mining databases on world wide web (2011)
6. K.B. Prakash, M.A.D. Rangaswamy, A.R., Raman performance of content based mining approach for multi-lingual textual data international J. Modern Res.(IJMER) **1**(1), 146–150
7. XXX
8. H. Soleimani-B, C. Lucas, B.N. Araabi, Recursive Gath–Geva clustering as a basis for evolving neuro-fuzzy modeling. *Evolv. Syst.* **1**, 59–71 (2010)

Effective Path Discovery Among Clusters for Secure Transmission of Data in MANET

P. Madhavan, P. Malathi and R. Abinaya

Abstract Ad hoc networks are new paradigm of wireless communication for mobile host. Owing to vulnerable nature of mobile ad hoc network (MANET), there are numerous security threats that disturb the development of MANET. In order to overcome such threats, an efficient anonymous routing protocol for MANETs is proposed. For this, secure neighbor should be discovered. After finding the secure neighbors, neighbors are clustered to form zones. For each group of clusters, cluster head (CH) is associated. Instead of sending the route request message to every neighbor nodes, the route request message would be send only to the CH. The CH would decide which all nodes should participate in the routing. The overhead for sending messages gets reduced as the number of route request messages (RREQ) sent is minimum. Among the different paths in the clusters, the shortest path will be determined. This protocol guarantees the security, anonymity, and high reliability of an established route.

Keywords Cluster · Mobile ad hoc network · Anonymous routing · Secure routing · Privacy preservation · Shortest path

1 Introduction

A wireless ad hoc network is a decentralized type of wireless network. The ad hoc network does not rely on a preexisting infrastructure, such as routers in wired networks or access points in managed (infrastructure) wireless networks. Instead,

P. Madhavan (✉) · R. Abinaya
Department of CSE, Sri Krishna College of Technology, Coimbatore, India
e-mail: madhrace@gmail.com

R. Abinaya
e-mail: r.p.abinaya@gmail.com

P. Malathi
Bharathiyar Institute of Engineering for Women, Deviyakurichi, Attur, Salem, India

each node participates in routing by forwarding data to other nodes, so the determination of which nodes forward data is made dynamically on the basis of network connectivity. In addition to the classic routing, ad hoc networks can use flooding for forwarding data.

Mobile ad hoc network (MANET) is the one type of self-configuring network that can change location and configure itself because MANETs are mobile which uses wireless connections to various networks. *MANET security goals*: MANETs are used to provide security services, such as authentication, confidentiality, integrity, anonymity, and availability, to mobile users. In order to achieve this goal, mobile node should be protected from malicious node.

In this paper, we make two contributions for secure data transmission in ad hoc networks. First, we give the salient features and drawbacks of anonymous routing protocol. Second, we present the salient features and drawbacks of secure routing protocol. Finally, we conclude that how routing can be enhanced in the proposed work.

2 Related Work

2.1 Secure Routing Protocol

MANETs are vulnerable to several attacks. Secure routing protocol is used to avoid threats and different attacks such as Sybil attack, rushing attack, and black hole attacks. It guarantees data integrity and confidentiality and ensures that the data to reach the correct destination.

A secure on-demand routing protocol is proposed in the paper “A Secure On-Demand Routing Protocol for Ad Hoc Networks” by Ariadne [1]. This protocol withstands node compromise and relies only on highly efficient symmetric cryptography. Ariadne prevents attackers or compromised nodes from tampering with uncompromised routes consisting of uncompromised routes and also prevents many types of denial of service attacks. This on-demand routing protocol allocates a fraction of the bandwidth along each link to each node. Ariadne can authenticate routing messages using one of the three schemes.

- Shared secret keys established between all pairs of nodes.
- Shared secret keys between communicating nodes combined with broadcast authentication.
- Digital signatures.

Drawbacks

- This protocol has high overhead in terms of the computational resources necessary for digital signature verification and in terms of its bandwidth requirements.

- In Ariadne, optimization of DSR is not secure. The resulting protocol is less efficient than the highly optimized version of DSR that runs in a trusted environment.

The secure efficient ad hoc distance vector (SEAD) protocol is proposed in this paper “Secure Efficient Ad Hoc Distance Vector (SEAD)” by Hu et al. [2] is based on the design of the destination-sequenced distance-vector (DSDV) routing protocol, which is a PMP, i.e., a proactive (table-driven) protocol. SEAD uses one-way hash chains for authentication and assumes the existence of a mechanism to distribute such hash chains. SEAD used a sequence number method for authenticating an entry in the routing update. In a routing update entry, the hash value corresponding to the sequence number prevents any node from advertising a route to some destination claiming a greater sequence number than that destinations own current sequence number.

Drawbacks

- Network overheads due to large number of advertisements it sends.
- Size of the advertisement is also large due to the addition of hash value in the route entry.

The secure ad hoc on-demand distance vector (SAODV) protocol is proposed in this paper “SAODV: A MANET Routing Protocol that can Withstand Black Hole Attack” by Lu et al. [3] is an RMP that consists of security extensions to the ad hoc on-demand distance vector (AODV) protocol. SAODV uses two mechanisms to secure AODV control messages: digital signatures for authentication of the static fields, i.e., non-mutable fields of a message and hash chains to secure the hop count information.

The authenticated routing for ad hoc networks (ARAN) is an RMP, i.e., on-demand routing protocol is proposed in this paper “ARAN: Analyzing security of Authenticated Routing Protocol” by Nagrath et al. [4] designed to provide end-to-end authentication message integrity and non-repudiation in ad hoc networks. ARAN assumes a preliminary certification process and requires every device to have a certificate. Such certificates are issued by a trusted certificate server and require a secure communicating channel between the server and the device.

2.2 Anonymous Routing Protocol

Anonymous routing protocol is used as a tool to protect informational privacy. Anonymity provides improvement in network performance, which can be evaluated in terms of packet delay, packet loss ratio, computational power required, and amount of data delivered versus amount of data transmitted.

Anonymous on-demand routing protocol is proposed in the paper “ANODR: ANonymous on Demand Routing with Untraceable Routes for Mobile Adhoc Networks” by Kong and Hong [5].

Salient features of ANODR

ANODR is to develop “untraceable” routes or packet flows in an on-demand routing environment. The anonymous route discovery process establishes an on-demand route between a source and its destination. Each hop en route is associated with a random route pseudonym. Based on route pseudonyms, data forwarding happens in the network with negligible overhead. The route pseudonymity approach allows us to “unlink” (i.e., thwart inference between) network member’s location and identity. So eavesdroppers can only detect the transmission of wireless packets stamped with random route pseudonyms. It is hard for them to trace how many nodes are in the locality, and who is the transmitter or receiver.

Drawbacks

- Efficiency
- Anonymity issue
- Trapdoor issue

Secure distributed routing protocol is proposed in the paper “SDAR: A Secure Distributed Anonymous Routing Protocol for Wireless and MANETs” by Boukerche et al. [6].

Salient Features of SDAR

SDAR protocol is to allow trustworthy intermediate nodes which are participate in the path construction protocol without jeopardizing the anonymity of the communicating nodes. To provide trust and anonymity for the path discovery, onion routing approach and trust management system are proposed. Trust management approach is used to select the routing path that meets certain trust requirements. The purpose of the system is to identify the malicious nodes in the network and avoid using the route establishment. Trust management approach is based on node trust level. Node trust level is defined as a cumulative value that is based on the past behavior of the node. Trust level of the node increases when the node behaves exactly as it supposed be; otherwise, it decreases when node misbehaves accordingly.

Drawbacks

- Trapdoor issue
- Scalability issue
- Security issue

Anonymous dynamic source routing protocol is proposed in the paper “AnonDSR: efficient anonymous dynamic source routing for MANETs” by Song et al. [7].

Salient features of AnonDSR

AnonDSR provide a strong security and anonymity protection and better scalability for MANETs. The new routing consists of three protocols. For secure and anonymous communication, shared secret keys and nonce are created between the source and destination. The second protocol uses the shared secret key and nonce to

create a trapdoor and employs an anonymous onion routing between the source and destination. In the last protocol, the source and destination use their session keys shared with the intermediate nodes to encrypt all communications with the cryptographic onion method.

3 Proposed Work

For secure data transmission, neighboring nodes should be discovered. These neighboring nodes are discovered using RSA algorithm. After finding the secure neighbors, route request messages (RREQs) are broadcasted to the destination through the neighboring nodes. Network overhead occurs by sending multiple RREQ messages. In order to avoid this overhead, neighboring nodes are clustered to form zones. For each group of zones (clusters), a CH is associated. Instead of sending the RREQ to every neighbor nodes, the RREQ would be send only to the CH. The CH would decide which all nodes should participate in the routing. The overhead for sending messages gets reduced as the number of RREQs sent is minimum. Among the different paths in the clusters, the shortest path will be determined. As a result, security threats would be prevented. Thus, the traffic gets reduced which would help to improve the overall performance of the system.

Figure 1 shows the system design.

3.1 Neighbor Node Discovery

In MANET, for a secure data transmission, neighbor nodes have to be find out. In order to find the neighbor nodes, neighbor discovery scheme is used.

- Regular-neighbor discovery
- Semi-neighbor discovery

3.1.1 Regular-Neighbor Discovery

The regular neighbor of a requesting user receives the neighbor discovery message from the requesting user directly and authenticate directly.

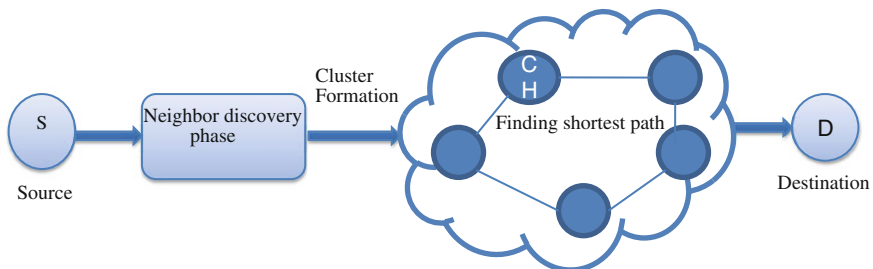


Fig. 1 System design

Steps

- User S (source) generates and broadcasts the neighbor discovery message to the participating nodes with nonce number and compute g^{rs} , where g is a generator of Z_p^* .
- The participating nodes (A, B) verify the message using source “S” public key and records in its neighbor candidate list.
- Source verifies the reply message from A by computing $K_{AS} = (g^{rA}) \bmod p$ and record user A as regular neighbor in neighbor list, where K_{AS} denotes the shared secret key of user S and A.
- Source generates the corresponding reply message and sends to user A. The user A verifies the message and removes S from the neighbor candidate list and adds S to the neighbor list.
- In a given time period, if the user A does not receive the reply message from “S,” then A retransmits the same message “ t ” time.

3.1.2 Semi-Neighbor Discovery

The semi-neighbor of a requesting user receives a neighbor discovery message directly, but can only indirectly authenticate through regular neighbors or semi-neighbors of the requesting user. After completing the regular-neighbor discovery phase, user B launches the semi-neighbor discovery phase to the users who are still in the B’s neighbor candidate list.

Steps

- User B generates and broadcasts the semi-neighbor discovery message to the participating nodes (S, A) with B’s broadcast key K_B^b , where K_B^b denotes the broadcast key of user B.
- The participating node A verifies the integrity of the message using K_B^b and replies the message with (A, B) common neighbor list and sign list to user B.
- User B verifies the reply message from user A and generates the corresponding reply message as $\{ID_s || ID_B || N_s || g^{rB} \bmod p\}$ to user A, where ID , N , and P represent the user ID, random nonce number, and prime number.
- User A received the message from user B and forwards the message to user S.
- User S verifies and generates the corresponding reply message by computing $K_{BS} = (g^{rB})^{rS} \bmod p$ and records user B as semi-neighbor in neighbor list S.

3.2 Cluster Formation

- Clusters are formed among the nodes in the network according to the distance between the nodes. A CH is associated with each group of clusters. The nodes discovered for packet transmission can send packets only to their CH instead of sending them for all the nodes.

- Clustering approach is used to address nodes heterogeneity to limit the amount of information propagated inside the network, thereby reducing the routing overhead and for speedy delivery of the packets.
- Among different clusters, one node that coordinates the cluster activities is CH. The CH monitors and identifies the intermediate nodes for packet transmission.
- The CH then distributes the packet inside the cluster and forwards them to be delivered to other CH or to their destination nodes. CH is the node having highest battery power in the network.
- The packets are routed from source node to the destination node via CH in less number of hops; thereby, it avoids the excessive overhead in the network and chooses the best optimal path during its transmission.

4 Results

Finally discovered the secure neighbors nodes using RSA algorithm, this is shown in Fig. 2. Discovering the Semi neighbor discovery scheme using RSA Algorithm, this is shown in Fig. 3.

Figure 4 shows the output for the cluster formation.

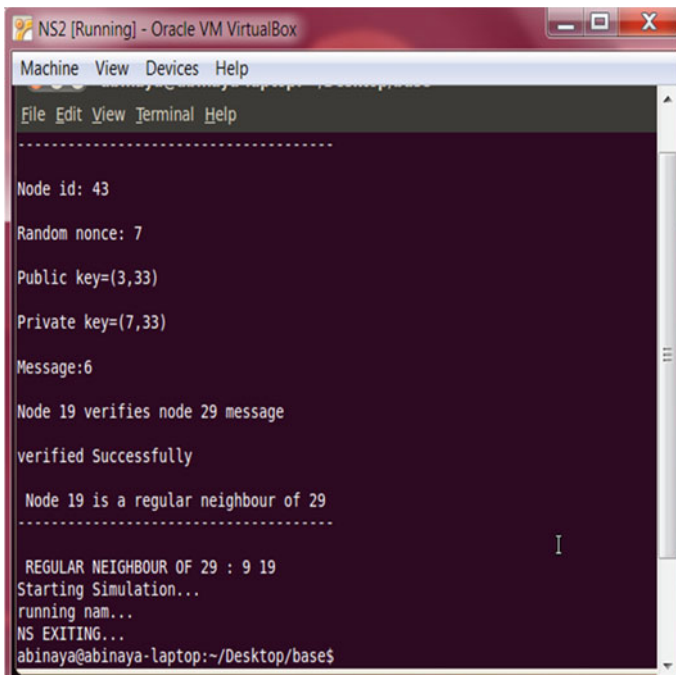


Fig. 2 Regular-neighbor discover

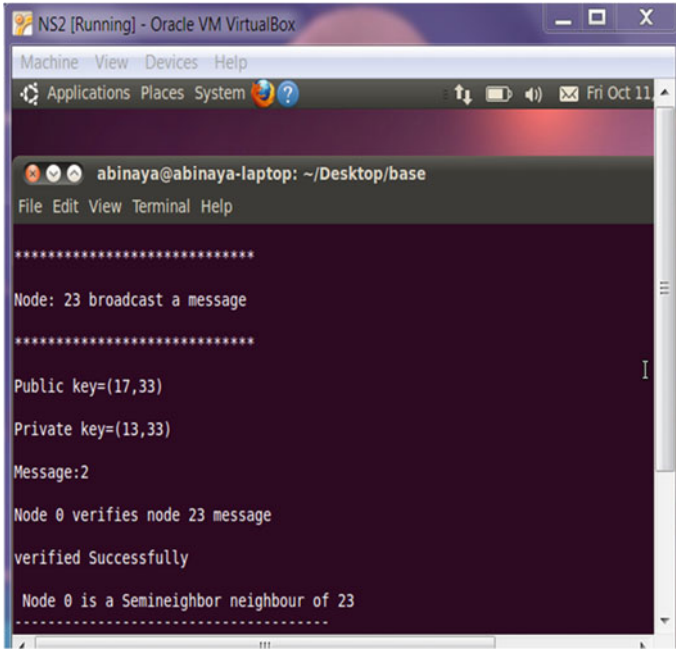


Fig. 3 Semi-neighbor discover



Fig. 4 Cluster formation


```
abinaya@abinaya-laptop: ~/Desktop/base
File Edit View Terminal Help
*****
Cluster Head : 24
Cluster Head : 27
Cluster Head : 16
Cluster Head : 13

ENTER SOURCE NODE :
1

ENTER DESTINATION NODE :
10

          PATH :
          *****

NODE   : 1
NODE   : 8
NODE   : 18
NODE   : 3
NODE   : 10
```

Fig. 5 Path discovery in cluster

Figure 5 shows the output for determining the intermediate nodes for the packet transmission.

5 Performance Evaluation

Figure 6 shows the calculation of packet loss.

Figure 7 shows the comparison of packet delivery ratio.

Figure 8 shows the comparison of network overhead.

Fig. 6 Calculation of packet loss

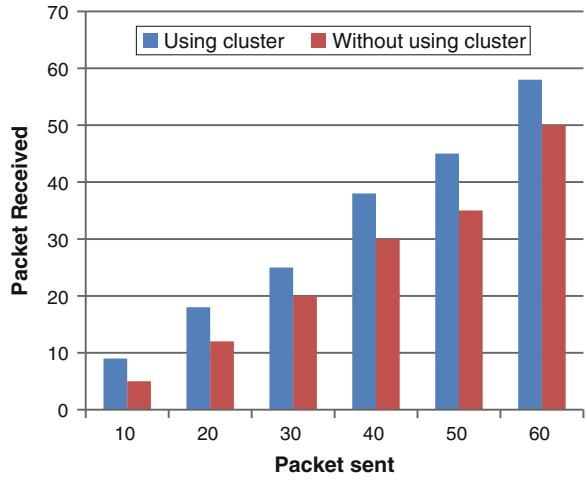


Fig. 7 Comparison of packet delivery ratio

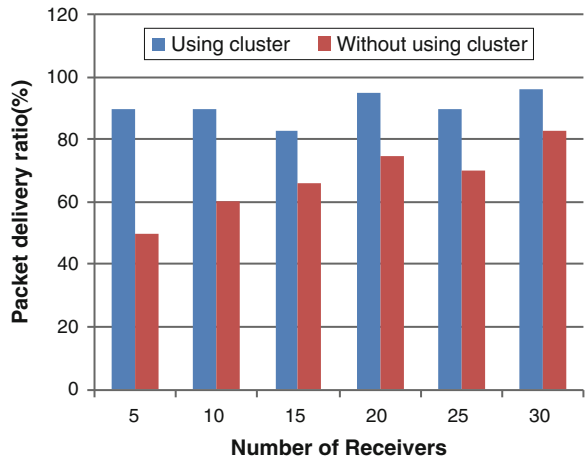
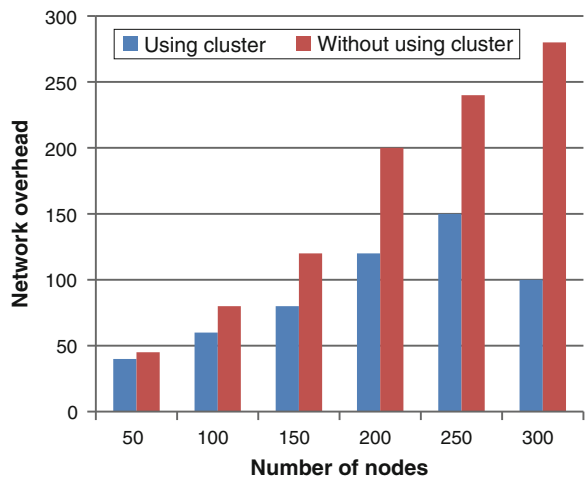


Fig. 8 Comparison of network overhead



6 Conclusion

MANET is a temporary network where every node joins and leaves a network. Still, ad hoc networks are subject to various types of attacks. Attack containment measures are used to minimize the effect of attacks. The proposed method ensures both the anonymity and security of the routing protocol with privacy preservation. Each user can identify as many neighbors as possible in its communication range by neighbor discovery scheme. This helps each user to obtain more resources from its neighbors. Anonymous and secure routing protocol provides secure data transmission, thus providing authentication, integrity, and anonymity.

References

1. Y.C. Hu, A. Perrig, D.B. Johnson, Ariadne: a secure on-demand routing protocol for ad hoc networks, in *8th annual international conference on mobile computing and networking (mobicom '02)* (2002)
2. Y.-C. Hu, D.B Johnson, A. Perrig, *Secure Efficient Ad Hoc Distance Vector (SEAD)* (Elsevier, 2003)
3. S. Lu, L. Li, K.-Y. Lam, L. Jia, SAODV: A MANET routing protocol that can withstand black hole attack, in *Computational Intelligence and Security, CIS '09*, vol. 2 (2009)
4. P. Nagrath, S. Mehla, B. Gupta, ARAN: analyzing security of authenticated routing protocol. *Int. J. Comput. Sci. Eng. (IJCSE)* **02**(03), 664–668
5. J. Kong, X. Hong, ANODR: anonymous on demand routing with untraceable routes for mobile adhoc networks, in *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (Computer Science Department University of California, 2003)
6. A. Boukerche, K. El-Khatib, L. Xu, SDAR: a secure distributed anonymous routing protocol for wireless and mobile ad hoc networks, in *29th Annual IEEE International Conference on Local Computer Networks* (University of Ottawa, 2004)
7. R. Song, L. Korba, G. Yee, *AnonDSR: Efficient Anonymous Dynamic Source Routing for Mobile Ad-Hoc Networks* (NRC Publications, 2005)
8. T. Sheltami, H.T. Mouftah, An efficient energy aware cluster head formation infrastructure protocol for MANETs, in *Eighth IEEE International Symposium on Computers and Communication* (2003)
9. X. Wang, H. Cheng, H. Huang, *Constructing a MANET Based on Clusters* (Springer Science +Business Media, New York, 2013)

Quality-of-Service Analysis of AOMDV and AOMDV-MIMC Routing Protocols for Mobile Ad hoc Networks

P. Periyasamy and E. Karthikeyan

Abstract Bandwidth scarcity is a major drawback in multi-hop ad hoc networks. When a single-interface single-channel (SISC) approach is used for both incoming and outgoing traffic, the bandwidth contention between nodes along the path has occurred as well as throughput is degraded. This drawback is overwhelmed by using multi-interface multi-channel (MIMC) approach as well as some of the quality-of-service (QoS) requirements have been enhanced. In this paper, we have applied MIMC approach to ad hoc on-demand multipath distance vector (AOMDV) routing protocol, called AOMDV-MIMC routing protocol, and its performance is compared with AOMDV routing protocol. The simulation results show the network lifetime, throughput, and packet delivery ratio of AOMDV-MIMC routing protocol have been tremendously improved than the AOMDV routing protocol.

Keywords SISC · MIMC · AOMDV-MIMC · Bandwidth scarcity · Network lifetime · Throughput · Packet delivery ratio

1 Introduction

In this modern world, wireless communication has become indispensable part of life. Research focuses on mobile ad hoc network (MANET), which is a collection of mobile devices by wireless links forming a dynamic topology without much physical network infrastructure such as routers, servers, access points/cables, or centralized administration. Each mobile device functions as router as well as node.

P. Periyasamy (✉)

Department of Computer Science and Applications, Sree Saraswathi Thyagaraja College,
Pollachi 642107, Tamil Nadu, India
e-mail: pereee@yahoo.com

E. Karthikeyan

Department of Computer Science, Government Arts College, Udumalpet 642126,
Tamil Nadu, India
e-mail: e_karthy@yahoo.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_56

511

The main characteristics of MANET are (i) dynamic topologies (ii) bandwidth-constrained links (iii) energy-constrained operation, and (iv) limited physical security [1, 2].

Most of the routing protocols in MANETs have been designed using a **single-interface single-channel (SISC)** approach. In this approach, single interface and single channel are commonly used for both incoming and outgoing traffic between nodes along the path. This leads to the bandwidth contention and throughput degradation issues. These issues can be tackled by using **multi-interface multi-channel (MIMC)** approach. The following are the major advantages of MIMC [3, 4] approach:

- (i) *Capacity Enhancement*: Sending and receiving of data packets by the forwarding nodes at the same time.
- (ii) *Load Sharing*: In order to increase robustness and lower latency, the traffic flow is distributed among the available connections.
- (iii) *Channel Failure Recovery*: Channel errors are possibly avoided because of multiple interfaces and multiple channels.

In this paper, we have applied multi-MIMC approach to ad hoc on-demand multipath distance vector (AOMDV) routing, called AOMDV-MIMC routing protocol, and its performance is compared with AOMDV routing protocol in terms of quality-of-service (QoS) requirements.

The rest of the paper is organized as follows: Sect. 2 briefly discusses the related work. In Sect. 3, we have presented multi-interface and multi-channel approach to AOMDV routing protocol, called **AOMDV-MIMC routing protocol**. In Sect. 4, the QoS metrics are given. In Sect. 5, the simulation and experimental results are discussed. In Sect. 6, the conclusions and future work are given.

2 Related Work

Many researchers have proposed many different approaches to MAC for utilizing multi-channel and multi-interface in MANETs. In [5], the authors proposed a centralized channel assignment scheme where traffic is directed toward specific gateway nodes in static networks. A hybrid channel assignment scheme [6] assigns some radios statically to a channel and some are dynamically changed their frequencies in the channel. A new channel assignment scheme [7] for utilizing multi-channels can reduce channel conflicts by removing hidden channel problem [8].

All the protocols in NS 2.34 have only SISC support because IEEE 802.11 a/b/g requires some modifications on MAC and link layer protocols in order to utilize multi-MIMC. In the MIMC [6, 9, 10] approach, the following are the designs of routing interface and channel assignments:

1. A solution for multi-interface that exploits multiple channels can be implemented on existing IEEE 802.11 hardware.

2. An interface assignment strategy using interface switching techniques simplifies the coordination among nodes through the utilization of available multiple channels.
3. A multiple-channel routing (MCR) scheme selects the routes with the highest throughput by accounting the cost of channel diversity and interface switching.

The modifications on MAC and link layer protocols were done by applying Kyasanur's and Vaidya's [6] interface assignment scheme because this scheme is more flexible and versatile among other schemes. Implementation of this scheme on MAC (medium access control) and LL (link layer) protocols is carried out using the technical report [3].

In [11], the optimal channel assignment and routing problem in wireless mesh networks is overwhelmed. In a distributed algorithm [12], a node has the number of available channels less than twice the number of network interfaces. In such case, channels are randomly assigned to network interfaces as there is a guarantee that a common channel can be found between any pair of nodes through the pigeonhole principle, otherwise skeleton assisted channel assignment scheme is used.

We propose several modifications on AOMDV routing protocol to utilize the multi-MIMC scheme efficiently, called **AOMDV-MIMC routing protocol**, to improve network performance.

3 AOMDV-MIMC Routing Protocol

3.1 AOMDV (Base Protocol)

AOMDV [13] is the extension of AODV [14] so as to eliminate the occurrence of frequent link failures and route breaks in highly dynamic ad hoc networks. It adds some extra fields in routing tables and control packets and follows the two rules during a route discovery phase in order to compute loop-free and link-disjoint multiple routes between source and destination. These rules are (i) a route update rule establishes and maintains multiple loop-free paths at each node and (ii) a distributed protocol finds link-disjoint paths. Link failures may occur because of node mobility, node failures, congestion in traffic, packet collisions, and so on.

There is no any common link among the multiple routes between a source and destination pair in the link-disjoint routes. To achieve loop freedom, every node maintains a variable called the advertised hop count. The advertised hop count is added in each RREQ (route request) or RREP (route reply), and in addition to the routing table, it has the usual fields that are used for AODV. The advertised hop count field of a node is set to the length of the longest available path to the destination expressed in terms of the number of hops if it initiates a RREQ or RREP with a particular destination sequence number and it remains unchanged until the associated destination sequence number is changed.

The loop-freedom rule says that if a node receives a RREQ (RREP) for a particular destination with a destination sequence number: (a) it should update its

routing information with the information obtained from the received RREQ (RREP) if the destination sequence number is higher than the one stored in its routing table; (b) it can re-send the received RREQ (RREP) when the advertised hop count in the RREQ (RREP) is greater than the corresponding value in its routing table if the destination sequence number is equal to the one stored in its routing table; and (c) it can update its routing table with the information contained in the received RREQ (RREP) when the advertised hop count in the RREQ (RREP) is less than the corresponding value in its routing table if the destination sequence number is equal to the one stored in its routing table.

For link disjointness, each node maintains a route list in its routing table for a particular destination and its route list contains the next hop, last hop, and hop count information for the destination. The next hop represents a downstream neighbor through which the destination can be reached. The last hop refers to the node immediately preceding the destination. The hop count is used to measure the distance from the node to the destination through the associated next and last hops.

The link disjointness among all the paths can be achieved if a node can ensure that those paths to a destination from itself differ in their next and last hops. Using this observation, AOMDV ensures link disjointness among multiple routes for the same source and destination pair and also adds a last hop field in each RREQ and RREP.

In AOMDV, all copies of an RREQ are examined for potential alternate reverse paths during route discovery. Upon receiving an RREQ, an intermediate node creates a reverse path if the RREQ satisfies the rules for loop freedom and link disjointness. Moreover, it checks whether it has one or more valid next hop entries for the destination. The intermediate node generates an RREP and sends it back to the source along the reverse path if such an entry is found. Otherwise, it rebroadcasts the RREQ. The destination follows the same rules for creating reverse paths if it receives RREQ copies. Unlike the intermediate nodes, it generates an RREP for every copy of RREQ that arrives via a loop-free path, for increasing the possibility of finding more disjoint routes.

3.2 Multi-interface Multi-channel Extension to AOMDV Protocol

Tables 1 and 2 show the route list structure of AOMDV and AOMDV-MIMC routing protocols, respectively. Using [3, 10], AOMDV-MIMC protocol has been developed by extending the AOMDV routing protocol.

AOMDV-MIMC routing protocol uses channel assignment and interface switching strategies in order to utilize multiple channels and multiple interfaces. If the **interface queue** associates with K channels and M interfaces, only one interface is fixed and the remaining interfaces are switchable as shown in Fig. 1.

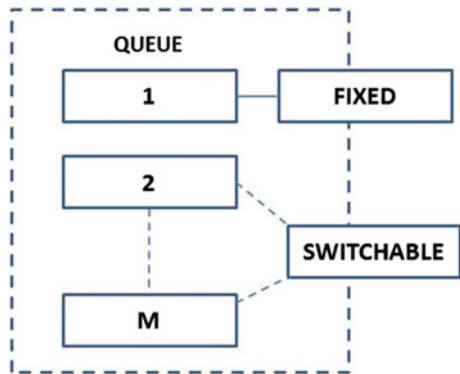
Table 1 Route list structure of AOMDV

hop_count1	next_hop1	last_hop1	expiration_timeout1
hop_count2	next_hop2	last_hop2	expiration_timeout2

Table 2 Route list structure of AOMDV-MIMC

hop_count1	next_hop1	last_hop1	expiration_timeout1	interface1
hop_count2	next_hop2	last_hop2	expiration_timeout2	interface2

Fig. 1 Illustration of the interface queue associated with K channels and M interfaces



The functions of AOMDV such as `command()`, `handle()`, `recvRequest()`, `forward()`, `sendRequest()`, `sendReply()`, `sendError()`, `sendHello()`, and `forwardReply()` have been modified to adopt with MIMC in order to improve the throughput and network’s lifetime. Each node in AOMDV-MIMC uses both fixed and switchable interfaces and multiple channels. In other words, each node uses one fixed interface and the remaining are switchable interfaces as shown in Fig. 2.

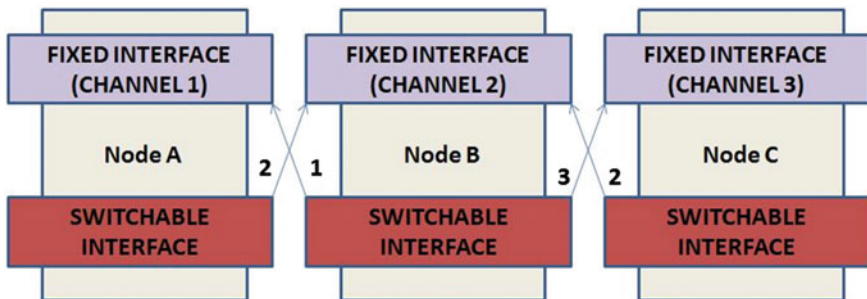


Fig. 2 Example of AOMDV-MIMC protocol interface switching operation with $K = 3$ channels and $M = 2$ interfaces

4 Quality-of-Service (QoS) Metrics

QoS metrics [15, 16] is a quantitative measure that can be used to evaluate any ad hoc routing protocol. The following metrics are considered in order to compare the performance of on-demand multipath routing protocols AOMDV and AOMDV-MIMC, respectively, in terms of variation in speed under random waypoint (RWM) in constant bit rate (CBR) traffic. The number of bits transferred per second through the traffic medium is called ‘**network load,**’ and the time taken by a node to choose the destination for packet delivery is called ‘**pause time.**’

- (i) **Packet Delivery Ratio** [15, 16]: It is the ratio of data packets delivered to the destination to those generated by the sources.
- (ii) **Throughput** [15, 16]: It is the number of bytes received successfully.
- (iii) **Normalized Routing Overhead** [15, 16]: It is the number of routing packets transmitted per data packet toward destination.
- (iv) **Average End-to-End Delay** [15, 16]: It is the average time of the data packet to be successfully transmitted across a MANET from source to destination. It includes all possible delays such as buffering during the route discovery latency, queuing at the interface queue, retransmission delay at the MAC, the propagation, and the transfer time.
- (v) **Network Lifetime** [17]: It is defined as the duration from the beginning of the simulation to the first time a node runs out of energy.
- (vi) **Total Energy Consumed by Nodes** [17]: It is the summation of the energy consumed by all nodes in the simulation environment, i.e., *energy consumed by a node = initial energy of that node – residual energy of that node.*

5 Simulation and Experiment

5.1 Simulation Model

The performance comparison of AOMDV and AOMDV-MIMC routing protocols is evaluated using NS 2.34 [18, 19, 20]. The AOMDV-MIMC uses 3 channels and 2 interfaces. The simulation is carried out for a network with 100 nodes and a dimension of 2200 × 600 m on continuous mobility (i.e., no pause) under RWM mobility model. The CBR with 40 maximal connections, 4 packets per second, and 512 bytes per data packet with maximum speed of 5, 10, 15, 20, and 25 are used. We set the initial energy per node as 100 Joules and transmitting as well as receiving power is 12.7. The two-ray-ground reflection model is used as propagation model. Simulations are run for 100 simulated seconds. The result of simulation is generated as trace files and the awk and perl script are prepared to analyze the trace files and produces the reports as shown in Fig. 3.

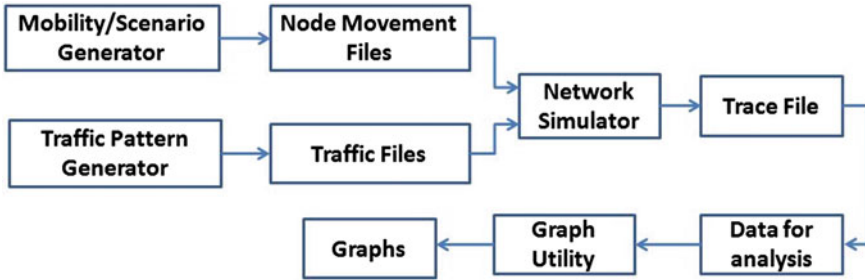


Fig. 3 Overview of the simulation model

5.2 Results and Discussions

The packet delivery ratio, throughput, and network lifetime of AOMDV-MIMC routing protocol is very much improved than the packet delivery ratio, Throughput and network lifetime of AOMDV are shown in Figs. 4, 5 and 8, respectively. Normalized routing overhead, average end-to-end delay, and total energy consumed by nodes of AOMDV-MIMC routing protocol are too high than the normalized routing overhead, average end-to-end delay, and total energy consumed by nodes of AOMDV, as shown in Figs. 6, 7 and 9, respectively.



Fig. 4 Packet delivery ratio (%)

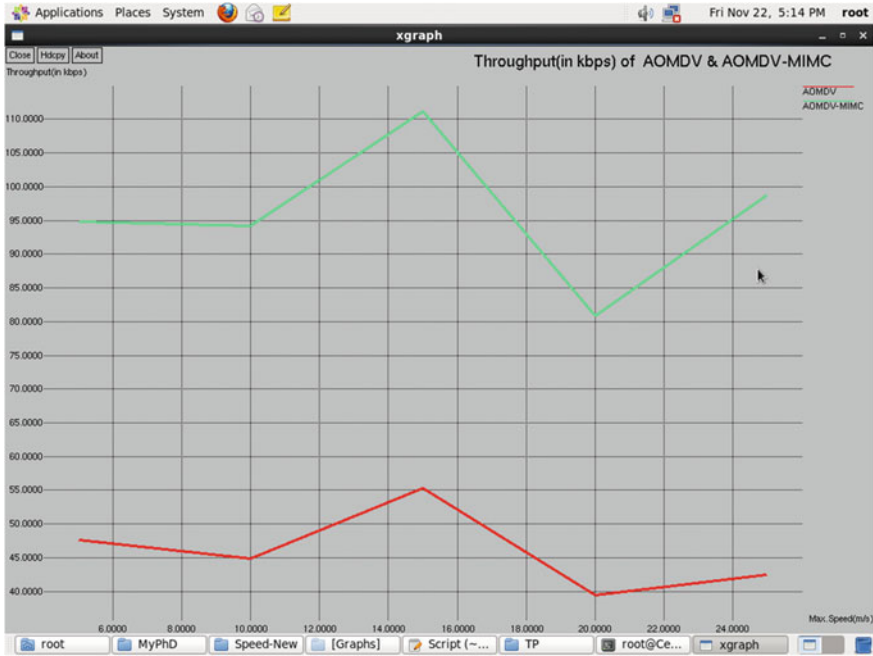


Fig. 5 Throughput (in Kbps)

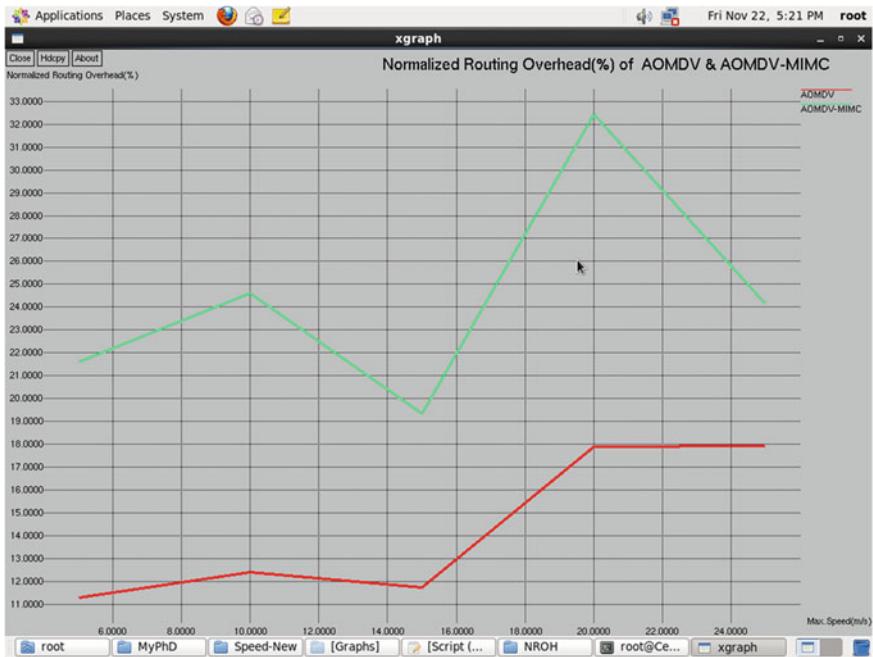


Fig. 6 Normalized routing overhead (%)

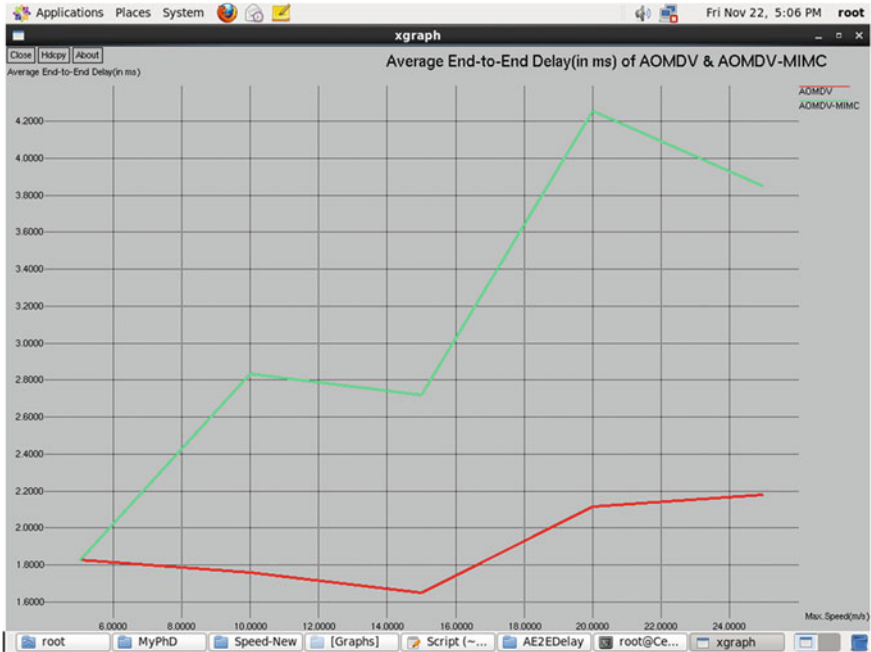


Fig. 7 Average end-to-end delay (in ms)

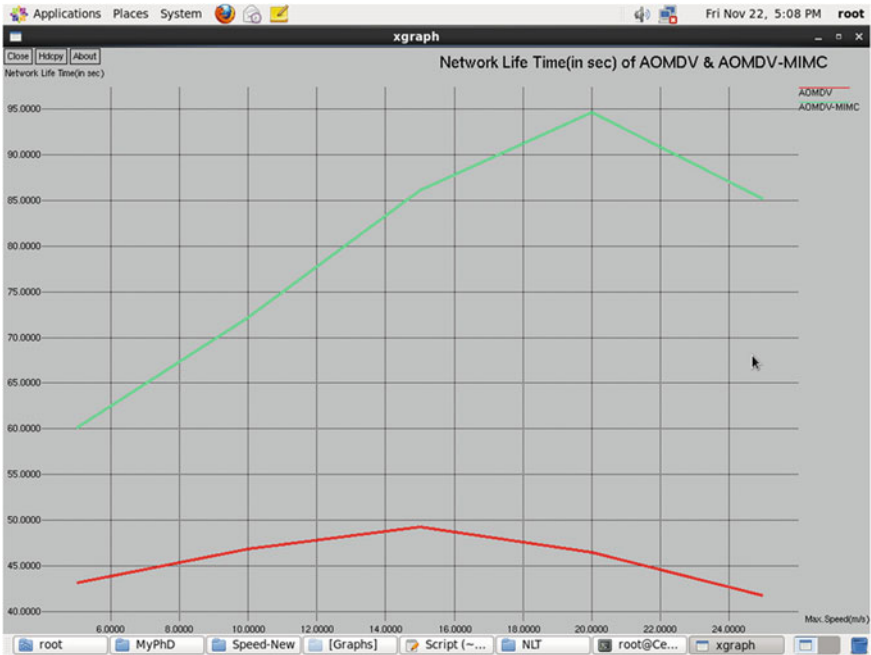


Fig. 8 Network lifetime

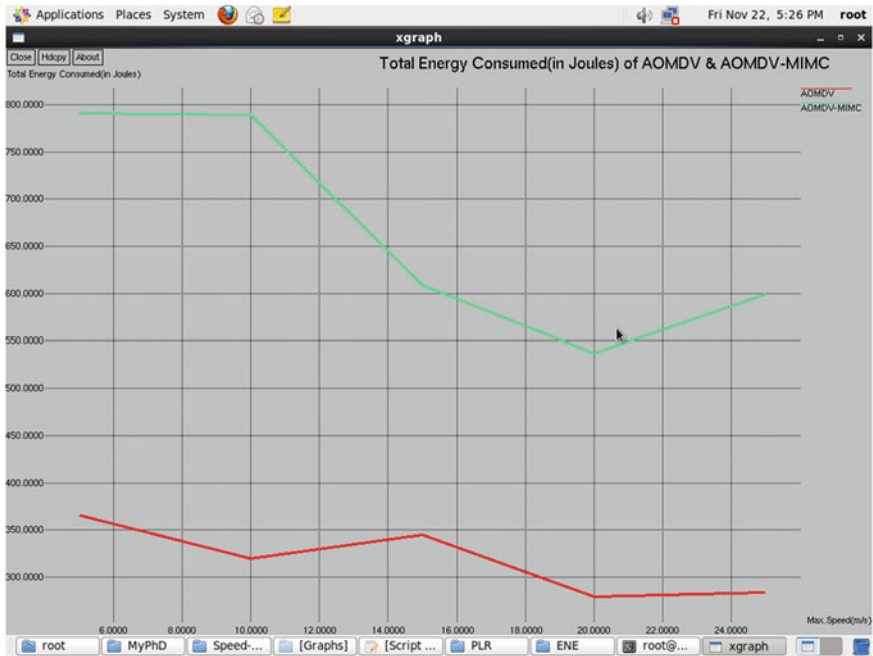


Fig. 9 Total energy consumed by nodes

6 Conclusions and Future Work

The packet delivery ratio, network lifetime, and throughput of AOMDV-MIMC routing protocol are very much improved than the packet delivery ratio, network lifetime, and throughput of AOMDV routing protocol. Total energy consumed by nodes, average end-to-end delay, and normalized routing overhead of AOMDV-MIMC routing protocol are too high than the total energy consumed by nodes, average end-to-end delay, and normalized routing overhead of AOMDV routing protocol. In future, we will concentrate on the reduction of either the average end-to-end delay or the total energy consumed by nodes or both in AOMDV-MIMC routing protocol because its performance is outstanding than the performance of AOMDV routing protocol.

References

1. E.M. Royer, C.K. Toh, A review of current routing protocols for ad hoc mobile wireless networks. *IEEE Pers. Commun.* **6**(2), 46–55 (1999)
2. M. Abolhasan, T. Wysocki, E. Dutkiewicz, A review of routing protocols for mobile ad hoc networks. *Ad Hoc Netw.* **2**(1), 1–22 (2003)

3. S.R.A. Calvo, J.P. Campo, *Adding Multiple Interface Support in NS-2* (University of Cantabria, 2007)
4. T.T. Luong, B.S. Lee, C.K. Yeo, Channel Allocation for Multiple Channels Multiple Interfaces Communication in Wireless Ad Hoc Networks. Lecture notes in computer science, vol. 4982 (Springer, Heidelberg, 2008), pp. 87–98
5. A. Raniwala, K. Gopalan, T. Chiueh, Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks. *Mobile Comput. Commun. Rev. (MC2R)* **8**(2), 50–65 (2004)
6. P. Kyasanur, N.H. Vaidya, Routing and interface assignment in multi-channel multi-interface wireless networks, in *Proceedings of IEEE WCNC*, vol. 4, pp. 2051–2056 (2005)
7. H. Nguyen, U. Nguyen, Channel assignment for multicast in multi-channel multi-radio wireless mesh networks. *Wirel. Commun. Mobile Comput.* **9**(4), 557–571 (2008)
8. C. Cherred, K. Pradeep, S. Jungmin, N.H. Vaidya, Multi-channel mesh networks: challenges and protocols. *IEEE Wirel. Commun.* (2006)
9. G. Zeng, B. Wang et al., Multicast algorithms for multi-channel wireless mesh networks, in *IEEE ICNP* (2007)
10. P. Kyasanur, N. Vaidya, Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks. *ACM SIGMOBILE MC2R* **10**(1), 31–43 (2006)
11. S. Avallone, I.F. Akyildiz, A channel assignment algorithm for multi-radio wireless mesh networks. *Comput. Commun.* **31**(7), 1343–1353 (2008)
12. M. Shin, S. Lee, Y. Kim, Distributed channel assignment for multi-radio wireless networks, in *IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, pp. 417–426 (2006)
13. M.K. Marina, S.R. Das, Ad hoc on-demand multipath distance vector routing. *Wirel. Commun. Mobile Comput.* **6**(7), 969–988 (2006)
14. S. Das, C. Perkins, E. Royer, Ad hoc on demand distance vector (AODV) routing, in *IETF RFC3561* (2003)
15. C.P. Agrawal, O.P. Vyas, K. Tiwari, Evaluation of varying mobility models & network loads on DSDV protocol of MANETs. *Int. J. Comput. Sci. Eng.* **1**(2), 40–46 (2009)
16. P. Periyasamy, E. Karthikeyan, Performance evaluation of AOMDV protocol based on various scenario and traffic patterns. *Int. J. Comput. Sci. Eng. Appl. (IJCSEA)* **1**(6), 33–48 (2011)
17. Y. Liu, L. Guo, H. Ma, T. Jiang, Energy efficient on demand multipath routing protocol for multi-hop ad hoc networks, in *IEEE 10th International symposium on Spread spectrum and applications, ISSSTA-08*, pp. 592–597 (2008)
18. V. Singla, P. Kakkar, Traffic pattern based performance comparison of reactive and proactive protocols of mobile ad-hoc networks. *Int. J. Comput. Appl.* **5**(10), 16–20 (2010)
19. K. Fall, K. Varadhan, *The ns Manual* (2012). University of Southern California, Information Sciences Institute (ISI). Available <http://www.isi.edu/nsnam/ns/ns-documentation.html>
20. NS-2 with Wireless and Mobility Extensions (2012). Available <http://www.monarch.cs.cmu.edu>

Particle Swarm Optimization-Based SONAR Image Enhancement for Underwater Target Detection

P.M. Rajeshwari, G. Kavitha, C.M. Sujatha and Dhilsha Rajapan

Abstract In the proposed work, particle swarm optimization (PSO) is employed to enhance SONAR images in spatial domain. A transformation function is used with the local and global data of the image to enhance the image based on PSO. The objective function considers the entropy of the image, number of edges detected, and the sum of edge intensities. PSO determines the optimal parameters required for image enhancement. PSO-based image enhancement is compared with median filter and adaptive Wiener filter both qualitatively and quantitatively. Mean square error (MSE) and peak signal-to-noise ratio (PSNR) are used as quantitative measures for the analysis of the enhanced images. MSE and PSNR analysis reveals that original details of the image are retained after enhancement. Based on the visual and quantitative analysis, it is considered that PSO-based technique provides better enhancement of SONAR images.

Keywords Particle swarm optimization · Mean square error · PSNR · Ant colony optimization · Artificial bee colony method

P.M. Rajeshwari (✉) · D. Rajapan
Marine Sensors Systems, National Institute of Ocean Technology, Chennai, India
e-mail: pmr@niot.res.in

D. Rajapan
e-mail: krd@niot.res.in

G. Kavitha
Madras Institute of Technology, Anna University, Chennai, India
e-mail: kavithag_mit@annauniv.edu

C.M. Sujatha
College of Engineering, Guindy, Anna University, Chennai, India
e-mail: sujathacm@annauniv.edu

1 Introduction

Image enhancement provides a better assessment of human view. It serves as an input for different military and civilian applications which include underwater target detection. Underwater target is generally imaged using SONAR and detected by magnetometer. Sonar image is made up of received echo data which are charted to intensities after performing signal processing. The image obtained from the SONAR may be corrupted due to noise and insufficient illumination. For better detection of objects, preprocessing techniques are required to enhance the image and give a clear understanding of the object or targets from the background. Image preprocessing is the initial step carried out, prior to segmentation. Histogram equalization is one of the enhancement methods in spatial domain, which attempts to scale the image contrast and broadens the histogram. The limitation of histogram equalization is that it sometimes fails to preserve the region of interest of the input image [1].

Image preprocessing also includes removal of noise from the image. SONAR images are corrupted by speckle noise, Gaussian noise, and salt and pepper noise [2]. Speckle noise reduction has been carried out using stationary wavelet transform and ant colony optimization (ACO) method [3]. Gaussian noise, Poisson noise, and salt and pepper noise have been eliminated using adaptive thresholding in framelet transform domain [4]. Al-Sbou [5] has proposed artificial neural network to tackle noise. Statistical features are obtained from noisy image and are used for training the network. Recently, many evolutionary algorithms such as particle swarm optimization (PSO), ACO, artificial bee colony method (ABC), and cuckoo search (CS) are employed along with conventional algorithm to enhance image. It has been found that CS and ABC could perform better preprocessing for satellite images [6].

PSO has been applied for contrast enhancement, and it has been reported that PSO-based method provided better results when compared to genetic algorithms [7]. Gorai and Ghosh [8] have also utilized PSO-based image enhancement for bench mark images. Zhang et al. [9] proposed an adaptive local enhancement algorithm based on PSO for side scan SONAR images.

In the present work, enhancement of SONAR images is carried out using PSO-based technique. The result obtained using PSO method is compared with median and adaptive Wiener filter methods both subjectively and quantitatively.

The paper is organized as follows: In Sect. 2, methodology of image acquisition and application of PSO algorithm for image enhancement is described. In Sect. 3, results and discussion are presented, and finally, in Sect. 4, conclusion of the work is presented.

2 Methodology of Image Acquisition

The image ‘table’ considered in this study is acquired at National Institute of Ocean Technology (NIOT). This image is generated by buried object-scanning SONAR system developed by NIOT. Table of 0.305×0.305 m is suspended at 4 m from the

surface of the tank at the Acoustic Test Facility of NIOT with tank dimension of $16 \times 9 \times 7$ m. Using data acquisition system, analog signals are acquired for 16 channels. All the channels are beam formed, and temporal image is obtained. For generation of a spatial image, beam-formed angles are varied. The other images, namely wreck, bridge support, and hurricane gate, are obtained from image gallery of M/s Edge Tech.

2.1 PSO Algorithm for Image Enhancement

Image enhancement is carried out in spatial domain. The images considered include buried object-scanning SONAR and side scan SONAR image. For the considered image of size $M \times N$, where M is the number of rows and N is the number of columns, a new intensity image is formed using the transformation [8] given by

$$g(i,j) = K(i,j)[f(i,j) - c m(i,j)] + m(i,j)^a. \quad (1)$$

where $g(i, j)$ is the new transformed image, $K(i, j)$ is the enhancement function which considers local and global data of the image, $f(i, j)$ is the actual input image, $m(i, j)$ is the local mean, and a and c are the parameters to be optimized. The local mean [8] of kernel size of $n \times n$ ($n = 3$) is defined as

$$m(i,j) = \frac{1}{n \times n} \sum_{i=1}^n \sum_{j=1}^n f(i,j). \quad (2)$$

The enhancement function $K(i, j)$ is defined as

$$K(i,j) = \frac{kD}{\sigma(i,j) + b}. \quad (3)$$

where k and b are the parameters to be optimized. D is the global mean [8] defined as

$$D = \frac{1}{M \times N} \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} f(i,j). \quad (4)$$

where σ is the standard deviation [8] of kernel size 3×3 defined as

$$\sigma(i,j) = \left\{ \frac{1}{M \times N} \sum_{i=1}^{i=M-1} \sum_{j=1}^{N-1} (f(i,j) - m(i,j))^2 \right\}^{0.5}. \quad (5)$$

Initially ‘ n ’ particles are created each with four parameters a , b , c , and k . The parameters a , b , c , and k are determined using PSO. Each particle is assigned random velocity within the limits reported [9]. The objective function considers the sum of intensity edges, entropy of the image, and number of edge pixels. The objective function [8] is defined as:

$$F(I_e) = \log(\log E(I_s)) \times \frac{\text{number of edges } (I_s)}{M \times N} \times H(I_e). \quad (6)$$

where I_e is the enhanced image, $E(I_s)$ the sum of pixel intensities of Sobel edge image I_s , no. of edges (I_s) is the number of the pixels, M is the number of rows, N is the number of column, and $H(I_e)$ is the entropy of the enhanced image [8] and is given by

$$H(I_e) = - \sum_{i=0}^{255} h_i \log_2(h_i). \quad (7)$$

where h_i is the probability of occurrence for i th intensity.

Personal best (P_{best}) and global best (G_{best}) are found by the value obtained from objective function fitness value. P_{best} provides the best solution achieved by the particle in each iteration. G_{best} is the best solution traced by the particle throughout all iterations of the swarm. After determination of G_{best} and P_{best} , particle updates its velocity and position. The update of velocity [8] is given by

$$V_i^{t+1} = WV_i^t + c_1 r_1^t [P_{\text{best}i}^t - X_i^t] + c_2 r_2^t [G_{\text{best}} - X_i^t]. \quad (8)$$

where V_i^t is the velocity of i th particle at iteration t , W is the weight inertia, c_1 and c_2 are acceleration constants used to level the cognitive and social component, respectively, r_1 and r_2 are random number between 0 and 1, X is current position of i th particle at iteration t , $P_{\text{best}i}$ is personal best of i th particle, and G_{best} is global best value of the group.

The new position of the particles [8] is given by

$$X_i^{t+1} = V_i^{t+1} + X_i^t. \quad (9)$$

where X_i^t is the position vector which contains parameters a , b , c , and k values. The updated velocity is added with the position vector to obtain the new position vector. Using these optimized values, a transformed image is generated. Then, Sobel operation is applied to the image. Using objective function, fitness value is evaluated. From that fitness value, P_{best} and G_{best} are determined for each particle. Using Eqs. 8 and 9, velocity and position updates are created. Iterations are repeated till objective function converges.

2.1.1 Optimization of PSO Parameters

For PSO-based image enhancement, parameters a , b , c , and k require optimization. In addition to that, an inertia weight ‘ W ’ is used in the algorithm. In this work, a decreasing linear weight [10] is used and defined as follows:

$$W = W_{\max} - \frac{W_{\max} - W_{\min}}{\text{iteration}} \times t. \quad (10)$$

where ‘ t ’ is the i th iteration. In the present work, W_{\max} is found to be 1.1 and W_{\min} is 0.1. c_1 and c_2 vary from 0 to 2. The values of c_1 and c_2 are chosen to be 2. Parameter ‘ a ’ ranges from 0 to 1.5, and ‘ b ’ varies from 1 to $D/2$, where D is the global mean of the image. Parameter ‘ c ’ ranges from 0 to 1, and ‘ k ’ ranges from 0.5 to 1.5.

2.2 Comparison with Other Methods

The other methods used for comparison include median and adaptive Wiener filters. Median filtering is a nonlinear operation [11] often used in image processing to reduce noise. Filtered output image is created by the median value of a 3×3 kernel. The 3×3 kernel is moved across the entire image, and center pixel is replaced by the median of 3×3 neighborhood pixels. Adaptive Wiener filter [12] filters a grayscale image that has been degraded by constant power additive noise. The filter uses a pixelwise adaptive Wiener method based on statistics estimated from a local neighborhood of each pixel. The various methods considered are compared based on mean square error (MSE) [13] and peak signal-to-noise ratio (PSNR) [13].

3 Results and Discussion

PSO-based enhanced technique is employed on image ‘table’ obtained from NIOT and on images from M/s Edgetech image gallery. PSO-based image enhancement is compared with median and adaptive Wiener filter.

3.1 Subjective Analysis

The ‘table’ image acquired at NIOT is shown in Fig. 1a. It is observed that object edges are not distinctive. Image enhancement by PSO is carried out for table image and is shown in Fig. 1b. Figure 1c, d shows the images subjected to median and adaptive Wiener filter, respectively. There is no visually prominent difference in these enhanced images. Enhancement could be detected with quantitative analysis only.

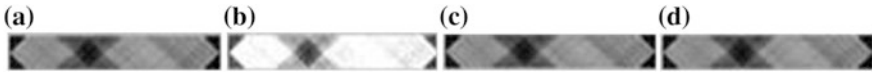


Fig. 1 a Original ‘Table’ image, b enhanced image using PSO, c median filter, and d adaptive Wiener filter

Figure 2a shows the histogram of the ‘table’ image. Contrast is less for this image as the maximum intensity value has extended only up to 200. Figure 2b depicts the histogram of PSO-based enhanced image. It is observed that the histogram has stretched to maximum intensity of 255 which in turn helps in segmentation when there are multiple objects present. From Fig. 1b, it is observed that the object alone is enhanced which is in dark shade and all the back ground is white which agrees with the histogram Fig. 2b.

Figure 3a shows the original image of wreck. The image enhanced by PSO method is shown in Fig. 3b. From this figure, it is evident that visually, contrast is enhanced when compared to original image. Figure 3c illustrates the image pre-processed by median filter and adaptive Wiener filter. The edges of the ship wreck are enhanced for Fig. 3c, d, but the image is slightly blurred.

Figure 4a, b shows the original image of bridge support and PSO-based enhanced image, respectively. From Fig. 4b, it is observed that the contrast of the image is enhanced and little debris lying near the bridge support can be seen clearly when compared to Fig. 4c, d which is preprocessed by median and Wiener filter, respectively.

Figure 5a, b demonstrates the original image of hurricane gate and PSO-based enhanced image, respectively. From Fig. 5b, it is noted that the edge intensities

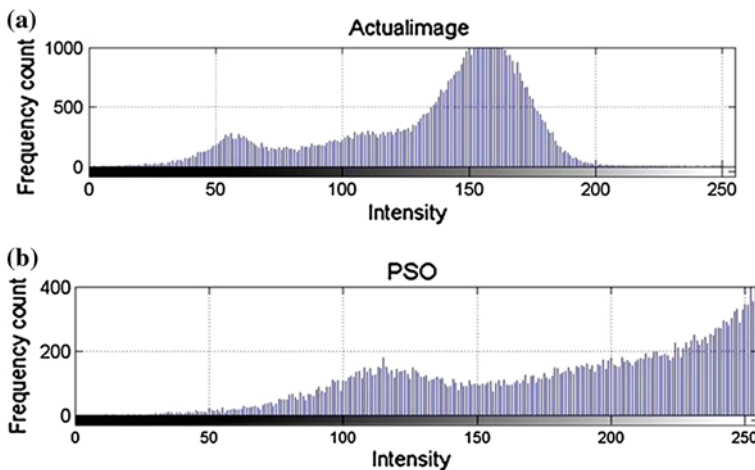


Fig. 2 a Histogram of original image ‘table’ and b histogram of enhanced image ‘table’ based on PSO

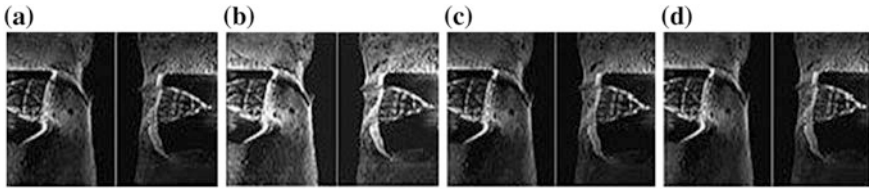


Fig. 3 **a** Original ‘wreck’ (image courtesy of M/s Edgetech), **b** enhanced image using PSO, **c** median filter, and **d** adaptive Wiener filter

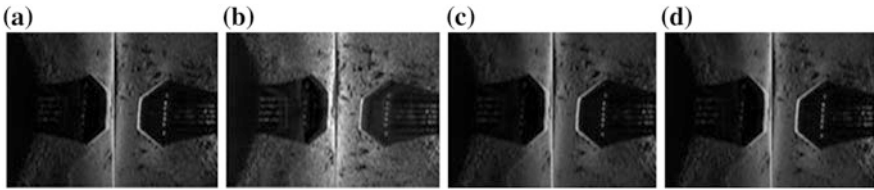


Fig. 4 **a** Original bridge support (image courtesy of M/s Edgetech), **b** enhanced image using PSO, **c** median filter, and **d** adaptive Wiener filter



Fig. 5 **a** Original ‘hurricane gate’ (image courtesy of M/s Edgetech), **b** enhanced image using PSO, **c** median filter, and **d** adaptive Wiener filter

have considerably increased. Blurring of edges is observed for Fig. 5c, d which is preprocessed by median and Wiener filter, and in addition to that, the ripples near the image hurricane support were blurred. Histogram analysis is also carried out for the other images considered and validated with the subjective analysis results.

3.2 Objective Analysis

Eye assessments of humans for determining image enhancement vary from person to person. Hence, quantitative analysis is also considered. MSE and PSNR are the two parameters used for quantitative analysis.

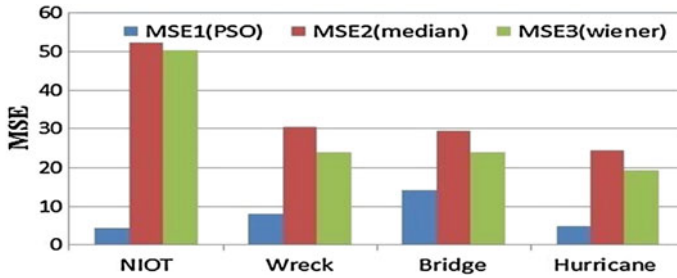


Fig. 6 Comparison of MSE for PSO-enhanced images, median filter, and adaptive Wiener filter

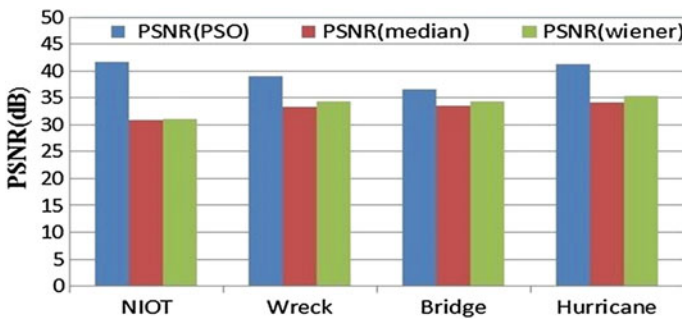


Fig. 7 Comparison of PSNR (dB) for PSO-based enhancement, median filter, and adaptive Wiener filter for different images

Figure 6 shows the comparison of MSE for four different enhanced images using PSO, median, and adaptive Wiener filtering techniques. Figure 7 shows the comparison of PSNR (dB) for the different images using the three techniques. From Figs. 6 and 7, it is very evident that MSE is less and PSNR is more for PSO-based enhancement method when compared to median and adaptive Wiener filters. MSE is less for table and hurricane gate images. Better enhancement is accomplished by optimizing the PSO parameters for all the images considered. PSNR values for images table, wreck, bridge support, and hurricane support are 42, 39, 36.5, and 41.1 dB, respectively.

4 Conclusion

An attempt has been made to enhance the image using PSO for buried object-scanning SONAR and side scan SONAR images. For optimization, 10 particles are initialized and 10 iterations are made. A linear decreasing weight function is utilized which ranges from 1.1 to 0.1. The results are compared with median filter and

Wiener filter preprocessing methods. The qualitative (subjective) analysis shows better performance of PSO-based enhancement compared to other methods. MSE- and PSNR-based quantitative analysis was performed on the images. MSE was found less. PSNR was maximum with PSO-based enhancement method when compared to the other methods.

Acknowledgments Authors would like to acknowledge M/s Edge Tech for image courtesy wreck, bridge support, and hurricane gate, and the Director, NIOT, for encouragement and permitting to publish the work.

References

1. P. Shanmugavadivu, K. Balasubramanian, K. Somasundaram, Modified histogram equalization for image contrast enhancement using particle swarm optimisation. *Int. J. Comput. Sci. Eng. Infor. Technol.* **1**, 13–27 (2011)
2. G. Padmavathi, P. Subashini, M. Kumar, S.K. Thakur, Performance analysis of non linear filtering algorithms for underwater images. *Int. J. Comput. Sci. Infor. Secur.* **6**, 232–238 (2009)
3. J. Alavandan, S. Santhosh Baboo, Enhanced speckle filters for SONAR images using stationary wavelet and hybrid inter- and intra-scale wavelet coefficient dependency. *Global J. Comput. Sci. Technol.* **12**, 12–19 (2012)
4. S. Sulochana, R. Vidhya, Image denoising using adaptive thresholding in framelet transform domain. *Int. J. Adv. Comput. Sci. Appl.* **3**, 192–196 (2012)
5. Y.A. Al-Sbou, Artificial neural networks evaluation as an image denoising tool. *World Appl. Sci. J.* **17**, 218–227 (2012)
6. V. Soni, A.K. Bhandari, A. Kumar, G.K. Singh, Improved sub-band adaptive thresholding function for denoising of satellite image based on evolutionary algorithms. *IET Sig. Process.* **7**, 720–730 (2013)
7. M. Braik, A. Sheta, Aladdin Ayesh: image enhancement using particle swarm optimization. *Proc. World Congr. Eng.* **01**, 696–701 (2011)
8. A. Gorai, A. Ghosh, Gray-level image enhancement by particle swarm optimization, in *World Congress on Nature and Biologically Inspired Computing*, pp. 72–76 (2009)
9. T. Zhang, L. Wan, Y. Xu, Y. Lu, Sonar image enhancement based on particle swarm optimisation, IN *3rd IEEE Conference on Industrial Electronics and Applications* (2008), pp. 2216–2221
10. J.C. Bansal, P.K. Singh, M. Saraswat, A. Verma, S.S. Jadon, A. Abraham, Inertia weight strategies in particle swarm optimization, in *Third World Congress on Nature and Biologically Inspired Computing* (2011), pp. 633–640
11. A.M. Amira, S.E.L. Rabaie, T.E. Taha, O. Zahran, F.E. Abd El-Samie, Comparative study of different denoising filters for speckle noise reduction in ultrasonic b-mode images. *Int. J. Image Graph. Sig. Process.* **2**, 1–8 (2013)
12. H.Y. Chai, E. Supriyanto, L.K. Wee, MRI brain tumor image segmentation using region-based active contour model. Latest trends in applied computational science, in *Proceedings of the 12th International Conference on Applied Computer and Applied Computational Science* (2013), pp. 36–41
13. A. Adeli, F. Tajeripoor, M.J. Zomorodian, M. Neshat, Comparison of the fuzzy-based wavelet shrinkage image denoising techniques. *Int. J. Comput. Sci. Issues* **9**, 211–216 (2012)

Intelligent Modeling and Optimization of ECM Process Parameters

T.M. Chenthil Jegan, D. Ravindran and M. Dev Anand

Abstract Electrochemical machining (ECM) is an unconventional process used for the machining of hard materials and metal matrix composites. In the present work, the artificial neural network trained with back-propagation algorithm is used for correlating the interactive and high-order influences of various machining parameters on the predominant machining factors. The operators' requirements cannot be satisfied by the machining parameters provided by ECM machine tool builders. The process parameters are then optimized using weighted sum particle swarm optimization. The fitness function for optimization is obtained from the developed model.

Keywords Electrochemical machining · Artificial neural network · Weighted sum particle swarm optimization

1 Introduction

Electrochemical machining (ECM) is a well-known process used for the manufacturing of various sophisticated parts such as turbine blades, hip joint implants, microcomponents, and many other applications. Different from the other machining

T.M. Chenthil Jegan (✉)

Department of Mechanical Engineering, St. Xaviers Catholic College of Engineering,
Kanyakumari, India

e-mail: optrajegan@yahoo.co.in

D. Ravindran

Department of Mechanical Engineering, National Engineering College,
Thoothukudi, India

e-mail: ravinec85@gmail.com

M. Dev Anand

Department of Mechanical Engineering, Noorul Islam Centre for Higher Education,
Kanyakumari, India

e-mail: anandpmt@hotmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_58

533

processes, in ECM, there is no contact between tool and work piece. The predominant response parameters considered are material removal rate (MRR) and surface roughness (SR). Ganesan et al. [1] determined the optimal machining parameters for continuous pro le machining with respect to the minimum production time. They set the practical constraints such as cutting force, power, dimensional accuracy, and surface finish. They used genetic algorithm and single objective PSO with crowding distance computation to find the machining parameters in CNC turning process. Prediction of some ECM process parameters using ANN is proposed in by Abuzied et al. [2]. They used multilayer feed-forward neural network and proved that ANN model was found to be in close agreement with the experimental data. Sen et al. [3] presented a hybrid neural network and non-dominated genetic algorithm approach for the multi-response optimization of the electro jet drilling process, where artificial neural network model was used to predict the response parameters of the process and then a genetic algorithm was applied to the trained neural network model to obtain the optimal process parameters, whereas Soleimaniehr et al. [4] developed an ANN for prediction of aluminum work piece SR in ultrasonic vibration-assisted turning. Navalertporn et al. [5] proposed neural network and bidirectional PSO for optimizing process parameters in tile manufacturing. The ANN was used to model the relationships between input process parameters and output quality characteristics, while the bidirected PSO served to obtain the optimal process parameter combinations. This approach allows multiple and conflicting objectives to be optimized simultaneously. Parsopoulos and Vrahatis [6] proposed a method which transforms the constrained optimization problem into a non-constrained optimization problem by adopting a non-stationary multi-stage assignment penalty function and then apply PSO to solve the converted problems. Swarm intelligence is also treated as an artificial intelligence (AI) technique based on the collective behavior in decentralized, self-organized systems. Optimization of multi-objective and conflicting objectives used in turning process is optimized using dynamic neighborhood particle swarm optimization ANN model. Multilayer feed-forwarded neural network is used for modeling, and the best solutions are obtained for various output parameters [7]. In this proposed work, these are used to investigate the predominant machining parameters such as current (C), voltage (V), electrolyte concentration (E), and feed rate (F). Its effects on MRR and SR are tested through a set of planned experiments. The optimal ECM process parameters are obtained using weighted sum particle swarm optimization (WSPSO).

2 Experimentation

ECM is the unconventional machining process, and here, the work piece acts as anode and tool material as cathode. NaCl is normally used electrolyte. One important point considered is anode and cathode is the electrically conductive materials. During machining flow of electrons takes place from cathode to anode and the machining is to be carried out. Electrolyte is used for carrying medium for

Table 1 Different levels of process parameters

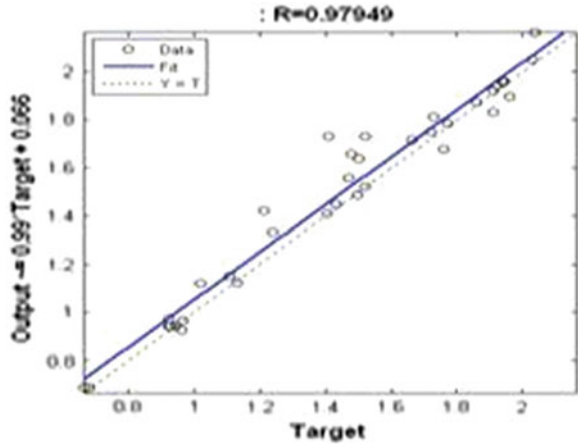
Factors	Levels				
	-2	-1	0	1	2
Current (A)	100	130	160	190	220
Voltage (V)	5	7	9	11	13
Electrolyte concentration (g/min)	5	8	11	14	17
Feed rate (mm/sec)	0.1	0.2	0.3	0.4	0.5

the flow of electrons from cathode to anode. In this research, anode is the aluminum-based metal matrix composite and tool material is made of copper. In the experiment, four machining parameters such as current (C), voltage (V), electrolyte concentration (E.), and feed rate (F), the specimens are tested for the evaluation of MRR and SR. The process parameters can be set prior to machining. The MRR is the volume of material removed from a work piece in a machining process per unit time. The SR is measured by a profile meter (3D-Hommelework) and it is expressed in μm . There are four input process parameters and the two output parameters that vary in a specified range. Hence, the number of experiments to be performed is decided based on the half fraction rotatable second-order design. The experimentation is carried out in 36 specimens with different level of process parameters. The specimen is made of B4C-reinforced Aluminum metal matrix composite. The variation of above-input parameters with +ve and -ve levels is shown in Table 1.

3 Modeling of ECM Process Parameters Using ANN

In the proposed work, for modeling applied voltage, current, electrolyte concentration, and feed rate are the four ECM process parameters considered for modeling. The number of input neurons used in the neural network layer is 4. The output neurons selected are 2 related to MRR and SR. The number of hidden units is determined by several factors. The proposed ANN is the multilayer feed-forward neural network (MLFFNN). MLFFNN has three layers as input layer, hidden layer, and output layer. In the proposed work, for modelling applied voltage, current, electrolyte concentration, and feed rate are the four ECM process parameters considered for modeling. 4-10-2 architecture MLFFNN is used for modeling. The proposed neural network used for modeling of ECM process parameter is shown in Fig. 1. It is a three-layer neural network with input unit $x_i = \{x_1, x_2, x_3, x_4\}$, hidden layer $h_j = \{h_1, h_2, \dots, h_{10}\}$, and output layer $y_k = \{y_1, y_2\}$. Each connection between nodes has a weight associated with it. In addition, there is a special weight (w) given as $w_{ij} = \{w_1, w_2, \dots, w_{10}\}$, where 10 denotes the number of neurons in the hidden layer that feeds into every node at the hidden layer, and a special weight (z) is given as $z_{jk} = \{z_1, z_2, \dots, z_{10}\}$ that feeds into every node at the output layer. These weights are called as bias, and it sets the thresholding values for the nodes. The sigmoid function is used as a thersholding function applied in hidden node and

Fig. 1 Regression plot for MRR



in the output node linear thresholding function is used. The network is first initialized by setting up all its weights to be a small random numbers between [0, +1]. The experimental results are normalized before used for training the network to increase the capability of the system. The network is trained and tested with 32 experimental data. The 4 input machining process parameters in training test are applied to the network and the output is calculated. The actual output (t) obtained is compared with the target output (y) and calculated by the experimentation, and the error is calculated by finding the difference between the target and the actual output. This error is then used mathematically to change the weights in such a way that the error will get minimized. The process is repeated again and again until the error is minimal. The mean square error function defined in Eq. (1) is used for training .The network is trained with back-propagation algorithm.

$$E = \frac{1}{n} \sum_{i=1}^n (y_i - t_i)^2 \tag{1}$$

The weights are updated using Eq. (2).

$$\Delta w_{ij}(t + 1) = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(t) \tag{2}$$

where η is the learning rate and α is the momentum factor. After training the network, the proposed system is tested with testing data set, and the various performances are analyzed.

4 Optimization of ECM Process Parameters Using Weighted Sum PSO

A single objective optimization algorithm provides a single optimal solution. Most of the multi-objective problems give rise to a set of optimal solutions instead of a single optimal solution. Multi-objective problems often involve conflicting objectives and the situation in which improvement in one objective may cause deterioration in solution. The classical approach to solve a multi-objective optimization problem is weighted sum approach [8]; a weight w_i is assigned to each normalized objective function $f(x)$ so that the problem is converted to a single objective problem with a scalar objective function as follows:

$$\begin{aligned} \text{Min } F(x) &= w_1f_1(x) + w_2f_2(x) + \dots + w_kf_k(x). \\ \text{where } \sum w_i &= 1. \end{aligned} \quad (3)$$

provides the weights. Solving a problem with the objective function (3) for a given weight vector $w = \{w_1, w_2, \dots, w_k\}$ yields a single solution, and if multiple solutions are desired, the problem must be solved multiple times with different weight combinations. The main advantage of the weighted sum approach is a straightforward implementation and it is computationally efficient [9]. Nontraditional and evolutionary algorithms are used nowadays to solve high dimension problems with multiple local optima. PSO is a stochastic optimization technique which is initialized with a population of random solutions and searches for the optimal by updating generations. In PSO, each particle in the population has a velocity $v_i(t)$, which enables it to y through the problem space. Therefore, each particle is represented by a position $x_i(t)$ and a velocity vector. Dimensions of position and velocity vectors are defined by the number of decision variables in the optimization problem. Modification of the position of a particle is performed by using its previous position information and its current velocity.

$$v_i(t+1) = wv_i(t) + c_1rand_1(Pbest_i - x_i(t)) + c_2rand_2(Gbest_i - x_i(t)) \quad (4)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (5)$$

The velocity update step in the PSO is stochastic due to the random numbers generated, which may cause an uncontrolled increase in velocity and therefore instability the search algorithm. In order to prevent this, velocities are limited to the dynamic range of the particle on each dimension. In this proposed work, PSO was designed for minimization problem for optimizing ECM process parameters. The main objective is to maximize MRR and minimize SR. Thus, to handle this

problem, the MRR and reciprocal of SR are used as objective functions. The problems can be defined as:

Minimize

$$f1 = \frac{1}{\text{MRR}(x)} \quad (6)$$

$$f2 = \text{SR}(x) \quad (7)$$

where, all decision variables are bounded within the experimental ranges given in Table 1. The steps involved in weighted sum PSO is discussed below.

Step 1: Initialization of PSO parameters

Performance of PSO depends largely on the setting of various parameters. In this work, the weighting functions and learning rate are adapted from Hu and Eberhart [10] where weight was chosen as $w = (0.5 + (\text{rand} = 2))$ and learning rates were taken as $c_1 = c_2 = 1.49445$. Population size = 30 particles and number of iterations = 100.

Initialize particles (30 for the swarm and 4 for G_{best} archive) with random position and velocity vectors.

Step 2: Evaluation of fitness function

Fitness function is the weighted sum value of MRR and SR. In our study, more significance is given to MRR. The weight assigned for MRR is 0.7 and for SR is 0.3

$$F(x) = (f1(x), f2(x)) \quad (8)$$

Step 3: Finding the Gbest and Pbest

Compare the fitness value of particles, select the lowest fitness value as Gbest value and its corresponding dimensions of particle are selected for current, voltage, feed rate, and electrolyte concentration. For the first iteration, Pbest value of all the particles will be same as the Gbest value.

Step 4: Updating the position of each particle

For every particle, velocity and new position of process parameters are calculated. For the first iteration, present position is compared with new position and a better one is selected as Pbest based on fitness function. The defined problem is a minimization problem, and hence, particle having minimum fitness function is assumed as better one.

Step 5: Termination

The process is continued for 100 iterations, and finally, the values stored in Gbest archive are taken as the optimized value.

5 Result Analysis and Discussion

The ANN model is used to correlate ECM process parameters and predominant output parameters. Correlation coefficient R can be used to determine how well the network output is closer to the desired output. The regression plot for MRR is shown in Fig. 1. From the plot, it is noticed that the residuals are closer besides the straight line, and the R value is 97.94. The regression plot for SR is shown in Fig. 2. From the plot, it is noticed that the residuals are also closer besides the t line and the R value is 98.25. The average error obtained for MRR and SR is 2.1 and 1.8, respectively. It is proved that the predicted values of MRR and SR have a close agreement with the actual output. The experimental value of MRR obtained and the ANN-predicted values are plotted in Fig. 3. The deviation between the proposed

Fig. 2 Regression plot for SR

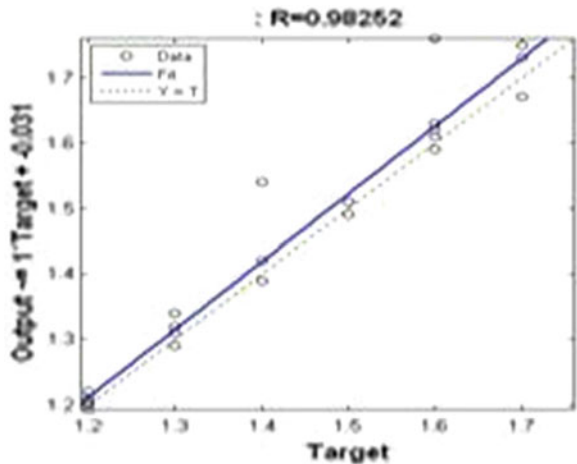
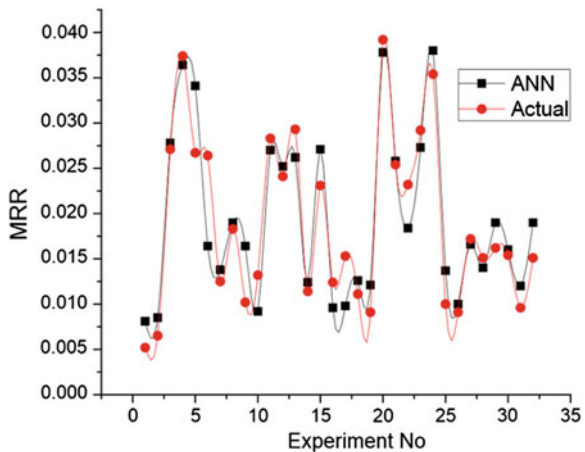


Fig. 3 Experimental versus predicted values of MRR



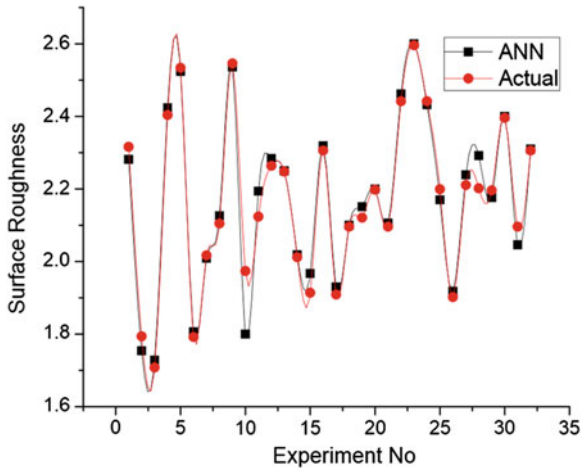


Fig. 4 Experimental versus predicted values of SR

model and the actual value lies in the acceptable level only. Experimental values versus ANN-predicted values of SR are shown in Fig. 4. From Fig. 3, the difference between the SR predicted by ANN and the experimental value is small. Thus, the SR evolved through ANN design can be used to successfully predict the SR values for any combination of ECM parameters within the range of experimentation. The process parameters are then optimized using weighted sum PSO. The control parameters in PSO were fixed suitably to obtain the best performance. The inertia weight is adjusted randomly between 0 and 1. It is found that the PSO with given control parameter produces better convergence and of optimal solution. The convergence plot is shown in Fig. 5. The PSO reaches good fitness function at iteration 50 and then there is no further improvement. Hence, 100 iterations are adequate for

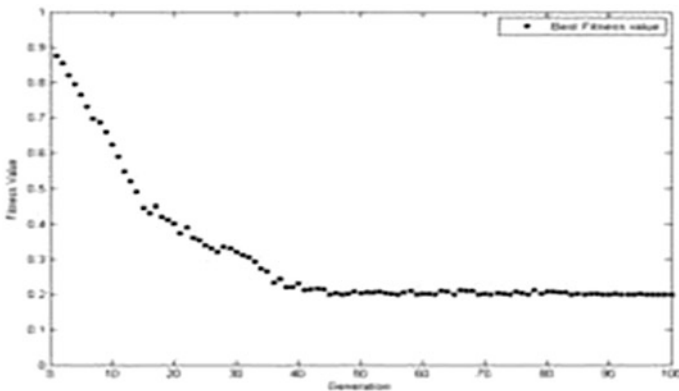


Fig. 5 Convergence plot for PSO

optimization. The optimal value obtained for MRR is 0.011 and SR is 2.102. The given MRR and SR predominant process parameters current, voltage, electrolyte concentration, and feed rate obtained are 196, 6, 13, and 0.34, respectively.

6 Conclusion

In the present study, a multi-input and multi-output ANN model is used for predicting MRR and SR in ECM machining process. The ANN with four input neurons, ten hidden neurons, and two output nodes are trained using back-propagation algorithm is for correlating the relationship between the ECM parameters such as current, voltage, electrolyte concentration, and flow rate and their effects on process parameters MR and SR. The predicted values obtained using proposed ANN model was found to be close to experimental value, and it is proved that the prediction accuracy is high. Hence, the developed model has been used as an objective function to define the parameters for optimization. The process parameters were optimized by weighted sum PSO, and it is noticed that the optimal parameters for ECM machining were obtained with less iterations and with good convergence.

References

1. H. Ganesan, G. Mohankumar, K. Ganesan, K. Ramesh Kumar, Optimization of machining parameters in turning process using genetic algorithm and particle swarm optimization with experimental verification. *Int. J. Eng. Sci. Technol.* **3** (2011)
2. H.H. Abuzied, M.A. Awad, H.A. Senbel, Prediction of electrochemical machining process parameters using artificial neural networks. *Int. J. Comp. Sci. Eng.* **4**(1), 125–132 (2012)
3. M. Sen, H.S. Shan, Electro jet drilling using hybrid NNGA approach, *Robot. Comp.-Integr. Manuf.* **23**, 17–24 (2007)
4. H. Soleimanimehr, M. J. Nategh, S. Amini, Modeling of surface roughness in vibration cutting by artificial neural network. *Proc. World Acad. Sci, Eng. Technol.* **40** (2009)
5. T. Navalertporn, N.V. Afzulpurkar. Optimization of tile manufacturing process using particle swarm optimization. *Swarm Evol. Comput.* **1**, 97–109 (2011)
6. K.E. Parsopoulos, M.N. Vrahatis, Particle swarm optimization method for constrained optimization problems. *Intell. Technol. Theor. Appl.* **76**, 214–220 (2002)
7. Y. Karpat, T. Zel, Multi-objective optimization for turning processes using neural network modeling and dynamic-neighborhood particle swarm optimization. *Int. J. Adv. Manuf. Technol.* **35**, 234–247 (2007)
8. B. Rubenstein-Montano, R.A. Malaga, Weighted sum genetic algorithm to support multiple-party multiple-objective negotiations. *IEEE Trans. Evol. Comput.* **6**(4), 366–377 (2002)
9. A. Konak, D.W. Coit, A.E. Smith, Multi-objective optimization using genetic algorithms: A tutorial. *J. Dalian Univ. Technol.* **91**, 992–1007 (2006)
10. X. Hu, R. Eberhart, Multi objective optimization using dynamic neighborhood particle swarm optimization, in *Proceedings of IEEE Swarm Intelligence Symposium* (2002), pp. 1404–1411

An Effective Automation Testing Framework for OATS Tool

Gobi Ramasamy and Sathishkumar Ramalingam

Abstract Oracle application test suite (OATS) is a test tool of Oracle. It is a very good integrated testing tool for Web applications, Web services, Oracle applications, and Oracle databases. The Oracle application testing suite is part of the Oracle Enterprise Manager product family and comprises the following tightly integrated products. They are Oracle load testing for scalability, performance and load testing, Oracle functional testing for automated functional and regression testing, and Oracle Test Manager for test process management, including test requirements management, test management, test execution, and defect tracking. OATS uses OpenScript platform. This paper discusses model-based test automation methods and tools referred to collectively as the Test Automation Framework that reduces the time and resources necessary to develop high-quality and high-assurance systems using OATS functional testing tool. Framework is named as Easy to Automate (Ez2Auto) framework. This OATS tool is newly available in market, and there is no established framework available in literature or ready to use in market.

Keywords OATS · Oracle application test suite · Open script · Automation test framework · Ez2Auto

1 Introduction

Framework is a set of standard guidelines to be used for better, reusable way of developing automation to make the automation tester's life easy [1]. This paper provides a brief on how the Test Automation Framework developed for our applications can seamlessly automate various types of applications with minimal

G. Ramasamy (✉) · S. Ramalingam
Bharathidasan University, Tiruchirappalli, India
e-mail: gobi86@gmail.com

S. Ramalingam
e-mail: satmce@gmail.com

effort. This increases the tester productivity by providing a robust and easy to use framework, which encapsulates most of the process internals of how to execute the test suite and facilitates the tester to remain focused on designing the test suite. Various automation test tools are available in market such as IBM RFT, HP's QTP, open source Selenium, and Watir. OATS mainly used for Web application, packaged oracle applications such as Siebel, Peoplesoft, E-business suite, and ADF applications [2]. This paper discusses the importance of framework, guidelines to be used, and implementation of Ez2Auto framework.

2 Automation Framework

Framework defines a set of guidelines standard for all phases of test automation life cycle: Requirement analysis, script design, execution, and reporting and maintenance [3]. A framework can be a wrapper around some complex internal architecture which makes it easy to use for the end user. It also enforces a set of standards for implementation. There is no standard set of guidelines available on developing a framework and what all considerations need to be taken during the development of the same. By default, none of the automation tools come with default framework. Maintenance cost is very high for the automation project which is implemented without automation framework and also standard is required when multiple team members are working on same automation test project to have common standard. Reusability is the main advantage of framework which helps the automation development easy.

3 Design Guidelines

Different types of framework available for automation project. This paper covers different aspect of a framework and key features it needs to have based on the requirements [3–5].

- Data-driven framework—Used when flow of the application remains constant, only the data changes. The data is provided by external medium such as Excel sheet and XML.
- Keyword-driven framework—This framework provides generic keywords that can be used with any type of application. It also provides abstraction from the type of automation tool used and type of being application tested, for example, it can test a similar Web and Windows application with the same test case.
- Hybrid framework—A hybrid framework is the one which takes advantages from both data-driven and keyword-driven frameworks. These frameworks do not implement generic keywords but implement business logic keywords based

on the application being tested, for example, login and logout could be the application-specific keyword that can be used.

- Do not reinvent the wheel—A framework should try and use the power of the automation tool rather than redefining the whole language by implementing new keywords. Developing a keyword-driven framework is time consuming and costly. A hybrid framework can be developed in a shorter time period and with less cost. The following parameters should be considered while designing a framework [3].
- Reusability—The framework should cut down the repetition of codes and make sure the reusability. Combining individual actions into business logic provides reusability. Example: Creating a function by combining the login activities such as entering username and password and clicking OK button.
- Backward compatibility—Framework should have basic backward compatibility feature and should be able to run your script at any version of your automation test tool.
- Version control—Automation script should be stored in change management tool in order to rollback to working version during the disaster. Various tools are available in market such as VSS, QC, and OATS test manager.
- Automation test environment—Automation should be considered as any other development project. Test scripts should be created and debugged in dedicated test environment. Once dry run or unit tested then it can be deployed in product environment for regression testing. Even few companies deliver their basic level automation test suite with their product or project.
- Configurable framework—Hard coding of values inside code is not a good practice. It will not work if there is a data change in the application. It should be externally configurable either in XML/Excel/database file instead of opening the tool to provide the input. Also you should have provision to configure what are the regressions test cases should be run for a testing cycle based on the impact analysis.
- Self configurable—In few cases, framework should have feature of self configurable like general pre-requisite setup for a test suite to run. Let us take an example of IE settings. Automation test suite can be run in any machine; we can keep a self configurable script to make the IE with expected settings.
- XPATH changes—Most common issues faced during automation are object identification changes. Framework should have provision to change the XPATH easily if the application is undergoing any change. This can be achieved by storing all object identification settings at a shared location.
- Execution—Framework should have options for users to execute based on their needs. It should have following features
 1. Execution of a individual test case
 2. Execution of a test batch (combination of tests)
 3. Re-execution of only failed test cases
 4. Execution of a test case/test batch based on result of another test case/test batch

5. There could be many other needs based on the project requirement. A framework might not implement all of them, but should be flexible enough to accommodate such requirements in future.
 - Status monitoring—A framework should allow monitoring the execution status in real time in some HTML report and should be capable of sending alerts in case of failure. This ensures quick turnaround time in event of a failure.
 - Reporting—Different applications have different reporting needs. Some require combined results for a test batch, and some require individual level test report for each test case in test batch. The framework should be flexible enough to generate required reports.
 - Debugging—Debugging takes a lot of time during automation, and hence, special care needs to be taken for this part. Keyword-driven frameworks which use external data source (such as a Excel spreadsheet) to read scripts keywords and process the same are difficult to debug, hence proper logs need to be maintained in order to debug easily.
 - Logging—Log generation is an important part of execution. It is necessary to generate debug information at various points in a test case. This information can help to find the problem area quickly and reduce the time to make a fix at the same time.
 - Usability—The framework should be easy for other automation testers/manual testers to learn and use. It is time consuming and costly to train a resource on a framework. A well-documented framework is easier to understand and implement. So, keeping the document up-to-date is an art of automation tester. Framework document is one of the greatest artifacts of automation testing.
 - Flexible—Framework should be flexible enough to accommodate any enhancements without impacting existing test cases.
 - Performance impacts—A framework should also consider the performance impacts of the implementation. A complex framework which increases the load time or execution time of scripts is never desirable. Techniques such as caching and compiling all code into single library while execution should be used to improve performance whenever possible.
 - Framework support tools—External tools can be developed to perform tasks that help in framework design. Some example are uploading scripts from local folder to Oracle Test Manager, associating library files to currently open scripts, and synchronizing local files with Oracle Test Manager.
 - Coding standards—Coding standards ensure scripts that are consistent, readable, and easily maintainable. Coding standards
 1. Naming convention for variables, subs, functions, file names, script names, etc. Example—`int_VarName` for integer, `fn_int_FuncName` for function returning integer
 2. Library, subs, functions comment header. This should include information such as version history, created by, last modified by, last modified date, description, parameters
 3. Object naming conventions such as `txtB_FieldName` for a text box.

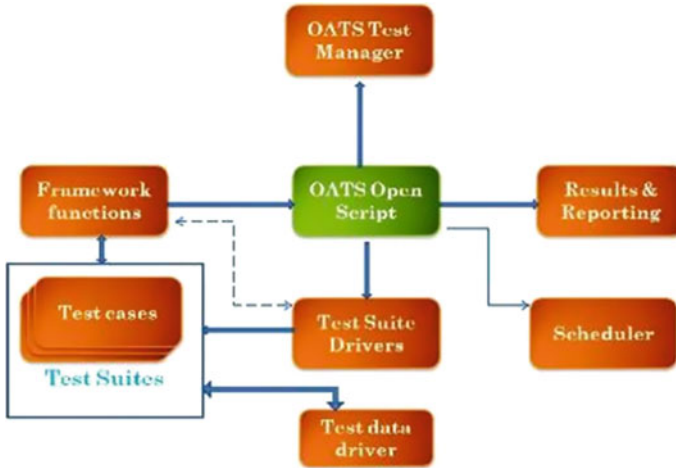


Fig. 1 Ez2Auto architecture

4 EZ2 Auto Framework

It is implemented using OpenScript and extended using Java code. Figure 1 depicts the various building blocks of Ez2Auto framework.

- **OpenScript:** OpenScript platform is used to combine the GUI with an eclipse-based Java IDE to support the automation testers. This platform used to record the functional test, customize according to their need, and replay features [4].
- **TestManager:** Oracle Test Manager is used to keep track of test plan, requirement, test assets (both manual and automation test), feature to run the test, reporting the issues, and result and dashboard to display the summary reports based on multiple criteria.
- **FrameworkFunctions:** This block talks about the functions being used for reusability and to maintain the standard in automation development within team which will simplify the automation tester's task. They are discussed in detail in the reference code sections.
- **TestSuite driver:** TestSuite is used to combine multiple automation test cases into single executable unit. TestSuite driver function is defined to initiate the execution. It can be controlled by the scheduler. It will get the input from TestSuite driver data input table to execute the list of test cases to be executed for this particular run based on the execute flag set against a tests case. This will help to execute the regression test cases according to your need.
- **Report/Result:** Report the result using the default OATS reporting mechanism as well as in the customized report mechanism available with generate HTML Report function as part of this framework.

- Scheduler: It plays an important role in the execution of the test suite automatically based on the settings defined for this test suite in the Oracle Test Manager.

5 Implementation

Implementation is done for the above architecture in OATS scripts. The following functions are designed as part of the architecture for the basic level of Web application testing in mind. It can be extended further based on requirements.

Functions implemented	Usage of functions
public void initialize()	To initialize the variables and necessary methods
public void run()	This is the place where the actual calls to the functions
public void frameworkFunctions()	Framework-specific reusable functions will be called in this functions, such as CreatingFile, DeletingFile, Login, Logout, and LaunchIE
public void testSuiteDriverFunctions()	This is most important function which calls the necessary test cases based on the input provided in the TestSuite data sheet
public void generateHTMLReport()	This helps to generate the customized HTML report
public void reportIssue()	This is used to log the necessary bug in the bug reporting tool
public void loadPropertyFile(String repository, String filePath)	This is to load the global variable file
public void element_ShowContextMenu (String elementName, String XPATH)	This function is used to display the menu
public void element_Click(String elementName, String XPATH)	This is used to click a Web element
public void element_doubleClick(String elementName, String XPATH)	This is written to do the double click operations
public void link_Click(String linkName, String XPATH)	This function is used to click on the link
public void textBox_SetText(String TextTobeEntered, String XPATH)	To set the text to a text box
public void textBox_SetPassword(String PwdToSet, String XPATH)	To set the password in encrypted mode
public void image_Click(String image-Name, String XPATH)	To click on the image
public void radioButton_Select(String radioButtonName, String XPATH)	To select the radio button

(continued)

(continued)

Functions implemented	Usage of functions
public void checkBox(String checkBox-Name, String XPATH, boolean Status)	To select/deselect the checkbox option
public void button_Click(String button-Name, String XPATH)	To click on the button
public void element_MouseOver(String elementName, String XPATH)	To mouse over the element to find out the tooltip or to do right click
public void SuccessLogger(String functionName, String Action)	To update successful steps in HTML file
public void FailLogger(String function-Name, String Action)	To report the failures in HTML file
public void finish()	End of the execution

6 Data Analysis

Ez2Auto framework is analyzed using various parameters. They are compared with other standard parameters of other automation frameworks available in market. Table 1 lists the value based on the survey/feedback from various automation testers. Figure 2 shows that Ez2Auto framework has good quality of design, data-driven feature and customer satisfaction. And at the same time, this framework is complex in nature. Once the framework is designed by an expert, automation tester/team can reuse the existing methodology defined in framework for consistency across the project. Though the initial cost is more in procuring the tool and designing framework, it would fetch high Return on Investment (ROI) for the regression testing in long run.

Table 1 Comparison of Ez2Auto framework with standard framework

Critical factors	Ez2Auto	Modularity	Keyword driven	Data driven
Data-driven approach	10	3	5	10
Parameterization	10	5	3	8
Application independent	7	5	7	7
Flexibility for changes	8	3	5	6
Complexity	5	5	3	4
Quality of the design	9	5	6	5
Maintenance	7	6	5	7
Duration to implement	3	6	9	8
Compatibility	8	7	5	7
Customer satisfaction	9	6	7	6

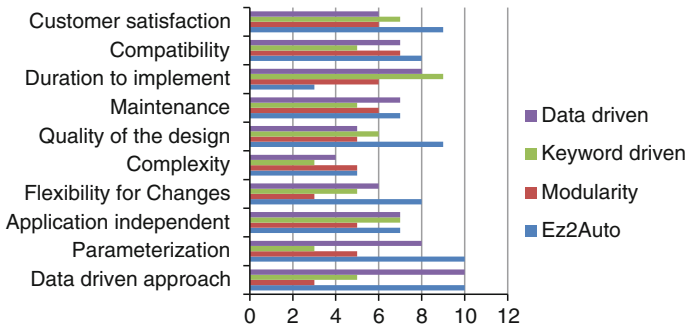


Fig. 2 Analysis of EZ2 automation framework with critical factors

7 Conclusion

Ez2Auto framework is designed using the basic principles of hybrid framework in mind which gives the modularity, keyword, and data-driven features. This framework helps the OATS automation testers to make robust automation test suite. Maintenance cost will be less when an OATS automation testing is done using Ez2Auto framework which gives very fruitful high ROI.

References

1. F. Wang, A test framework based on a web. in *IEEE/ACIS 11th international conference on computer and information science*, 2012
2. Kumar, Achieving higher ROI by implementing hybrid test automation framework (Data and Keyword driven). in *9th International software testing conference, QAI*, 2009

Multimodal Biometric Authentication System Based on Score-Level Fusion of Palmprint and Finger Vein

C. Murukesh, K. Thanushkodi, Padmanabhan Preethi
and Feroze Naina Mohamed

Abstract Multimodal biometrics plays a major role in our day-to-day life to meet the requirements with the well-grown population. In this paper, palmprint and finger vein images are fused using normalization scores of the individual traits. Palmprint features extracted from the discrete cosine transform (DCT) are classified by using multi-class linear discriminant analysis (LDA) and self-organizing maps (SOM). Finger vein identification is designed and developed by using repeated line tracking method to extract the patterns. A multimodal biometric authentication system integrates information from multiple biometric sources to compensate for the limitations in performance of each individual biometric system. These systems can significantly improve the recognition performance of a biometric system apart from catalyzing population coverage, impeding spoof attacks, increasing the degrees of freedom, and reducing the failure rates.

Keywords Multimodal biometrics · DCT · Multi-class LDA · SOM

1 Introduction

In recent years, multimodal biometrics has gained the substantial attention to all organizations and more than one biometric is fused together. Vein patterns serve a highly secured authentication system over other biometrics. It is noninvasive, reliable, and well accepted by users [1]. Preprocessing of finger vein images yields a better quality image by removing the noise and increasing the image contrast [2].

C. Murukesh (✉) · P. Preethi · F.N. Mohamed
Department of EIE, Velammal Engineering College, Anna University, Chennai
Tamilnadu, India
e-mail: pcmurukesh@gmail.com

K. Thanushkodi
Akshaya College of Engineering and Technology, Anna University, Coimbatore
Tamilnadu, India

Acquisition of infrared finger vein image using various LED contains not only the vein patterns but also irregular shading produced by the various thicknesses of the finger bones and muscles [3]. The finger vein pattern from the unclear images is extracted by using line tracking, which starts from various positions. A person retrieval solution using finger vein can be accomplished by searching an image in the database in a reasonable time [4]. A wide line detector for feature extraction can obtain precise width information of the finger vein and improve the inferences of the extracted feature from low-quality image [5]. The finger vein patterns extracted by using gradient-based threshold and maximum curvature points are applied to neural network to train and test the quality of system [6, 7]. Extracting the finger vein patterns regardless of vein thickness or brightness is necessary for accurate personal identification [8]. Face and finger vein biometric authentication system at multi-level score-level fusion is very efficient to reduce the false rejection rate [9]. Multiple features like texture (gabor), line, and appearance (PCA) features extracted from the palmprint images are fused using particle swarm optimization techniques to improve the performance [10–12]. The wavelet-based fusion technique is suggested to fuse extracted features as it contains wavelet extensions and uses mean–max fusion method to overcome the problem of feature fusion [13].

The rest of this paper is organized as follows: Sect. 2 describes the palmprint recognition system. Section 3 highlights the feature extraction algorithm for finger vein authentication system. Score-level fusion is discussed in Sect. 4. Section 5 provides experimental results of the proposed system, and Sect. 6 offers the conclusion.

2 Palmprint Authentication

This paper designs the most efficient, high-speed method for palmprint recognition and also develops an algorithm for the palmprint recognition system which formulates an image-based approach, using the 2-dimensional discrete cosine transform (2D-DCT) for image compression and a combination of multi-class linear discriminant analysis (LDA) and self-organizing map (SOM) neural network [7] for recognition purpose.

2.1 Image Compression

The image compressed using 2D blocked discrete cosine transform (DCT) is applied with a mask, and high coefficients in the image are discarded.

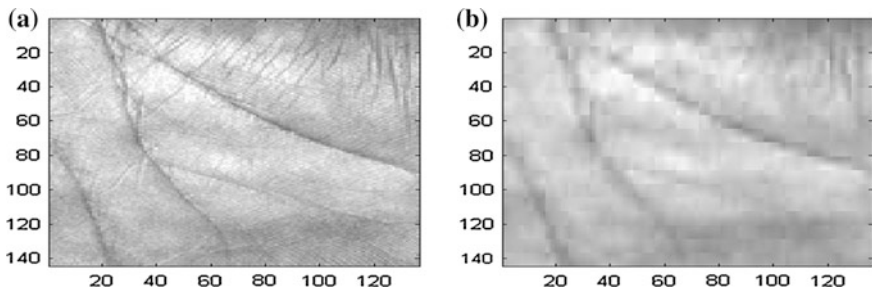


Fig. 1 a Palmprint image and b compressed image using 2D-DCT

The 2D-DCT is defined as

$$X[k_1, k_2] = \alpha[k_1]\alpha[k_2] \sum_{n=0}^{N_1-1} \sum_{n=0}^{N_2-1} x[n_1, n_2] \cos\left(\frac{\pi(2n_2 + 1)k_1}{2N_1}\right) \cos\left(\frac{\pi(2n_2 + 1)k_2}{2N_2}\right) \tag{1}$$

for $k_1 = 0, 1, \dots, N_1-1$ and $k_2 = 0, 1, \dots, N_2-1$

The 2D-IDCT is given by

$$x[n_1, n_2] = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \alpha[k_1]\alpha[k_2]X[k_1, k_2] \cos\left(\frac{\pi(2n_2 + 1)k_1}{2N_1}\right) \cos\left(\frac{\pi(2n_2 + 1)k_2}{2N_2}\right) \tag{2}$$

For $n_1 = 0, 1 \dots N_1 - 1$ and $n_2 = 0, 1 \dots N_2 - 1$

Mathematically, the DCT is perfectly reversible and there is no loss of image definition until coefficients are quantized. Since the DCT algorithm is used for JPEG image compression, the input image is firstly divided into 8×8 blocks and each block is quantized separately by discarding redundant information [14]. The receiver decodes the quantized DCT coefficients of each block separately and computes the 2D-IDCT of each block. The resultant palmprint image shown in the Fig. 1b is compressed image, which is blurred due to the loss of quality, evidently showing the block structure. The image is reshaped into single column and fed into neural network.

2.2 Multi-class LDA

FLDA tries to find a mapping from the high-dimensional space to a low-dimensional space in which the most discriminant features are preserved [15, 16]. It achieves by minimizing the variation within the same class and maximizing the variation between classes. The between-class scatter matrix is given by

$$S_B = \sum_{i=1}^n \lambda_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (3)$$

Consider that each pattern in the learning set belongs to one of the n patterns (m_1, m_2, \dots, m_n). From the patterns given above, the within-class scatter matrix is defined by

$$S_W = \sum_{i=1}^n \sum_{X_k \in m_i} (X_k - \mu_i)(X_k - \mu_i)^T \quad (4)$$

where μ_i is the mean of class i , m_i is the number of cases, and the superscript T indicates a transpose action. The objective of FLDA is then to find U_{opt} maximizing the ratio of the between-class scatter to the within-class scatter.

$$U_{opt} = \underset{w}{\operatorname{argmax}} \left| \frac{U^T S_B U}{U^T S_W U} \right| \quad (5)$$

Finding the maximum U_{opt} could be tricky, but fortunately, it is known that the solution can be found in a relatively simple method.

$$S_B U - Q S_W U = 0 \quad (6)$$

where Q is known as a diagonal matrix and its elements are the eigenvalues. The column vectors of matrix U are eigenvectors corresponding to the eigenvalues.

2.3 Self-Organizing Feature Maps

The principal goal of SOM is to transform a signal pattern of arbitrary dimension into one- or two-dimensional discrete map and to perform this transformation accordingly in a topologically ordered fashion [17]. The output nodes are connected in an array (usually 1 or 2 dimensional). Randomly choose an input vector x and determine the winning unit of the output node.

$$|w_i - x| \leq |w_k - x| \forall k \quad (7)$$

The winning node weights are updated by

$$w_k(\text{new}) = w_k(\text{old}) + \mu N(j, k)(x - w_k) \quad (8)$$

Thus, units close to the winners as well as the winners themselves have their weights updated appreciably.

3 Finger Vein Authentication

The patterns of veins were extracted by combining two segmentation methods, which include morphological operation and maximum curvature points in image profiles. The finger vein patterns were acquired by passing near-infrared light through the finger vein. The result is an image of the unique patterns of veins, as dark lines can be captured by a sensor placed below the finger.

3.1 Feature Extraction and Matching

The finger vein features are extracted by repeated line tracking method. Let the initial value of the current tracking point of the pixel (x_c, y_c) is (x_s, y_s) . R_f is the set of pixels within the finger's outline, and T_r is the locus space. D_{lr} and D_{ud} are the parameters that prevent the tracking point and are determined by

$$D_{lr} = \begin{cases} (1, 0) & (\text{if } R_{nd}(2) < 1) \\ (-1, 0) & (\text{otherwise}) \end{cases} \tag{9}$$

$$D_{ud} = \begin{cases} (1, 0) & (\text{if } R_{nd}(2) < 1) \\ (0, -1) & (\text{otherwise}) \end{cases} \tag{10}$$

$R_{nd}(n)$ is uniform random number between 0 and n

The detection of the dark line direction and movement of the tracking point is determined by the set of pixels N_c .

$$N_c = T_c \cap R_f \cap N_r(x_c, y_c) \tag{11}$$

$N_r(x_c, y_c)$ is the set of neighboring pixels of (x_c, y_c) , selected as follows:

$$N_r(x_c, y_c) = \begin{cases} N_3(D_{lr})(x_c, y_c) & (\text{if } R_{nd}(100) < p_{lr}); \\ N_3(D_{ud})(x_c, y_c) & (\text{if } p_{lr} + 1 \leq R_{nd}(100) < p_{lr} + p_{ud}) \\ N_3(x_c, y_c) & (\text{if } p_{lr} + p_{ud} \leq R_{nd}(100)), \end{cases} \tag{12}$$

$N_3(D)(x, y)$ is the set of three neighboring pixels of (x_c, y_c) whose direction is determined by the moving direction attribute D .

$$N_3(D)(x, y) = \left\{ (D_x + x, D_y + y), (D_x - D_y + x, D_x - D_y + y), (D_x + D_y + x, D_y + D_x + y) \right\} \tag{13}$$

Parameters p_{lr} and p_{ud} are the probability of selecting the three neighboring pixels in the horizontal or vertical direction. The line evaluation function reflects the depth of the valleys in the cross-sectional profiles around the current tracking point:

Fig. 2 Extraction of finger vein patterns



$$V_i = \max_{(x_c, y_c) \in N_c} \left\{ \begin{array}{l} F(x_c + r \cos \theta_i - \frac{W}{2} \sin \theta_i, y_c + r \sin \theta_i + \frac{W}{2} \cos \theta_i) \\ + F(x_c + r \cos \theta_i + \frac{W}{2} \sin \theta_i, y_c + r \sin \theta_i - \frac{W}{2} \cos \theta_i) \\ - 2F(x_c + r \cos \theta_i, y_c + r \sin \theta_i) \end{array} \right\} \quad (14)$$

Let W is the width of the profiles, r is the distance between (x_c, y_c) and the cross section, and θ_i is the angle between the line segments $(x_c, y_c) - (x_c + 1, y_c)$ and $(x_c, y_c) - (x_b, y_i)$. The current tracking point (x_c, y_c) is added to the locus position table T_c . The total number of times the pixel (x, y) has been in the current tracking point in the repetitive line tracking operation is stored in the locus space, $T_r(x, y)$. Therefore, the finger vein pattern is obtained as chains of high values of $T_r(x, y)$. The patterns of veins shown in Fig. 2 are extracted using repeated lines tracking and by iterative process every minute details of finger vein are taken into account.

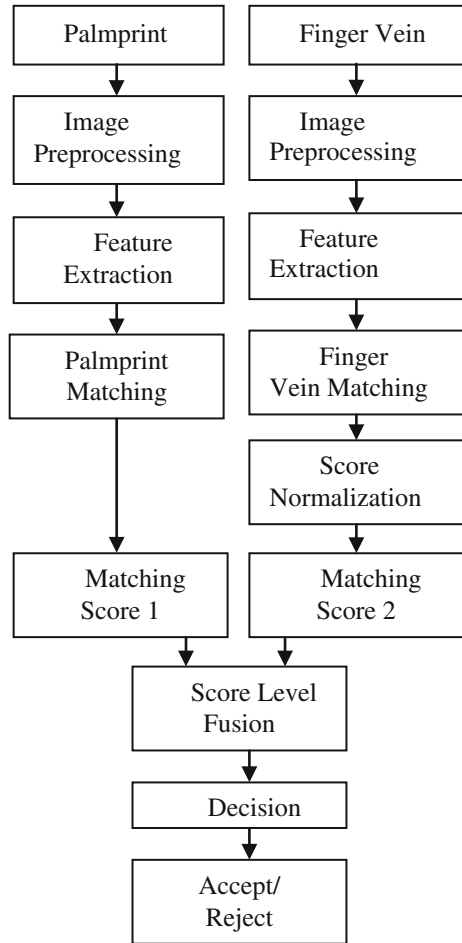
4 Score-Level Fusion

The matching scores of palmprint recognition is obtained by finding the minimum absolute deviation for each palmprint image. The matching scores of finger vein recognition are normalized using Z-score normalization technique. The mean (μ) and standard deviation (σ) are estimated from a given set of matching scores.

The normalized scores are given by

$$S'_K = \frac{S_K - \mu}{\sigma} \quad (15)$$

Fig. 3 The block diagram of the proposed multimodal biometric system



The palmprint and finger vein scores are fused by weighted fusion method.

$$S = w_1s_1 + w_2s_2 \tag{16}$$

Let s_1 and s_2 are palmprint matching score and finger vein matching score, w_1 and w_2 are the weights assigned to both the traits, and S is the fusion score (Fig. 3).

5 Experimental Results and Discussions

Experiments have been conducted on homogeneous multimodal database consisting of palmprint and finger vein images acquired from 50 subjects, and each subject has provided 5 palmprints and 5 finger vein images. Palmprint images are acquired

Table 1 Absolute minimum deviation of palmprint images

No. of subjects	Minimum deviation
1	4
2	8
3	13
4	17
5	23

using a high-resolution digital camera, and finger vein images are obtained using a CCD camera. The ROI of the center part of the palmprint image is extracted, and the image is compressed with 2D-DCT. The coefficients are reshaped and are fed to multi-class LDA. The output of multi-class LDA of different persons is made into single array and given as input to neural network. The image in the training database which is the closest match by the SOM neural network for the input palmprint image is found by finding the minimum absolute deviation as shown in Table 1.

The vein patterns are extracted using repeated line tracking, and score is obtained to fuse the process. Normalization scores of finger vein images from different subjects are formulated in Table 2. After obtaining the scores from both the modalities, palmprint and finger vein traits are fused using score-level fusion.

Table 3 shows false accept rate (FAR) and recognition rate determined for the proposed techniques. The score-level fusion of palmprint and finger vein images has the recognition rate of 98.5 % with 2 % FAR. The proposed multimodal biometric system overcomes the limitations of individual biometric systems and also meets the response time as well as the accuracy requirements.

Table 2 Normalization scores of finger vein images

No. of subjects/samples	1	2	3	4	5
1	74.5689	75.6421	75.2389	75.8999	74.1256
2	82.5671	82.9795	82.1256	82.0145	82.4789
3	86.2356	86.1005	86.0005	86.1856	86.1458
4	73.4566	73.0025	73.1255	73.0189	73.5809
5	90.1289	90.1478	90.1456	90.1236	90.1006

Table 3 Recognition performance of the proposed system

Traits	FAR %	Recognition Rate %
Palmprint	6	94.5
Finger vein	4	96
Palmprint + Finger vein	2	98.5

6 Conclusion

The proposed method bypasses the need to perform score normalization and choosing optimal combination weights for each modality. In this sense, the proposed solution is a principled and general approach that is optimal when the matching score distributions are either known or can be estimated with high accuracy. Palmprint authentication is implemented using LDA and SOM for feature classification and 2D-DCT for image compression. Finger vein is authenticated by using repeated line tracking for feature extraction, and matching is done with the template created. Once identification is done, minimum deviation for palmprint and matching score for finger vein are calculated and are fused using scoring level. Error rate is reduced, and it provides accurate results.

References

1. X. Li, S. Guo, The fourth biometric—vein recognition. *Pattern Recogn. Tech. Technolo. Appl.* **24**, 626 (2008)
2. D. Hejtmankova, R. Dvorak, A new method of finger veins detection. *Int. J. Bio-Sci. Bio-Technol.* **1**, 11 (2009)
3. N. Miura, A. Nagasaka, Takafumi miyatake.: feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Mach. Vis. Appl.* **15**, 194–203 (2004)
4. N. Miura, T. Miyataket, A. Nagasakat, Automatic feature extraction from non-uniform finger vein image and its application to personal identification. *IAPR Workshop on Machine Vision Applications.* (2002)
5. B. Huang, Y. Dai, R. Li, D. Tang, W. Li, Finger-vein authentication based on wide line detector and pattern normalization. *International Conference on Pattern Recognition.* (2010), pp. 1269–1273
6. I. Malik, R. Sharma, Analysis of different techniques for finger-vein feature extraction. *Int. J. Comput. Trends Technol. (IJCTT).* **4**(5) (2013)
7. A.N. Hoshyar, R. Sulaiman, A.N. Houshyar, Smart access control with finger vein authentication and neural network. *J. Am. Sci.* **7**(9) (2011)
8. J.H. Choi, W. Songa, T. Kima, S.-R. Leeab, H.C. Kim, Finger vein extraction using gradient normalization and principal curvature. *Proceedings of SPIE, Image Processing: Machine Vision Applications II.* (2009), pp. 7251
9. I.R. Muhammad, Multimodal face and finger veins biometric authentication. *Sci. Res. Essays* **5**(17), 2529–2534 (2010)
10. G. Yang, X. Xi, Y. Yin, Finger vein recognition based on (2D)2 PCA and metric learnin. *J. Biomed. Biotechnol.* **2**, (2012)
11. K. Krishneswari, S. Arumugam Intramodal feature fusion based on PSO for palmprint authentication. *ICTACT J. Image Video Process.* **02**(04) (2012)
12. P. Tamil Selvi, N. Radha, Palmprint and Iris based authentication and secure key exchange against dictionary attacks. *Int. J. Comput. Appl.* **2** (11) (2010)
13. K. Krishneswari, S. Arumugam, Intramodal feature fusion using wavelet for palmprint authentication, *Int. J. Eng. Sci. Technol.* **3**(2011)
14. D.V. Jadhav, R.S. Holambe, Radon and discrete cosine transform based feature extraction and dimensionality reduction approach for face recognition. *Signal Process.* **88**, 2604–2609 (2008)

15. T. Connie, A. Jin, M. Ong, D. Ling, An automated palmprint recognition system. *Image Vision Comput.* **23** (2005)
16. X. Wu, D. Zhang, Fisher palms based palmprint recognition. *Pattern Recogn. Lett.* **24**(15), 2829–2838 (2003)
17. A. Hussein A. Al-Timemy, A robust algorithm for ear recognition system based on self organization maps. *1st Regional Conference of Engineering Science NUCEJ* (special issue). **11**(2) (2008)

Synergistic Clinical Trials with CAD Systems for the Early Detection of Lung Cancer

G. Vijaya and A. Suhasini

Abstract *Background* Nowadays, lung cancer catches the attention of medical and social communities in the recent years because of its high frequency allied with difficult treatment. Developing an effective computer-aided diagnosis (CAD) system for lung cancer is of great clinical importance and can increase patient's chance of survival. A systematic screening of lung cancer using computed tomography (CT) images is reliable enough to find lung tumors in their early stages. *Materials and Methods* We conceded the available lung cancer images and its database to preprocess so as to achieve more quality and accuracy in our experimental results. After identifying the size and grade of the lung tumors along with the attributes collected from the patient's history, we can endure classification. Significant frequent patterns are revealed using Vote, SMO, IBk, multilayer perceptron (MLP), J48, ZeroR, and Naïve Bayes classifier. *Results* Finally, J48 classifier outperforms, which yields 99 % of correctly classified instances. Our main objective is to predict the lung cancer image to be fit in benign or malignant class.

Keywords CAD · Vote · SMO · J48 · ZeroR · Naïve Bayes · IBk · MLP

1 Introduction

Lung cancer is the uncontrolled growth of anomalous cells in one or both lungs. As they grow, the abnormal cells form tumors and interfere in the functioning of lungs [1]. Large numbers of people in India and worldwide have lung cancer. Most of them do not even know that they have it. Death is foreseeable. So the ability to

G. Vijaya (✉) · A. Suhasini
Department of Computer Science and Engineering, Annamalai University,
Chidambaram 608002, Tamil Nadu, India
e-mail: viji.pooshan@gmail.com

A. Suhasini
e-mail: suha_babu@yahoo.com

predict lung cancer in its early stage plays a vital role in the diagnosis process. This paper proposed an effective lung cancer prediction system using data mining techniques. Medical diagnosis is very subjective because of the attempt to identify a disease or disorder, and the opinion given by the physicians matters the diagnosis. The number of studies shows that the diagnosis of one patient can differ significantly if it is examined by different doctors or the same doctor at different times [2].

The rest of the paper is organized as follows: Introduction to the fundamental issues of lung cancer in data mining is given in Sect. 2. Section 3 outlines how the lung cancer images are classified using the knowledge discovery process. Section 4 shows the experimental results and its discussion. Finally, Sect. 5 concludes the problem.

2 Background

Most of the medical image mining method marks as segmentation or image feature extraction process for mining information from images. Computational effort increases, when the labeling can be done only after segmenting the scanned images.

Image segmentation refers to the process of splitting an image into discrete regions by grouping together neighborhood pixels based on some predefined similarity criterion. Several literature surveys depict segmentation works very well in lung cancer diagnosis [3].

Segmentation algorithms [4–9] are based on one of the two basic properties of intensity values: discontinuity and similarity. Discontinuity is to partition the image based on abrupt changes in intensity, such as edges in an image. Similarity is based on partitioning the image into regions that are similar according to a predefined criterion. Histogram thresholding approach falls under this category.

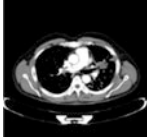
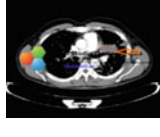
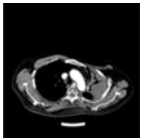
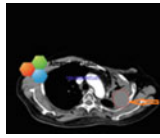
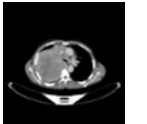
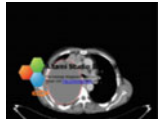
The authors in [10] evaluated neural networks and association rule mining techniques to detect early lung cancer and to classify it by using X-ray chest films. Classification can be done in two categories: normal and abnormal. The normal ones are those characterizing a healthy patient. The abnormal ones include types of lung cancer.

The authors in [11] analyze the lung cancer data available from the SEER program with the aim of developing accurate survival prediction models for lung cancer using data mining techniques. Furthermore, they have developed an online lung cancer outcome calculator for estimating the risk of mortality.

3 Methodology

Multislice computer tomography (MSCT) has revealed to improve lung nodule detection sensitivity in comparison with classical radiography [12–15], but it is associated with the disadvantage of large number of CT slices that have to be

Table 1 Identification of lung mass and their extent in lungs

Original image	After tumor identification	Measurement of volume (in cm)
		2.2
		8.4
		25.2

reviewed by the radiologists when searching for lung nodules. Thus, computer-aided diagnosis (CAD) systems are hosted aiming to assist radiologists in the lung nodule detection task.

Here, the lung mass is automatically calculated using Altami Studio software. Table 1 shows the identification of lung tumor and its measurement. Lung tumor is recognized from the arrow mark in Table 1.

3.1 Dataset

The dataset used in this study consists of 418 cases of CT images collected from different cancer research institutes in India. The location of all nodules in CT images of all patients was confirmed by three radiologists. The criteria for inclusion of radiographs in the database were as follows: (1) absence of suspicious nodules that were not confirmed by CT confirmation, (2) no more than one nodule per patient, and (3) absence of nodules with margins that could not be confirmed by radiologists. All the lung CT images are in the format of DICOM with the size of 512×512 . Among the 418 cases, 357 have malignant and 65 have benign tumors.

3.2 Classification Patterns

The classification patterns used in this study were as follows: J48 decision tree [16], Naïve Bayes classifier, sequential minimal optimizer (SMO) [17], Vote [18], IBk [19], multilayer perceptron (MLP) [20], and ZeroR.

3.3 Lung Cancer Dataset Attributes

Some of the demographic attributes used for the diagnosis are age, gender, marital status, smoking, living area, occupation, and doing exercise. The diagnosis attributes used are as follows: symptoms, tumor grade, and tumor size. The final outcome attribute specified is as follows: malignant (1)/benign (0).

The dataset was run through the Weka 3.7 runtime Java environment in which the above-mentioned classification algorithms were used to classify the data and to predict the accuracy of the classified data.

4 Results and Discussion

4.1 Performance Rating Using Confusion Matrix

Confusion matrix is a tool that is used to measure the performance of the classifier. It represents how far the classifier labels correctly and how far it mislabels the classes. It is a two-row, two-column table that represents the number of true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values. In our planned work, we have specified the correctly classified malignant images as true positive (TP).

Sensitivity measures the proportion of malign tumors correctly identified as malign, whereas specificity measures the proportion of benign tumors correctly identified as benign. Accuracy measures the probability or proportion of tumors correctly classified. Table 2 depicts the classification of lung cancer dataset using different algorithms.

Table 2 Classification of lung cancer dataset using different algorithms

Algorithms	TP	FP	TN	FN	SPEC	SENS	ACC
J48	352	0	63	3	1	0.9915	0.9928
SMO	347	2	61	6	0.9682	0.9831	0.9808
Naïve Bayes	335	10	55	18	0.8461	0.949	0.933
MLP	332	19	48	29	0.7164	0.9196	0.901
IBk	324	13	46	35	0.779	0.902	0.885
ZeroR	351	0	67	0	0	0.8397	0.8397
Vote	330	12	20	56	0.625	0.855	0.8373

TP true positive, *FP* false positive, *TN* true negative, *FN* false negative, *SPEC* specificity, *SENS* sensitivity, *ACC* accuracy

4.2 Performance Analysis by 10-Fold Cross-Validation

In 10-fold cross-validation, the original sample is randomly partitioned into 10 subsamples. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining nine subsamples are used as a training data. The cross-validation process is then repeated 10 times (the folds), with each of the 10 subsamples used exactly once as the validation data. Then, the 10 results from the folds can be averaged to produce a single estimation.

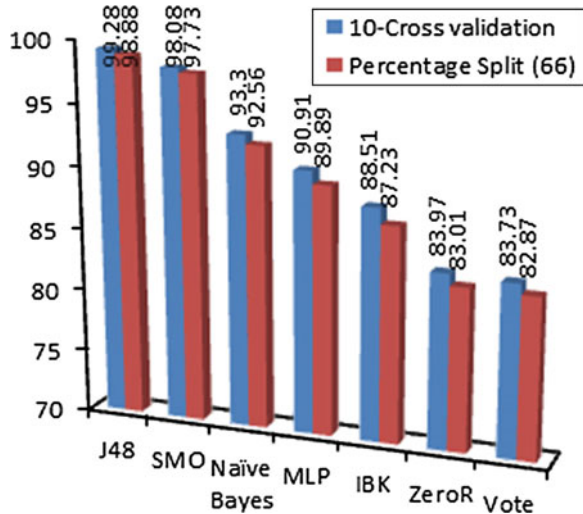
4.3 Performance Analysis Using Percentage Split (66 %)

Rating of the performance can also be computed using percentage split technique. Here, 66 % of the total dataset can be used as the training set, and the remaining 34 % is used as the test set.

4.4 Result

The final outcome of the various classification techniques was depicted using 10-fold cross-validation and percentage split (66) and it is shown in Fig. 1.

Fig. 1 Graph for various classifications based on 10-fold cross-validation and percentage split (66)



5 Conclusion

The anxiety in the diagnosis of lung cancer is overcome in this paper. This paper offers a CAD system for the early detection of lung cancer. For assessment, we have used 418 CT images along with the patients' history as the dataset for classification. In the first phase of the proposed model, automatic extraction of lung volume from the CT images. In the second phase, this can be classified by using different data mining classifiers, along with the details collected from the patients. The best classifier is evaluated based on confusion matrix, 10-fold cross-validation, and percentage split. The experimental result depicts the advantage of the proposed CAD system for the early detection of lung cancer. The research shows that J48 outperforms the entire classification algorithm, followed by SMO and Naïve Bayes classifier.

References

1. E. Donald, Introduction to data mining for medical informatics. Clin. Lab. Med. 9–35 (2008)
2. A.P. Dhawan, Medical image analysis. IEEE Press Ser. Biomed. Eng (2003)
3. S.G. Armato III, H. MacMohan, Automated lung segmentation and computer-aided diagnosis for thoracic CT scans. Int. Congr. Ser. 977–982 (2003)
4. E.M. van Rikxoort, B. de Hoop, M.A. Viergever, M. Prokop, B. van Ginneken, Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. Med. Phys. 2934–2947 (2009)
5. C. Lei, L. Xiaojian, Z. Jie, C. Wufan, Automated lung segmentation algorithm for CAD system of thoracic CT. J. Med. Coll. PLA. 215–222 (2008)
6. J. Zhou, S. Chang, Q. Liu, D. Metaxas, B. Zhao, M.S. Ginsberg, L.H. Schwartz, Automatic detection and segmentation of large lung cancers from chest CT images. First Int. Workshop Pulm Image Process. 165–174 (2009)
7. E.D.O. Nunes, M.G. Perez, in *Medical Image Segmentation by Multilevel Thresholding Based on Histogram Difference*. 17th International Conference on Systems, Signals and Image Processing (2010)
8. I. Levner, H. Zhang, Classification-driven watershed segmentation. IEEE Trans. Image Process. **16**(5), 1437–1445 (2007)
9. Z.S. Zubi, R.S. Saad, in *Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer*. Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases (2011) pp. 32–37
10. A. Agrawal, S. Misra, R. Narayanan, L. Polepeddi, A. Choudhary, A lung cancer outcome calculator using ensemble data mining on SEER data (2011)
11. A. El-Baz, G.M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, A. Elnakib, A. Soliman, B. Abdollahi, Computer-aided diagnosis systems for lung cancer: challenges and methodologies. Int. J. Biomed. Imaging. 2013 **46**, (2013)
12. D.E. Midthun, J.R. Jett, S.J. Swensen et al., Evaluation of nodules detected by screening for lung cancer with low dose spiral computed tomography. Lung Cancer **41**(Suppl 2), S14 (2003)
13. T. Sobue, N. Moriyama, M. Kaneko et al., Screening for Lung Cancer with low-dose helical computed tomography: anti-lung cancer association project. J. Clin. Oncol. **20**, 911–920 (2002)

14. K. Okada, D. Comaniciu, A. Krishnan, Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multi-slice CT. *IEEE Trans. Med. Imaging* **24**(3), 409–423 (2005)
15. J. Han, M. Kamber, *Data Mining Concepts and Techniques*. 2nd edn. The Morgan Kaufmann Series in Data Management Systems. (2006)
16. J.C. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Microsoft Research, Technical Report MSR-TR-98-14 (1998)
17. J. Kittler, Combining classifiers: a theoretical framework. *Pattern Anal. Appl.* **1**(1), 18–27 (1998)
18. D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms. *Mach. Learn.* **6**, 37–66 (1991)

A Unified Framework for Network Bandwidth and Link Latency Detector Based on Cloud Computing

S. Suguna and A. Suhasini

Abstract In general, users often do not know which organizations or services obtain the right to use and may store, utilize, or redistribute their data when sensitive data have been released to a cloud service. The research field of usage control deals with such troubles by enforcing constraints on the usage of data after it has been revealed and is therefore principally important in the cloud environment. Usually, existing solutions apply cryptographic methods to maintain sensitive user data confidential against untrusted servers, by disclosing data decryption keys particularly to the authorized persons. In doing so, on the other hand, these solutions unavoidably bring in heavy computation overhead on the data owner for key distribution and data management when fine-grained data access control is needed and as a result do not perform well. To avoid these problems in this work, we propose customized bivariate parametric detection mechanism (cbPDM) that utilizes a sequential probability ratio test, permitting for control over the false-positive rate while examining the trade-off between detection time and strength of an anomaly and also the packet delivery rate. The method is examined using the bit-rate signal-to-noise ratio (SNR) metric, which is an effective metric for anomaly detection. This enhanced detection method does not need or try to model the full traffic patterns. First, the anomaly detection controls aggregate traffic, devoid of flow separation or deep packet inspection. After that, unlike prior anomaly detection approaches, our method computerizes training and does not require hand-tuned or hard-coded parameters. After that, we make use of both the packet rate and the sample entropy of the packet size distribution guides to guarantee robustness against false positives, consequently overcoming one of the traditional problems of anomaly detection methods. From the comparison results we can see the proposed method, which is better than existing methods due to its rate of packet delivery and bit-rate values.

S. Suguna (✉) · A. Suhasini

Department of Computer Science and Engineering, Annamalai University, Chidambaram,
Tamil Nadu, India
e-mail: karmeljino@gmail.com

A. Suhasini

e-mail: suha_babu@yahoo.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_62

569

Keywords Network bandwidth • Link latency detector • Bivariate parametric detection mechanism anomaly detection • Customized bivariate parametric detection mechanism • Cloud computing

1 Introduction

Cloud computing is used to share data and application via a common center, which uses the Internet and shares resources to preserve data. Security is an imperative concern since cloud has many benefits and thus has many clients. In addition, controls on data in the cloud computing environment comprise the governance policies set in place to make sure that data can be relied on, and with this the integrity, reliability, and confidentiality of the data must be beyond censure. This holds for cloud providers as well. For instance, suppose that you are using a cloud service for word processing where the documents you generate are stored by way of the cloud provider. These documents are owned by your company and you are supposed to control access to them. Nobody else should be able to access them without your permission, other than possibly a software bug that lets other users access the documents. A malfunctioning access control is the reason for any privacy violation that you are desperate to make sure does not ensue.

To make database as a far-famed cloud service, needs are ascertained from completely different aspects; these aspects are composed of chiefly two domains; user-centric needs and supplier-centric needs. As elements of the primary domain, the user wants handy application programming interfaces (API) with less governance and maintenance, which might give high performance in terms of throughput, quantifiability, latency, accessibility, and dependability, whereas the second domain targets work handling, effective resource allocation, information security, value profit analysis, valuation schemes, and therefore the arrangements to satisfy user's service-level agreements. Most of the wants in the second domain have associate degree influence over their fastened infrastructures that motivate the suppliers to style value-effective cloud storage infrastructures and services. Cloud storage service refers to the act of storing information in remote databases maintained by third-party service suppliers, rather than storing in native storage devices. There are many different cloud storage systems. Some have specific focus like storing Internet e-mail messages or digital footage.

Others are obtainable to store all styles of digital information. At its most simple level, a cloud storage system needs only one information server connected to the Web. A consumer (e.g., a mortal subscribing to a cloud storage service) sends copies of files over the Web to the information server that then records the knowledge. When the consumer needs to retrieve the knowledge, he accesses the information server through a Web-based interface. The server then either sends the files back to the consumer or permits the consumer to access and manipulate the files on the server itself. Cloud storage systems usually think about many

information servers. As computers often need maintenance or repair, it is vital to store identical information on multiple machines. This introduces redundancy. Although cloud storage systems cannot guarantee shoppers non-redundancy, they might access their information at any given time. Most systems store identical information on servers that use different power providers. In this way, shoppers can access their information even if one of the power provider fails. Challenges such as effective resource utilization, huge information handling (handling exabyte or zettabyte of information per second), information privacy, and quicker transition are key problems in cloud database.

This chapter deals with the concern of how to manage and control a big data process during migration, for example, a long-distance data transfer, cloud cost modeling, cloud energy measurement and optimization, application modeling for cloud migration, disaster recovery/backup, clouds to support e-science, and cloud service architectures. The progress is based on penetrating into two principles.

- Probes are sent among nodes so as to compute the probe response delay and drop rate on each link.
- Adaptive probe trials are executed in each node for the detection of anomalous network behavior.

Based on the expected probe response delay and therefore the expected drop rate, probe tests and inquiring intervals are autonomously tailored to the present network conditions on individual links. To cut back the communication overhead, we use two different intervals for probe tests and individual probes. For detection of communication faults, a probabilistic threshold is used to attain reliable fault detection with few false positives. Reconciling probe tests with square measure performed in every node checks the supply of adjacent nodes and links. From the discovered probe response delays, overlapping applied mathematics models square measure are compared to find and adapt to long shifts within the expected response delays.

The approach that we use has two kinds of network anomalies:

- Communication faults,
- Shifts in commonly discovered probe response delays.

When a quest takes a look at an affiliation that fails, a communication fault has been detected and fault localization method that supported node collaboration is initiated with the aim of pinpointing the fault to a link or node. Shifts within the commonly discovered probe response delay on a link square measure are detected if the previous and current latency models disagree with one another. The police investigation node can in this case report the latency shift on the link and send word to the neighboring node to cut back management message overhead. In case all links between nodes have detected latency shifts on all its connections, more or less at the same time, the node can report associate degree alarm regarding the present state. The main aim of our work is to develop the changed quantity constant detection mechanism (cbPDM), which is totally passive, incurs no further network overhead, and operates on mixture traffic.

Furthermore, this work suggests that it is possible to find anomalies and attacks that support mixed traffic at network edges, and not only close to attack victims. Our detection methodology that employs the consecutive likelihood quantitative relation takes a look at (SPRT), a time-adaptive detection technique for the two mixed traffic options we consider: packet rate and packet size. Combining the SPRTs for these two options ensures that the cbPDM is powerful against false positives and also maintains fast detection capabilities. We introduce the bit-rate signal-to-noise ratio (SNR), which is found to be an efficient metric for analysis and is superior to the antecedent-projected packet SNR metric. Our rule conjointly performs comparably to or higher than a particular set of existing detection schemes and this may be shown as a compared graph. The rest of the paper is organized as follows. Section 2 contains the work related to the small print of the earlier approaches and ways. Section 3 explains the previous approach and Sect. 4 elaborates the clarification of the projected ideas and ways. In Sect. 5, the analysis result and therefore the comparison result are explained. Section 6 contains the conclusion and future work.

2 Related Work

The information held within the cloud is owned by another person or organization aside from the cloud owner. This information is also valuable, therefore, it ought to be secure enough in order that nobody may have access to it apart from the licensed person. The information that holds on in a cloud surrounding is handled by external parties and so it is known as outsourced data. The information area unit holds on in such a way therefore to build it freelance of geographic location, to cut back the price to take care of necessities for storage like hardware and package. The advantage over the cloud is that the use is primarily based on valuation and also the prepared availability of the resources, without regard for its maintenance. However, as everything has its own pros and cons, cloud too has some drawbacks. The main problem is the protection and privacy of the information held in the cloud surroundings. This information in a cloud area unit if handled by untrusted parties can end in insecurity of information. To resolve this downside, one needs to take measures to keep this information secure. There exist several security measures for information that holds on cloud.

Abadi et al. [1] mentioned the constraints and opportunities of deploying information management problems on these rising cloud computing platforms (e.g., Amazon Web Services). They speculate that giant-scale information analysis tasks, call support systems, and application-specific information marts square measure additional seemingly to require advantage of cloud computing platforms than operational, transactional information systems a minimum of at the start. Another question of concern is a way to balance the trade-offs between fault tolerance and performance. Increasing fault tolerance generally suggests that rigorous checks must inform intermediate results, although this typically comes at a performance

value. Kelbert et al. [2] highlighted the distributed aspects of usage management. Policy specification and evolution also as social control among single systems' square measure are out of the scope. Whereas the underlying issues, planned approaches, and expected results square measure additional general, they are going to be applied to cloud services and evaluated exploitation real-world use case situations.

Yu et al. [3] planned a theme to realize this goal by exploiting KPABE and unambiguously combining it with techniques of proxy re-encryption and lazy re-encryption. Moreover, our planned theme will alter the information owner to delegate most of the computation overhead to powerful cloud servers where confidentiality of user access privilege and user secret key responsibility may be achieved. Formal security proofs show that our planned theme is secure beneath customary scientific discipline models. Chow et al. [4] planned to increase management measures from the enterprise into the cloud through use of trustworthy computing and applied scientific discipline techniques. These measures ought to alleviate a lot of today's problems in cloud computing and, they believe, have the potential to produce demonstrable business intelligence advantages to cloud participation. Hausenblas et al. [5] raised the question as to what extent NoSQL systems may be accustomed method joining information in a typical cloud computing setup. The additional general question of the suitable information management infrastructure for distributed information or science clouds is out of scope of the work at hand.

Suryawanshi et al. [6] planned a novel extremely localized information responsibility framework to keep track of the particular usage of the user's information within the cloud. This approach permits the information owner to not solely audit his content, and conjointly enforce sturdy back-end protection if required. In addition, one in every option of labor is that it allows the owner to audit even those copies of data created that do not belong to him. Madhavapeddy et al. [7] identified a number of perquisites for reconfigurable computing systems within the cloud and pick out many situations created attainable with vast cloud-based computing capability. Chakravarty et al. [8] used single-end controlled offered information measure estimation tools and a colluding network entity that may modulate the traffic destined for the victim. This threat can become more apparent and correct as future networks and hosts participate in higher end-to-end output circuits. Weinman et al. [9] propose that the cloud will force larger attention on network situations.

This new domain has the potential to extend performance, while optimizing economical resource allocation and utilization; however, it conjointly reveals new challenges in frameworks and tools of international value and performance improvement. Recently, however, a variety of sensible attacks against such obscurity systems have been planned. These attacks generally leverage a small range of compromised network entities to partly or totally expose information of a few users of such systems [11]. Moreover, it is assumed that consistent trailing of users of such obscure systems is impractical for large-scale users and organizations, as a result of the shortage of many ubiquitous adversaries. They measure this assumption for gifted and future high-speed obscurity networks. Moreover, common low-latency anonymizing systems as well as Tor cannot guarantee consistent

quality of service. This is often attributable to a mixture of the scientific computation, queuing delays, and traffic planning and learning [11]. Curiously, Evans et al. [10] explained the techniques, which conjointly degrade the adversaries' ability to successfully track anonymous users wherever low volume of traffic causes traffic analysis techniques to be less correct.

To qualify for these specific needs, strategies need to be accommodative to varied conditions facilitating analysis of abnormal behavior, and ideally operate in a distributed manner. Previous work indicates that anomaly detection and fault localization square measure typically support network traffic analysis, like traffic identification, signature matching, signal analysis, applied mathematics analysis, etc., excluding acting analyses of network traffic and active inquiry, which is another approach to anomaly detection and fault localization. As an example, Rish et al. [12] proposed a technique that uses probe choice to find and localize network faults with reduced traffic overhead. Alternative strategies for probe choice square measure proposed by Natu and Sethi et al. [13] support probe stations that monitor a collection of alternative nodes on a path. Several of these strategies were square measure developed for centrally managed networks with faster network topologies and instrumentation, probably hoping on selected nodes for watching. However, in dynamic ad hoc networks with varied topology and network instrumentation, distributed strategies for autonomous fault handling square measure are convenient.

3 Distributed Adaptive Fault detection Algorithm

In this section, is developed a distributed approach to adaptative anomaly detection and cooperative fault localization. The method used is predicated on parameter estimation of gamma distributions obtained by measurement response delays (or two-way link latency) through searching. The aim is to adaptively learn the expected latency link from each node so that manual configuration effort is decreased. Rather than specifying algorithmic rule parameters in time intervals and specific thresholds for traffic deviations thought as abnormal, parameters area unit here is fixed either as a price or as a fraction of the expected probe response delay. In this section, we describe the extension of the prevailing approach to adaptative fault handling. With the exception of detective work communication faults, our model is here extended to incorporate detection of shifts in determined network latencies. Shifts in native network latencies are symptoms of, wrong instrumentality, malicious activities, misconfiguration, or variable user behavior.

Having the ability to capture such events will, increase the potency of network management and fault handling. Further, we describe the event of an applied math learning approach with holograph properties for autonomous adaptation to long-run latency variations. Thus, the extended approach will observe latency shifts on individual links and adapt to the new "regime," while forgetting older observations step-by-step. The adaptive distributed diagnosis for fault detection (ADSD) rule is that the initial utilization of system-level identification theory has been enforced.

Before we discuss the specification of the rule, we have to clarify the ideas of “test” and “testing round” and “diagnosis latency” used in the distributed system-level identification literature. A “test” can be merely a node i causal message to node j to elicit some data. If the response is correct and on time, then node i evaluates node j as fault free. Otherwise, node j is faulty. The construct of testing spherical plays a vital role in expressing the identification latency (or capturing the time complexity) of a distributed identification rule.

A “testing round” is outlined because the amount of your time during which each fault-free node within the system has tested another node as fault-free, and has obtained diagnostic data from that node, or has tested all alternative nodes as faulty. In other words, the length of a “testing round” includes the time taken by node i to seek out a fault-free node j or value all the nodes as imperfect. For instance, assume a node i at time t start its sixth look at execution, and finds nodes $i + 1$, $i + 2$ as faulty and $i + 3$ as fault-free at time t_4 . At this point, node i stops testing. On the other hand, node $i + 3$ at time t starts its sixth check execution and finds node $i + 4$ as fault-free at time t_2 and so stops testing. Though the number of days and also the variety of tests for node i and node $i + 3$ to seek out a fault-free node area unit are different, we still say that nodes i and $i + 3$ performed their tests within the same testing spherical, that is, sixth testing spherical.

“Diagnosis latency” is outlined because of the time from the detection of a fault event to the time all the fault-free nodes properly diagnose the event. In the following, our interest is in identification latency when the last fault event has occurred. The approach is predicated on searching for two functions. First, probes area unit sent between nodes so as to live the probe response delay and drop rate on every link. Second, adaptative probe tests area unit performed in every node for detection of abnormal network behavior. Therefore the expected drop rate supported the expected probe response delay and, probe tests and searching intervals area unit autonomously custom-made to the present network conditions on individual links. To scale back communication overhead, they use two different intervals for probe tests and individual probes. For detection of communication faults a probabilistic threshold is employed to realize reliable fault detection with few false positives. Adaptative probe tests area unit were performed in every node to check the supply of adjacent nodes and links. From the collected probe response delays, overlapping applied math models area unit were compared to observe and adapt to long-run shifts within the expected response delays.

The approach that we use has two kinds of network anomalies: communication faults and shifts in ordinarily determined probe response delays. Once an enquiry to check an affiliation fails, a communication fault has been detected and fault-localization method supported node collaboration is initiated with the aim of pinpointing the fault to a link or node. A shift within the ordinarily determined probe response delay on a link is detected if the previous and current latency models dissent considerably from one another. The detective work node can report the latency shift on the link and advise the neighboring node to scale back management message overhead. In case all links between nodes have detected latency shifts on all its connections a lot of or less at the same time, the node can report associate

alarm regarding the present state. The detection of latency shifts generates a burst of alarms till the training model has converged to the new latency regime. The algorithmic rule presently uses easy thresholds to stop causing over one alarm per detected latency shift.

4 Proposed Scheme

In this section, we derive the SPRTs for the packet rate and packet size options that square measure the first parts of the cbPDM. The cbPDM operates on a simplex sampled time series of combination network traffic. The constant models used to derive the cbPDM are not representative of general net traffic, however, square measure is chosen to differentiate between the presence-of-anomaly and background-only hypotheses. A classical SPRT assumes identified and constant model parameters. In reality, such parameter values are not invariably accessible, and therefore, we tend to take into account a generalized chance magnitude relation check (GLRT), defined as $G_{N(X)} = \prod_{k=1}^N \frac{p(x_k, \hat{\Theta}_1|H_1)}{p(x_k, \hat{\Theta}_0|H_0)}$ where we use the notation $p(x_k, \hat{\Theta}_i|H_i)$ to denote exchange actuality values of the model parameters $\hat{\Theta}_i$ of the conditional probability density $p(x_k|H_i)$ with their maximum-likelihood (ML) approximations $\hat{\Theta}_i$. To create the generalized SPRT, the calculable parameters square measure are substituted into the check kind as antecedently delineated. Especially, we continue taking observations if $A < G_{N(X)} < B$ and make a decision choosing H_0 or H_1 if $G_{N(X)} \leq A$ or $G_{N(X)} \geq B$, respectively. When applying the GLRT, the model parameters linked with either or both densities may be anticipated. We take on the information $\hat{\theta}_i = \hat{\theta}|H_i$ to indicate the estimate $\hat{\theta}$ of the parameter θ when H_i is true. In this, for similarity of presence-of-anomaly and background-only hypotheses, the individual model parameters are approximated by means of observations in the SPRTs for both the characteristics. Particularly, the model parameters are informed using nonoverlapping casements. We at the outset utilize fixed-size windows for mutual hypotheses; a 1-s sliding window guarantees that sufficient data are being gathered to obtain good approximates of the background and attack parameters, which indicated $M_{\text{init}} = N_{\text{init}} = 1$ s. To estimate H_1 , the offset window is utilized. At any time the SPRT crosses the lower threshold obtaining the absence of an attack, the ASN (average sample number) function is calculated under hypothesis H_0 , along with the update window size reset to $M = \min\{\mathbb{E}_0(N), M_{\text{init}}\}$. Likewise, when an attack is discovered by the cbPDM, the length of the update window for the H_1 parameters is reset to $N = \min\{\mathbb{E}_0(N), N_{\text{init}}\}$.

The cbPDM combines the SPRTs of the packet rate and packet size options. For completion of the bPDM, we first recall that the bPDM should be at first deployed within the absence of an abnormality. Formerly, the initial parameter estimates are computed, succeeding observations square measure will not update the parameter

estimates for each hypotheses and reckon the chance ratios. For each of the SPRTs, the chance magnitude relation is updated. The continual change in the chance magnitude relation and also H_0 and H_1 parameters, calculable using a field variety of samples, obviates the need for a priori data of the background or baseline parameters. Throughout the operation of the bPDM, if just one of the SPRTs (packet rate or packet size) crosses the higher threshold B, then we declare an opening warning and continue computing the chance magnitude relation when resetting the corresponding SPRT.

For instance, a rise in the packet rate without any significant modification in the sample entropy of the packet size distribution could also be attributable to a standard nonmalicious increase in traffic. Thus, an attack is asserted, providing that an initial warning is followed by the opposite SPRT crossing the higher threshold, i.e., we declare an attack, providing that each packet rate and packet size SPRTs “coincidentally” crosses the higher threshold. Requiring the SPRTs to cross the higher threshold at a similar sample is too restrictive, being the equivalent of msec accuracy; therefore, we define the “hold time,” $\tau_{Hp} = 0.1$ s, and require the SPRTs to cross the higher threshold at intervals τ_{Hp} of different samples. Consequently, a false positive occurs once each SPRT coincidentally crosses the higher threshold and there is no anomaly gift within the traffic. The salient options of cbPDM operation square measure are highlighted in Fig. 1. Different from the cbPDM operation delineated, the packet rate may even be used singly to notice anomalies. However, this can lead to a significant variety of false positives being declared, since most legitimates will increase in traffic that would be flagged as attacks.

In the cbPDM framework, the quantity of packet arrivals within the interval is $\left[\frac{i}{p}, \frac{(i+1)}{p}\right]$. Let S_i denote the set of distinct packet sizes that arrive during this interval, and q_i denote the proportion of packets of size j to the overall range of packets within the same interval. Thus, the sample entropy y_i is computed as $y_i = -\sum_{j \in S_i} \log q_j$. The sample entropy is modeled using the Gaussian distribution given by $p(y|H_i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[-\frac{1}{2\sigma_i^2}(y - \mu_i)^2\right]$ for each of the background ($i = 0$) and attack ($i = 1$) hypotheses. Thus, the log-likelihood magnitude relation (LLR), in the given observations, is specified as $\log L(y) = a_2 \sum_{i=1}^N y_i^2 + \sum_{i=1}^N y_i + a_0$ where $a_2 = \left(\frac{1}{2\sigma_1^2}\right) - \left(\frac{1}{2\sigma_0^2}\right)$, $a_1 = \left(\frac{\mu_1}{\sigma_1^2}\right) - \left(\frac{\mu_0}{\sigma_0^2}\right)$, and $a_0 = N \left[\left(\frac{\mu_0^2}{2\sigma_0^2}\right) - \left(\frac{\mu_1^2}{2\sigma_1^2}\right) + \log\left(\frac{\sigma_0}{\sigma_1}\right)\right]$.

As in the case of the GPD/sGPD hypothesis check, the model parameters in the case of the sample entropy square measure are calculable in real-time exploitation the sliding and growing update windows. Since the sample entropy is modeled using Gaussian distribution, the parameter estimators for μ and σ^2 for each of the hypotheses are the sample mean \bar{x} and sample variance s^2 , given by $\bar{x} = \frac{1}{M} \sum_{i=1}^M x_i$ and $s^2 = \frac{1}{M-1} \sum_{i=1}^M (x_i - \bar{x})^2$ using the individual update windows. The ensuing SPRT needs that we still take additional observations if $\log(A) < \log G(y) < \log(B)$ where $G(y)$ is the generalized chance magnitude relation related to the packet size SPRT. The form of $\log L(y)$ is $\log G(y)$, but the constants a_2 , a_0 , and a_1 are defined in terms of the parameter estimates and

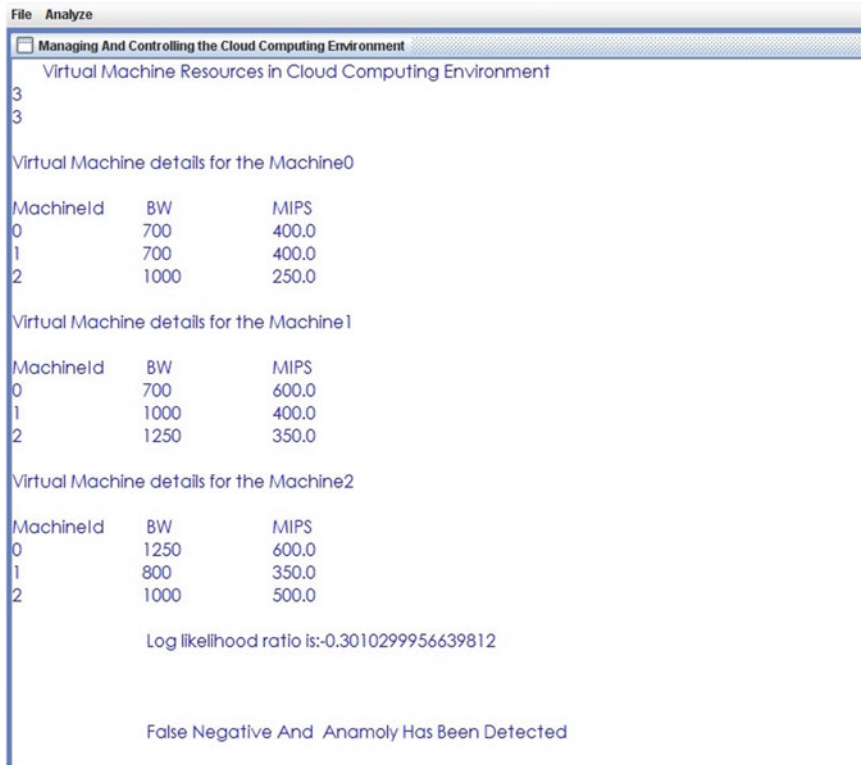


Fig. 1 Network bandwidth and link latency detector

$\{\hat{\mu}_0, \hat{\sigma}_0^2\}$ and $\{\hat{\mu}_1, \hat{\sigma}_1^2\}$ instead of the true parameter values. Given two options, ideally we might compute a joint density to work out one quantity SPRT. However, given the mixed nature of the two options (discrete packet arrivals and continuous entropies), computing this joint density seems to be noncompliant. Instead, we currently describe our cbPDM rule that effectively combines the two SPRTs to yield anomaly detection mechanism that features a low chance of false positives.

Pseudocode of cbPDM

Input: Number of nodes, number of VMs, and number of links

Output: Anomaly detection and false positive

Start Comp_SPRTs(LLR)

if(SPRT LLR $\frac{1}{2} < \text{Log}(A)$)

then Reset LLR $\frac{1}{2}$

Update(H_0)

else

Check (SPRT (LLR 1, LLR 2))

if SPRT (LLR 1, LLR 2) $> \text{Log}(B)$

```

Stop Update( $H_0$ )
else Comp_SPRTs(LLR)
check  $\rightarrow H_1$  is true
if true
display Anomaly
else
display False Positive

```

5 Experimental Results and Discussion

In this section, we compare the packet SNR to the bit-rate SNR and we find that the latter may be a simpler example of anomaly strength for this application. The packet SNR is defined as follows:

$$\text{pkt}_{\text{rate}} = \frac{\text{No. of attack packets}}{\text{No. of background packets}} \quad (1)$$

Given that each metrics square measure equivalent in the case of attacks uses fixed-size attack packets, however, as the packet SNR yields unreasonable end in the case of the sensible opponent, we conclude that the bit-rate SNR is a good metric for comparison and analysis and is better than the packet SNR. The graph shows that the packet rate for the planned system is simpler than the present system. In Figs. 2 and 3, the proposed system is suddenly decreased during 10,000–15,000 packet rate compared to the existing algorithm, which shows that the packet SNR is equivalent to the bit-rate SNR in the case of attacks with fixed-size packets.

The cbPDM uses each packet size each as a feature for detection and to cut back false positives. We currently contemplate the sensible resister state of affairs, whereby the assailant constructs AN attack whose distribution of packet sizes makes an attempt to match that of the background traffic. For this purpose, we produce a collection of sensible resister artificial attacks whereby the attack stream uses a continuing bit rate, however, with a distribution of packet sizes that is drawn from the bimodal distribution. We recall that the packet size distribution of nominal net traffic has been characterized as principally bimodal, a result valid by AN examination of our background trace information.

We observe the theoretical time to detect tendencies as within the experimental cases: The time to detect decreases because the bit-rate SNR will increase. Within the theoretical case, we find that the observation time is associated with mathematical function of the bit-rate SNR; lower rate attacks take a significantly longer time to detect than higher rate ones. The theoretical detection times square measure is below the empirical times since there is no notion of cross traffic, or interaction between the packets from the background and attack streams, as toughened in very real virtual machines. An analogous movement is seen within the case of

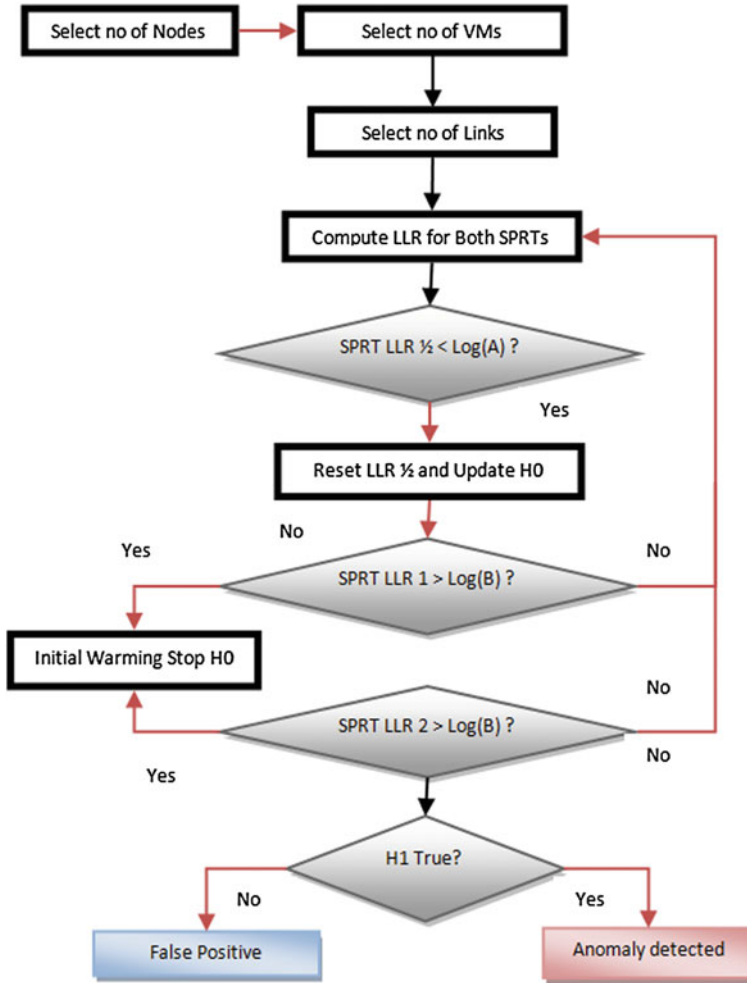


Fig. 2 Customized bivariate parametric detection mechanism

experimental information; however, a rigorous fit cannot be performed as a result of the tiny set of averaged information points.

Thus, we find that attacks with higher bit-rate SNR values square measure detected additional quickly for the simulated and emulated attacks, which is in line with what is foreseen by the underlying theoretical model. The essential principles of detection theory show that the time to detect a signal in noise is related to the SNR. On the other hand, for anomaly detection, here is no clear notion of what an appropriate SNR measure would be. We gift the bit-rate SNR metric that is defined as

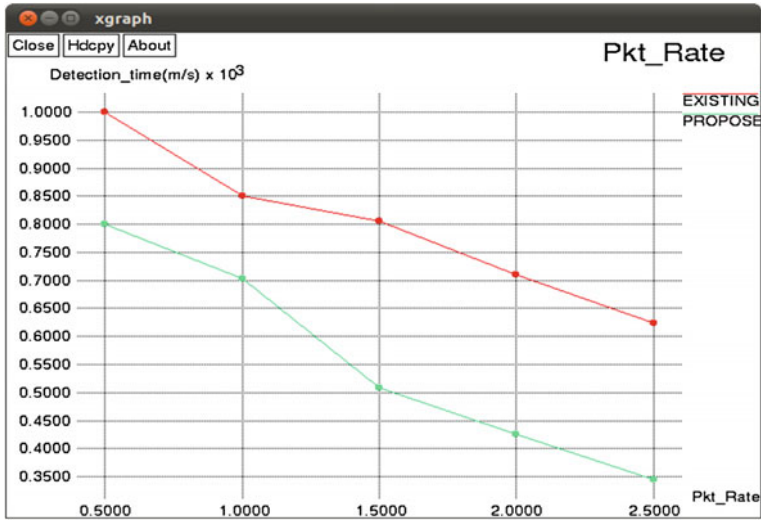


Fig. 3 Packet rate comparison

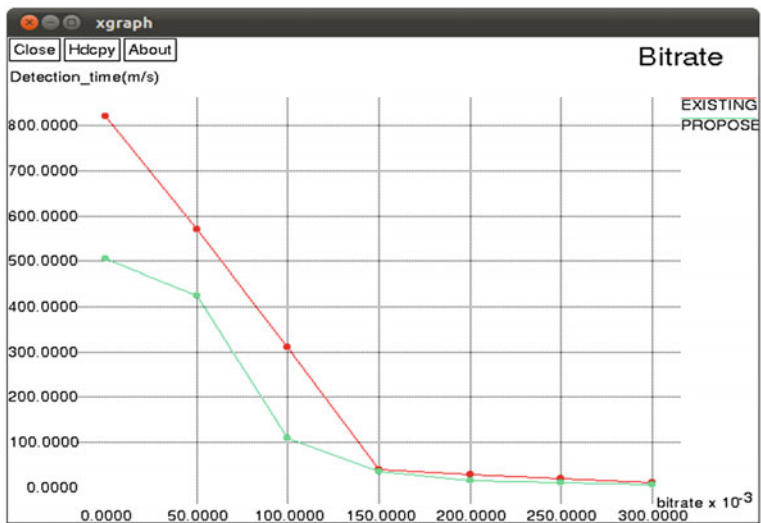


Fig. 4 Bit-rate comparison

$$bit_{rate} = \frac{\text{Anomalous traffic level}}{\text{Background traffic}} \tag{2}$$

Given that each metrics square measure is equivalent in the case of attacks that use fixed-size attack packets, although the packet SNR yields unreasonable end in

the case of the sensible opponent, we conclude that the bit-rate SNR is a good metric for comparison and analysis and is better than the packet SNR. The graph shows that the bit rate for the planned system is simpler than the present system (Fig. 4).

6 Conclusion and Future Work

We have employed the customized bivariate parametric detection mechanism (cbPDM) that might observe anomalies and low-rate attacks for a few seconds. This advance permits the period estimation of model parameters and only needs 2–3 s of background-only traffic for coaching. Integrating the packet rate and packet size options permits us to observe anomalies in encrypted traffic and avoid state-intensive flow pursuit since our methodology does not use flow-separated traffic; what is more, combining these same two options conjointly eliminates most false positives. Our planned theme permits the knowledge owner to hand over most computation-intensive tasks to cloud servers while not revealing data contents or user access privilege information. Consequently, our planned theme can serve as a perfect candidate for knowledge access control within the rising cloud computing surroundings. By contrast, existing access control schemes in connected areas both lack measurability and do not offer adequate proof of knowledge confidentiality. In future, confidentiality of user access privilege and user secret key responsibility is achieved.

Acknowledgments This paper is sponsored by the University Grants Commission of India, under the National Fellowship Program Grant no. TAM—24467.

References

1. D.J. Abadi, Data management in the cloud: limitations and opportunities (2009)
2. F. Kelbert, Data usage control for the cloud. Technische University at Munchen (TUM)
3. S. Yu, C. Wang, K. Ren, W. Lou, Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing. Dept. of ECE, Illinois Institute of Technology
4. R. Chow, P. Golle, M. Jakobsson, R. Masuoka, J. Molina, Controlling data in the cloud: outsourcing computation without outsourcing control
5. M. Hausenblas, R. Grossman, A. Harth, P. Cudre-Mauroux, Large-Scale Linked Data Processing: Cloud Computing To the Rescue? University of Chicago and Open Cloud Consortium
6. S.P. Suryawanshi, A.M. Bagade, Secure data processing in cloud computing. *Int. J. Comput. Appl.* **76**(5), 0975–8887 (2013)
7. A. Madhavapeddy, S. Singh, Reconfigurable Data Processing for Clouds (University of Cambridge, Cambridge)
8. S. Chakravarty, A. Stavrou, A.D. Keromytis, Traffic analysis against low-latency anonymity networks using available bandwidth estimation

9. J. Weinman, Network implications of cloud computing, in Technical Symposium at ITU Telecom World (ITU WT) (2011)
10. N. Evans, R. Dingleline, C. Grothoff, A practical congestion attack on tor using long paths, in *Proceedings of the 18th USENIX Security Symposium (USENIX Security)*, (2009), pp. 33–50
11. J. Reardon, I. Goldberg, Improving tor using a TCP-over-DTLS tunnel, in *Proceedings of 18th USENIX Security Symposium* (2009)
12. I. Rish, M. Brodie, S. Ma, N. Odintsova, A. Beygelzimer, G. Grabarnik, K. Hernandez, Adaptive diagnosis in distributed systems. *IEEE Trans. Neural Netw.* **16**, 1088–1109 (2005)
13. M. Natu, A.S. Seti, Efficient probing techniques for fault diagnosis, in *2nd International Conference on Internet Monitoring and Protection (ICIMP 2007)* (2007) pp. 2085–2090
14. R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, J. Molina, Controlling data in the cloud: Outsourcing computation without outsourcing control, in *Proceedings of ACM Workshop on Cloud Computing Security* (2009)
15. J. Park, R. Sandhu, Towards usage control models: Beyond traditional access control, in *Proceedings of 7th ACM Symposium on Access Control Models and Technologies* (2002)
16. A. Pretschner, M. Hilty, F. Schutz, C. Schaefer, T. Walter, Usage control enforcement: present and future. *IEEE Secur. Priv.* **6**(4) (2008)
17. M. Harvan, A. Pretschner, State-based usage control enforcement with data flow tracking using system call interposition, in *Proceedings of 3rd International Conference on Network and System Security* (2009)
18. T. Wuchner, A. Pretschner, Data loss prevention based on data driven usage control, in *Proceedings of 23rd IEEE International Symposium on Software Reliability Engineering* (2012)
19. A. Pretschner, E. Lovat, M. Buchler, Representation-independent data usage control, in *Proceedings of Conference on Data Privacy Management* (2011)
20. B. Katt, X. Zhang, R. Breu, M. Hafner, J.-P. Seifert, A general obligation model and continuity-enhanced policy enforcement engine for usage control, in *Proceedings of 13th ACM Symposium on Access Control Models and Technologies* (2008)
21. P. Kumari, A. Pretschner, J. Peschla, J.-M. Kuhn, Distributed data usage control for web applications: A social network implementation, in *Proceedings of 1st ACM Conference on Data and Application Security and Privacy* (2011)
22. G. Ateniese, K. Fu, M. Green, S. Hohenberger, Improved proxy re-encryption schemes with applications to secure distributed storage, in *Proceedings of NDSS'05* (2005)
23. S.D.C. di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, P. Samarati, Over-encryption: Management of access control evolution on outsourced data, in *Proceedings of VLDB'07* (2007)
24. V. Goyal, O. Pandey, A. Sahai, B. Waters, Attribute-based encryption for fine-grained access control of encrypted data, in *Proceedings of CCS'06* (2006)
25. R. Bianchini, R. Buskens, Implementation of on-line distributed system level diagnosis theory. *IEEE Trans. Comput.* **41**(5), 616–626 (1992)
26. Elias Procopio Duarte Jr, Takashi Nanya, A hierarchical adaptive distributed system level diagnosis algorithm. *IEEE Trans. Comput.* **47**(1), 34–45 (1998)
27. M.-S. Su, K. Thulasiraman, A. Das, Multilevel adaptive distributed fault location in a network of processors, in *Proceedings of the Allerton Conference on Communication, Control and Computing* (2001)
28. F.P. Preparata, G. Metze, R.T. Chien, On the connection assignment problem of diagnosable systems. *IEEE Trans. Electr. Comput.* **16**, 848–854 (1967)

Development of Concatenative Syllable-Based Text to Speech Synthesis System for Tamil

B. Sudhakar and R. Bensraj

Abstract This paper addresses the problem of improving the intelligibility of the synthesized speech in Tamil text-to-speech (TTS) synthesis system. The human speech is artificially generated by speech synthesis. The normal language text will be automatically converted into speech using TTS system. This paper deals with a corpus-driven Tamil TTS system based on the concatenative synthesis approach. Concatenative speech synthesis involves the concatenation of the basic units to synthesize an intelligent, natural sounding speech. In this paper, syllables are the basic unit of speech synthesis database and the modification of syllable pitch by timescale modification. The speech units are annotated with associated prosodic information about each unit, manually or automatically, based on an algorithm. An annotated speech corpus utilizes the clustering technique that provides way to select the suitable unit for concatenation, depending on the minimum total joint cost of the speech unit. The entered text file is analyzed first, this syllabication is performed based on the linguistics rules, and the syllables are stored separately. Then, the syllable corresponding speech file is concatenated and the silence present in the concatenated speech is removed. After that, discontinuities are minimized at syllable boundaries without degrading the quality. Smoothing at the concatenated syllable boundary is performed, changing the syllable pitches by timescale modification.

Keywords Linear predictive coding · TTS synthesis system · Corpus-driven Tamil TTS system · PCM

B. Sudhakar (✉) · R. Bensraj
Department of Electrical Engineering, Annamalai University, Annamalai Nagar,
Chidambaram, India
e-mail: balrajsudhakar@gmail.com

R. Bensraj
e-mail: bensraj_au@rediffmail.com

1 Introduction

Over the past years, there has been an immense development in speech technologies. Among the applications of speech technology, the automatic speech production, which is referred to as text-to-speech (TTS) system, is the most natural sounding technology. TTS synthesis is the process of converting ordinary orthographic text into speech signal which is indistinguishable from human speech [1–10]. It can be widely classified into front end and back end as shown in Fig. 1. The conversion of natural language text to a structured linguistic representation is associated with front end. From the raw text, this front end identifies a sequence of segments called target segments. These target segments have different features estimated from the text. The back end is referred as the second part of the system which modifies these target segments into a speech waveform.

There are two main methods used for speech production. These methods are format synthesis and concatenation synthesis illustrated in [11]. The format synthesizer utilizes a simple model of speech generation and a set of rules to generate speech. While these systems can achieve enhanced intelligibility, their naturalness is typically low, since it is very tedious to perfectly describe the process of speech produced in a set of rules. The TTS has been the main research focus in automatic speech production in Indian languages nowadays. Some of TTS systems for Indian languages like Hindi, Telugu, Tamil, and Bengali have been developed using the unit selection and festival framework in [2, 6]. Listeners are able to clearly perceive the message with little attention and act on synthesized speech of a command correctly and without perceptible delay in noisy environments. Although many TTS approaches, the intelligibility, naturalness, comprehensibility, and recall ability of synthesized speech is not good enough to be widely accepted by users. There is still considerable rule for further improvement of performance of the TTS production system.

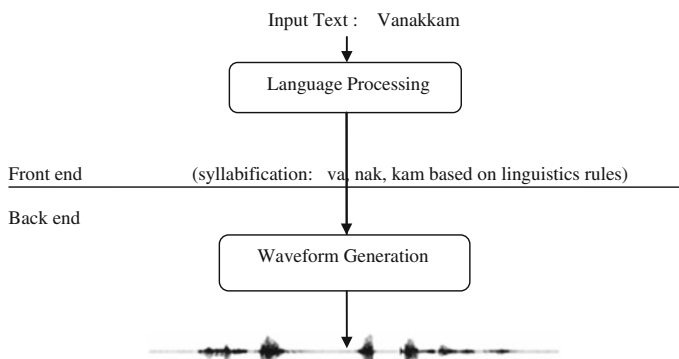


Fig. 1 Parts of speech synthesis system

This paper proposed corpus-driven TTS system. In this paper, the concatenative-based approach is used to produce desired speech through prerecorded speech waveforms. Over the past decades, this proposal was very complicated to implement because of limitation of computer memory. With the advancements in computer hardware and memory, a large quantity of speech corpus can be stored and utilized to produce high-quality speech signal for a given text. Thus, the synthesized speech preserves the naturalness and intelligibility. Here, the given input text is analyzed first. Based on the linguistics rules, syllabication is performed. The syllables are called basic speech units. The repository of these units is created with its prosodic information.

The pitch value for syllable is changed by performing timescale modification. During the synthesis, these units are selected and concatenated with lowest joint cost. After performing concatenation, the waveform is smoothed at concatenation joints using linear predictive coding (LPC). The rest of this paper is organized as follows. In Sect. 2, the concatenative speech synthesis system and syllabification rules for Tamil are discussed. In Sect. 3, the proposed Tamil speech-to-text synthesis system is described. In Sect. 4, the quality test is discussed. Finally, in Sect. 5, synthesized speech waveforms and conclusions are provided.

2 Concatenative Speech Synthesis

Concatenative speech synthesis utilizes phones, diphones, syllables, words, and sentences as basic units. Based on selecting, these units from the database speech are synthesized, called as a speech corpus. Many researches have been made, selecting each separate unit as the basic unit. When phones are selected as basic units, for Indian languages, the size of the database will be less than 50 units. Database may be small, but phones give very poor co-articulation data across neighboring units, thus falling to model the dynamics of speech sounds. Diphones and triphones as basic units, it will minimize the discontinuities at the concatenation points and captures the co-articulation effects. But a single example of each diphone is not enough to generate precious quality speech. So this paper presented a syllable as a basic unit. Indian languages are syllable centered, where pronunciations are based on syllables. For Indian language intelligible speech synthesis, a syllable can be the best unit.

The general form of Indian language syllable is C^*VC^* , where C is a consonant, V is vowel, and C^* indicates the presence of 0 or more consonants. There are 18 consonants and 12 vowels in Tamil languages. There are defined set of syllabification rules formed by researchers, to generate computationally reasonable syllables. Some of the rules used to perform grapheme to syllable conversion [12] are as follows:

- Nucleus can be vowel (V) or consonant (C)
- If onset is C, then nucleus is V to yield a syllable of type CV
- Coda can be empty of C
- If character after CV pattern is of type CV, then the syllables are split as CV and CV
- If the CV pattern is followed by CCV, then syllables are split as CVC and CV
- If CV pattern is followed by CCCV, then the syllables are split as CVCC and CV
- If the VC pattern is followed by the V, then the syllables are split as V and CV
- If the VC pattern is followed by CVC, then the syllables are split as VC and CVC.

The following new rules have been implemented in this paper to implement grapheme to syllable conversion

- If character after CV pattern is of type CV, then the syllables are split as CVCV
- Similarly, if character after CV pattern is of type CVCV, then the syllables are split as CVCVCV
- If the CV pattern is followed by CVC, then syllables are split as CVCVC
- If the CV pattern is followed by CCV, then syllables are split as CVCCV.

This paper proposed the following recommended combinations to achieve the best acceptable synthesis:

- Monosyllables at the beginning of a word and bisyllables at the end.
- Bisyllables at the beginning of a word and monosyllables at the end.
- Monosyllables at the beginning and trisyllables at the end of a word.
- Trisyllables at the beginning and monosyllables at the end of a word.

3 Proposed Tamil TTS Synthesis System

3.1 Text Analysis

To implement the proposed TTS system, the MATLAB 2012 has been used. In text analysis, first stage is text normalization, then performing removal of punctuations such as double quotes, full stop, and comma. A pure sentence is synthesized at the end of text analysis. Then, all the abbreviations present in the input text are expanded and also unwanted punctuation like (:, ; '\$') are removed to avoid confusion and not to give any disturbance in the naturalness of the speech.

The next step in the text normalization is normalizing non-standard words like abbreviations and numbers. The next stage in the text analysis is sentence splitting. In this stage, the given paragraph will be split as sentences. From these sentences, words are separated out. The last stage is romanization which is the representation of written words with a roman alphabet. In this system, romanized form of Tamil word/syllables is generated.

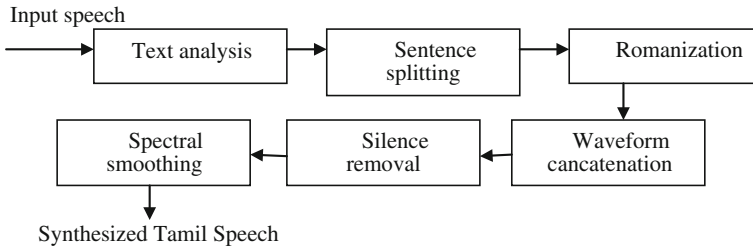


Fig. 2 Block diagram of proposed Tamil text-to-speech synthesis system

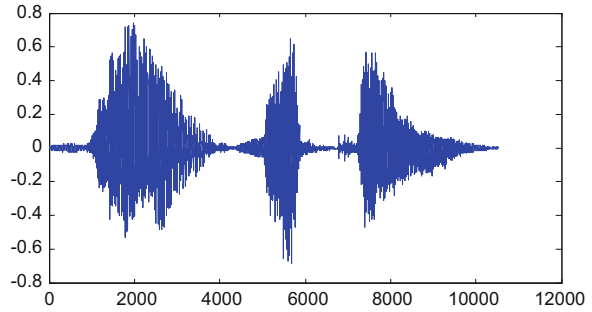
3.2 *Speech Corpus*

Building a speech corpus for Indian languages is a difficult task than that of English speech corpus. Prosodic information such as pitch, duration, and intonation prediction has to be done in the corpus development stage itself, and some more information has to be specified with the basic speech units after storing them in the corpus. The problems such as mispronunciation, untranscribed speech units, phrase boundary detection, and pronunciation variants are to be identified and addressed. For corpus creation, we selected one person for recoding these basic units, who has uniform characteristics of speaking, pitch rate, and energy profile and developed speech corpus in [2]. The digitized speech signal with sampling rate of 16 kHz and 16-bit resolution (pulse code modulation uncompressed data format) is proposed in [2]. The speech wave files are saved according to the requirement. The speech wave files corresponding to the Tamil words are named according to their corresponding romanized names. The words collected comprises dictionary words, commonly used words, Tamil newspapers, and story books, also different domain such as sports, news, literature, and education for building unrestricted TTS initiated in [2] (Fig. 2).

3.3 *Waveform Concatenation*

In the final stage of the concatenation process, the required syllables are retrieved from the corpus based on the text analysis and arranged to produce the speech. Then, all the arranged speech units are concatenated using a concatenation algorithm. The main problem in concatenation process is that there will be glitches in the joint. These are removed in the waveform smoothening stage. The concatenation process combines all the speech files which are given as an output of the unit selection process and then made into a single speech file.

Fig. 3 Resulting concatenated waveform after performing silence removal



3.4 Spectral Smoothing

The timescale modification is carried out for each syllable to produce individual smoothness for syllable in Tamil TTS. The timescale modification is used to change the pitch value for Tamil syllable. Praat software is used to calculate the duration value for each syllable [13]. Smoothing at concatenation joints are performed using LPC. The LPC is used for representing the spectral envelope of a digital signal of speech using the information of a linear predictive model. It is one of the most powerful methods for encoding enhanced quality speech at a low bit rate and gives extremely accurate estimates of speech features in [14, 15]. Now, we are getting the improved quality speech for the given input text. It can be played and stopped anywhere needed. The main aim of the proposed scheme is to achieve good naturalness in output speech. Figure 3 shows the smoothed output waveform.

4 Quality Test

For developing a Tamil TTS, we have considered 700 sentences for recording the speech corpus. These are selected from various domains from newspaper, Wikipedia, news broadcast, and story books. After formulating the data, speech corpus will be recorded in a studio environment with a suitable trained speaker. Recorded 700 sentences of speech corpus are classified into two parts: (i) part-1 contains 600 sentences for building the training corpus and (ii) part-2 contains 100 sentences for evaluating the established Tamil TTS system. Five speakers' voices are selected for constructing prototype TTS systems. From this proposed work, the following performance are inferred that the speech of five speakers with respect to (i) the quality of the synthesized speech (ii) Variations in natural prosody and (iii) the perceptual distortion with respect to prosodic and spectral modifications.

Evaluation of quality of the synthesized speech is carried out by subjective measures. Intelligibility and naturalness are estimated from the listening tests. Tests are conducted with 25 research scholars in the age group of 23–35 years. The subjects have sufficient speech knowledge for proper assessment of the speech

Table 1 Instructions to evaluators

Score	Subjective perception
1	Poor speech, with distortion and very low intelligibility
2	Poor speech with distortion and intelligibility
3	Good speech with less distortion and intelligibility
4	Very good speech quality with less naturalness
5	As good as natural speech

Table 2 Mean opinion scores for three sets

Test set	MOS
Set-1	4.01
Set-2	3.25
Set-3	2.97

signals, as all of them have taken a full semester course on speech technology. Speech utterances corresponding to the test sets are synthesized using the developed Tamil TTS system. Each of the subjects was given a pilot test about perception of speech signals by playing the original speech samples of the test files. Once they are comfortable with judging, they are allowed to take the tests. The tests are conducted in the laboratory environment by playing the speech signals through headphones. In the test, the subjects were asked to judge the distortion and quality of the speech. Subjects are asked to assess the quality and distortion on a 5-point scale for each of the sentences. The 5-point scale for representing the quality of speech and the distortion level is given in Table 1.

For evaluating the quality of synthesized speech generated from the developed TTS system, three sets of test utterances are considered. Each set consists of 20 sentences: Set-1: All the words are available from the training data, but the entire word sequences are not present in the training data and Set-2: 50 % of the words available in training corpus. Set-3: none of the words are available in the training corpus. Table 2 shows the MOS scores for the three test sets. From Table 2, it is observed that MOS for Set-1 is more compared to Set-2 and Set-3 as all the words of sentences in Set-1 are present in database. So it provides better performance compared to other two sets.

5 Results and Conclusion

In this proposed work, a speech synthesis system has been designed and implemented for Tamil language. A database has been created from various domain words and syllables. Syllable pitch modification is performed based on timescale modification. The speech files present in the corpus are recorded and stored in PCM

format in order to retain the naturalness of the synthesized speech. The given text is analyzed, and syllabication is performed based on the rules specified. The desired speech is produced by concatenative speech synthesis approach such that spectral discontinuities are minimized at unit boundaries. It is inferred that the produced synthesized speech is preserving naturalness and good quality based on the subjective quality test results. The final output speech file is stored in the specified location in the system for further analysis.

References

1. M. Macchi, T.R. Bellcore, Issues in text-to-speech synthesis, in *Proceedings IEEE International Joint Symposia on Intelligence and Systems* (1998), pp. 318–325
2. N.P. Narendra, K.S. Rao, K. Ghosh, R.R. Vempada, S. Maity, Development of syllable-based text to speech synthesis system in Bengali. *Int. J. Speech Technol.* **14**(3), 176–181 (2011)
3. C. Pornpanomchai, N. Soontharanont, C. Langla, N. Wongsawat, A dictionary-based approach for Thai text to speech (TTTS), in *Proceedings 3rd International Conference on Measuring Technology and Mechatronics Automation*, vol. 1 (2011), pp. 40–43
4. M.N. Rao, S. Thomas, T. Nagarajan, H.A. Murthy, Text-to-speech synthesis using syllable like units, in *National Conference on Communication, IIT Kharagpur* (2005), pp. 227–280
5. D.H. Klatt, Review of text-to-speech conversion for English. *J. Acoust. Soc. America* **82**(3), 737–793 (1987)
6. S. Majji, A.G. Ramakrishnan, Festival based maiden TTS system for Tamil Language, in *Proceedings 3rd Language and Technology Conference* (2007), pp. 187–191
7. S.P. Kishore, A.W. Black, Unit size in unit selection speech synthesis, in *Proceedings Eurospeech* (2003)
8. N.S. Krishna, P.P. Talukdar, K. Bali, A.G. Ramakrishnan, Duration modeling for Hindi text-to-speech synthesis system, in *Proceedings of International Conference on Spoken Language Processing (ICSLP 04)* (2004)
9. A. Hunt, A. Black, Unit selection in a concatenative speech synthesis system using a large speech database, in *Proceedings of IEEE International Conference Acoustic, Speech, and Signal processing*, vol. 1 (1997), pp. 373–376
10. R.J. Utama, A.K. Syrdal, A. Conkie, Six approaches to limited domain concatenative speech synthesis (2006)
11. S.D. Shirbahadurkar, D.S. Bormane, R.L. Kazi, Subjective and spectrogram analysis of speech synthesizer for Marathi TTS using concatenative synthesis, in *Recent Trends in Information, Telecommunication and Computing (ITC)* (2010)
12. S. Saraswathi, T.V. Geetha, Design of language models at various phases of Tamil speech recognition system. *Int. J. Eng. Sci. Technol.* **2**(5), 244–257 (2010)
13. K.P. Mohanan, T. Mohanan, Lexical phonology of the consonant system in Malayalam, in *Linguistic Inquiry*, vol. 15 (The MIT Press, Cambridge, 1987)
14. K. Panchapagesan, P.P. Talukdar, N.S. Krishna, K. Bali, A.G. Ramakrishnan, Hindi text normalization, in *5th International Conference on Knowledge Based Computer Systems (KBCS)* (2004)
15. T. Charoenporn, A. Chotimongkol, V. Sornlertlamvanich, Automatic romanization for Thai, in *Proceedings of the 2nd International Workshop on East-Asian Language Resources and Evaluation* (1999)

Design of Low-Power Blink Detector for Minimally Invasive Implantable Stimulator (SoC) Using 180 nm Technology

J. Joselyn Priyadarshini and S. Ravindrakumar

Abstract Facial palsy is a form of neurological problem that results in loss of the ability to blink. At present, treatments for the ocular complications that results from facial palsy are severely lacking. Neuromuscular electrical stimulation (NMES) is found to be a better solution in restoring eyeblink. NMES is the elicitation of muscle contraction using electrical impulses. Many works are going on in designing stimulator circuits at PCB level and in some lower technologies like 600nm, 350nm etc. We propose a stimulator chip that can stimulate blink in the palsied eye in coordination with the non-palsied eye. Here, the EMG signal is first detected for blink. This blink detector block is implemented using AND logic/comparator. The output of this stage drives the output stage to deliver the required stimulating current. This output stage is implemented using charge pump to deliver the stimulation current in contrast to the controller and V/I converters used in the previous works. It is found that blink detector using dynamic AND gate consumes less power of 2.612 μ W when used in stimulator chip that is used to restore blinking in paralyzed eyelid. This paper is implemented in 180nm technology using Cadence Virtuoso.

Keywords Neuromuscular electrical stimulation · AMD · Retinitis pigmentosa · EMG signal · PMOS

1 Introduction

The fast advances in micro-fabrication and silicon technologies result in the application of engineering principles and design concepts to medicine and biology for healthcare purposes (e.g., diagnostic or therapeutic) to close the gap between

J. Joselyn Priyadarshini (✉) · S. Ravindrakumar
Department of Electronics and Communication Engineering, Chettinad College of Engineering and Technology, Anna University, Chennai, Karur, Tamil Nadu, India
e-mail: jjoselyn.pres@gmail.com

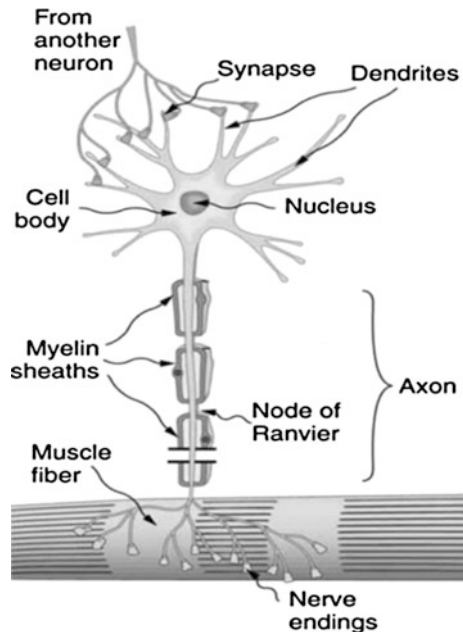
S. Ravindrakumar
e-mail: gsravindrakumar7@gmail.com

engineering and medicine. Prominent biomedical engineering applications include the development of various diagnostic and therapeutic medical devices ranging from clinical equipment to micro-implants.

1.1 Neuroscience

Of all the parts that make up the human body, the nervous system is by far the least understood complex system. Messages that the body uses to influence thoughts, senses, movements, and survival are directed by nerve impulses transmitted across brain tissue to the rest of the body. Neurons are the basic functional unit of the nervous system. Neurons have special electro-chemical properties that allow them to process information and transmit them to the muscle fibers to execute action (Fig. 1). Disconnection in transmission occurs due to damage of neurons during heavy injury or due to stroke. This results in paralysis. When this paralysis results in damage of facial nerves, they loss the ability to convey facial expression and produce eyeblink. The use of electrical stimulation to restore the blinking is found to be effective because of its convenience, naturalness, and commercial viability.

Fig. 1 Nerves controlling muscle action



1.2 Existing NMES with SoC

Active Books for cauda equine: The Active Books overcomes the limitation of conventional nerve root stimulators which can only support a small number of stimulating electrodes due to cable count restriction through the dura [2]. Instead, a distributed stimulation system with many tripole electrodes can be configured using several Active Books which are addressed sequentially.

Implantable retinal prosthetic device: Retinitis pigmentosa (RP) and age-related macular degradation (AMD) result in profound loss of vision through degeneration of light-sensing photoreceptor cells (rods/cones). It has been demonstrated that the retina of patients affected with RP and AMD can be stimulated by charge-balanced biphasic current pulses to elicit perception of vision [3].

Cochlear implant: Worldwide, nearly 100,000 people have received cochlear implants to date to overcome severe to profound sensorineural hearing impairments. An interface providing multi-point micro-stimulation and position sensing has been developed for a cochlear prosthesis, integrating a MEMS-based electrode array with signal processing electronics [4].

Diaphragm pacemaker: It has provided respiratory assistance for individuals with higher level, respiration-compromising injuries. Although mechanical ventilation provides respiratory support, it distorts the voice, limits mobility, and increases infection risks. Using NMES to stimulate diaphragmatic contractions, called phrenic nerve pacing, allows users to minimize ventilator use.

ENTERRA neurostimulator: It is implanted under the skin, usually in the lower abdominal region. Two insulated wires called leads are implanted in the stomach wall muscle and then connected to the neurostimulator. It sends mild electrical pulses through the leads to stimulate the smooth muscles of the lower stomach. This may help to control the chronic nausea and vomiting caused by gastro paresis.

INTERSTIM neurostimulator: It offers a potential bladder and bowel control mechanism for individuals with spinal cord injury. It stimulates sacral nerves to achieve desired effects. It can be used to treat urinary incontinence in individuals with complete and incomplete spinal cord injury.

2 Literature Survey

2.1 Facial Palsy

Patients suffering from facial nerve damage experience substantial disfigurement and dysfunction due to the loss of ability to convey facial expression and produce eyeblink. The preeminent ophthalmic complication in seventh nerve palsy is loss of the blink response, which predisposes patients to corneal exposure and dry eye complications. The loss of function of the orbicularis oculi muscle causes a loss of

Table 1 A decision matrix of the different methodologies of blink stimulation

Stimulations	Safety	Naturalness	Blink latency	Convenience	Commercial viability
Electrical	7	8	9	8	7
Mechanical	6	5	8	6	6
Chemical	4	6	7	5	3

the corneal “squeegee” effect, resulting in failure of the lacrimal pumping mechanism. Consequentially, these patients often suffer from conjunctivitis, exposure keratitis, corneal ulceration, and ultimately loss of vision and the eye. In addition, patients also suffer disfigurement due to a loss of the lower lid orbicularis oculi muscle function and gravitational action on the lower face producing paralytic ectropion and lower eyelid retraction. Paralytic upper eyebrow ptosis also inhibits the ability to convey facial expression. Disfiguration is responsible for high rates of depression, anxiety, and social isolation in facial paralysis patients. Given the lack of effective therapies, complications related to the loss of eyeblink and visions are profoundly disturbing for these patients. Comparison of different blink stimulation methods [5] available is given in Table 1.

2.2 Survey of Circuits Used in on Chip Stimulator

Blink detector—Tripple-threshold algorithm: The processed EMG signal is first digitized using a 10-bit ADC with a sampling rate of 2 kHz. A triple-threshold algorithm was designed and implemented in an internal 8051 MCU of the same transceiver chip. Compared with quiescent state, the EMG signal of blink state has much larger amplitude and duration around 0.3 s. Hence, three different kinds of thresholds (*Thred amp*, *ThredWinLth*, and *ThredWinNum*) are applied to the sampled EMG signal one after another. The first threshold *Thred amp* is used to judge the amplitude of the digitized EMG signal. If the amplitude of the sampled signal exceeds the value of *Thred amp*, the counter, *Data counter*, increases by 1. The other two thresholds, *ThredWinLth* and *ThredWinNum*, are used to determine the duration of signal surpassing *Thred amp* to eliminate the interference of accidental spikes of the signal which may incur misjudgment. Within a window length of 30 data points, once the number of data surpassing *Thred amp* in amplitude, which is recorded by the counter *Data counter*, exceeds *ThredWinLth*, and such window appears three times (*ThredWinNum*) in series, a blink is confirmed. Then, the MCU will trigger onetime stimulation to the paralyzed eyelid by sending a set of parameters to the stimulator. This can also be implemented using FPGA [6].

Blink detector—Using comparator: The preamplifier stage is employed in order to reduce the offset and prevent kickback from the regenerative feedback to the sensitive input signal. The bias current is set to 1 μ A. In this design, the drain

terminals are directly connected to Vdd during the reset phase when ϕ latch is low. Two inverters are placed at the latch output in order to restore the outputs to logic high and low values [7].

3 Proposed Work

Our main aim is to design a low-power blink detector for stimulator chip to restore blinking in the unilaterally facial paralysis. The proposed work mainly consists of two blocks. They are blink detector and output stage (Fig. 2).

3.1 Signal Conditioning

The EMG signal is obtained from the healthy eye using electrodes. It is about $100 \mu\text{v}$ and is firstly pre-amplified by an instrumentation amplifier to reject common mode interference, then filtered by a 4th-order Chebyshev 200–500 Hz band-pass filter, and further magnified by a non-inverting amplifier stage. The signal is finally processed by a peak detector to generate the envelope of the signal to facilitate blink detection. Our design does not involve all these steps. We make use of the peak envelope obtained after these steps as input for our design. This peak envelope signal is amplified to 1.8 Vp so that it can be used as input voltage in 180nm technology.

3.2 Blink Detector

This blink detector decides the occurrence of blink. In the existing methods, it is implemented using microcontroller. Our blink detector is constructed using a AND gate and comparator. The AND gate is used both in static and dynamic logic for comparative study. In this logic both inputs must be high for the output to become high which in turn represents the presence of blink. The dynamic logic has the

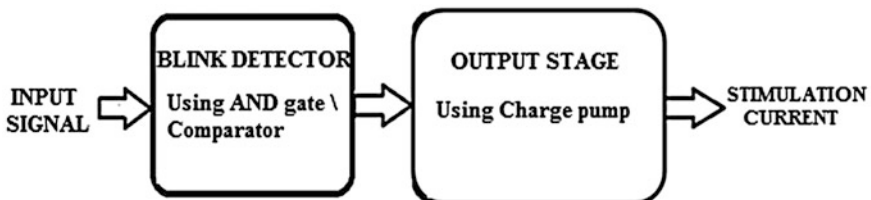


Fig. 2 Block diagram of stimulator chip (SoC)

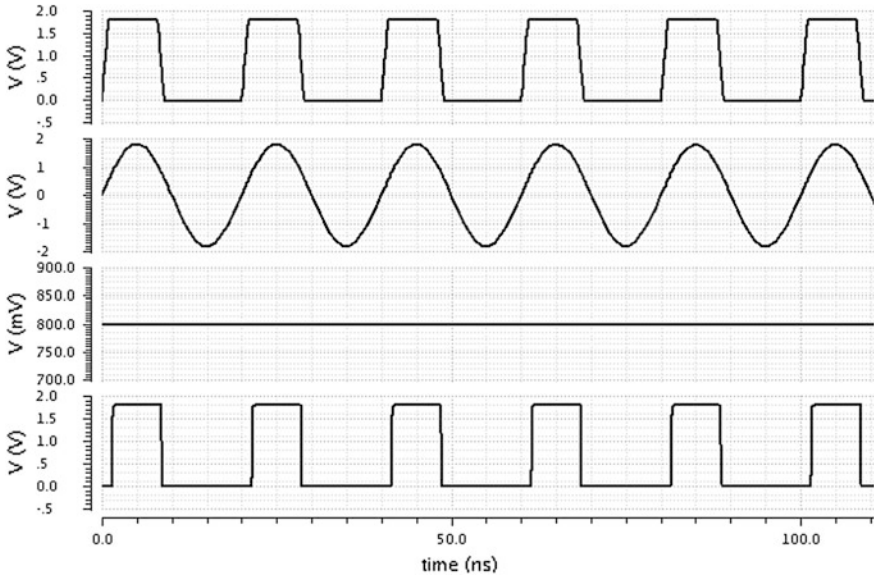


Fig. 3 Waveform showing blink detection when input exceeds a threshold value

Table 2 Comparative analysis of different blink detectors

Block	Number of transistors	Current delivered (μ A)	Total power
Comparator 1	17	58.284	1.732865 mW
Comparator 2	29	277.1802	3.403427 mW
Static AND	6	59.355	1.010785 mW
Dynamic AND	6	78.68907	864.5257 μ W

advantage of simple and fast design [1]. But its operation and design are more involved than those of its static counterpart, due to an increased sensitive to noise (Fig. 3). Comparator is also used to detect blink. In this, whenever the input signal exceeds the reference voltage, the output will be $+V_{sat}$ else it will be $-V_{sat}$. Two different comparators are used for comparative analysis. In comparator 1, the cross-coupled inverter pair reduces the parasitic capacitance to achieve high comparison speed. The extra switching transistor short circuits the latch’s output to half power supply. This helps in shifting the output to the corresponding digital levels easily. This also increases the speed and performance of the comparator. In comparator 2, the addition of extra buffer stages removes noise resulting in low offset. Though it has more transistors, it delivers more current compared to other blink detectors. On comparison, it is found that the dynamic AND logic comparator though delivers less current works with low power 864.5257 μ W at 100 MHz. Thus when designing a complete stimulator chip, blink detector block using dynamic AND logic can be used to obtain a low power stimulator chip (Table 2).

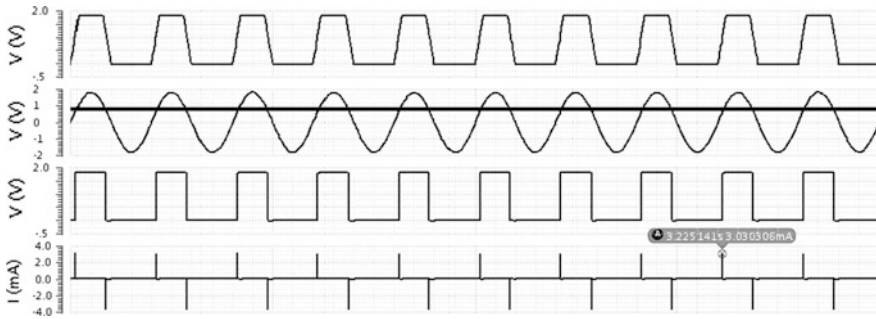


Fig. 4 Output waveform of stimulator chip

Table 3 Comparative analysis of stimulator chip using different blink detectors

Blink detectors	P_{avg} for 4 s	I_{L+} (mA)	I_{L-} (mA)	Area (mm ²)
Comparator 1	1.153 mW	3.00843	-3.38482	0.3284125
Comparator 2	804.9 μ W	3.09073	-3.78561	0.3745
Static AND	185.5 μ W	2.97799	-3.42127	0.31518
Dynamic AND	2.612 μ W	3.03036	-3.41	0.351631

4 Results and Conclusion

The above analyzed blink detectors are tested with the output stage for their functionality. For the blink detector stage, the input signal is a sine wave of 1.8 Vp and reference voltage is taken as 800 mv. The frequency of clock is taken as 2.5 Hz with 30 % duty cycle. The charge pump is used in the output stage to deliver 3 mA of current (Fig. 4). It is used to boost the current delivered by the blink detector stage. In this, PMOS are connected in series with capacitors of high value connected in parallel in each node. It uses two clocks of opposite phase. Instead of clocks, the output from the blink detector stage is used as clock here so that the current is delivered when the blink is detected. Here, the anodic current is delivered at positive edge and cathodic current is delivered at negative edge resulting in biphasic current. Since the current is delivered as soon as the blink is detected, the latency is in nanoseconds. The blink detector stage and output stage are connected using buffer stage. From the analysis, it is found that the dynamic AND blink detector used in a stimulator chip at 2.5 Hz consumes less power of 2.612 μ W which is about 70 % more efficient than other blink detectors in the context of consumption of less power. Also, it delivers the required current of 3.03036 mA and occupies moderate area of about 0.351631 mm² (Table 3).

References

1. J.P. Uyemura, *Introduction to VLSI Circuits and Systems* (Wiley, New York, 2002)
2. X. Liu, A. Demosthenous, D. Jiang, N. Donaldson, Active Books. The design of an implantable stimulator that minimizes cable count using integrated circuits very close to electrodes. *IEEE Trans. Biomed. Circuits Syst.* **6**(3), 216–227 (2012)
3. M. Sivaprakasam, W. Liu, M.S. Humayun, J.D. Weiland, A variable range bi-phasic current stimulus driver circuitry for an implantable retinal prosthetic device. *IEEE J. Solid-State Circ.* **40**(3), 763–771 (2005)
4. J. Wang, M. Gulari, K.D. Wise, An integrated position-sensing system for a MEMS-based cochlear implant. in *IEEE International Electron Devices Meeting* (2005)
5. K. Chen, T.C. Chen, K. Cockerham, W. Liu, Closed-loop eyelid reanimation system with real-time blink detection and electrochemical stimulation for facial nerve paralysis. in *IEEE International Symposium on Circuits and Systems* (2009)
6. X. Yi, J. Jia, S. Deng, S.G. Shen, Q. Xie, G. Wang, A blink restoration system with contralateral EMG triggered stimulation and real-time artifact blanking. *IEEE Trans. Biomed. Circuits Syst.* **7**(2), 140–148 (2013)
7. F. Shahrokhi, K. Abdelhalim, D. Serletis, P.L. Carlen, R. Genov, The 128-channel fully differential digital integrated neural recording and stimulation interface. *IEEE Trans. Biomed. Circuits Syst.* **4**(3), 149–161 (2010)

Energy- and Trust-Based AODV for Quality-of-Service Affirmation in MANETs

Sridhar Subramaniam and Baskaran Ramachandran

Abstract A mobile ad hoc network (MANET) is a wireless network proficient of self-directed actions, and nodes communicate with each other without the use of infrastructure. Nodes are nomadic, and topology can be very dynamic. In MANETs, the provision of quality-of-service (QoS) guarantee is more challenging. Therefore, it is important that routing protocols incorporate QoS metrics. A trust-based system can be used to track misbehaving nodes and isolate them from routing. In this paper, a trust-based reliable AODV is presented where trust is calculated for nodes that participate in routing. Energy is also introduced for nodes. Trust and energy levels of the nodes are considered before they are selected for routing. A threshold value is defined, and nodes are considered for routing only if its trust and energy levels are higher than threshold. The work is implemented and simulated on NS-2. The simulation results have shown improvement on QoS metrics.

Keywords Mobile ad hoc network · AODV protocol · QoS metrics · PDR · Delay and throughput

1 Introduction

MANET is an extremely confronted network environment due to its special qualities such as decentralization, dynamic topology, and neighbor-based routing. They do not rely on existing infrastructure to support communication. Each mobile node

S. Subramaniam (✉)

Department of Computer Applications, S.A. Engineering College,
Chennai 600077, India
e-mail: ssridharmca@yahoo.co.in

B. Ramachandran

Department of Computer Science and Engineering, CEG, Anna University,
Chennai 600025, India
e-mail: baskaran.ramachandran@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_65

601

acts as an end node when it is the source or destination of a communication and forwards packets for other nodes when it is an intermediate node of the route. Mobile ad hoc network [1] is a system of wireless mobile nodes that self-organizes itself in dynamic and temporary network topologies.

There have been many ad hoc routing protocols, which fall into several categories: proactive routing protocols such as DSDV, OLSR, and TBRPF, and on-demand routing protocols such as DSR, AODV, and SSA. Proactive routing protocols have little delay for route discovery and are robust enough to link breaks and obtain a global optimal route for each destination. However, their routing overhead is also high. On-demand routing protocols are easy to realize, and their overhead is low. But routes in on-demand routing protocols are easy to break in the case of topology variations. In AODV [2], node does not have any information about other nodes until a communication is needed. By broadcasting HELLO packets in a regular interval, local connectivity information is maintained by each node. Local connectivity maintains information about all the neighbors.

Recent QoS solutions are planned to operate on trusted environments and totally assume the participating nodes to be cooperative and well behaved [3, 4]. Providing different quality of service levels in a persistently changing environment is a challenge because unrestricted mobility causes QoS sessions to suffer due to recurrent path breaks, thereby requiring such sessions to be re-established over new paths.

The inherent freedom in self-organized mobile ad hoc networks introduces challenges for trust management, particularly when nodes do not have any prior knowledge of each other. The concept of trust originally derives from social sciences and is defined as the degree of subjective belief about the behaviors of a particular entity [5]. Trust-based routing mechanism helps to identify and eliminate misbehaving nodes in MANET and performs an efficient and effective routing. Energy for nodes is also considered while routing since nodes may drain out of energy levels. Energy is announced by the proposed AODV protocol that checks for energy levels of nodes before taking part in routing.

2 Literature Survey

A security-enhanced AODV routing protocol called reliant ad hoc on-demand distance vector routing (R-AODV) [6] uses a modified trust mechanism known as direct and recommendations trust model and then incorporates it inside AODV. This enhances security by ensuring that data do not go through malicious nodes that have been known to misbehave.

A framework for estimating the trust between nodes in an ad hoc network based on quality-of-service parameters using probabilities of transit time variation, deleted, multiplied and inserted packets, and processing delays is used to estimate and update trust [7]. A schema is formed via direct and indirect approach to compute trust value among anonymous nodes [8]. To evaluate trust values, the

parameters like reputation, knowledge, observation, and context were used. The trust schema that is built is used to allow resource to be shared among trusted nodes.

A trust model introduced in the network layer leads to a secure route between source and destination without any intruders or malicious nodes in the network [9]. This trust-based routing protocol concentrates both in route and node trust. The analysis is based on the comparison of two energy-based mechanisms called E-AODV, an energy consumption rate-based routing protocol, and F-AODV, a cross-layer-based routing protocol [10]. The trends and the challenges on designing cross-layer communication protocols for MANETs are investigated.

A novel energy saving energy routing protocol ES-AODV [11] is presented. Nodes made use of the HELLO message mechanism in AODV and reduced energy consumed by inserting intermediate node iteratively. A routing protocol [12] adds a field in request packet and also stores trust value indicating node trust on neighbor based on level of trust factor. This scheme avoids unnecessary transmit of control information, thus efficiently utilizing channels, and also saves node power.

3 Proposed Work

Routing in mobile ad hoc networks is pretentious due to the dynamic nature of nodes. But still nodes communicate with each other and exchange data within the available nodes on the network. The architecture of the proposed work is shown in Fig. 1.

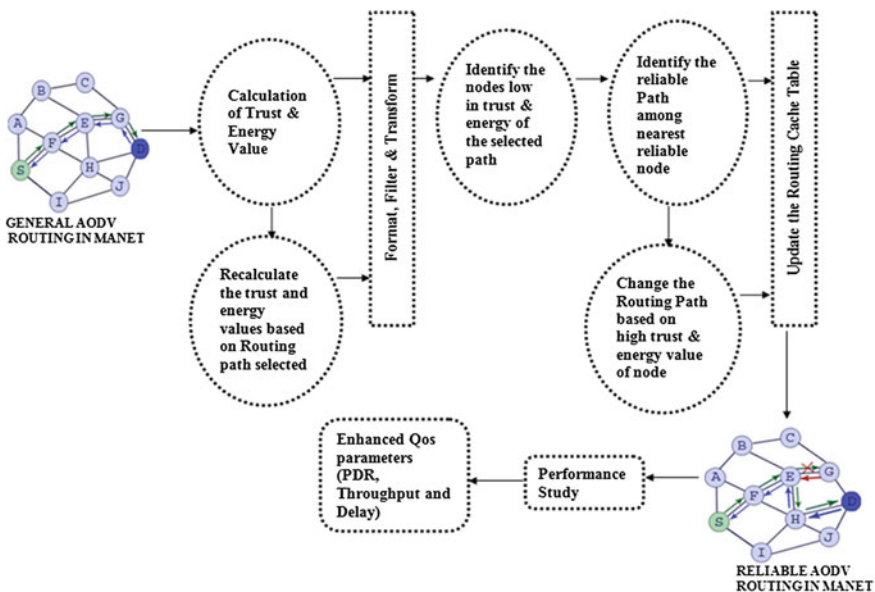


Fig. 1 Architecture of proposed trust- and energy-based AODV routing in MANET

Table 1 Trust value calculation parameters

Count type	RREQ	RREP	Data
Success	Qrs	Qps	Qds
Failure	Qrf	Qpf	Adf

The trust-level value calculation is based on the parameters shown in the Table 1. The count field describes about two criteria—success and failure—which describes whether the transmission was a successful transmission or a failure. RREQ and RREP are the route request and route reply, respectively, which are exchanged between nodes in the network. Data refers to the payload transmitted by the node in the routing path.

The parameter q_{rs} is defined as the query request success rate which is calculated based on number of neighboring nodes that have successfully received (RREQ) from the source node which has broadcasted it; q_{rf} is defined as the query request failure rate which is calculated based on number of neighboring nodes that have not received the query request; q_{ps} is defined as the query reply success rate which is calculated as successful replies (RREP) received by the source node which has sent the RREQ; and q_{pf} is defined as the query reply failure rate which is calculated based on the number of neighboring nodes which have not sent the replies for the query request received. q_{ds} is defined as the data success rate calculated based on successfully transmitted data, and q_{df} is defined as data failure rate calculated based on data which have failed to reach destination.

$$Qr = \frac{q_{rs} - q_{rf}}{q_{rs} + q_{rf}} \quad (1)$$

$$Qp = \frac{q_{ps} - q_{pf}}{q_{ps} + q_{pf}} \quad (2)$$

$$Qd = \frac{q_{ds} - q_{df}}{q_{ds} + q_{df}} \quad (3)$$

$$TL = T(RREQ) * Qr + T(RREP) * Qp + T(DATA) * Qd \quad (4)$$

where Qr, Qp, and Qd are intermediate values that are used to calculate the node request rate, reply rate, and data transmission rate. TL is the trust-level value, and T (RREQ), T(RREP), and T(DATA) are time factorial at which route request, route reply, and data are sent by the node, respectively. Trust-level value (TL) is calculated for each node during routing and is checked against the threshold value (average of trust values of the nodes that take part in routing). If lesser than threshold, there is a possibility for this node to be marked as misbehaving node for the current transmission.

The node energy level also plays a very crucial role in MANET routing. Node is selected for routing only if its energy level is greater than the threshold value (average of energy values of the nodes that take part in routing). If energy level is

not sufficient, the proposed protocol selects an alternate path to carry on routing successfully using reliable nodes. Energy calculation is based on nodes' sending and receiving rate.

4 Evaluation Results

The performance of proposed AODV protocol is analyzed using NS-2 simulator. The network is designed using network simulator with a maximum of 50 nodes. Results are obtained from this simulation applying both general AODV and proposed AODV protocols. In the proposed AODV protocol PDR, throughput is increased and delay is reduced compared to the general AODV. The results obtained are shown in Table 2. The traditional AODV is affected due to the existence of misbehaving nodes, which results in low PDR and also causes the delay to increase.

Figure 2 shows the snapshot of the simulation which contains 50 nodes. Node 1 is the source, and node 34 is the destination. The initial routing path is node 1 → node 5 → node 20 → node 28 → node 34. Figure 3 shows the misbehaving nodes (marked in red circle); thus, routing takes an alternate path avoiding misbehaving node. The new path is node 1 → node 6 → node 4 → node 15 → node 20 → node 29 → node 33 → node 34. Figure 4 indicates how the proposed AODV protocol has shown a good decrease in delay when compared to the general AODV. Figure 5 shows the increase in PDR when compared with the general AODV.

Table 2 Comparison of result with 50 nodes

Protocol	PDR	Delay	Throughput
General AODV	70.05	0.84472	114,559.87
Proposed AODV	96.68	0.02685	483,236.55

Fig. 2 Snapshot of 50 nodes in a MANET

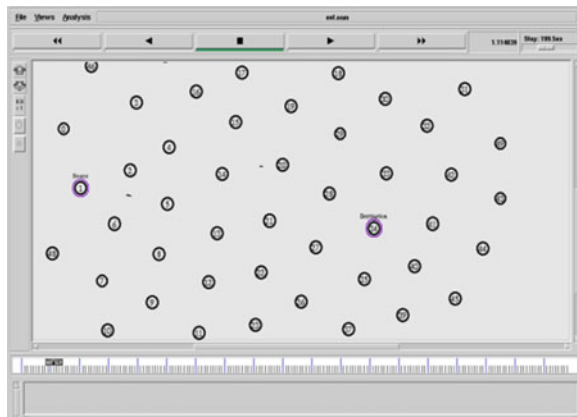


Fig. 3 Snapshot of misbehaving nodes

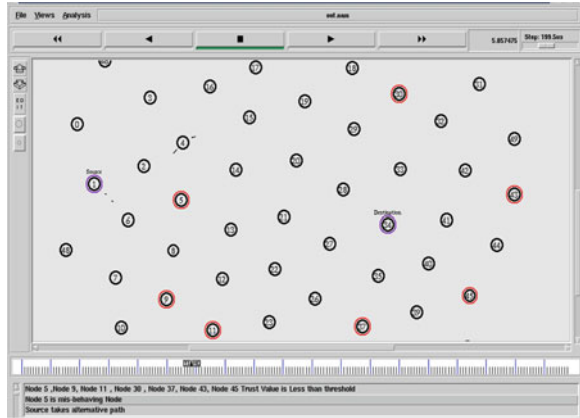


Fig. 4 Comparison of general AODV and proposed trust AODV delay

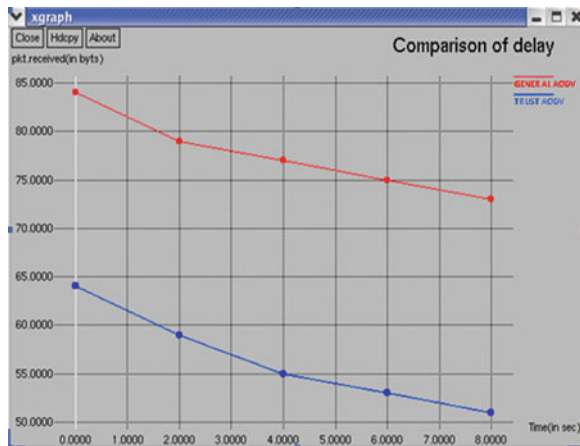
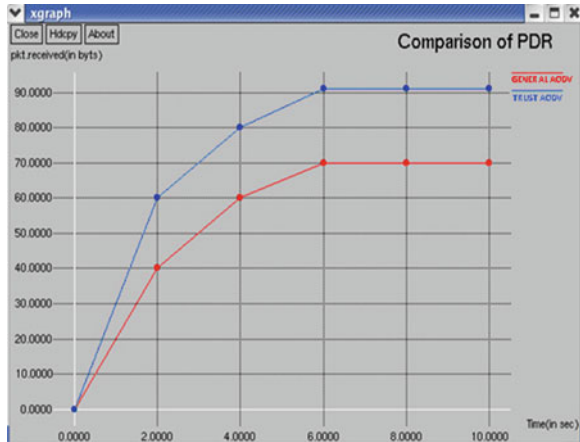


Fig. 5 Comparison of general AODV and proposed trust AODV PDR



5 Conclusion

A trust-based AODV protocol is proposed. Trust values are calculated for each node in the path. Energy is announced, and only nodes with high energy and trust levels are used for routing. The proposed protocol isolates misbehaving and packet dropping nodes in the path. Proposed protocol improves reliability in routing and shows good improvement in QoS metrics like PDR, delay, and throughput.

References

1. G. Kortuem, J. Schneider, D. Preuit, T.G.C. Thompson, S. Fickas, Z. Segall, When peer-to-peer comes face-to-face: collaborative peer-to-peer computing in mobile ad hoc networks, in *1st International Conference on Peer-to-Peer Computing* (2001), pp. 75–91
2. C.E. Perkins, E.M. Royer, S.R. Das, *Ad hoc On-Demand Distance Vector (AODV) Routing*. IETF Internet Draft: draftietf-manet-aodv-05.txt (2000)
3. Y. Hu, *Enabling secure high-performance wireless ad hoc networking*. PhD Thesis (2003)
4. M. Ilyas, *The Handbook of Wireless Ad Hoc Network* (CRC, 2003)
5. K.S. Cook (ed.), *Trust in Society*, vol. 2 (Russell Sage Foundation Series on Trust, New York, 2003)
6. H.S. Jassim, S. Yussof, A routing protocol based on trusted and shortest path selection for mobile ad hoc network, in *IEEE 9th Malaysia International Conference on Communications* (2009)
7. D. Umuhoza, J.I. Agbinya, Estimation of trust metrics for MANET using QoS parameter and source routing algorithms, in *The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications* (2007)
8. A.A. Bakar, R. Ismail, J. Jais, Forming trust in mobile ad hoc network, in *2009 International Conference on Communications and Mobile Computing* (2009)
9. A.M. Pushpa, Trust based secure routing in AODV routing protocol, in *IEEE international conference on Internet Multimedia Services Architecture and Applications (IMSAA)* (2009)
10. L. Romdhani, C. Bonnet, in *9th IFIP International Conference on Mobile Wireless Communications Networks* (2007), pp. 96–100
11. X. Wang, Q. Liu, N. Xu, *Fourth International Conference on Natural Computation*, vol. 5, (2008), pp. 276–280
12. R.S. Mangrulkar, M. Atique, Trust based secured adhoc on demand distance vector routing protocol for mobile adhoc network, in *Sixth IEEE International Conference on Wireless Communication and Sensor Networks (WCSN)* (2010)

Classification of Remote Sensing Image Based on Different Similarity Measures

Kartik Shah, Shantanu Santoki, Himanshu Ghetia and D. Aju

Abstract Advanced wide field sensor (AWiFS) is a multi-spectral camera used to capture image from IRS-P6 (Indian remote sensing) satellite. Iterative self-organizing data analysis technique (ISODATA) is one of the most frequently used unsupervised classification algorithms. There are too many techniques available for classifying an image. In this paper, we will use similarity-based techniques to classify an image. Then, we will compare the result of each similarity measure classification techniques. We will use normalized difference vegetation index (NDVI) values to classify the images.

Keywords Unsupervised classification · Similarity measures · AWiFS image · ISODATA

1 Introduction

IRS-P6 is a multiple sensor platform and also called as RESOURCESAT-1. It was launched by Indian PSLV-C5 on Oct. 17, 2003, from Satish Space Center. The platform contains Linear Imaging Self-Scanning System (LISS III), LISS IV, and AWiFS sensors. The application of AWiFS is in agriculture, land (vegetation), water management, solid earth, etc. AWiFS is an improved version of WiFS of IRS 1C/1D. It operates in four spectral bands as shown in Table 1.

K. Shah (✉) · S. Santoki · H. Ghetia · D. Aju
School of Computing Science and Engineering, VIT University, Vellore 632014, India
e-mail: kartikshah@mail.com

S. Santoki
e-mail: shantanu.santoki@gmail.com

H. Ghetia
e-mail: himanshu.ghetia@gmail.com

D. Aju
e-mail: daju@vit.ac.in

Table 1 AWiFs Bands

Band	Spectral band (μm)	Resolution (m)
1	0.52–0.59	56×56
2	0.62–0.68	56×56
3	0.77–0.86	56×56
4	1.55–1.70	56×56

Table 2 LISS III bands

Band	Spectral band (μm)	Resolution (m)
2	0.52–0.59	23×23
3	0.62–0.68	23×23
4	0.77–0.86	23×23
5	1.55–1.70	23×23

Table 3 LISS IV bands with its mode

Band	Band	Spectral band (μm)	Resolution (m)
Panchromatic		0.50–0.59	5.8×5.8
Multispectral	2	0.52–0.69	5.8×5.8
	3	0.62–0.68	5.8×5.8
	4	0.77–0.86	5.8×5.8

Similarly, LISS III can be operated in the following four bands (Table 2).

LISS IV operates in 2 modes called panchromatic and multispectral modes. It can be shown by Table 3

The normalized difference vegetation index (NDVI) is a measure for vegetation cover of land surface. Vegetation is different from other land surfaces because it absorbs the red wavelengths of sun and reflects the near-infrared wavelengths. NDVI takes values between -1 and 1 . Rock, sand, or snow contain NDVI values 0.1 or below. Moderate values 0.2 – 0.3 represent shrubs and grasslands. 0.5 or more NDVI values indicate dense vegetation like forest. NDVI value can be computed using two band values red (visible) band and near-infrared (invisible) band values by the following formulae.

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}$$

Classification of image is a process of classifying pixels to particular class based upon available information. We can use multispectral data for classification based upon the numerical data which comes under pattern recognition. The image classification techniques are mainly divided into two types: supervised classification and unsupervised classification. This classification may also include features of the

soil type and land surface elevation that are not derived from the image. A pattern is thus a set of measurements on the chosen features for the individuals to be classified. So the classification methods are sometimes called as pattern recognition.

Supervised Classification: Supervised classification technique requires the training data for each class based upon which it can classify the image. It will check the nearest point and classify into that class. Training data are used for classification and are called as supervised classification. Three steps involved in supervised classification are as follows:

1. The training stage: The analyst will identify the interested area and will create the training data for particular classes which are interested.
2. The classification stage: The 'unknown' class data are now classified based upon the training data available for classification. This is important stage of supervised classification.
3. Final output stage: The result of classification stage can be used in many ways like graphs, maps, analysis, etc.

Unsupervised Classification: These algorithms do not compare with training data, but it computes the unknown data vectors and divides them into classes based upon properties they have. In this algorithm, we need not to know the number of clusters. Algorithm is capable of splitting and merging clusters. It works as follows in two steps:

1. Cluster centers are placed randomly, and pixels are assigned to clusters based upon the shortest distance to the center. It is compared with all the centers.
2. The standard deviation within each cluster and distance between cluster centers is generated.

2 Previous Work

For analysis of space time structure of vegetation, multiple year data sequences of satellite images give us a source of data for analysis [1]. We can get value of NDVI for that image [2]. Also we can get the value of enhanced vegetation index (EVI) [3]; these values can be used for study in functional and structural analysis of an image [4–6]. We can use the complete classification method as follows [7].

Figure 1 shows the basic classification process [7]. In classification process, first, the optical and radar images are considered for preprocessing. In preprocessing, the image is gone from image transformation, mosaicking, and data fusion. Then, it is sent to feature extraction and image classification stage. During this stage, the part of image is labeled. Then, the final stage is post-processing in which data are passed from single-pixel region cleaning, mosaicking, and data fusion. From studies, it is found that variation in temporal MODIS vegetation index is connected with temperature and climatic conditions [8–11]. The authors also found that the variation from inter-annual in NDVI and EVI values for particular eight-day periods was

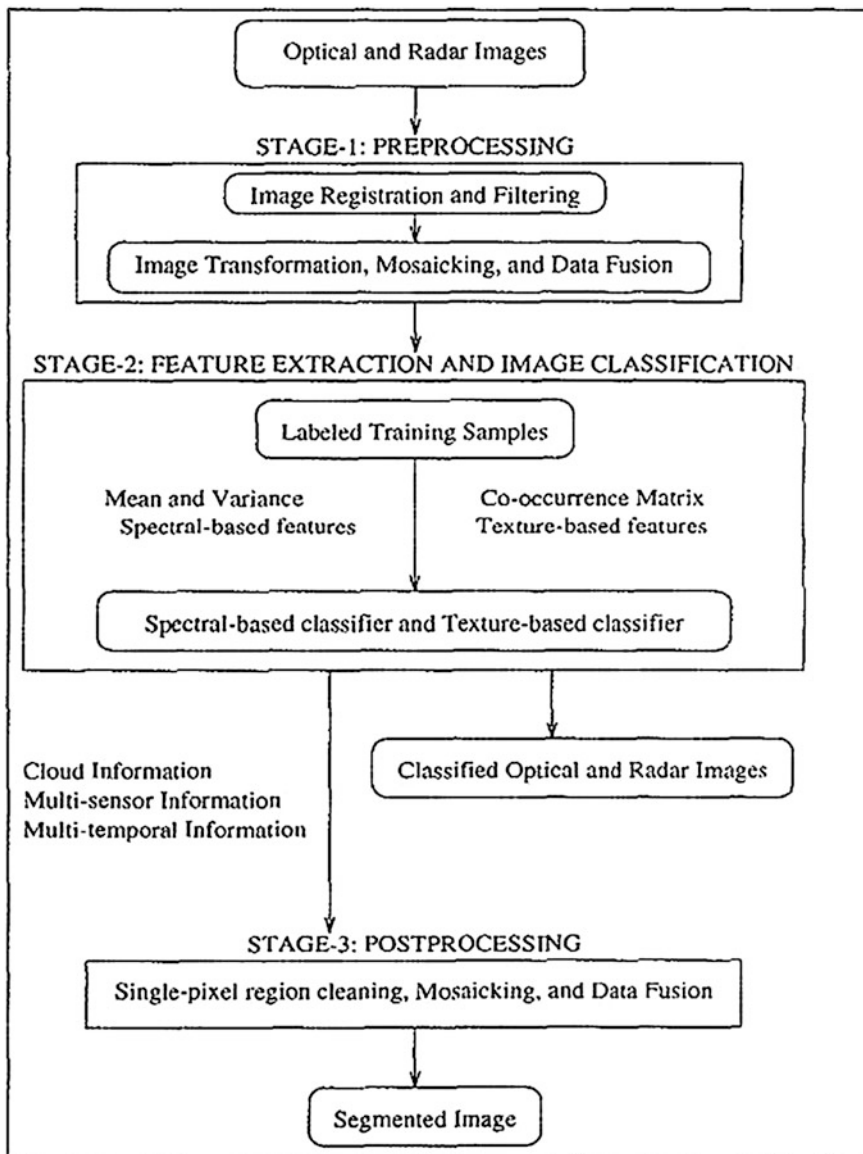


Fig. 1 Classification process [7]

correlated with the phenological indicators [8]. It is also found that vegetation covers of different moisture conditions have different variation patterns in the time series of the MODIS EVI values [12]. Template matching is a basic method for finding the presence or the absence of objects and to check whether it is available in image or not. In this search, the image is compared pixel by pixel, all with the other

image, and at each position of the template, the closeness of the template to the area covered by it is measured. The location which is very closer is considered as matching of the object detected [13]. Template matching is one of the simple techniques used from past many decades. It is a basic technique for image as it can answer too many questions related to image [14].

3 Proposed Methodology

In this paper, we are going to propose a methodology called similarity-based measures for image classification process. We are going to take different similarity measures to classify an image into particular reference class. Some of the similarity measures are as follows (Table 4).

Before using this similarity measures, we will obtain the reference image and unknown image each having four bands (red, green, blue, and near infrared) which need to be classified. Now, we will calculate the NDVI value by using the following algorithm:

For each i in range

$$\text{ndvi_lower_part} = (\text{nir_value}[i] + \text{red_value}[i])$$

$$\text{ndvi_upper_part} = (\text{nir_value}[i] - \text{red_value}[i])$$

$$\text{ndvi} = \text{ndvi_upper_part} / \text{ndvi_lower_part}$$

Algorithm 1: NDVI calculation algorithm

After collecting NDVI values, in our method, first we will take an unknown class which will contain the data which need to be classified and another class will contain the labeled class which will be reference for the unknown class. Then, we will apply the similarity measure, and based upon the best match with the reference class, we can compare it. The output will be the label (or class) for the unknown class. The reference class is shown in Table 5.

Table 4 Different similarity measure

Euclidean distance	$\sqrt{\sum_j [\text{Abs}(p - x_j)^2 + \text{Abs}(q - y_j)^2 + \text{Abs}(r - z_j)^2]}$
Manhattan distance	$\sum_j [\text{Abs}(p - x_j) + \text{Abs}(q - y_j) + \text{Abs}(r - z_j)]$
Squared euclidean distance	$\sum_j [\text{Abs}(p - x_j)^2 + \text{Abs}(q - y_j)^2 + \text{Abs}(r - z_j)^2]$
Bray curtis distance	$\sum_j \frac{[\text{Abs}(p-x_j) + \text{Abs}(q-y_j) + \text{Abs}(r-z_j)]}{[\text{Abs}(p-x_j) + \text{Abs}(q+y_j) + \text{Abs}(r+z_j)]}$
Canberra distance	$\sum_j \frac{\text{Abs}(p-x_j)}{p+x_j} + \frac{\text{Abs}(q-y_j)}{q+y_j} + \frac{\text{Abs}(r-z_j)}{r+z_j}$
Chessboard distance	$\text{MAX}[\text{Abs}(p - x_j), \text{Abs}(q - y_j), \text{Abs}(r - z_j)]$

Table 5 Reference class

2/10/12	8/11/12	10/12/12	15/01/13	18/02/13	Reference class
137.89	103.56	102.22	132.44	165.11	1
157.56	106	92.33	127.33	168	2
115.56	93.67	113.89	147.22	155.44	3
133.56	97	105.67	129.44	156.44	4
142.33	96.67	104.78	117.11	162.22	5
96.22	116.44	153.89	158	131.89	6
103.56	108.44	141	156.78	141.11	7
102.56	111.33	161	156.11	129.78	8
100.44	123.11	156.44	154.33	105.78	9
97.56	113.89	167	157	131	10
116.11	111.89	156.89	126.67	136.78	11
71.44	66.22	53	60.33	82.67	12
78.78	76.44	72.67	81.11	72.33	13
79.22	71.67	61	60.89	83.11	14
74.11	75	74.33	77	83.44	15
115.78	102.44	107	115.44	101.33	16
106.33	101.11	100.22	99.56	96.67	17
142.89	120.44	114.56	124.44	104.11	18
160.78	127.44	140	125.89	110.89	19
156.22	119.89	136.78	126.56	107.11	20

The unknown cluster file is shown in Table 6. Now, we will apply each formula on Tables 5 and 6. In our process, first we will take each record of Table 6 and will use the formulae for each record of Table 5 and then will find minimum value out of all 20 references. The minimum value class will be the class of that unknown class. Now, we will use the same process for all records of Table 6 and then apply all similarity measures. Based upon the similarity measure formulae, we can find which algorithm gives nearer result by mapping the result into graph. After applying algorithm to each clusters, the final result is shown in Table 7 (shown in Result and Analysis). And the comparison of these algorithms is done in Fig. 2 using line chart. We have considered one or two records from 30 records to plot in the line chart.

Algorithm 1.2 Proposed Vertex Cover Algorithm.

4 Results and Analysis

After applying algorithms on Tables 2 and 3 data, we will get the output as Table 4 data, which shows the labeling of each unknown clusters. We can see the result generated by Canberra distance algorithm, which shows unusual labeling. So, we can say that this similarity measure cannot be used for classification process, while

Table 6 Unknown class

02/10/12	8/11/12	10/12/12	15/01/13	18/02/13	Cluster no.
101.08	97.52	99.38	99.32	96.68	1
98.02	98.76	107.31	114.17	114	2
111.27	102.21	104.16	103.45	99.16	3
120.83	102.11	101.52	101.34	96.72	4
113.84	106.15	114.93	109.1	104.16	5
118.87	103.95	105.48	114.88	105.48	6
122.4	105.64	107.34	105.11	100.39	7
127.24	109.16	112.93	110.01	103.11	8
129.1	108.79	111.91	115	118.06	9
133.49	107.05	107.34	105.43	100.51	10
124.16	111.64	127.29	112.01	107.55	11
136.69	111.73	115.45	111.03	103.95	12
137.39	115.34	123.04	117.11	107.6	13
125	107.67	112.04	130.95	106.06	14
147.98	116.01	118.19	114.94	106.42	15
148.8	120.06	129.4	119.98	109.63	16
132.57	117.35	133.23	128.07	116.56	17
109.54	107.69	129.03	124.4	110.52	18
154.63	125.12	136.91	127.33	113.83	19
115.86	102.86	110.23	127.31	123.44	20
122.84	106.89	123.99	117.64	134.59	21
116.49	101.92	105.73	108.41	125.69	22
132.6	110.6	116.32	133.8	129.49	23
131.6	104.35	105.87	114.2	138.3	24
100.26	98.56	111.87	135.86	130.27	25
110.98	99.49	105.68	118.76	140.9	26
121.86	100.12	104.58	116.85	155.17	27
129.77	103.12	108.87	130.61	146.54	28
149.77	106.27	104.45	116.63	155.11	29
117.88	105.63	115.24	147.86	116.77	30

for other similarity measures, we can say that the classes are nearly similar. Chessboard similarity measure also gives better result compared to Canberra distance. Euclidean and squared Euclidean similarity measures provide approximately same classification. Manhattan and Bray Curtis similarity measures also provide similar results in most of the cases. As a part of analysis, consider the Fig. 2, which shows the comparison of algorithm after classification. In this figure, we have considered unknown class number 20 for analysis. In the graph shown in Fig. 2, we have considered unknown class 20 and the reference class result of six similarity measures. From Fig. 2, it is clear that the Canberra distance is not providing proper result and remaining algorithms provide the nearer result.

Table 7 Output after classification

Cluster_	Euclidean	Manhattan	Sq. euclidean	Bray curtis	Canberra	Chessboard
1	14	17	17	17	8	14
2	16	16	16	16	10	16
3	17	17	17	17	16	17
4	17	17	17	17	1	17
5	16	16	16	16	11	16
6	16	16	16	16	11	16
7	16	16	16	16	11	16
8	16	16	16	16	11	16
9	16	16	16	16	11	16
10	16	16	16	16	11	16
11	16	16	16	16	11	20
12	18	18	18	18	11	18
13	18	18	18	18	11	18
14	16	16	16	18	7	18
15	18	18	18	18	20	18
16	20	20	20	20	19	20
17	18	20	18	20	8	20
18	16	16	16	16	7	18
19	19	20	19	20	8	19
20	16	16	16	16	3	16
21	4	11	4	11	11	11
22	16	16	16	16	3	16
23	18	4	18	4	8	18
24	4	4	4	4	3	5
25	3	1	3	3	7	4
26	1	1	1	4	3	4
27	4	4	4	4	3	4
28	4	4	4	4	3	4
29	5	5	5	5	5	5
30	16	3	16	3	7	18

5 Implementation

To calculate the NDVI value of an image, we can go for algorithm shown in Algorithm 1. It can be implemented in Python language using GDAL library. We can add the GDAL library in Python, and then, we can calculate the NDVI values from image. The algorithms shown in Table 1 can be implemented using C# language in .NET technology. It is shown in Fig. 3. The coding part is as explained below:

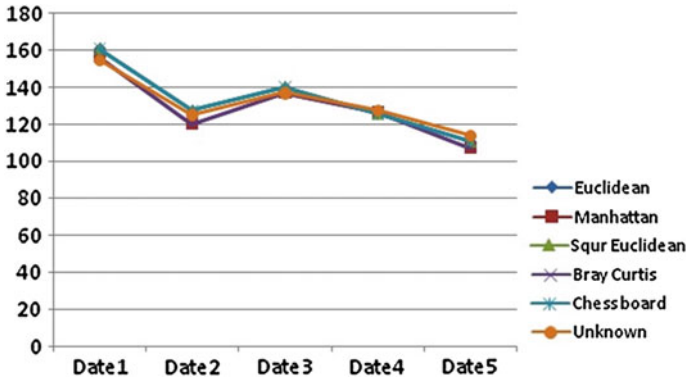
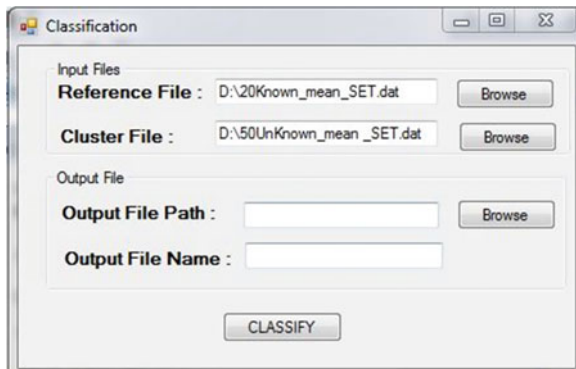


Fig. 2 Analysis of unknown class 20

Fig. 3 Implementation



```

for each i
    for each j
        for each k
            CODE
    
```

The code part will be different for each algorithms. The code part is as follows for each algorithm:

Euclidian

```

sum+=Math.Pow(Math.Abs(Math.Round(float.Parse(sArray2[j*
(width - 1) + i + k]) - float.Parse(sArray1[j* (width - 1) + j + k]), 2)),2);
sum2 = Math.Sqrt(sum);
    
```


Manhattan

```
sum+=Math.Abs(Math.Round(float.Parse(sArray2[i*
(width - 1) + i + k]) - float.Parse(sArray1[j* (width - 1) + j + k]), 2));
```

Squared Euclidian

```
sum+=Math.Pow(Math.Abs(Math.Round(float.Parse(sArray2[i*
(width - 1) + i + k]) - float.Parse(sArray1[j* (width - 1) + j + k]), 2)),2);
```

Bray Curties

```
sum+=Math.Abs(Math.Round(float.Parse(sArray2[i * (width - 1) +
i + k]) - float.Parse(sArray1[j * (width - 1) + j + k]), 2));
sum2+=Math.Abs(Math.Round(float.Parse(sArray2[i * (width - 1) +
i + k]) + float.Parse(sArray1[j * (width - 1) + j + k]), 2));
sum3=sum/sum2;
sum1[j] = sum3;
```

Canberra Distance

```
sum += (Math.Abs(Math.Round(float.Parse(sArray2[i * (width - 1) + i + k]) -
float.Parse(sArray1[j * (width - 1) + j + k]), 2))) / (Math.Round(float.Parse(sAr-
ray2[i * (width - 1) + i + k]) + float.Parse(sArray1[j * (width - 1) + j + k]), 2));
```

Chessboard Distance

```
sum += Math.Abs(Math.Round(float.Parse(sArray2[i * (width - 1) + i + k])
- float.Parse(sArray1[j * (width - 1) + j + k]), 2)); if (sum > max)
{max = sum;}sum1[j] = max;
```

6 Conclusion and Future Work

From the analysis and results, we can understand that Canberra distance similarity measure cannot be used for labeling of cluster because it provides false result, while the chessboard distance measure may provide wrong result sometimes. Other algorithms provide nearer results, and those can be used for classification process. These measures are distance-based similarity measures, and the technique is unsupervised classification. We can extend these algorithms by adding concept of standard deviation and can have more accuracy.

7 Future Work

In this proposed algorithm, we are using depth first search (DFS) algorithm to find the articulation points of the graph. This algorithm takes $O(V + E)$ time to compute the articulation points. We can go for some other techniques to find the articulation

points which can take less time compared to DFS. Also we can find some other techniques which always provide the exact solution to optimal solution and also some more techniques which can be used to provide nearer solution. It is also possible to find algorithm which runs faster than this algorithm.

Acknowledgments The authors would like to thank the School of Computer Science and Engineering, VIT University, for giving them the opportunity to carry out this research.

References

1. E. Swinnen, F. Veroustraete, Extending the SPOT-VEGETATION time series (1998–2006) back in time with NOAA-AVHRR data (1985–1998) for Southern Africa. *IEEE Trans. Geosci. Remote Sens.* **46**(2), 558–572 (2008)
2. C.J. Tucker, Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **8**(2), 127–150 (1979)
3. C.O. Justice, E. Vermote, J.R.G. Townshend, R. Defries, D.P. Roy, D.K. Hall, V.V. Salomonson, J.L. Privette, G. Riggs, A. Strahler, W. Lucht, R.B. Myneni, Y. Knyazikhin, S. W. Running, R.R. Nemani, Z.M. Wan, A.R. Huete, W. van Leeuwen, R.E. Wolfe, L. Giglio, J.P. Muller, P. Lewis, M.J. Barnsley, The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Trans. Geosci. Remote Sens.* **36**(4), 1228–1249 (1998)
4. P. Jonsson, L. Eklundh, Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Trans. Geosci. Remote Sens.* **40**(8), 1824–1832 (2002)
5. B.D. Wardlaw, S.L. Egbert, J.H. Kastens, Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sens. Environ.* **108** (3), 290–310 (2007)
6. L.A. Méndez-Barroso, J. Garatuzza-Payán, E.R. Vivoni, Quantifying water stress on wheat using remote sensing in the Yaqui Valley, Sonora, Mexico. *Agric. Water Manag.* **95**(6), 725–736 (2008)
7. A. Murni, A.K. Jain, J. Rais, framework for multi-date multisensor image interpretation. *IEEE IGARSS.* **3**, 1851–1854 (1996)
8. P.R. Bajgiran, Y. Shimizu, F. Hosoi, K. Omasa, MODIS vegetation and water indices for drought assessment in semi-arid ecosystems of Iran. *J. Agric. Meteorol.* **65**, 349–355 (2009)
9. A.R. Huete, G. Ponce, Satellite observed shifts in seasonality and vegetation—rainfall relationships in the south-west USA. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **38**, 775–777 (2010)
10. D. Song, P. Guo, H. Sheng, Spatial distribution pattern of MODISNDVI and correlation between NDVI and meteorology factors in Shandong province in China. *Piers Online* **4**, 191–196 (2008)
11. A. Yuhua, L.A. Scuderi, MODIS-derived NDVI characterization of drought-induced evergreen dieoff in western North America. *Geograph. Res.* **47**, 34–45 (2008)
12. Y. Zheng, H. Qiu, in *Mapping urban landuse types in Los Angeles using multi-date moderate-resolution imaging spectroradiometer vegetation image products*. Proceedings of Second International Workshop on Earth Observation and Remote Sensing Applications (2012)
13. L. Prasad, S.S. Iyengar, High performance algorithms for object recognition problem by multiresolution template matching (1995)
14. J.A. Mikhail, Faster image template matching in the sum of the absolute value of differences measure. *IEEE Trans. Image Process.* **10**(4) (2001)

A Start to Fail Frequency Technique for Detecting Hardware Trojan

Sharmila Durai, Prasanna Kumar and Srinivasan Ramasamy

Abstract Rapid development of portable devices plays a major role in communication sector, which has led to a significant concern for cryptology. Cryptographic circuits are vulnerable to various side-channel attacks by the hardware Trojan horses (HTHs). HTHs are the malicious design modification that either alters the actual behavior of the design or reveals the secret information. In this work, we proposed a technique which is elusively for detection of hardware Trojans based on Start to Fail (STF) frequency concept. We inferred that this detection technique is efficient based on the experimental results observed, using ISCAS benchmark circuits and scalable encryption algorithm (SEA).

Keywords GDSII · Hardware Trojan horses (HTH) · HTH detection and insertion · Scalable encryption algorithm (SEA) · Path delay

1 Introduction

Increasing globalization and economic market trends in the field of semiconductor design and fabrication process has driven most hardware manufacturers to out-source their IC fabrication for the purpose of cost effectiveness, but this makes easier for an attacker to compromise the IC supply chain, and also, ICs are more

S. Durai (✉)

Department of Electronics and Communication Engineering, RMD Engineering College, Chennai, India

e-mail: dsharmila25@gmail.com

P. Kumar · S. Ramasamy

Department of Electronics and Communication Engineering, RMK Engineering College, Thiruvallur, India

e-mail: prasannavlasi@yahoo.com

S. Ramasamy

e-mail: srs.ece@rmkec.ac.in

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_67

621

prone to malicious activities and modifications. The Defense Advanced Projects Research Agency (DARPA) has provided industry with a subjective assessment snapshot of the relative risk in each of the phases, as shown in Fig. 1 [1]. These vulnerabilities have raised threats in area of sensitive and security critical applications such as defense, military, and transportation security systems. Not only it is difficult to control the outsourcing manufacturing facilities due to certain market requirements such as to reduce the product design and development time, lowering the cost and reducing the manufacturing time, for example to decrease cycle time.

We use several auto-placement and routing tools which is hard to avoid, and these tools due to their less optimal nature leave lot of unused space in the chip which an hacker to easily embed malicious circuits called hardware Trojan horses (HTHs) or simply Trojans. Karri et al. [2] compiled MI statistics from an accumulation of examples at the 2008 Embedded Systems Challenge, mentioned earlier, along with data from previously published MIs.

These HTHs are usually composed of two main parts: (i) trigger and (ii) payload. Activation mechanism of a Trojan will be referred as ‘Trojan Trigger.’ And part of the circuit or functionality affected by the trigger referred as ‘Trojan payload.’ The basic classification of Trojan activation mechanisms and payloads was discussed by Chakraborty et al. in [3].

This Trojans either affect the functional behavior or reveal the secretive information; in worst cases, it can move up to the extent of even destroying the system. Hardware threats pertain to three of the key aspects in security: availability (denial of service), confidentiality (information leakage), and integrity (data tampering).

There are several approaches for the detection of Trojan 8 which can be broadly classified into two, namely (i) Full Trojan activation, this approach requires test vectors to identify Trojans, by comparing the actual expected responses. But this is a difficult task in case of real design implementation; moreover, the property of Trojan gets hidden in the functional verification. Whereas the second approach (ii) side-channel analysis (SCA) is reliable, this approach detects Trojan based on certain parameters such as power and delay (parameters independent of functionality). The authors in [4–6] use transient power as a side-channel signal to detect Trojans, but a

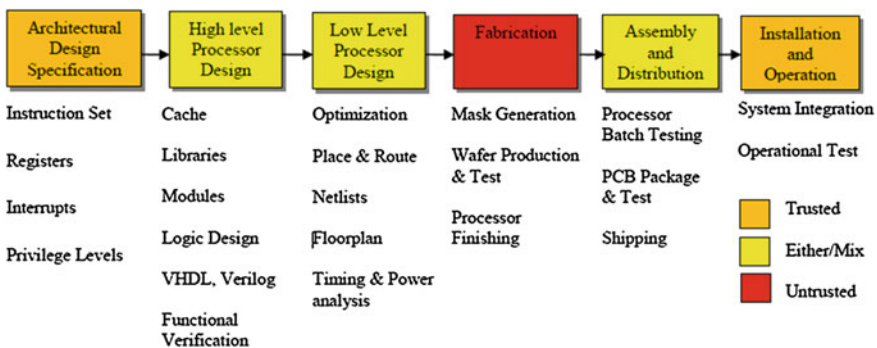


Fig. 1 Processor design life cycle, with DARPA risk assessment

partial activation of Trojans is still required. The transitions for partial activation may be very difficult to generate on some hard-to-activate nets in the circuit. The authors in [7, 8] measure path delays to detect changes caused by Trojans. Although effective, only critical paths in [7] are measured, limiting detection of Trojans inserted on noncritical paths. The authors in [8] use one additional shadow register and one comparator to measure each path delay in the circuit.

Out of two techniques in SCA, the delay method is efficient comparing with power, due to the reason power method requires the Trojan to be activated either fully or partially. But it is not so in case of delay method. Based on these challenges,

1. Requires maximum detection coverage of Trojan that needs to be placed and distributed on various critical paths present.
2. Low-cost path delay measurement.

Proves existing methods are inefficient; hence, we propose a technique based on Start to Fail (STF) frequency.

2 Background

2.1 Impact on Path Delay Due to Trojan

Sometimes, Trojan inserted in the may be hidden and undetectable while considering path delays alone. To avoid this uncertainty, we consider nodes along with path delays. Depending on the action and activation, we classified Trojans into three types: (i) only trigger, (ii) only payload, and (iii) trigger and payload.

For a design the Trojan trying to change the design, a payload will be inserted at the output of gate node. An example of Trojan is shown in Fig. 2.

To show the effect of Trojan on path delay, we inserted a Trigger along with payload in a design and performed simulation using Faraday 90 nm technology library. We inserted a Trojan on critical path of the design as shown in Fig. 2 (Table 1).

Fig. 2 An example of Trojan payload

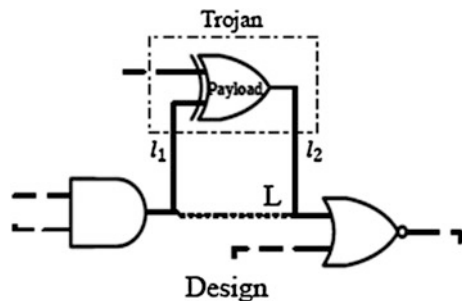


Table 1 Trojan impact on delay

	Long 1 and 2 (ps)
Without Trojan	5,099
With Trojan	5,226
Increased delay	127

In genuine design, the sensitized path passes through L node. Additional delay due to propagation delay of payload and the interconnection path's capacitances (I1 and I2) are unavoidable. If the Trojan triggered, the additional delay due to trigger and interconnections further will increase capacitance value. Based on this concept, we proposed 'STF frequency' technique in next section.

2.2 Start to Fail Frequency

In previous delay-based methods, transition delay fault (TDF) and path delay fault (PDF) have been used to detect Trojan, which gives only the increased slack delay. The major limitation of this method is whenever Trojan is in the shortest path (T_s) and the delay due to Trojan along with shortest path is lesser than the critical path delay (T_c) leaves Trojan undetected.

For example, assume that the delay of critical path is greater than the sum of shortest path and Trojan delay. Then, the Trojan on shortest path can be undetectable if we consider only the path delay as in previous delay-based methods. The critical path delay will hide the Trojan on shortest path.

In order to overcome the advantages of previous methods, we proposed a technique that depends on concept STF frequency which targets on the shortest path affected by Trojans. The basic idea of STF frequency is applying test patterns at different frequencies range from low to high, and this is widely practiced in industry called 'speed binning.' Some of the paths which have the delay greater than the current clock period will fail when frequency increases. The obtained fail clock frequency indicates the delay of the sensitized by the test vectors.

For example, assume that the three paths in Fig. 3a can be trace by the test vectors. The clock period is varied from f_j to f_4 , and the step size is Δf as

$$\Delta f = f_{i+1} - f_i \quad (1)$$

$$\Delta = \frac{1}{f_{i+1}} - \frac{1}{f_i} \quad (2)$$

As shown in Fig. 3b, f_j is the 'circuit functional frequency' and f_4 is the 'maximum allowable frequency.' Assume that the path 1-3 will propagate correct values between f_j and f_j . Hence, the STF frequency of 1-3 path exists between f_1 and f_2 . Similarly, the STF frequencies for paths 1-2 and 2-3 are existing between f_3 - f_4 and

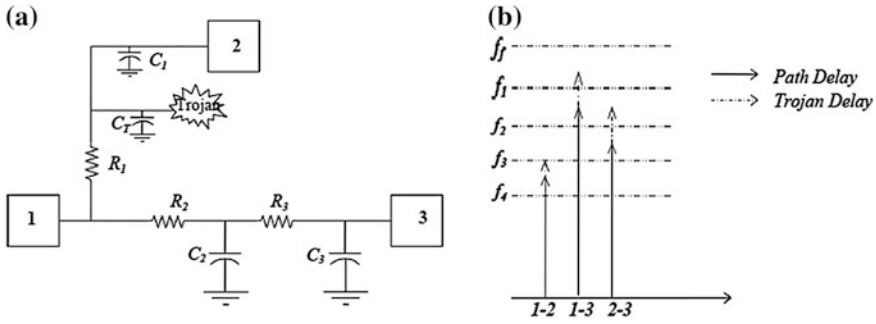


Fig. 3 a An example circuit. b Start to Fail frequency concept

f_2 – f_3 , respectively. Due to the Trojan, there exists a capacitance (C_T) which further increases delay of the paths. Increased delay will decrease the operating frequency which causes the circuit will fail less than the Trojan free STF frequency. Path **1–3** will fail before f_1 which indicates the existence of Trojan. In this technique, the step size (Δf) plays the major role. If the step size is larger, then the frequency change due to Trojan will hide the existence of Trojan. For example, assume that the Trojan delay is less than Δt for the path **1–2** in our example. Then, the frequency change still exist between f_3 and f_4 experienced the circuit is Trojan free.

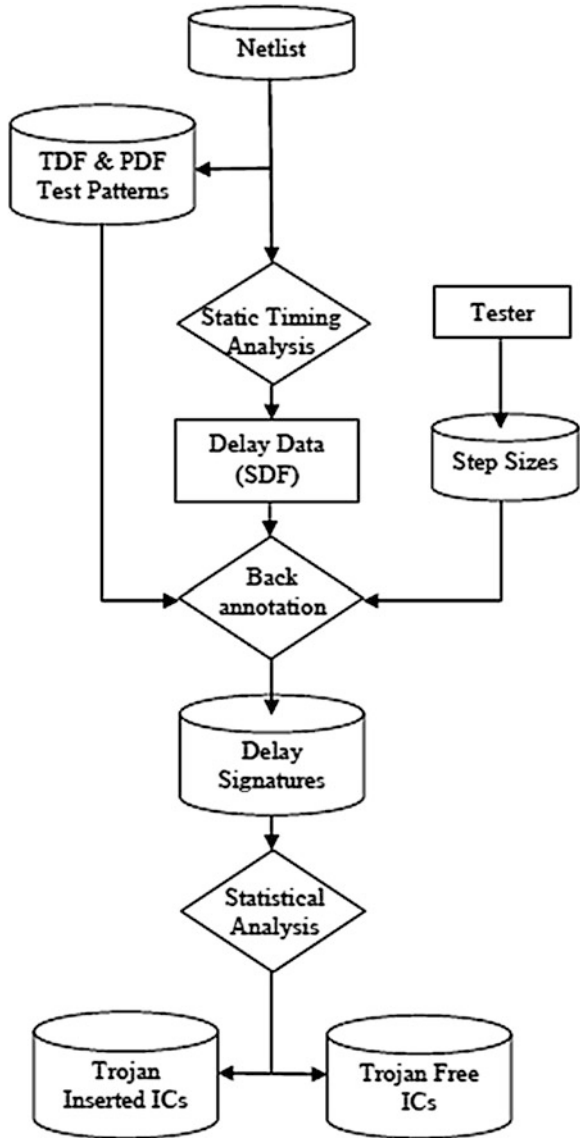
3 Trojan Detection Methodology

In this work, we consider both the nodes on short and long paths, but focused mainly on short paths in terms of frequency along with path delay. Signature generation methodology is shown in Fig. 4.

3.1 Test Pattern Generation

Test vectors generated by TDF and PDF have been taken into account. Particularly, we have taken TDF test patterns because of the reason the payload affects the nodes by either making the gates slow to fall or slow to rise. The advantage of TDF the number of faults is linearly relative to nodes which lead to test cost reduction by having patterns for available nodes alone. ATPG processes are matured recently [9]. Some of the test pattern generation techniques particularly for hardware Trojan detection are proposed recently [10].

Fig. 4 Start to Fail frequency procedure



3.2 Selection of Step Size

Based on the step size (Δf) selection, the efficiency of proposed method will depend. The smaller the step size, more effective the Trojan detection technique. But if the step size is small, then the test time and data volume will be high.

3.3 Node Selection

After the determination of step size of frequencies, the paths are long paths if the delay of the sensitized paths larger than maximum frequency. Remaining paths are short paths. In Fig. 3, path **1–2** is the shortest path. The delay of long paths will be decided by their STF frequency. Sensitized short-path patterns are still useful in generating transitions for power-based Trojan detection [5, 6].

In order to sensitize the short paths, we took a parameter from the power equation ($P = C_L^2 \bar{V} DDf$), i.e., frequency.

3.4 Methodology

The proposed methodology is similar to conventional TDF testing; the only difference is we need to consider frequency also. The test patterns will be applied at different frequencies from low to high depending on step size.

1. **Signature Generation:** With the help of pass and fail frequencies, we found STF frequency for each pattern. This technique guarantees that all sensitized long paths will fail at a particular frequency. Hence, the coverage of nodes is proportional to the TDF test coverage.

We used a test pattern generation technique proposed in 9 to obtain maximum coverage.

2. **Statistical Analysis:** The data obtained in this technique are of high dimensionality which has to be reduced. Hence, we use one of the multivariate statistical techniques named ‘principle component analysis (PCA)’. PCA method is based on the correlation techniques between different dimensions of the data. It is more efficient if the available data are linear.

The convex hull of a set of points X is the smallest convex region containing of all the points of X . Quick hull algorithm is widely used for its computation speed and efficiency [17].

4 Scalable Encryption Algorithm

In order to test efficiency of our STF frequency technique in hardware Trojan detection, we implemented this technique on scalable encryption algorithm (SEA). SEA is a parametric block cipher for resource constrained systems (e.g., sensor networks, RFIDs) that has been introduced in [11]. It was initially designed as a low-cost encryption/authentication routine (i.e., with small code size and memory) targeted for processors with a limited instruction set (i.e., AND, OR, XOR gates,

word rotation, and modular addition). The algorithm takes the plaintext, key, and bus sizes as parameters. In this section, we give a complete description of SEA algorithm and its loop implementation.

4.1 Basic Operations and Parameters

$SEA_{n,b}$ operates on various text, key, and word sizes. It is based on a Feistel structure with a variable number of rounds and is defined with respect to the following parameters:

N	Plaintext size, key size
B	Processor (or word) size
$n_b = \frac{n}{2b}$	Number of words per Feistel branch
n_r	Number of block cipher rounds

As only constraint, it is required that n is a multiple of $6b$. For example, using an 8-bit processor, we can derive a 96-bit block ciphers, denoted as $SEA_{96,8}$ [11].

SEA is based on a limited number of operations denoted as follows:

1. Bitwise XOR \oplus
2. Addition mod 2^b \boxplus
3. 3-bit substitution box $S := \{0,5,6,7,4,3,1,2\}$
4. Word rotation R on n_b -word vectors:

$$R : y_{i+1} = x_i, \quad 0 \leq i \leq n_b - 2 \quad y_0 = x_{n_b-1}$$

5. Bit rotation r on n_b -word vectors:

$$\begin{aligned} r : y_{3i} &= x_{3i} \ggg 1, \\ y_{3i+1} &= x_{3i+1}, \\ y_{3i+2} &= x_{3i+2} \lll 1, \quad 0 \leq i \leq n_b - 2 \end{aligned}$$

where \ggg and \lll represent the cyclic right and left shifts inside a word.

4.2 Generic Loop Architecture

Loop architecture implementation of SEA introduced in [12]. Unlike in [11], this loop architecture will support both encryption and decryption and execute one round per clock cycle. In this implementation, the round function and key schedule do not share any resources. This loop architecture has benefits for FPGAs compared to [11] architecture. The structure of generic loop architecture of SEA is referred in [11].

5 Experimental Results

We verify the efficiency of our STF frequency technique by inserting different types of Trojans on SEA chiper. After synthesis, the SEA circuit has equivalent area of 9237 NAND gates. The maximum operating frequency is 253 MHz. Layout was completed with Cadence SOC Encounter. The Trojan gates were inserted and routed in unused spaces.

In this work, we do not concern the type and size of the Trojan. We focus on the number of payloads and triggers. We inserted more number of Trojans, whereas we consider only three Trojans. All Trojans are constructed using minimum-sized gates and place very nearer in order to decrease the interconnection delay which makes the Trojan hard to detect. The three Trojans were constructed as follows. Trojan 1 has 2-bit comparator as a trigger and two payloads. This type of Trojan just alters the two output bits that depend on triggering. The activation probability of this Trojan is very less to escape the functional testing. Trojan 2 is same as the Trojan 1 but at different location. And Trojan 3 has two triggers and a single payload. Unlike Trojan 1 and Trojan 2, Trojan 3 leaks the plaintext data to the output.

We also verify the efficiency of our when Trojan and design from different process corners. For example if the design is from typical corner and inserted Trojan from fast corner. We produced simulation results for all possible nine combinations, three from each process corner (fast, typical, and slow).

The maximum frequency, functional frequency, and step size in our simulation are 550, 253 MHz, and 10 ps, respectively. We generate 200 fingerprints for 100 Trojan-free ICs and 100 Trojan ICs using the methodology proposed in Fig. 4. This

Table 2 Experimental results for different combinations of design—Trojan process corners

Combinations of process corners	Start to Fail frequency of a particular design (MHz)	Decrease in frequency due to Trojan (MHz)
<i>Design library—fast</i>		
Trojan free	197	–
Fast Trojan	193	4
Typical Trojan	194	3
Slow Trojan	190	7
<i>Design library—typical</i>		
Trojan free	196	–
Fast Trojan	191	5
Typical Trojan	193	3
Slow Trojan	188	8
<i>Design library—slow</i>		
Trojan free	162	–
Fast Trojan	158	4
Typical Trojan	159	3
Slow Trojan	156	6

results for different combinations of process corners of Trojan 1. The results showed that our STF frequency technique is efficient for different combinations of process corners when the step size (Δt) is minimum.

From the results, we can estimate the worst and best conditions in Trojan detection point of view. For SEA chip, the Trojan is hard to detect when the design and Trojan from two combinations of process corners: one when design and Trojan from fast corner and second design from typical and Trojan from fast corners. That means both are the worst combinations in designer point of view where Trojans may escape by this STF frequency method. For the remaining combinations, the Trojan can easily detect because the distance between convex hull and dots is long enough (Table 2).

6 Conclusion

In this work, we used STF frequency concept to generate the finger prints using the critical and noncritical path delays to detect hardware Trojans in ICs. We proved that our method is efficient by statistical methods when compared to the previous delay-based detection methods which are only capable to detect Trojans on critical paths. Simulation results show that our method is effective under any operating conditions.

References

1. B. Sharkey, DARPA TRUST in Integrated Circuits Industry Day Brief (2007)
2. R. Karri, J. Rajendran, K. Rosenfeld, M. Tehranipoor, Trustworthy hardware: identifying and classifying hardware Trojans. *IEEE Comput. Mag.* **43**, 39–46 (2010)
3. R.S. Chakraborty, S. Narasimhan, S. Bhunia, in *Hardware Trojan: Threats and Emerging Solutions. High Level Design Validation and Test Workshop*, HLDVT2009 (2009), pp. 166–171
4. D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, B. Sunar, in *Trojan detection using IC fingerprinting*. Proceedings IEEE Symposium Security Privacy (SP) (2007), pp. 296–310
5. R. Rad, M. Tehranipoor, J. Plusquellic, A sensitivity analysis of power signal methods for detecting hardware Trojans under real process and environmental conditions. *IEEE Trans. Very Large Scale Integr.* **18**(12), 1735–1744 (2009)
6. H. Salmani, M. Tehranipoor, J. Plusquellic, in *New design strategy for improving hardware Trojan detection and reducing Trojan activation time*. Proceedings of IEEE International Symposium on Hardware-Oriented Security Trust (HOST) (2009), pp. 66–73
7. Y. Jin, Y. Makris, in *Hardware Trojan detection using path delay fingerprint*. Proceedings IEEE International Workshop on Hardware-Oriented Security Trust (HOST) (2008), pp. 51–57
8. J. Li, J. Lach, in *At-speed delay characterization for IC authentication and Trojan horse detection*. Proceedings of IEEE International Workshop on Hardware-Oriented Security Trust (HOST). (2008) 8–14

9. G. Xu, A. Singh, in *Achieving high transition delay fault coverage with partial DTSFF scan chains*. Proceedings of IEEE International Test Conference (ITC) (2007), pp. 1–9
10. L. Fang, L. Li, Z. Li, A practical test pattern generation technique for hardware Trojan detection. *Elektrotechniski vestnik, J. Electr. Eng. Comput. Sci.* (2013)
11. F.-X. Standaert, G. Piret, N. Gershenfeld, J.-J. Quisquater, in *SEA: a scalable encryption algorithm for small embedded applications*. Proceedings of CARDIS (2006), pp. 222–236
12. F. Mace, F.-X. Standaert, J.-J. Quisquater, in *ASIC implementations of the block cipher SEA for constrained applications*. Proceedings of RFIDSEC'07 (2007), pp. 103–114
13. M. Tehranipoor, F. Koushanfar, A survey of hardware Trojan taxonomy and detection. *IEEE Des. Test Comput.* (2010), pp. 10–25
14. S. Adee, The hunt for the kill switch. *IEEE Spectr.* **45**(5), 34–39 (2008)
15. C. Bradford Barber, D.P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls. *ACM Trans. Math. Software.* **22**(4), 469–483 (1996)

A Novel Approach Privacy Security Protocol Based SUPM Method in Near Field Communication Technology

S. Kannadhasan, M. Isaivani and G. Karthikeyan

Abstract For the past few years, so many mobile terminals have been released into market with NFC which is abbreviated as near-field communication. The NFC together with smart devices has made to improve the effective utility range of NFC. Specifically, NFC electronic payment is expected to take the place of credit cards in e-payment. For this, it is essential to direct the attention of the security issues in NFC. Presently, the security standards in NFC make use of user's public key at a fixed value in key agreement process. The message's relevancy occurs at the public key of NFC. NFC is used for gathering the message based on the malicious attacker. This leads to the infringement of privacy of user. The planned work presents conditional privacy protection method based on pseudonyms to overcome the aforementioned problems. The users can communicate with other parties based on some set of rules by sending the conditional privacy-preserved protocol data unit (PDU) through NFC terminals. The proposal works well in decreasing the update cost and computation overhead by considering the merit of the physical characteristics of NFC.

Keywords SE · Key agreement · NFC

S. Kannadhasan (✉)

Department of ECE, Raja College of Engineering and Technology,
Madurai, Tamilnadu, India
e-mail: kannadhasan.ece@gmail.com

M. Isaivani

Embedded System Technologies, Raja College of Engineering and Technology,
Madurai, Tamilnadu, India
e-mail: isainec06@gmail.com

G. Karthikeyan

Department of ECE, St. Michael College of Engineering and Technology,
Kalayar Koil, Sivagangai, Tamilnadu, India
e-mail: Gkkbeg@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_68

633

1 Introduction

NFC is an emerging close-range, low-bandwidth wireless communication technology with coverage distance of around 4 in. and operates at the speed range between 106 and 424 kbps at 13.56 MHz frequency. It is backed by the NFC forum which makes NFC to be tightly integrated with modern consumer electronics like smart phones and operating systems such as Windows 8 and Android in Table 1. Through NFC technology, it is feasible to combine the interface of a reader and a smart card in a single device.

NFC is a type of short-range wireless communication technology which has the coverage distance up to 4 in. and operates at the speed range that varies from 106 to 424 kbps at 13.56 MHz operating frequency. The application of NFC together with smart devices includes access control, consumer electronics, health care, transport, data exchange, discovery, connection, e-payment, and ticketing. NFC is one-to-one communication which lets to the ease of identification for the users. NFC is used in payment applications performs sporadic communication which is pseudonym used to generate the one time ID. Additionally, there is no need to store a large number of pseudonyms because there is so much of time before next payment. Also, the users can check whether the communication is done properly via each other's device since communication is progressed with the target in front of our eyes. Presently, the NFC makes use of public keys that are exchanged between users for secret communication with key agreement process. The public key received from the Certificate Authority uses a fixed value. Attacker can easily acquire the public key to create the new profiles of users using key agreement process. MITM attack, eavesdropping, and data modulation are the possible threats of NFC. For the purpose of applying the NFC in electronic payment, we have to address the security standard. The proposed work incorporates privacy protection method based on pseudonyms to protect the user's privacy. Conditional privacy is also provided to identify the users and can be checked by the trusted third party (TTP) to solve problems if necessary. Conditional privacy protocol data unit (PDU) also defined that the users can make use of the protected PDU to receive the personalized services and use conditional privacy PDU to conceal the needed information.

Table 1 Comparison of NFC with other wireless communication technologies

S. No.	Parameters	NFC	ZigBee	Bluetooth	WLAN
1.	Coverage distance	Around 4 in.	<100 m	<100 m	<250 m
2.	Speed	106–424 kbps	<250 kbps	<2.1 MbPS	<866.7 MbPS
3.	Size of the network	2	2–64	8	2007
4.	Operating frequency	13.56 MHz	868/915 MHz 2.40 GHz	2.4 GHz	2.4/5 GHz

2 Related Work

In VANET (Vehicular Ad-Hoc Networks), the vehicle that uses PACP scheme should register with the motor vehicle department using its identity to get a ticket. Then, it communicates with road side unit by using the ticket to receive tokens that are specifically used to generate pseudonyms for anonymous broadcast communication with other vehicles [1]. NFC was used for Secure Element (SE) with connecting the hardware for the passive module. Public key infrastructure (PKI) can be deployed using X.509 Certificates for authentication between the voucher issuer and all the users. The SE on NFC phones consists of mifare 4 k area and a Java card. But there are some limitations are used Java Card is limited freedom JAVA Card 2.2.1 developers have and timing constraints and maximum APDU (Advanced Pro Series Digital Variable Ignition System) size of 256 bytes [2]. NFC is tightly integrated with smart phones, devices and operating systems sharing of text identity verification websites or setting up Bluetooth/Wi-Fi connections for large file transfers can be possible in NFC touching the NFC-enabled devices together. However, eavesdropping, denial of service, and data modification relay attack are the possible threats that NFC faces are addressed recently [3]. The NFC specifies the secure channel and shared Secret Services for NFCIP and Protocol data units and protocol for those services. Peer NFC-SEC entities should exchange NFC-SEC-PDUS through NFC service access points confirming the NFC-SEC protocol over NFC-SEC connection in order to support the NFC-SEC services. NFC-SEC-PDU consists of NFC-SEC protocol control information and a single WFC-SEC-SDU. NFC-SEC Standard specifies with two SSE services to establish the shared secret between two peers. NFC-SEC users can utilize the discretion, and SCH is used to provide a secure channel [4]. In multiple pseudonym-based method (MuPM), if user A requests for pseudonym from trusted service manager (TSM), TSM produces pseudonym set which depends on the generated data and transmits it to user A who performs SSE protocols by picking one of the pseudonyms to be sent. The revocation list manages the used pseudonym. It is simple, and user's authentication can be easily performed.

3 Important Elements and Features in NFC

Pseudonyms are randomly changing IDs received from TTP. It is composed of public key, private key, and a certificate [5]. Through pseudonym, the anonymity of user can be assured and they are authenticated as normal users through a certificate pseudonym and actual IDs of users are stored by the TTP to reveal the anonymity if any problem occurs [6]. Since it needs additional cost for storage and communication, recent studies have been conducted to generate pseudonym without the help of TTP.

TSM is an institution which safely transfers customer's mobile financial data to financial institutions [7]. It serves as Certificate Authority and Registration Authority [8].

Secure element is a security area which stores vital data such as financial information, authentication information, and service applications safely [9]. The storage features and secure domain are included in all types of implementation.

One-to-one communication is easy for one of the communicating party to identify with whom it communicates [10].

Near-field communication enables users to check whether the communication is conducted in a proper manner through each other's device since it is done in front of eyes of users [11].

Sporadic communication enables the use of one-time ID such as pseudonym for payment, and since there is so much time before next payment, there is no need to store a large amount of pseudonyms in advance [12].

4 Proposed Method

4.1 SuPM: Self-updateable Pseudonym-Based Method

The protocol design process of NFC can be configured in such a way can update pseudonym without communicating with TSM only used for keeping track of the message constructor.

User A concatenates RA' , RA'' , NA and sends it to user B. User B obtains $R'A$ & $R''A$ by decompressing RA' and RA''

$$RA' = SA RA = SA EA H$$

$$RA'' = SA EA RS + RA = SA EA ES H + EA H$$

According to the algorithm, user B cannot find that $R'A$ is RA which is the message made by same person, even if user B knows RA . Similarly, in $R''A$, user B cannot find $SA EA RS$ since user B does not know $SA EA$. So user B cannot remove $SA EA RS$ from $R''A$ and cannot identify that the message was constructed by user A. But TSM can recognize the message originate by the following decryption in Fig. 1.

$$R''A - ES RA = (SA EA (ES H)) + RA - ES(SA/EA H = RA)$$

Two users can get the common values by exchanging messages of $R'A$ and $R'B$. Multiplication of private key and the random of value are $R'A$ and $R'B$ is performed in order to obtain the same value

$$Q = SA EA R'B = SA EA (SB EB H) = SA EA SB EB H = SB EB (SA EA H) \\ = SB EB RA$$

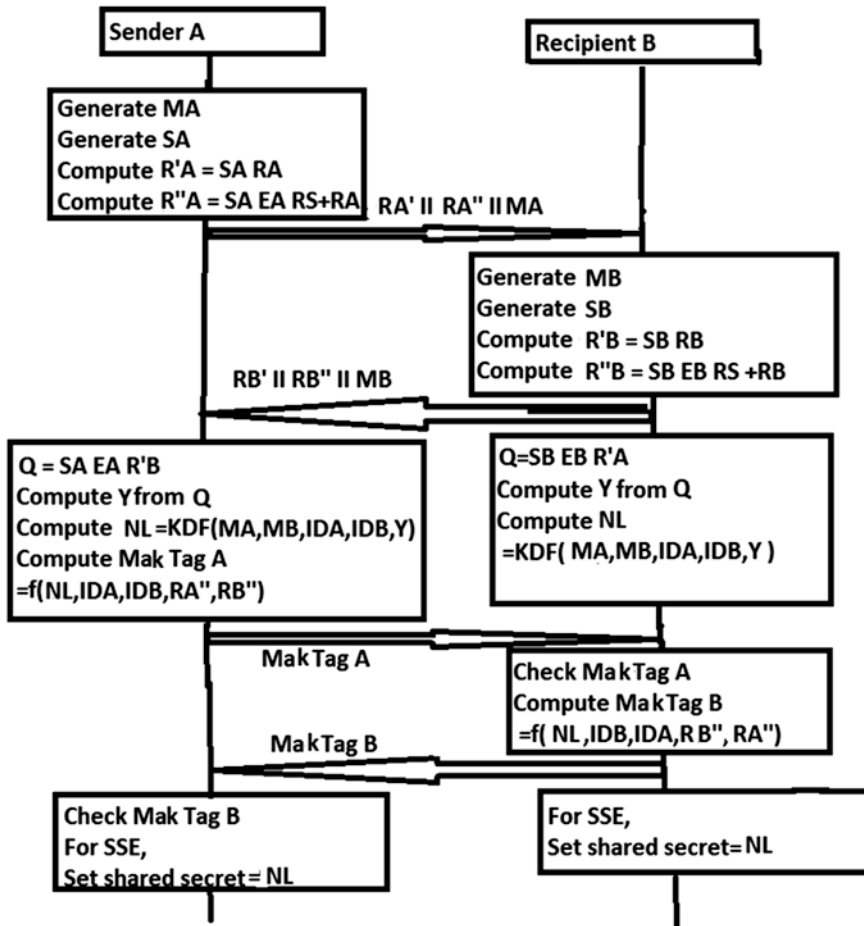


Fig. 1 Proposed key agreement and confirmation protocol

Two users can obtain a shared secret value Y by taking x coordinate value at point P (Table 2). We compare with the existing protocols, and $R'A$ and $R'B$ can replace RA and RB , the existing public keys. The pseudonym consists of

$$PN = \text{Public Key} + \text{Encrypted Private Key} + \text{ID of TSM} + \text{Signature}$$

If they are attackers, they cannot decrypt the information of sender. The planned work proposes privacy protection methods based on pseudonyms to protect privacy of users. It provides conditional privacy to identity of users can be verified by the TTP to resolve disputes if necessary.

Table 2 Acronyms of key agreement

II	Concatenation symbol
Mx	Nonce of user x
IDx	Random ID of user X for the activation of transport protocol
RX, RX', RX''	Compressed elliptic curve public key of user
Rx, R'x, R''x	Elliptic curve public key of user
Ex	Elliptic curve private key of user
H	Elliptic curve base point
KDEF	Key derivation function
Mak Tagx	Key verification tag received from X
NL	Shared secret key
Y	Unsigned integer
Sx	Random integer generated by user X

4.2 Advantages

In the proposed method, pseudonyms are generated, so there is no hacking using public keys, legal recipient only knows the key to decrypt the data, and others may get the data, but they cannot decrypt.

5 Performance and Simulation Result

5.1 Performance

a. Storage for Pseudonyms

The size of single pseudonym can be calculated as follows:

$$\begin{aligned} \text{Size of PN} &= \text{Public key} + \text{Encrypted Private Key TID of TSM} + \text{Signature} \\ &= 1200 \text{ bits} \end{aligned}$$

If there are 1000 numbers of pseudonyms, the needed memory is 146.484 kbytes. Though it is not a big deal, updated environment is limited due to the billing charges of mobile device.

b. Computation Time

The multiplication, square, and inverse operations are represented as MU, SQ, and IN, respectively, the required time to add two points is $a3 = a1 + a2$, and the needed computation time for doubling is $a2 = 2a1$. A total of 288 times of doubling operation is needed which is increased by 96 times compared to NFC-SEC (Table 3).

Table 3 Computation time

Operation	Format	Computation time
Doubling	T (2a1)	2MU + 2SQ + IN
Addition	T (a1 + a2)	2MU + SQ + IN

c. More Transference Time

The proposed method needs each user to transfer points on the elliptic curve in the key agreement process. It requires 4.59 ns, and transfer of 3.628 ns is additionally needed when compared to the standard method which requires only 2.727 ns.

d. Comparison

MuPM method requires more storage space for the maintenance of pseudonyms, overhead of managing revocation list, and communication cost for the issuance of pseudonyms. On the other hand, the proposed SuPM method needs not any communication cost for the issuance of pseudonyms, yet it requires additional computation time and transfer time.

5.2 Simulation Result

In Fig. 2, two ARM LPC 2103 processors and two LCD displays each for sender and receiver are shown. Based on the public key, encrypted private key, and user id, sender computes the pseudonym value and it is displayed in the sender’s display.

In Fig. 3, two ARM LPC 2103 processors and two LCD displays each for sender and receiver are shown. The receiver receives the pseudonym which is created by the sender, and it is displayed in the receiver’s display.

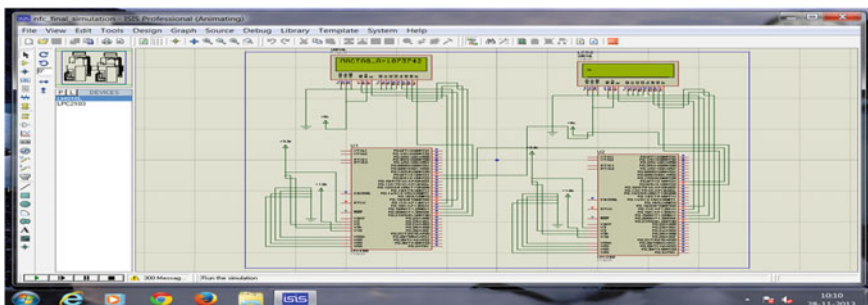


Fig. 2 Sender generating pseudonym

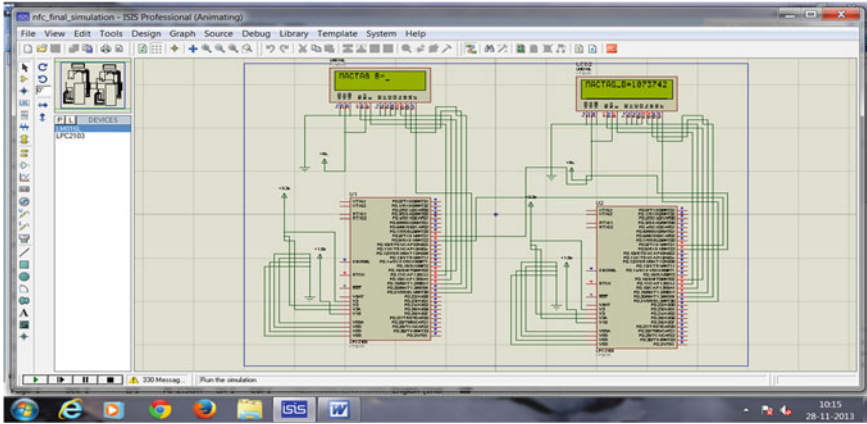


Fig. 3 Receiver receiving pseudonym

In Fig. 4, two ARM LPC 2103 processors and two LCD displays each for sender and receiver are shown. The receiver generates the pseudonym value by using public key, encrypted private key, and user id, and it is shown in the receiver's display.

In Fig. 5, two ARM LPC 2103 processors and two LCD displays each for sender and receiver are shown. The sender obtains the pseudonym which is created by the receiver, and it can be viewed on the sender's display.

Figure 6 shows that after verifying both users' pseudonyms to each other, the data transfer will begin, which indicates that authentication is done successfully since legal recipient only knows the key to decrypt the data, and others may get the data, but they cannot decrypt.

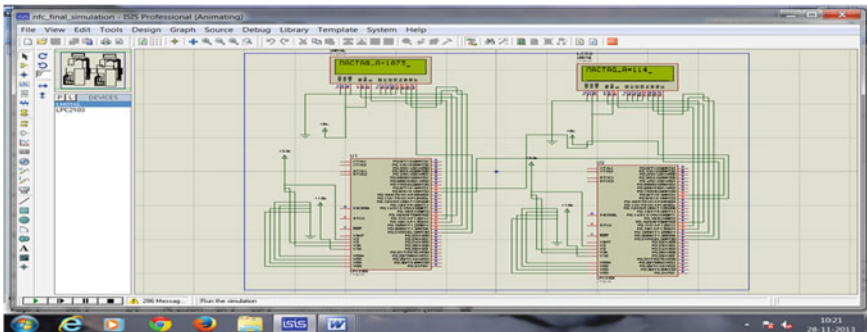


Fig. 4 Receiver computing pseudonym

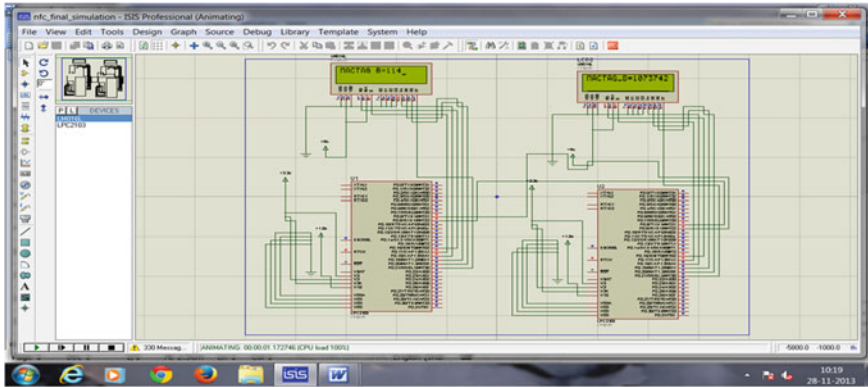


Fig. 5 Sender receiving pseudonym

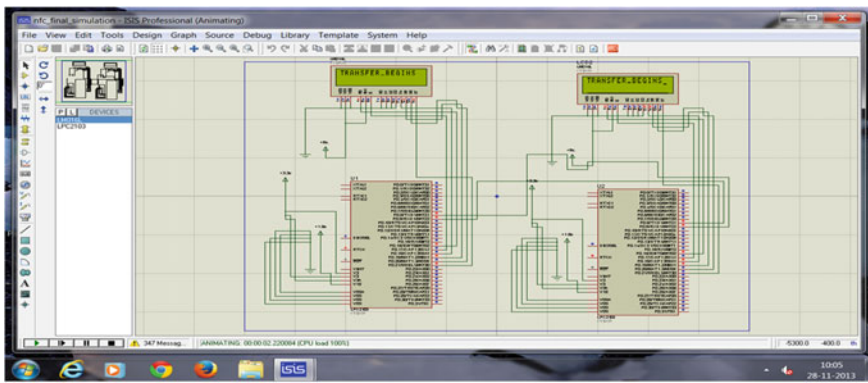


Fig. 6 Successful authentication

6 Conclusion and Future Work

The various NFC enabled devices for the purpose of e-payment market using NFC to be activated. If there is any leak of the users' transaction information, it will compromise user's privacy. The planned work proposes conditional privacy protection to solve the aforementioned problems. The method uses updateable pseudonyms and the updating is based on the long-term public key issued from TSM and storing the long-term public key in the SE safe management can be achieved. In future, this will be implemented in hardware using ARM 2103 processor, NFC reader, and smart phone.

References

1. S.J. Moses, P.A.C. Angelin, Enhancing the privacy through pseudonymous authentication and conditional communication in vanets. *Int. J. Eng. Sci.* **2**(7), 45–49 (2013)
2. G. Van Damme, K.K. Wouters, *Practical Experiences with NFC Security on Mobile Phones*
3. U.B. Trottmann, *NFC—Possibilities and Risks. Matthias Wachs Seminar Future Internet WS2012*
4. H. Eun, H. Lee, J. Son, S. Kim, H. Oh, Conditional privacy preserving security protocol for NFC applications, in *IEEE International Conference on Consumer Electronics (ICCE)*, (2012), pp. 380–381
5. J.-H. Lee, J. Chen, T. Ernst, Securing mobile network prefix provisioning for NEMO based vehicular networks. *Math. Comput. Model.* **55**(1), 170–187 (2012)
6. Juniper Research, *NFC Mobile Payments and Retail Marketing—Business Models and Forecasts*, (2012), pp. 2012–2017
7. J. Yu, W. Lee, D.-Z. Du, Reducing reader collision for mobile RFID. *IEEE Trans. Consum. Electron.* **57**(2), 574–582 (2011)
8. D. Huang, S. Misra, M. Verma, G. Xue, PACP: an efficient pseudonymous authentication-based conditional privacy protocol for VANETs. *IEEE Trans. Intell. Transp. Syst.* **12**(3), 736–746 (2011)
9. R.J. Hwang, Y.K. Hsiao, Y.F. Liu, Secure communication scheme of VANET with privacy preserving, in *Proceedings of the 2011 IEEE 17th International Conference on Parallel and Distributed Systems (ICPADS 2011)*, (2011), pp. 654–659
10. J.C.M. Teo, L.H. Ngho, H. Guo, An anonymous dos-resistant password-based authentication, key exchange and pseudonym delivery protocol for vehicular networks, in *Proceedings of the 2009 International Conference on Advanced Information Networking and Applications (AINA 2009)*, (2009), pp. 675–682
11. D. Eckhoff, C. Sommer, T. Gansen, R. German, F. Dressler, Strong and affordable location privacy in VANETs: Identity diffusion using time-slots and swapping, in *Proceedings of the 2010 IEEE Vehicular Networking Conference (VNC 2010)*, (2010), pp. 174–181
12. Gartner, *Market Insight: The Outlook on Mobile Payment. Market Analysis and Statistics* (2010)
13. E. Haselsteiner, K. Breitful, *Security in Near field Communication (NFC)—Strengths and Weaknesses. RFIDSec 2006* (2006)

Author Biographies



S. Kannadhasan received his BE degree from Anna University, Chennai, in the year 2009. He received ME degree from Anna University in June 2013. He serves as an Assistant Professor at Raja College of Engineering and Technology. His research interests include digital image processing, wireless sensor networks, and embedded systems. He is a member of the Institution of Engineers and Indian Society of Technical Education.



M. Isaivani received her BE degree in electronics and instrumentation engineering from Anna University in the year 2006. She is currently doing her ME in embedded system technologies. Her research interests include control engineering, digital image processing, and cryptography.



G. Karthikeyan received his BE and ME degrees from Anna University, Chennai, in the year 2009 and 2012. He is currently working as an Assistant Professor in St. Michael College of Engineering and Technology from June 2012. His research interests include wireless communication and wireless sensor networks.

Gabor Transform for the Time–Frequency Localization of Impulse Faults in a Transformer

N. Vanamadevi, S. Santhi and M. Arivamudhan

Abstract Signal transformation is one of the most important signal analysis techniques that are widely used in fault detection applications. Certain features in a signal are clearly observed in frequency domain than in time domain. It is well known that Fourier transformation is applied to analyze a stationary signal, and the time–frequency transformation is required for its counterpart namely the non-stationary signals since it preserves both frequency and time information of the signal. In this paper, we propose the application of Gabor transform to determine the time of occurrence of impulse faults that are likely to occur in a transformer winding through a simulation work and validate its candidature as superior to Fourier analysis. The results are encouraging that the method can be adopted as a better signal analysis tool for impulse fault detection and localization.

Keywords Signal transformation · Frequency response analysis method · Time-scale analysis · Gabor transform · Wigner–Ville distribution

1 Introduction

Transformers play a vital role in power utility systems and hence require proper monitoring to avoid major economic loss in case of any failure. As windings are the heart of the transformer, assessing the integrity of the winding is an essential task. Impulse testing of transformer is a routine test to demonstrate dielectric integrity of

N. Vanamadevi · S. Santhi (✉) · M. Arivamudhan
Department of Instrumentation Engineering, Annamalai University, Annamalai Nagar,
Chidambaram, Tamil Nadu, India
e-mail: santhi.sathyamurthy@gmail.com

N. Vanamadevi
e-mail: vanamadevinraju@gmail.com

M. Arivamudhan
e-mail: aumaei@gmail.com

windings of a transformer, and the test procedure is described in standards such as IEC 60076 Part IV, 2002 [1]. Detection of failure of a transformer during impulse testing is possible by observing the variations in the winding current signature through the transfer function method [2]. Among the various faults that a transformer may suffer, series and shunt faults are considered to be important and hence need a suitable diagnostic system that detects, quantifies, and locates these faults at an incipient stage to avoid any major failure [3].

In frequency response analysis (FRA) method, the detection of a fault condition is made by observing the changes in the resonant frequencies that characterizes the winding [4]. Methods based on timescale analysis have been proposed for the detection and classification of impulse faults [5–10]. However, timescale analysis is suitable if the signal contains transients of varying scale, and hence, an approach based on time–frequency analysis using Gabor transform is attempted in this work. The algorithm to implement the signal analysis technique using Gabor transformation is implemented through MATLAB coding and is validated for a simulated non-stationary signal. Further, in this paper, a method based on the use of Gabor transform for the detection of time of occurrence of series and shunt faults that may occur in a transformer is proposed and validated through simulation. The simulation work is carried out by considering the lumped parameter model of a transformer winding with parameters as mentioned in [11].

2 Objectives and Scope

The objectives of the present work are to understand the need for time–frequency analysis of real-world signals and then to develop a signal analysis method using Gabor transform for the detection and localization of the winding faults that may occur in a transformer winding when it is subjected to impulse test. The scope of the work is limited to simulation study of lumped parameter model of a transformer winding using a circuit simulation package. The series and shunt faults in the winding model are simulated at known time instants, and the objective is to develop a method based on time–frequency analysis with increased sensitivity of detection.

3 Time–Frequency Analysis

Earlier signal analysis strategies were based upon time, frequency, or scale. For certain tasks such as edge detection, time-domain analysis is sufficient. But the inherent periodicity within the signal regions makes us to perform frequency transformation. Frequency or spectral analysis is a global approach that suffers signal interpretation difficulties when the oscillations of our interest exist only within a limited signal region. Signal scale analysis is effective when the structure of a signal contains transients of varying scale. A time–frequency signal transform

is a class of mixed domain signal analysis that combines traditional Fourier transform of the signal with a time location variable [12]. This results in a two-dimensional transformed signal with an independent frequency variable and an independent time variable. Gabor transform and short time Fourier transform are linear methods, whereas Wigner–Ville distribution (WVD) is a nonlinear method, and it is the original time–frequency signal analysis technique. WVD does not rely on a separate window function; instead, it uses a bilinear term that involves the original signal. Though WVD has a better spectral resolution, the drawback of it is the presence interference terms that are hard to overcome.

4 Gabor Transform

Gabor transform is a most intuitive and historically prior time–frequency technique. It is a class of short time Fourier transform that employs a Gaussian window. Even though the Gaussian window is not truly finite in extent and decays very fast, for practical computation aspects, a level of significance can be chosen such that it serves the purpose of localization of signal values [13]. A signal could be expressed as a summation of mutually orthogonal time-shifted and frequency-shifted Gaussian function. The Gabor transform of a signal $x(t)$ is given by

$$X(\mu, \omega) = \int_{-\infty}^{\infty} x(t) e^{(t-\mu)^2/2\sigma^2} e^{-j\omega t} dt \quad (1)$$

where t and ω are the time and frequency variables. μ and σ are the mean and standard deviation of the Gaussian window function. The parameter μ represents the center of the window, while σ refers to the width of the window. If σ changes while ω remains fixed, it results in scale-based signal analysis. However, since the interest now is the time–frequency analysis, it is decided to use the practical implementation expression for the transform. Accordingly, the Gabor transform of a signal $x(t)$ is also represented as follows.

$$G_x(t, f) = \int_{-\infty}^{\infty} e^{-\pi(\tau-t)^2} e^{-j2\pi f\tau} x(\tau) d\tau \quad (2)$$

where t , f , and τ are the time, frequency, and time shift variables, respectively. The equation indicates that Gabor transform is used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. The equation is interpreted in the sense that the function to be transformed is first multiplied by a Gaussian function and the resulting function is then transformed with a Fourier transform to derive the time–frequency analysis.

5 Signal Analysis Using Gabor Transform

In order to appreciate the effectiveness of Gabor transform, a non-stationary signal is to be simulated and analyzed through Gabor transform algorithm. The following steps provide direction to implement Gabor transform algorithm.

Choose Gaussian window function of finite length

1. Put the window on top of the signal at $t = 0$
2. Truncate the signal using this window
3. Compute the FT of the truncated signal, save
4. Slide the window to the right by a small amount
5. Go to step 3, until window reaches the end of the signal.

For each time location where the window is centered, a different FT is obtained. Hence, each FT provides the spectral information of a separate time slice of the signal, providing simultaneous time and frequency information. In order to elucidate that Fourier transform does only global analysis of a signal and do not preserve time information, a non-stationary signal that contains spectral components at different time instants is simulated using MATLAB coding. Figure 1 shows the simulated non-stationary signal consisting of the frequency component 1 Hz over the time range -6 to 0 s and 2 Hz over the time range $0-6$ s. Figure 2 shows the Fourier transform of the simulated signal, and it is observed that the magnitude spectrum indicates only the frequency components 1 and 2 Hz and the time of occurrence of those spectral components are not indicated. The caveat is that standard Fourier technique that depends on knowledge of the entire time-domain extent of the signal.

Now, to illustrate the localized analyzing feature of Gabor transform, a Gaussian window as shown in Fig. 3 is simulated and the time–frequency transformation of the above-mentioned non-stationary signal is determined through implementation

Fig. 1 Simulated non-stationary signal

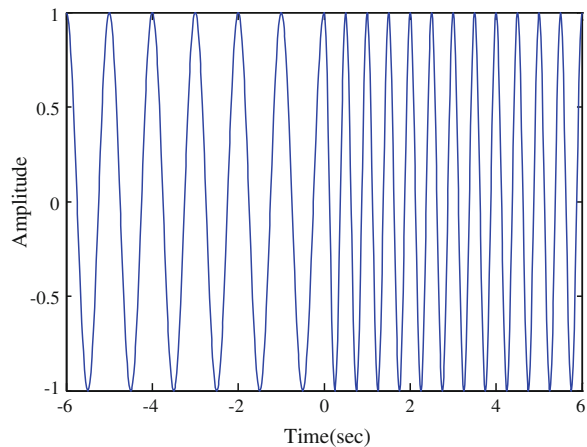


Fig. 2 Fourier-transformed non-stationary signal

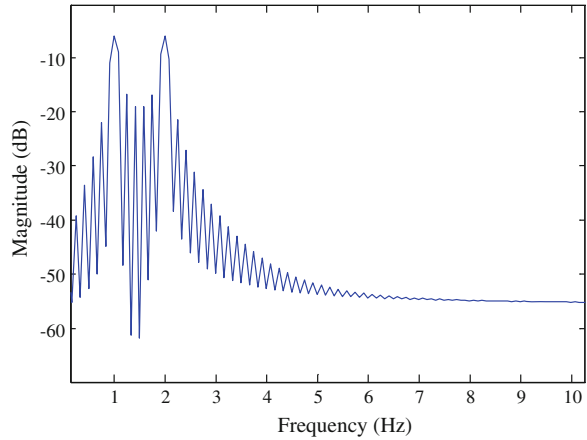
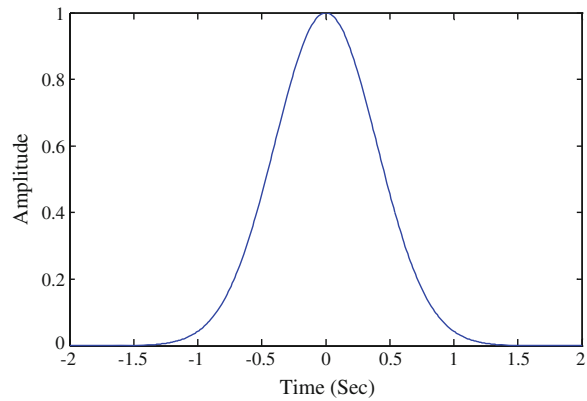


Fig. 3 Gaussian window



of Gabor transform algorithm using MATLAB coding. Figure 4 shows the time–frequency distribution of the simulated non-stationary signal, and it is evident that the time range of the spectral features is indicated clearly.

To elucidate further, a logarithmic chirp signal as shown in Fig. 5 is simulated and analyzed using Gabor algorithm. The logarithmic chirp contains frequency components ranging from 1 to 10 Hz, increasing logarithmically with a start time of 0 s and end time of 10 s. Figure 6 shows the Fourier transform of the logarithmic chirp that contains 1–10 Hz, and obviously only the spectral contents of the signal are indicated. The time of occurrence of the spectral components and the increasing trend in the frequency of the signal are not visible in the frequency domain.

The chirp signal is then Gabor transformed to time–frequency domain and is shown in Fig. 7. The dark color indicates the unused frequency band in the signal. The time–frequency distribution clearly indicates the spreading of spectral contents

Fig. 4 Gabor-transformed simulated non-stationary signal

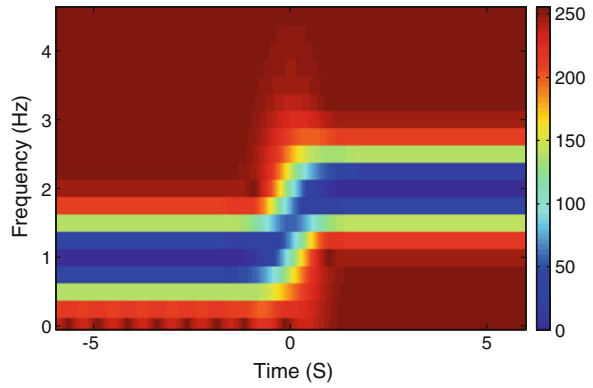


Fig. 5 Simulated logarithmic chirp signal (start at 1 Hz end at 10 Hz)

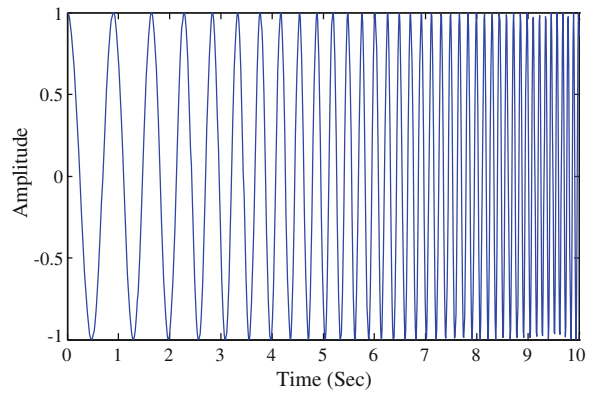
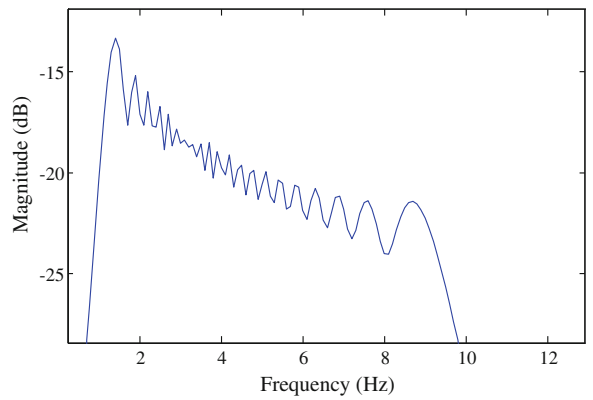


Fig. 6 Fourier-transformed logarithmic chirp signal



in the time range over which the signal is defined. Also the increasing frequency trend of the signal is evident. The Gabor coefficient values of the respective spectral components are the energy density values and are displayed with various colors.

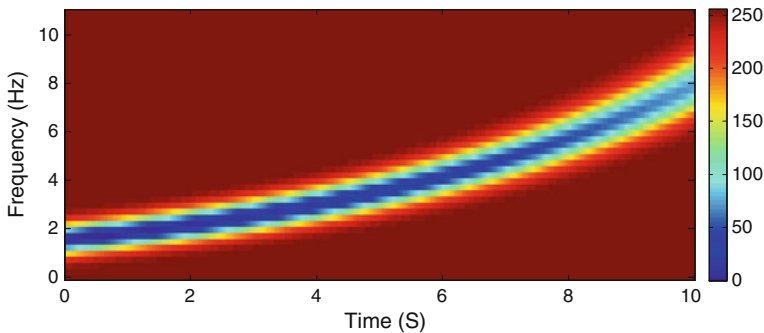


Fig. 7 Gabor-transformed logarithmic chirp signal

6 Impulse Fault Detection Using Gabor Transformation

All transformers are demonstrated for their dielectric integrity through impulse test. The above-described simulation exercise has provided sufficient theoretical foundation and practical motivation for the use of Gabor transform for time–frequency signal analysis. A method of power system transients and harmonics analysis using Gabor transform has been shown [14]. In the present work, to demonstrate its application for winding fault detection, a lumped parameter model of a transformer winding is considered for simulation using circuit simulation package. The model parameters such as inductances, ground capacitances, series capacitances, and mutual inductances of layer winding are chosen as mentioned in [11] and have been used to carry out the simulation. The inductance of each section is 0.5 mH, the ground capacitance of each section is 0.4 nF, and the series capacitance of each section is 0.1 nF

The mutual inductances for the layer windings are $M_{1-2} = 0.25$ mH, $M_{1-3} = 0.167$ mH, $M_{1-4} = 0.125$ mH, $M_{1-5} = 0.10$ mH, $M_{1-6} = 0.0833$ mH, $M_{1-7} = 0.071$ mH, $M_{1-8} = 0.0625$ mH, $M_{1-9} = 0.0556$ mH, $M_{1-10} = 0.050$ mH and $M_{1-2} = M_{2-3} = M_{3-4, \dots} = M_{9-10}$, $M_{1-3} = M_{2-4} = M_{3-5, \dots} = M_{8-10}$, $M_{1-4} = M_{2-5} = M_{1-2} = M_{3-6, \dots} = M_{7-10}$ and so on.

Figure 8 shows the lumped parameter model of a transformer winding with model parameter values as mentioned above referring to no fault (NF) condition. The winding current I_w of the model through a current-viewing resistor R is simulated after subjecting it to impulse excitation similar to standard lightning impulse of 1 V amplitude and 1.2/50 μ s front time and fall time, respectively.

A fault in the winding is reflected as change in the electrical parameters, namely inductance, series capacitance (C_s), or shunt capacitance (C_g). In order to simulate a series fault (SF) or shunt fault (SHF), changes are made in C_s or C_g . To facilitate signal analysis using Gabor, a SF is introduced at 0.1 ms by effecting change in series capacitance in the fourth section (middle of the winding) of the lumped

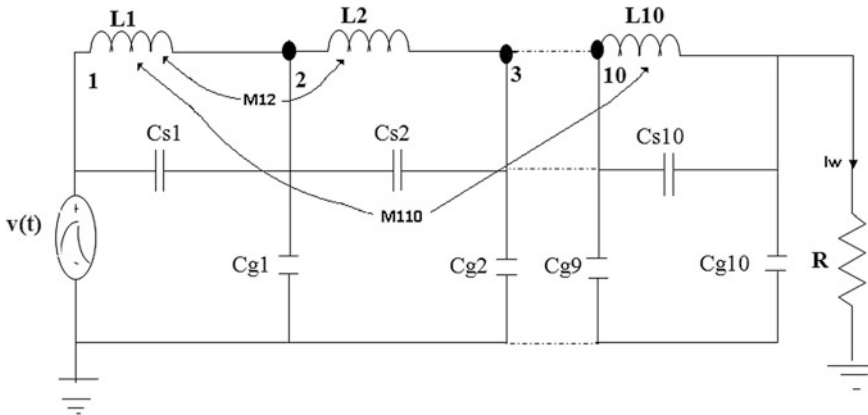


Fig. 8 Lumped parameter model of transformer winding

parameter model of the winding and winding current is computed using circuit simulation package. Similarly, a SHF is introduced by varying the ground capacitance in the fourth section and winding current due to SHF is computed. The simulated winding currents under NF, SF, and SHF in the fourth section are shown in Fig. 9 with their Fourier-transformed constituents in Fig. 10 only as sample records.

It is observed from the Fourier-transformed winding current under NF condition that the characteristic resonant frequencies of the layer winding are 58, 119, 175, 221, 257, 284, 303, 316, and 323 kHz. Presence of a fault is indicated in the frequency domain through shift in resonant frequencies along with its corresponding change in magnitude. To demonstrate the sensitivity of Fourier-based analysis, the Fourier coefficients corresponding to each resonant frequency under NF and the faulty conditions are extracted and the percentage change in the coefficient value with respect to NF condition is calculated. Table 1 indicates the sensitivity of FFT to detect SF and SHF. It is observed that the resonant frequencies 284 and 316 kHz exhibit larger variation both in terms of frequency shift and magnitude change for the case of SF. Similarly, the resonant frequencies 221 and 316 kHz show larger variation in frequency shift and magnitude change for the case of SHF. Now, to obtain the time of occurrence of these spectral features, the winding current is Gabor transformed and Fig. 11 shows the time–frequency representation of the winding current under NF condition. The time–frequency distributions of winding current with SF- and SHF-introduced condition are shown in Figs. 12 and 13.

The Gabor coefficient values are normalized with reference to 256 color scale. Comparison of the Figs. 11, 12 and 13 reveals that it is possible to identify the time of occurrence of the oscillations of our interest in the signal. The detection sensitivity of Gabor transform can be appreciated by extracting the Gabor coefficient

Fig. 9 Winding current under NF and with fault condition

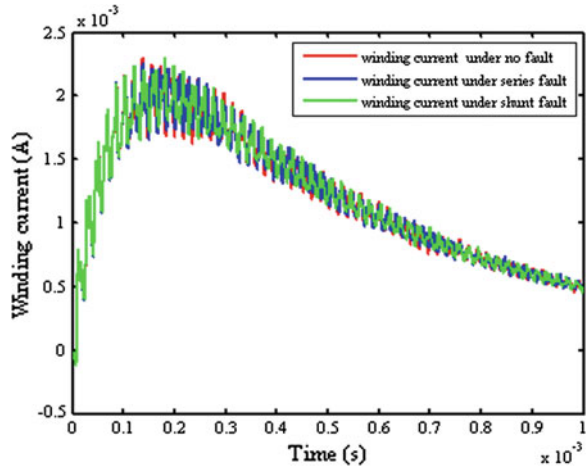
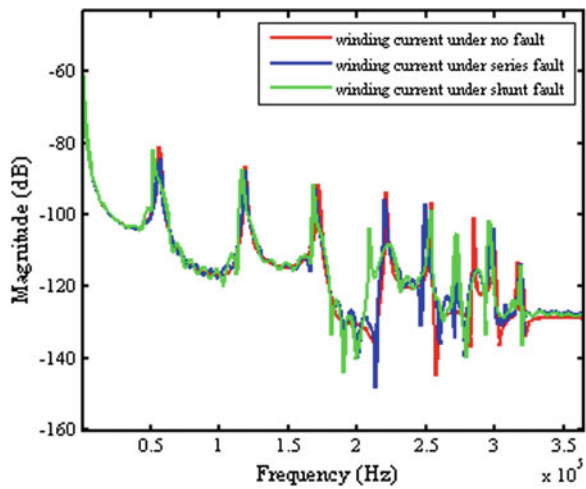


Fig. 10 Winding current in frequency domain under NF and with fault condition



values corresponding to the resonant frequencies at time instants after 0.1 ms. Table 1 indicates the detection sensitivity of Gabor transform-based signal analysis, and it is evident that Gabor transform-based analysis is more sensitive than FFT analysis since the percentage change in magnitude for SF and SHF with respect to NF are high compared to their FFT constituents.

Table. 1 Comparison of Fourier transform and Gabor transform analysis

Resonance frequency of winding in kHz			Percentage change in Fourier coefficient with respect to NF		Percentage change in Gabor coefficient with respect to NF	
NF	SF	SHF	SF	SHF	SF	SHF
58	56	52	0.61	2.11	53.12	6.51
119	119	116	1.44	1.00	27.60	69.22
175	169	168	2.86	1.46	11.33	13.10
221	220	209	2.43	11.04	16.62	22.33
257	249	253	0.39	1.90	28.86	51.12
284	272	272	12.49	4.46	26.78	21.94
303	297	295	1.69	4.42	28.79	38.83
316	307	313	10.72	9.30	12.21	41.17
323	318	318	6.89	6.72	29.44	48.75

Fig. 11 Gabor-transformed winding current under NF condition

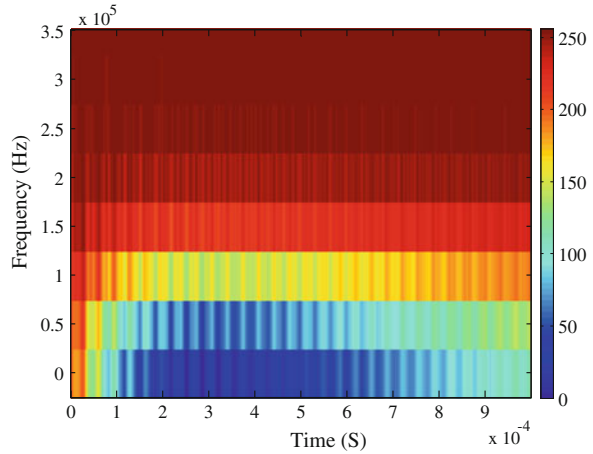
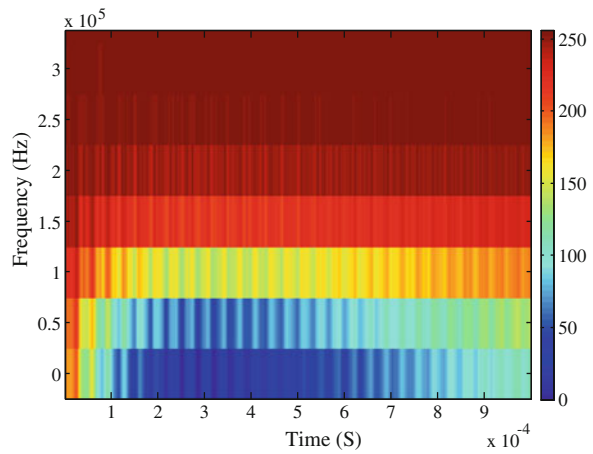


Fig. 12 Gabor-transformed winding current under SF condition



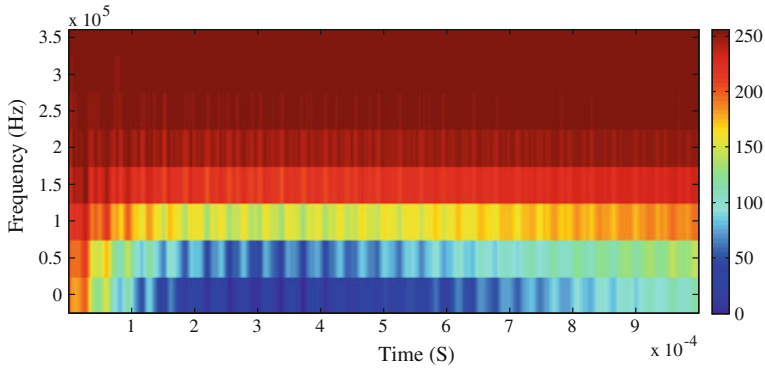


Fig. 13 Gabor-transformed winding current under SHF condition

7 Conclusion

The need and application of Gabor transform, as a signal analysis tool for the detection of impulse faults that are likely to occur in transformer, is formulated in this paper. The drawback of Fourier transform for the analysis of a non-stationary signal was demonstrated through simulation. It is justified through simulation study that the global analyzing feature of Fourier transform does not retain the time information of a signal. Hence, the Gabor transform was applied to this synthetic signal to validate the effectiveness of time–frequency transformation. Next, the application of Gabor transform for transformer winding fault detection was demonstrated through simulation of the lumped parameter model of transformer winding using circuit simulation package and MATLAB coding. The series and shunt faults have been introduced at known time, and the changes in the winding current after the time of fault condition are analyzed using Gabor algorithm. The coefficients corresponding to the characteristic resonant frequencies after 0.1 ms have been extracted, and the percentage change is computed with respect to NF condition. The results are satisfactory and encourage that Gabor transform provides better detection sensitivity than Fourier analysis. The results corresponding to SF and SHF in one section alone are provided as sample records for page limit, but the results with other fault conditions and simultaneous fault conditions have also been obtained and found satisfactory. Also it offers sufficient signal analysis strategy that could further be extended for automated impulse fault detection and classification.

References

1. IEC 60076—Part V, Power transformers—Ability to withstand short circuit. IEC, Geneva, Switzerland (2000)
2. R. Malewski, B. Poulin, Impulse testing of power transformer using the transfer function method. *IEEE Trans. Power Delivery* **3**, 476–489 (1988)

3. S. Arunkumar, V. Sandeep, S. Shankar, M. Gopalakrishnan, K. Udayakumar, V. Jayashankar, Impulse testing of power transformers—a model reference approach. *IEE Proc. Sci. Meas. Technol.* **151**, 25–30 (2004)
4. Huseyin Akcay, S.M. Islam, B. Ninness, Identification of power transformer models from frequency response data: A case study. *Signal Process.* **68**, 307–315 (1998)
5. P. Purkait, S. Chakravorti, Pattern classification of impulse faults in transformers by wavelet analysis. *IEEE Trans. Dielectr. Electr. Insul.* **9**(4), 555–561 (2002)
6. C.K. Roy, J.R. Biswas, Studies on impulse behaviour of a transformer winding with simulated faults by analog modeling. *IEE Proc. Gener., Transm. Distrib.* **141**(5), 401–412 (1994)
7. N.P. Kumar, J. Amarnath, K.D. Shrivastava, B.P. Singh, Identification of winding faults in power transformers by low voltage impulse test and neutral current method using wavelet transform approach. Annual conference report on electrical insulation and dielectric phenomenon. pp. 140–143, 2005
8. L. Satish, Short-time fourier and wavelet transforms for fault detection in power transformers during impulse tests. *IEEE Proc. Sci. Meas. Technol.* **145**, 77–84 (1998)
9. A. Bhoomaiah, G. Sreelatha, P. Appala Naidu, M. Mohan Rao, B.P. Singh, Wavelet technique for noise separation in the neutral current of a power transformer during impulse test. Annual report conference on electrical insulation and dielectric phenomena, pp. 577–580, 2005
10. N. Vanamadevi, S. Santhi, Impulse fault detection and classification in power transformers with wavelet and fuzzy based technique. *Recent Advancements in System Modeling Applications*. Lecture Notes in Electrical Engineering, vol. 188 (Springer, Berlin, 2013) pp. 261–273
11. L. Satish, A. Jain, Structure of transfer function of transformers with special reference to interleaved windings. *IEEE Trans. Power Delivery.* **17**, 754–760 (2002)
12. L.B. Almeida, The fractional Fourier transform and time frequency representations. *IEEE Trans. Signal. Process.* **42**, 3084–3091 (1994)
13. R.L. Allen, D.W. Mills, *Signal Analysis Time, Frequency, Scale and Structure*. (IEEE Press, Wiley, New York, 2004)
14. S.J. Huang, C.L. Huang, C.T. Hsieh, Application of GABOR transform technique to Supervise power system transients and harmonics. *IEE Proc. Genr. Transmn. Distrib.* **143**, 461–466 (1996)

A Modified Priority-Based Multischeduler (PBMS) for Optical Network

A. Adaikalam, S. Manikandan and V. Rajamani

Abstract In an optical network, packet scheduling is a big challenge over the communication. For the requirement to achieve high-speed communication, the network communication society needs the novel scheduler to schedule the variable length of packets in apropos method. The previous scheduling algorithms support mostly fixed size of packet, and it provides low throughput, high waiting time, and high latency. The proposed modified scheduler priority-based multischeduler (PBMS) scheduling algorithm overcomes these disadvantages through the following three methodologies: (i) prioritizer, (ii) splitter, and (iii) Round robin (RR) method; the prioritizer provides priority to the variable length of packets, queue splitter fixes the appropriate queue where the packet has been processed, and RR method assigns a fixed time unit to process the packets in queue with round drip mode. This paper proves the feasibility of our scheduler and comparative experimental results.

Keywords Priority-based multischeduler (PBMS) · Turnaround time (TT)

A. Adaikalam (✉)
ECE, MCA, ECE Department, Anna University, BIT Campus,
Chennai, Tamil Nadu, India
e-mail: adaikalam211@rediffmail.com

S. Manikandan
RMD Engineering College, Kavaraipettai, Chennai, India
e-mail: manidindigul@rediffmail.com

V. Rajamani
Vel Tech MultiTech Dr. Rangarajan Dr. Sakunthala Engineering College,
Avadi, Chennai, India
e-mail: rajavmani@gmail.com

1 Introduction

The need of scheduling is to minimize the resource starvation which will also require high throughput and low latency. In the recent years, the cell scheduling algorithms were present for fixed size of packet. Earlier, a high-performance variable length of packet scheduling algorithm was proposed for efficiently switching variable length of packets, and the performance of the proposed scheme is evaluated in terms of the packet latency. The result of that proposed work is overall performance is better than the conventional scheduling algorithms [1]. The proposed scheduling algorithm uses a prioritizer, queue splitter, and a round robin (RR) scheduler. The main advantages of using priority-based multischeduler (PBMS) packet scheduling algorithm are as follows: It can reduce switching complexity and provide high throughput, less latency, minimum waiting time, and good response time. Thus, we could meet the requirement for avoiding out-of-sequence delivery [2]. If differentiated or guaranteed quality of service is offered, as opposed to best effort communication, weighted fair queuing may be utilized. If the information is sent as it is without fixing the packet size, then the delay in time is reduced as well as the performance is increased and reduction in memory space. So it leads to the transfer of variable length of packets in networks [3].

- (A) First in–first out (FIFO) simply queues processes in the order that they arrive in the ready queue. Scheduling overhead is minimal. Throughput can be low. Turnaround time, waiting time, and response time can be high for the same reasons above that no prioritization occurs; thus, this system has trouble meeting process deadlines. It is based on queuing if all packets are queued in a single queue, and the sender transmits data packets using FIFO rule; the transmission failure of head-of-line (HOL) packet will prevent other packets in the FIFO queue from being transmitted but gives high latency, low throughput, and low response time [4].
- (B) (1) Non-priority scheduling services both control and data packets in FIFO order. We include this scheduling algorithm to contrast with the effect of giving high priority to control packets. (2) Priority scheduling gives high priority to control packets. It maintains control packets and data packets in separate queues in FIFO order. Currently, this scheme is used in most comparison studies about mobile ad hoc networks [5].
- (C) RR scheduling; balanced throughput between FCFS and SJF, shorter jobs are completed faster than in FCFS and longer processes are completed faster than in SJF. Starvation can never occur, since no priority is given. RR scheduling maintains per-flow queues, and each flow queue is allowed to send one packet at a time in RR fashion [6]. In that paper, they determined only on packet sequence delivery. Examine result of our literature work is need novel scheduler for scheduling that will support the variable length of packets. In our work, we tried to achieve overall solution like less latency and high throughput for transmitting the variable length of packets in network.

2 Proposed Scheduler

(A) Scheduling variable length of packets

In general, it is not common to use variable length of packet scheduling for switching due to high complexity on low throughput. Fixed length of packet needs the fixed amount of time to transmit or process, but it is not in variable length of packet. Here, transmission or processing time depends on packet size [7]. Queues in the network increase if the flow sends at a faster rate. Queues empty as the flow sends at a slower rate. The virtual clock approach also provides a way of monitoring whether a flow is exceeding. In general, in a variable length of packet consisting of two parts header and data, the header is 20–60 bytes in length and contains information (20–65536 bytes) essential to routing and delivery.

(B) Arrangement of packets in sequence

Emulation of output queuing (OQ) switches using combined input–output queuing (CIOQ) switches has been studied extensively in the setting where the switch buffers have unlimited capacity. They discuss setting where the OQ switch and the CIOQ switch have finite buffer capacity $B \geq 1$ packets at every output. The number of packets ahead of any packet in an input queue is incremented at most $1 + [B - 1/s - 1]$ times, irrespective of whether it is eventually dropped or transferred to the output. Thus, the buffer occupancy at any input port never exceeds $1 + [B - 1/s - 1]$ newly arriving packets to preempt packets already in the switch buffers [8].

t_n : The arrival time of the generic packet n ; d_n : the delay assigned to the generic packet n ; l_n : the duration of the generic packet n ; a_n : the time at which the first bit of the generic packet; b_n : the time at which the last bit of the generic packet; n : is scheduled to leave the switch. It is obvious that

$$b_n = a_n + l_n \quad \text{and} \quad a_n = t_n + d_n \quad (1)$$

We deem the above statements and proof are used in my works so that the processing time (T_i) is directly proportional to total number of packet size (P_i)

$$(T_i \propto P_i) \equiv T_i = K \log (P_i) \quad (2)$$

where ‘ K ’ and ‘ a ’ are nonzero constants. When the packets enter into scheduling process, time and out-of-sequence delivery have been traced back from the proposed algorithm called priority-based multischeduling (PBMS) with less latency; here, we are introducing very less amount of an interval or sleep time (S_T) between circular queues. In order to avoid rejecting all the packets that arrive out of order, rearrangement of buffer per connection may be implemented at the destination. Out-of-order packet delivery can be solved by providing enough buffer resources at destination node to store all packets that do not arrive in order.

(C) Multischeduling with packet Analyzation

Considering a set of jobs, each job is defined as a set of operations and denotes the processing time of operation. The objective is to schedule the jobs on the processors so as to minimize the make span, or maximum finish time of the schedule, subject to the following constraints. C1 All operations of job are executed on the same processor. C2 The operations cannot be simultaneously executed, for all. C3 A processor may execute at most one operation at any time instant. Constraints C1–C3 define a set of compatibility constraints among the different operations. Specifically, two operations are said to be incompatible if either operations are compatible. Incompatible operations cannot be executed simultaneously. Furthermore, the operations of a job can be executed on processor in any order. The scheduling problem defined above can be logically decomposed into two sub-problems. Because of constraint C1, the first subproblem is to assign each job to a processor, meaning that all operations will be executed. Given this assignment of jobs to processors, the second one is to schedule the operations on their assigned processors so as to minimize the make span, while also satisfying the compatibility constraint C2 and the processor constraint C3. This decomposition leads to a natural way of solving the scheduling problem [9]. Based on the idea of proposed scheduled first part, i.e., scheduling variable length of packets, scheduler allows packets in different circle queues. The demand of the queue is formed based on the total number of packet. In each circle queue having the weight limit, the arriving packets join into appropriate queue. Therefore, all the packets are processed without any packet loss.

T_i = individual packet processing time.

$CQ_{(n)}$ = number of circular queue (depends on the total number of packets between circular queue. S_T = sleep time between circular queue. $T_{cq(n)}$ = total circular queue time. Total processing time of multischeduler

$$P_{cq}(n) = \sum_{n=1}^{\infty} T_{cq}(n) \quad (3)$$

$$T_{cq}(n) = \sum_{i=0}^n T_i + \sum_{T=0}^1 S_T \quad (4)$$

S_T = Negligible amount of sleep time; time and space complexity of the algorithm is better compared to other algorithm because the sorting of packet is not processed separately.

(D) Packet reordering and scheduling

To measure the effect of packet reordering on application throughput when packets are reordered in a backbone link carrying multiple flow types [10]. Reorder rates it

is evident that the bandwidth drop rate increased significantly. The nested-DRR algorithm proposed modifies the deficit round robin (DRR) scheduler by creating a nested set of multiple rounds inside each DRR round [11]. We delve into scheduler: Scheduler receives number of packets from any network layer that will give an input of scheduler PBMS and output of PBMS given an input of optical switch based on load information packets is transmitted.

3 Proposed Algorithm: Priority-Based Multischeduler (PBMS)

```

// packets entering into the scheduler

L <- assumed total number of packet size
i,j <- Variable length of packet
for j <- i+1 to L
if ps[i] > ps[j]//packets are distributed in queue
based on the packet size
for i<-0 to L

$$Tps = \sum_{i=0}^n p v_i$$

Pv<- Variable length of packets
TPS<- Total Packet size
if Ps[i] <= limit
Limit <- Process limit of queue based on TPS
//packet enter into the new queue q1
ST<- Introducing an Interval between circular queue
if ps[i] > limit and ps[i]<=limit
//Priority Based Multi scheduling
if ps[i]>t//Packets is processed with specified time
quantum and remaining are preempted and put back in the
ready queue.
else
//Packet is completely processed
i <- i+1
wt[i] <- wt[i-1] + et[i-1]
while i<L
for i <- 0 to L
for j=i+1 to L
if pn1[i] <- pn1[j]
found <- j
if found != 0
wt[i] = wt[found] - (count * t)
count <- 0
found <-0//Total waiting time is calculated.

```

3.1 Performance Analysis

Our proposed scheduler has been simulated by using NS2 with number of processors, and to analyze the performance of PBMS, the java toolkit is used, each having different turnaround time, latency, throughput, and response time, whose result will be cited in Figs. 1, 2 and 3.

Fig. 1 Analysis of waiting time with basic scheduling algorithm

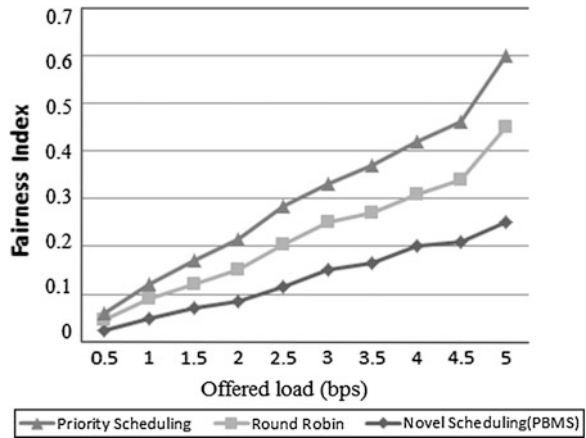


Fig. 2 Analysis of throughput with basic scheduling algorithm

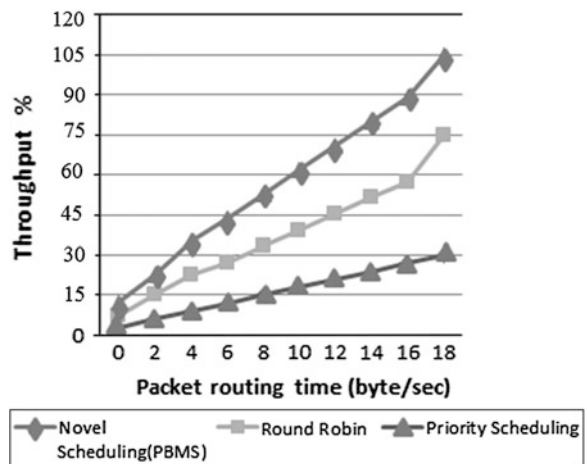
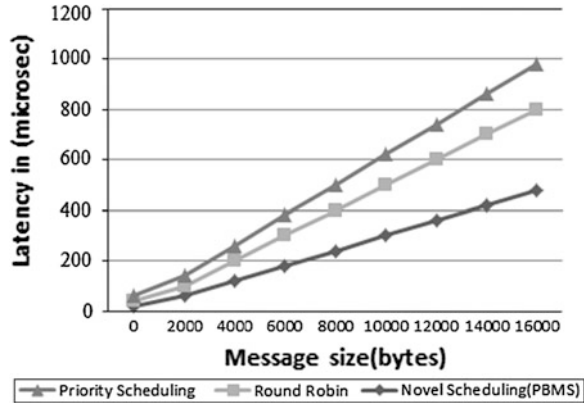


Fig. 3 Analysis of turnaround time with basic scheduling algorithm



3.1.1 Performance Metrics

We have taken the well-known metrics of scheduling algorithm like latency, throughput, turnaround time, and response time to analyze the performance of various scheduling algorithm. We formulated the problem as a delay-controlled splitting minimization problem and applied a genetic algorithm to provide a near-optimal solution for multicasting in WDM mesh network. We also measured the average waiting time of PBMS to prove our proposed algorithm is better to compare other algorithm.

4 Experimental Results

In this section, we compare the proposed PBMS algorithm to the priority and RR scheduling; the two basic scheduling algorithms are considered, where first is priority scheduling; the traffic rate is increased above 600 kbps at fairness index of 1,250, and second is slightly less compared to the PBMS.

4.1 Comparison of Waiting Time

The first and second plot shows the RR; priority scheduler waiting time for different size of packet, waiting time for each individual packet is calculated. The delay is caused by waiting time for optical packet transmission time spent in the output buffers, Fig. 1 shows the waiting time in the case of basic algorithm (priority, RR scheduling) and PBMS. Using PBMS algorithm, the waiting time is constant up to (1 bytes); this situation changes above 1.5 bytes; therefore, the PBMS has minimum waiting time when compared with other basic algorithm.

4.2 Comparison of Throughput

Routing wavelength assignment–spectrum allocation (RWSA) problem in flexible grid networks with maximum throughput. The result of throughput is defined as the number of packets that gets scheduled per time unit, for maximum throughput configuration; as illustrated in Fig. 2, the throughput provided by PBMS is from 45 to 99 %, and the maximum throughput value of the RR and priority scheduling is respectively low. Thus, we conclude that achieves maximum throughput by a PBMS.

4.3 Comparison of Latency

As shown in Fig. 3, PBMS continues to differentiate the latency in micro sec with basic scheduling algorithm (RR, priority scheduling). Here, the basic algorithm priority scheduling sends the packet 4000 bytes with the transmission time of 200–300 μ s, but in the case of PBMS, it needs 50–100 μ s for the transmission of the same packet size.

The message size (bytes) is increased, and the total time slightly increased over the basic scheduling algorithm.

5 Conclusion

In this paper, we propose a modified PBMS scheduler for packet scheduling; it supports the variable length of packets. In this paper, we tried and proved our proposed scheduler gives better results to schedule the variable length of packets, and the network-simulated and experimental results are shown above; in future, we plan to extend this work to optical switching systems.

References

1. S.H. Moon, D.K. Sung, in *High-performance variable-length packet scheduling algorithm for IP traffic*, vol. 4. Global Telecommunication Conference, GLOBECOM'01 (2001), pp. 305–701
2. S. Jaiswal, G. Iannacone, C. Diot, J. Kurose, D. Towsley, in *Measurement and classification of out-of-sequence packets in a tier-1 IP backbone*, vol. 2. Proceedings of INFOCOM (2003), pp. 1199–1209
3. I. kotuliak, T. Atmaca, in *Optical networks access node performance*. Proceedings of ELMAR-2004 (2004), pp. 418–423

4. H. Xia, Z. Zeng, in *Congestion based opportunistic packet scheduling algorithm with variable size packets support in Ad hoc Networks. Wireless communications networking and mobile computing (WiCOM), 2010*. 6th International Conference on 23–25 Sept 2010, pp. 1–5
5. M. Elhaddad, R. Melhem, in *On the emulation of finite-buffered output queued switches using combined input–output queuing*, vol. 5218. Proceedings of 22nd International Symposium DISC (2008), pp. 197–221
6. B.-G. Chun, M. Baker, Evaluation of packet scheduling algorithms in mobile ad hoc networks. *Mobile Comput. Commun. Rev.* **6**(3), 36–49 (2002)
7. Z. Hu, X.W. Li, B. Liu, Fixed length switching vs variable length switching in input queued IP switches, IP operations and management, 2004, in *Proceedings of IEEE Workshop (2004)*, pp. 117–122
8. F. Xue, Z. Pan, H. OSA, J. Yang, J. Yang, OSA Cao, K. Okamoto, S. Kamei, B. Venkatesh Akella, S.J. Yoo, Design and experimental demonstration of a variable-length optical packet routing system with unified contention resolution. *J. Lightwave Technol.* **22**(11), 2570–2581 (2004)
9. B.C. Chatterjee, N. Sarma, P.P. Sahu, Priority based routing and wavelength assignment with traffic grooming for optical networks. *Opt. Commun. Netw. IEEE/OSA J.* **4**(6), 480–489 (2012)
10. M. Laor, L. Gendel, The effect of packet reordering in a backbone link on application throughput. *Netw. IEEE* **16**(5), 28–36 (2002)
11. W.K. Jaspheer, A. Raj, Packet scheduling algorithms in different wireless networks a survey. *Int. J. Eng. Res. Technol. (IJERT)* **1**(8), 1–6 (2012)

Comparative Analysis of Digital Watermarking in Discrete Wavelet Transform and Mojette Transform

Chandini Rajeev and K.P. Girish

Abstract Watermarking is the technique to insert some kind of ownership information in any digital media like image, audio, or video. This enables the owner of the media to claim the ownership against any illegal use or claim of false ownership. In this paper, the watermarking is performed with gray image in transform domain to compare discrete wavelet transform (DWT) and discrete Mojette transform based on the quality measures like peak signal-to-noise ratio (PSNR) and mean square error (MSE). The proposed work also deals with a review of the various attacks on the watermarked image to alter the watermark, to remove, and to degrade the quality of the watermark in both cases.

Keywords Watermark · Mojette domain · Wavelet domain · PSNR

1 Introduction

The development of information technologies created the necessary demand for the copyright protection of digital media, since the media can be reconstructed and manipulated through different attacks. Watermarking describes the techniques that are used to convey information in a hidden manner. This information is required to be robust against intentional removal by malicious parties. In contrast to cryptography where the existence but not the meaning of the information is known, watermarking aims to hide the existence of the information from any potential eavesdropper altogether [1].

C. Rajeev (✉) · K.P. Girish
TIFAC CORE in Cyber Security, Amrita Vishwa Vidyapeetham University,
Coimbatore, India
e-mail: chandinir2@gmail.com

K.P. Girish
e-mail: girikalam@gmail.com

In this paper, the invisible watermarking is performed in the image, after transform it into wavelet domain and Mojette domain. In wavelet domain, the cover image is decomposed into three levels by discrete wavelet transform, so the approximation subbands with low-frequency components and 9 detail subbands with high-frequency component. Then, the watermarking components are embedded into cover image. The watermark is embedded into the selected group of wavelet transform coefficients, varying the watermark strength according to the subband level and the group where the corresponding coefficients reside [2]. After embedding the watermark, inverse wavelet transform will be calculated.

In second case, the Mojette phantoms are used as the mark and the watermarking is performed in the Mojette domain. The phantoms are image that has the projections for all the given angles at the zeros, which are filled with meaningful characters. The method calculates the block size of the phantom and the angles that are used to take projections. It uses the pixel bin correspondence at each step. When adding the image with Mojette phantoms, the projection will not change [3]. Instead of the cover image, the marking is performed in the projections.

The digital images are easy to manipulate and modify for ordinary people. This makes it more and more difficult for a viewer to check the authenticity of a given digital image [1]. To demonstrate the robustness of the proposed scheme, the peak signal-to-noise ratio (PSNR) is used to estimate the quality between watermarked images in both domains, after performing some image processing operations.

The rest of this paper is organized as follows: Sect. 2 describes the related works and Sect. 3 discusses the proposed system. The attacks on the watermarked image will be explained in the remaining section, and finally, the conclusion will be presented.

2 Related Work

The watermark is embedded in the perceptually significant regions of the image because of the fact that human eyes are less sensitive to noise in the texture region than in the smooth areas. There are different approaches for watermarking is developed based on this principle.

2.1 Spatial Domain

It is based on the substitution of LSB plane of the original image with the given watermark since LSB bits are visually insignificant. In a digital image, the data can be inserted into every bit of an image to hide messages in less perceptible parts of an image [4]. The feature of LSB includes that it is simple and easy to implement. In terms of watermark capacity, the spatial domain is the worst place to insert a

high-capacity watermark. The frequency domain offers higher capacity. It is fragile and is distorted easily by common image processing operations and other attacks.

2.2 Transform Domain

Transform domain embeds a message by modifying the coefficients of the cover image as opposed to the pixel values. In discrete Fourier transform domain (DFT), the watermark is embedded into the magnitude of the DFT coefficients and it is modified according to the human visual system [5]. In discrete cosine domain, the coefficients are selected and modified by the watermark. The image is separated into parts of different frequencies [6]. The less important frequencies are discarded in the quantization process, where actual compression of the image occurs. Only important frequencies are used to retrieve the original image in the decompression process [7]. Watermarking in the wavelet transform domain is a problem of embedding watermark in the subbands of the cover image. There are four subbands in each level of wavelet decomposition; they are low–low pass subband (LL), high–low (horizontal) subband (HL), low–high (vertical) subband (LH), and high–high (diagonal) pass subband (HH). Subsequent level of wavelet transformation is applied to the LL subband of the previous one [7].

3 Proposed System

The invisible image watermarking scheme is proposed based on DWT and Mojette transform domain. In wavelet domain, the cover image is decomposed into three levels by DWT, so the approximation subbands with low-frequency components and 9 detail subbands with high-frequency component. The watermarking components are embedded into cover image, which is decomposed by wavelet. The watermark is embedded into the selected group of wavelet transform coefficients. From these new coefficients, IDWT will be calculated. The overview of the proposed system is shown in Fig. 1.

Then, in the second case, for a given cover image, the Mojette phantoms are embedded as the watermark. First of all, design the information to be carried in the watermark. The Mojette transform is a kind of discrete Radon transform, which needs many angles. For an image at each angle, a group of projections will be acquired. Every projection is called as a bin [8].

The original image is a gray-level image of 512×512 pixels. Perform the wavelet decomposition, which results in the image which is decomposed into different levels to get the multiresolution decomposition. The threshold value will be calculated for the subband. The watermark is embedded until the whole coefficients are calculated. Then, find out the inverse wavelet transform and the output will be the watermarked image.

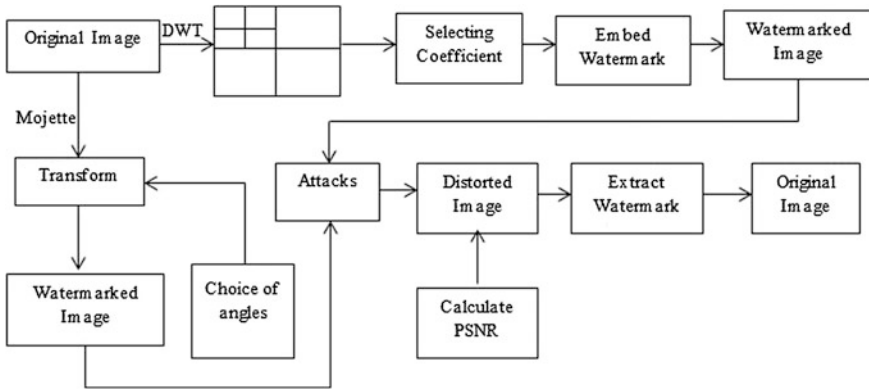


Fig. 1 Architecture of proposed system

The watermarking in Mojette domain deals with the Mojette phantoms. The watermarking is performed in the projections instead of the cover image. The watermark is constructed using the phantoms, which are images which has the proper projections for all the given angles are zeros. When the image is adding with this, there will not be any change to the projections. Depending upon the information, the watermark is constructed. Then, merge the phantom blocks and the zero blocks. If there exists any information, then the blocks are filled with the watermark, otherwise filled it by zeros.

The embedding scheme uses the cover image and the watermark information. The scheme requires calculating the angles to take the projections of the image. The Mojette phantoms and the watermark will result into the watermark. Then, embed the watermark in the cover image. The watermarked image is then divided into small blocks and takes the Mojette transform for the calculated angles. Thus, the projections will be obtained. After performing the watermarking, the PSNR values will be calculated between the wavelet-watermarked image and Mojette-watermarked image.

4 Performance Evaluation

The different watermark attacks have different coefficients to detect. Some of the attack requires one coefficient which includes Gaussian noise and salt and pepper noise. Then, check the PSNR values between the unmodified watermark image and the attacked watermarked image.

For embedding of watermark in the original image, the value of scaling factor k is varied from 1.5 to 0.6 by keeping q constant and the best result is obtained for $k = 0.98$ for both three- and two-level wavelet transform. As the value of k is decreased further to 0.2, the watermarked image becomes darker and finally



Fig. 2 Attacked watermarked images

Table 1 Calculated PSNR values

Attack	PSNR
No	46.6889
Gaussian noise	20.0888
Median filter	26.7381
Sharpened	23.5651
Salt and pepper noise	23.57
Contrast	19.1679

becomes invisible. Figure 2 shows the watermarked image using three-level discrete wavelet transform for different values of k .

Table 1 shows the results after performing watermarking in the image after wavelet transform.

5 Conclusion

The PSNR value is calculated between the unmodified watermarked image and the attacked watermarked image. The result shows that the watermarking in wavelet domain is better than that in the Mojette domain. They show the robustness to the attacks.

References

1. C.I. Podilchuk, E.J. Delp, Digital watermarking: algorithms and applications. *IEEE Signal Proc. Mag.* **18**(4) (2001)
2. C. Song, S. Sudirman, M. Merabti, Recent advances and classification of watermarking techniques in digital images, in *Proceedings of Post Graduate Network Symposium* (2009)
3. J.P. Guedon, N. Normand, The Mojette transform: the first ten years, in *Discrete Geometry for Computer Geometry* (2008)
4. G. Kaur, K. Kaur, Image watermarking using LSB. *Int. J. Res. Comput. Sci. Softw. Eng.* **3**(4) (2013)
5. A. Poljicak, L. Mandic, D. Agic, Discrete fourier transform–based watermarking method with an optimal implementation radius. *J. Electron. Imaging* **20**(3) (2011)
6. P. Sharma, S. Swami, Digital image watermarking using 3 level discrete wavelet transform, in *Conferences on Advances in Communication and Control Systems* (2013)
7. A. Kingston, F. Autrusseau, Lossless image compression via predictive coding of discrete radon projections. *Sig. Process: Image Commun.* **23**(4), 313–324 (2008) (IRCCyN lab, Elsevier)
8. G. Eason, B. Noble, I.N. Sneddon, Radon and Mojette projections equivalence for tomographic reconstruction using linear systems. **A247**, 529–551 (1999)

A Three Factor Authentication System for Smartcard Using Biometric, Visual Cryptography and OTP

Akhitha S. Kumar and K.P. Girish

Abstract In this paper, we use an authentication system that combines Biometrics, Visual cryptography and One-Time password (OTP). We propose a basis matrix construction method for generating shares in visual cryptography technique. Generated shares are in mutually orthogonal groups so that only shares from the same group can reveal the original image. The system provides the user a share of the biometric data and keeps the other share in the server itself. The security of our scheme is based on biometric verification using visual cryptography and hash of a OTP. Therefore, the scheme is appropriate to be used in applications such as smart card and device authentication systems.

Keywords Biometrics · Visual cryptography · Optical character recognition

1 Introduction

Authentication is one of the most serious security concerns in online services such as those for public utilities, shopping, banking etc. Authentication of customers has now gained significant importance due to the continuous growth in online services in addition to the increased use of alternate delivery channels by banks. Intruders use vulnerabilities in authentication software to gain unauthorized access into remote computer systems. Smartcard authentication method is currently by a token and the knowledge of a secret to establish the identity of an individual. But, a token can be lost, stolen and a secret can be forgotten or disclosed to an unauthorised person. Biometrics does not suffer from the disadvantages of the traditional methods. Smartcard systems require a method for authentication to address the

A.S. Kumar (✉) · K.P. Girish

TIFAC CORE in Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: akhithaskumar@gmail.com

K.P. Girish

e-mail: girikalam@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_72

673

problem of card theft and to enable controlled access to the functionalities of the card [1]. Authentication systems are often categorized by the number of factors that they incorporate. The three factors considered as the cornerstone of authentication are:

- Something you know (for example, password)
- Something you have (for example, an ID or a cryptographic key)
- Something you are (for example, a voice print or other biometric)

Authentication systems that incorporate all three factors are stronger than systems that only incorporate one or two of the factors. In this paper, we propose an authentication system comprising of the above mentioned factors namely user PIN, smart card and biometric. Section 2 describes related visual cryptography techniques in authentication systems. Proposed algorithm for virtual cryptography and authentication system are explained in Sect. 3.

2 Related Works

There are several approaches for biometric data security using visual cryptography [7]. This section begins with a brief description of some of the related works in visual cryptography applications and later explains the preliminaries and definitions.

One of the proposed variant of VCS application using threshold visual cryptography (k, n) scheme requires the to be split into n shares. One share is stored in the user's ID card and the remaining $n - 1$ shares are stored in the server. The value of ' k ' is a secret so that any k shares can be used to reveal the original image. From the reconstructed image, minutiae is extracted for biometric comparison [2].

Another (k, n) secret sharing scheme includes the additional capability of steganography [3, 6] and authentication. In this method, the biometric image is split into shares which are hidden in camouflage images using steganography. These camouflage images are watermarked by the use of parity bit-checking. This helps to detect false or tampered participant share before executing recovery process. The biometric image is recovered from k or more authenticated share images. Existing authentication systems using biometrics have implemented several variations of visual cryptographic technique. Our contribution is a combination of biometric, visual cryptography and one time password, thus providing additional security.

2.1 Preliminaries and Definitions

This section describes the threshold visual cryptography share generation technique and definitions. Secret sharing, introduced by Blakley and Shamir, is a scheme to encode a secret into n shares to be distributed to a set of n -participants such that

Table 1 Notations in VCS construction

Symbol	Symbol meaning
S^0	$n \times m$ matrix corresponding black pixel '0'
S^1	$n \times m$ matrix corresponding white pixel '1'
S_i^0	i th row of the matrix S^0
S_{ij}^0	Boolean 'OR' of S_i^0 and S_j^0
$H(V)$	Hamming weight of a vector V
C^0	All matrices obtained by permuting columns of S^0
C^1	All matrices obtained by permuting columns of S^1

only specified subsets of the participants may reconstruct the secret. In a (k, n) threshold secret sharing scheme, any subset of at least k -participants may reconstruct the secret. Naor and Shamir introduced VCS scheme to split a secret image into multiple shares [4]. Secret image is recovered using the concept of secret sharing.

2.1.1 Construction of Visual Cryptographic Share (VCS)

The construction of VCS is completely determined by using the concept called basis matrices. Usually two basis matrices S_0 and S_1 , are required to encode the secret image. When encoding a white (resp. black) pixel in the secret image, the process randomly permutes the columns of S^0 (resp. S^1), and then chooses the i th row of the permuted matrix to fill into the corresponding positions of the i th share. After all pixels in the secret image are encoded, n -shares are formed. Obviously, each share has the size ' m ' times as that of the original image. The parameter ' m ' is called pixel expansion. The notations used in VCS construction are shown in Table 1.

Suppose the n -participants or shares are labeled as $1, 2, \dots, n$. To share a black pixel, our application randomly chooses one of the matrices in C^0 , and to share a white pixel, it randomly chooses one of the matrices in C^1 . Given a Boolean matrix A , let A_i denote the i th row of A and let A_{ij} denote the Boolean "OR" of rows A_i and A_j . Also, let $H(V)$ be the number of 1's in a Boolean vector V . We assume that the secret image consists of a collection of black and white pixels and represent a black and a white pixel by 0 and 1 respectively.

2.1.2 Definition

A VCS, with n participants and pixel expansion m , is defined by two $n \times m$ Boolean basis matrices S^1 and S^0 , respectively, for white and black pixels, such that

- (a) S_i^1 and S_i^0 are equal up to a column permutation, i.e., $H(S_i^1) = H(S_i^0)$, $1 \leq i \leq n$,
- (b) $H(S_{ij}^1) > H(S_{ij}^0)$, $1 \leq i < j \leq n$.

For any $i < j$ the quantity ξ_{ij} is called **relative contrast** for the recovery of image by participants i and j and is given by $\xi_{ij} = 1/m \left(H(S_{ij}^1) - H(S_{ij}^0) \right)$

Let $A = \{x_1, x_2, \dots, x_n\}$ be a set with n -elements. $I_k = \{A_{k1}, A_{k2}, \dots, A_{kn}\}$ is a collection of k -subsets of A and is called a block with the property that each element of A occur in exactly k blocks of I_k . The basis matrices S^1 and S^0 are constructed from I_k as follows where S^0 is derived from S^1 by replicating any row of S^1 by n times. For example, let $A = \{1, 2, 3, 4, 5\}$ be a set with 5 elements. Then

$$I_1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}.$$

$$I_2 = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{1, 5\}\}.$$

C^{11} and C^{01} be the set of all column permutations of S^{11} and S^{01} respectively. Similarly C^{12} and C^{02} be the set of all column permutation of basis matrices S^{12} and S^{02} respectively. From each collection, we get the image by stacking any two of the shares.

3 Our Contribution

The architecture of the proposed system and the two phases in the authentication process are explained below.

3.1 Architecture

In the proposed approach, the biometric image is divided into 2 shares. One share is stored in the server database while the other share is embedded in the smart card. There are various operations performed at the server side to validate the user identity. The registration and authentication processes are described below. The proposed system uses optical character recognition (OCR) technique to interpret the user identity from the biometric image. OCR algorithm relies on a set of learned characters and optimisation algorithms [8]. Fingerprint Recognition is implemented with minutiae pattern matching algorithm [5]. A finger minutia is a fingerprint ridge ending, or a ridge bifurcation which are characteristics that make each fingerprint unique.

The general block diagram for the proposed system is given in the Fig. 1.

3.1.1 Registration

In the registration phase, the fingerprint of the user is captured by a scanner, enhanced, and converted into a template. The fingerprint is divided in small sectors, and the minutiae are extracted and stored for further comparison. The application

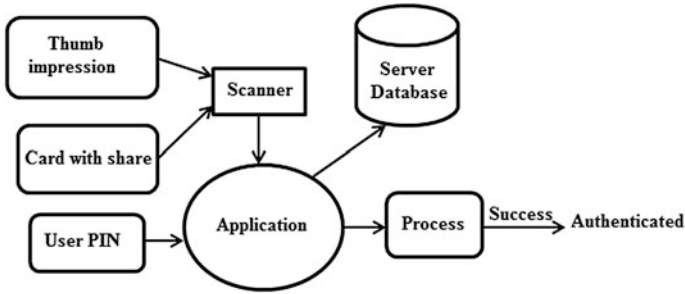


Fig. 1 Proposed Authentication system

generates a unique identity and a corresponding unique PIN number for a newly registered user. The identity sequence is inscribed on the biometric image which is split into two shares using visual cryptography. One share of the biometric image is embedded in a smart card and the other share is stored in the server database. On successful completion of the registration process, the user is provided the unique PIN number and the smart card with the VCS share.

3.1.2 Authentication

Authentication process includes reconstruction of the two shares to obtain the original biometric template. Biometric and user PIN are matched and all further sessions are encrypted to provide security. In this phase, the user inputs his thumb impression and the smart card share which are scanned by a scanner and sent to the server. Each step in the registration and authentication protocols are represented in Fig. 2.

The major steps to be done during authentication are:

- (1) The application superimposes the user share and the corresponding share from the server database to reconstruct the original biometric.
- (2) Fingerprint minutiae are extracted from the generated biometric template and are matched with the minutiae stored in the server during user registration.
- (3) If biometric matches, the unique user identity can now be analysed from the reconstructed image using OCR.
- (4) The PIN number of the user corresponding to the recognized ID is obtained and is matched with the PIN provided by the user.

The authentication phase is completed with the above verifications. The steps involved in the computation of session key are:

- (1) Server generates a random sequence of the one-time password (OTP) using seed.
- (2) This OTP sequence combined with the hash of the user share is further hashed which is used as the session key for all further transactions in the current session.

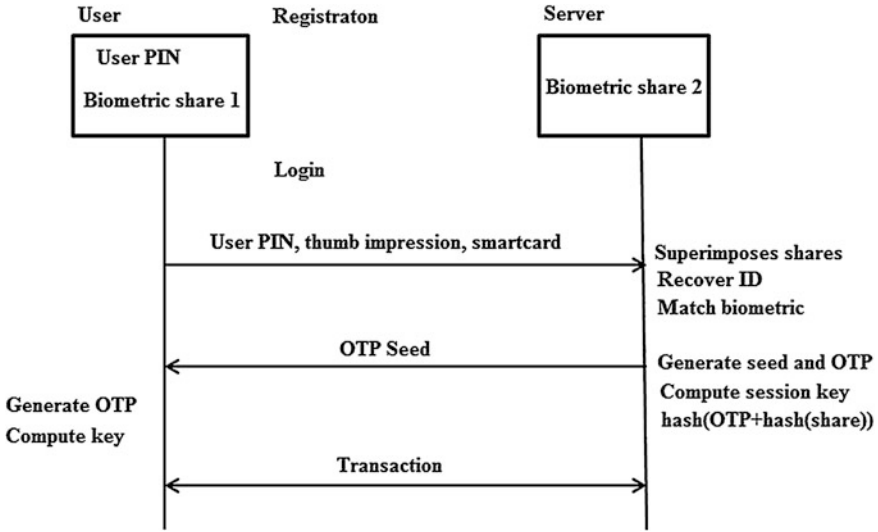


Fig. 2 Registration and Login process

- (3) User obtains the same sequence with the seed and computes the same hash sequence as the session key.

Since the same session key is generated by the user as well as the server, all further transactions are encrypted using this session key. Replay attacks can be avoided since OTP is used for session key generation. The main advantage of the proposed system is reduced computational complexity. This is achieved since the fundamental operations included in the system are matrix operations of image processing.

4 Conclusion and Future Work

This paper focuses on a biometric based authentication protocol using visual cryptography and OTP for smartcard systems. The security of our system is ensured by the combination of biometric, visual cryptography and OTP. Imprinting identity number on the biometric image before splitting it into shares allows the system to recognize the identity after image reconstruction during authentication. Tampering of any share causes the system to fail since the identity will not be visible. Further, user's thumb impression is required as one of the inputs to prevent an adversary from logging on to the system merely with the knowledge of user PIN and possession of the smart card. Also, session key generation using OTP prevents replay attacks.

The proposed idea can be enhanced by storing shares from same group in different servers and embedding both user shares which are from different groups, into the smartcard. For example, consider group A with shares A1, A2, A3, A4 etc., group B with shares B1, B2, B3, B4 etc., and 3 different servers to store the biometric data. Server 1 database stores A1 and B1, server 2 stores A2 and B2 and server 3 stores A3 and B3. Both the shares A4 and B4 are embedded in the card and the share for authentication is chosen randomly for each session. In this system, additional security is provided since biometric data is not leaked with a compromise of any single server.

References

1. L. Rila, C.J. Mitchell, Security protocols for biometrics-based cardholder authentication in smartcards, in *Proceedings of Applied Cryptography and Network Security, First International Conference, ACNS*,(2003)
2. R. Mukeshi, V.J. Subashini, Fingerprint based authentication system using threshold visual cryptographic technique, in *IEEE-International Conference on Advances In Engineering, Science And Management*,(2012)
3. Chang-Chou Lin, Wen-Hsiang Tsai.: Secret image sharing with steganography and authentication. *J. Syst. Softw.* **73**(3), 405–414 (2004)
4. M. Naor, A. Shamir, Visual cryptography, in *Advances in Cryptology Eurocrypt 1994*, vol.950 (Springer, Heidelberg, 1995), p.1-12
5. J. Abraham, P. Kwan, J. Gao, Fingerprint matching using a hybrid shape and orientation descriptor, *State of the art in Biometrics*, (2011), p.25-56
6. N. Agrawal, M. Savvides, Biometric data hiding: A 3 factor authentication approach to verify identity with a single image using steganography, encryption and matching, in *Proceedings of Computer. Vision.and Pattern Recognition Workshop*, (2009) p.85–92
7. A. Rossand, A. Othman, Visual cryptography for biometric privacy, *IEEE Transactions on Information Forensics Security*, **6**(1) (2011)
8. K. Safronov, I. Tchouchenkov, H. Worn, Optical character recognition using optimisation algorithms, in *Proceedings of Computer Science and Information Technologies*, (Ufa, 2007)

Identifying Sound of RPW In Situ from External Sources

Betty Martin, P.E. Shankaranarayanan, Vimala Juliet and A. Gopal

Abstract Over the last decade, speech recognition has been used in the field of security system, gender identification for automatic speech recognition, pattern recognition, biometrics, voice finger, dragon naturally speaking, etc. In the recent past, lots of research works are being carried out in these fields. The proposed research work also deals with one such interesting system, wherein the characteristics of the sound generated by red palm weevil (RPW) for recognition of their presence in the palm in a nondestructive way is done. For this work, a text-independent identification system makes use of feature extraction and feature matching technique. The sound of RPW recorded is compared against external sources for easy detection. Out of the several techniques available for feature extraction and comparison, mel-frequency cepstral coding (MFCC) technique has been utilized for feature extraction and the comparison is being carried out using vector quantization (VQ).

Keywords Speech recognition · Red palm weevil · Mel-frequency cepstral coding · Vector quantization · RB signal

B. Martin (✉) · P.E. Shankaranarayanan
Sathyabama University, Chennai, India
e-mail: bettymartin1205@gmail.com

V. Juliet
SRM University, Chennai, India

A. Gopal
CEERI, Chennai, India

1 Introduction

1.1 Overview of the Technique Involved in Identification of RPW

Figure 1 shows overview of the identification system employed [1–3]. The analog sound signal emanating from red palm weevil (RPW) is converted to discrete signal which undergoes mel-frequency cepstral coding (MFCC) feature extraction. Training matrices [4] for each input file are formed from available MFCC matrices obtained. The training matrices are then utilized to obtain codebooks [5] which will serve as references for each input file after vector quantization (VQ) is applied on the sound input. Now, input vectors are compared with each codebook using Euclidean distance criterion [4, 5]. The MFCC matrix is matched with all available codebooks stored. The codebook that returns the lowest quantization error belongs to the sound whose voice is contained in audio input file. Each subsystem will be discussed in brief in the forthcoming section. MFCC is comparatively better than other feature extraction methods.

The subsystem conditions the raw sound signal and prepares it for further digital manipulations and analysis. In this system, necessary signal conditioning and analog-to-digital conversion are done. The input is the raw analog sound signal which undergoes processing and outputs a digitized conditioned sound signal as one vector containing all sampled values. During analysis, the digitized vector of sample values stored will be framed into overlapping blocks. To extract a smooth output of the acoustic signal, each time the frame is scaled by a suitable window function and then transformed to frequency domain using FFT. Each block will be windowed to minimize spectral distortion and discontinuity [7, 8]. The MFCC feature extraction system is shown in Fig. 2.

Window function is given by $w(n)$, where $0 \leq n \leq N - 1$. If N is equal to number of samples in each frame, then $y(n)$ can be calculated as in Eq. (1). For $x(n)$ = input signal, $y(n)$ = Output signal, $w(n)$ = Hamming window, then the result of windowing signal is as shown below,

$$y(n) = x(n) * w(n). \quad (1)$$

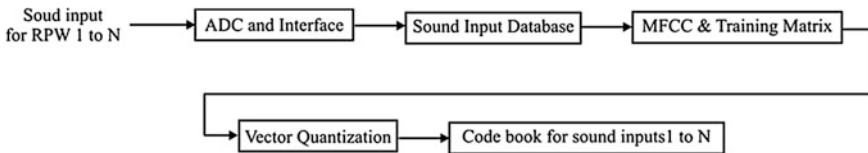
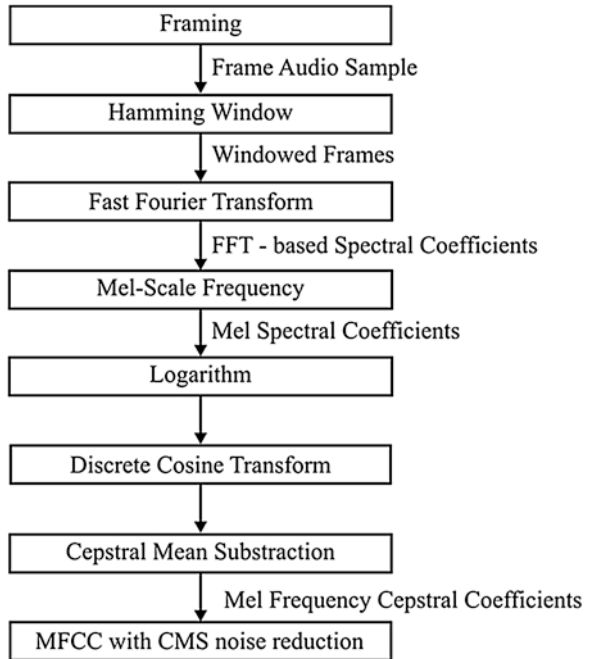


Fig. 1 Overview of the proposed identification system

Fig. 2 MFCC feature extraction system



The coefficient of a Hamming window is computed from Eq. (2)

$$w(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N}\right) \quad 0 \leq n \leq N \tag{2}$$

N represents the width, in samples of a discrete time window function $w(n)$, and n is an integer with values $0 \leq n \leq N - 1$. The window length $L = N + 1$ (+ve integer). The next process converts each frame of N samples from the time domain into frequency domain. The FFT is a fast algorithm to implement discrete Fourier transform which is defined on the set of N samples X_n as given in Eq. (3).

$$X_k = \sum_{n=0}^{N-1} (x_n e^{-j2\pi nk/N}), \quad \text{where } k = 0, 1, 2, \dots, N - 1 \tag{3}$$

X_k is the signal to be obtained in frequency domain. X_n is the signal in time domain which is to be transformed. The result after this step is referred as spectrum. In the MFCC extraction, the calculation of the mel cepstrum is same as the real cepstrum except that the mel cepstrum frequency scale is wrapped to mel scale. Mel scale depends on the study of observing pitch or frequency. Thus, while the actual frequency is measured in Hz, a subjective pitch is measured on a scale called mel scale [9]. The mel-frequency scale is a linear frequency spacing below 1,000 Hz and logarithmic spacing above 1,000 Hz. According to mel frequency, width of

triangular filters varies and so the log total energy in a critical band around the center frequency is calculated. After warping, we get number of coefficients from Eq. (4)

$$m = 2595 \log_{10} \left(\frac{f}{700} + 1 \right) = 1127 \log_e \left(\frac{f}{700} + 1 \right) \quad (4)$$

where f is actual frequency in Hz. In the final step, log mel spectrum has to be converted back to time, and the result is called mel-frequency cepstral coefficient [10]. As mel cepstral coefficients are real numbers, they may be converted to time domain using discrete cosine transform [11]. First, the time domain signal $S(n)$ is transferred into frequency domain by a M point discrete Fourier transform. The resulting energy spectrum can be represented as in Eq. (5).

$$|S(k)|^2 = \left| \sum_{n=1}^M S(n) e^{(-j2\pi nk/M)} \right|^2 \quad (5)$$

where $1 < k \leq M$, $S(k)$ is the signal to be obtained in frequency domain, and $S(n)$ is the signal in time domain. Finally, discrete cosine transform is taken on log filter bank energies, and MFCC coefficient C_n can be written as in Eq. (6).

$$C_n = \sum_{k=1}^K (\log \bar{S}_k) \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{k} \right] \quad (6)$$

where $n = 1, 2, \dots, K$. C_n = MFCC coefficient and n = integer. The number of mel cepstrum coefficient is chosen as 16. This set of coefficient is called acoustic vector. These vectors represent the characteristics of the sound input signal of RPW. The acoustic vectors extracted from input signal provide a set of training vectors, and this process undergoes clustering to build a specific vector quantization codebook using the training vectors [12].

1.2 Vector Quantization and Codebook

The acoustic vectors are obtained in the above section form input to compression system where it results in codebook. The VQ data compression was chosen for research work due to its ease in its implementation and accuracy. Vector quantization which is otherwise called pattern matching quantization is often used to map multidimensional vector space into a finite set of values of lower dimension. The transform is usually done by using codebook obtained by simulation [13]. Figure 3 shows the flow chart of the experimental setup. The threshold is the value obtained from the average of the Euclidean distance from input of 42 samples comprising of external sources and infested sound. In this work, the average distortion factor

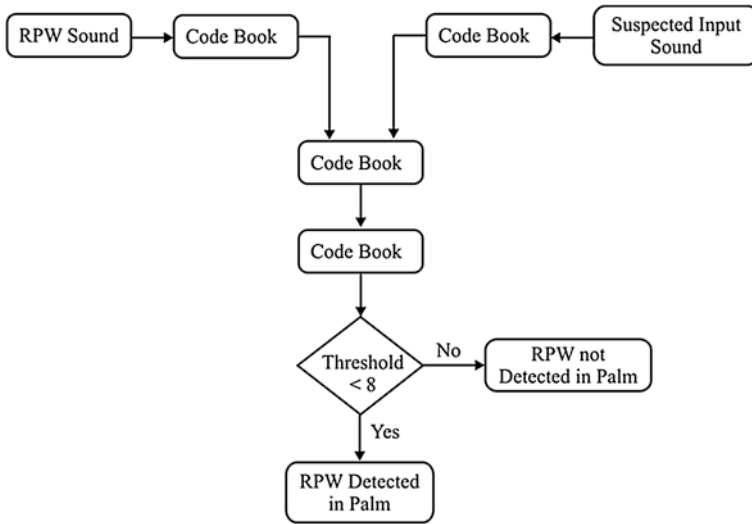


Fig. 3 Flow chart to detect RPW

obtained is 8. If the Euclidean distance is lesser than the given threshold, then the detection of RPW in palm is confirmed. If the Euclidean distance is greater than the given threshold, then detection of RPW in palms fails.

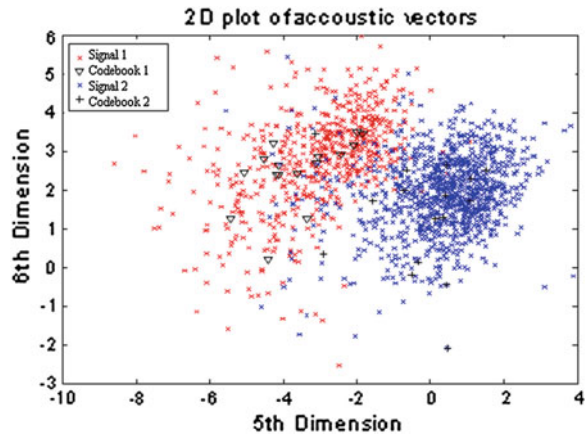
2 Experimental Methodology

The sound recorded by the digital voice recorder taped on to palms in the field is enrolled into the database and stored as inputs as described. These recordings contained sounds of human whispers, moving vehicle, wind, and RPW, if available. From the comparison, the study proved that there exists no correlation between the reference signal of RPW and the noninfested one. But the other suspected ones exhibited changes when cross-correlated as can be seen in the following discussions.

2.1 Comparison of RPW Signal with Noninfested Signal

Initially, the noninfested signal is compared with the known infested signal that is taken as reference from laboratory recording. Signal 1 represents a RPW signal in red color, and signal 2 stands for a noninfested signal in blue color. Figure 4 shows that the acoustic space used in 2D by the two signals is spread out without any overlap. The space used by signal 1 is not used by the other signals. The code words

Fig. 4 Codebook of acoustic vectors for RPW with noninfested signal (∇ symbol—RPW and + symbol—noninfested)



of each signal are separated, and there is no overlap between the two code words which are shown by inverted triangles and plus signs. This confirms that the signal 2 used in this test does not contain the RPW signal. The Euclidean distance between the two signals is calculated. From the analysis made, the threshold set is 8.000. The distance between the code words used in comparison phase of two signals is computed. The difference between the two signals is used to make recognition decision. Here, as the distance 15.5 is greater than the threshold, the verdict simply announces the absence of RPW.

2.2 Comparison of Known RPW Signal with Unknown RPW Signal

The test signals considered for comparison are between two RPW signals. Signal 1 represents the known RPW laboratory signal in red color, and signal 2 stands for another RPW signal in blue color. Figure 5 shows that the acoustic space used in 2D by the two signals is overlapping mostly. The space used by signal 1 is used by the other signals. The code words of each signal are almost overlapping, and there exists overlap between the two code words. The Euclidean distance between the two signals is calculated as 4.4597 which is lesser than the threshold, and the verdict simply announces the detection of RPW.

2.3 Comparison of RPW Signal with RB

In this work, the probability of test signals considered for the detection of RPW is RPW and RB signal taken from laboratory recording. Signal 1 represents the RPW

Fig. 5 Codebook of acoustic vectors for known RPW with unknown RPW signal (∇ symbol-RPW and + symbol—RPW)

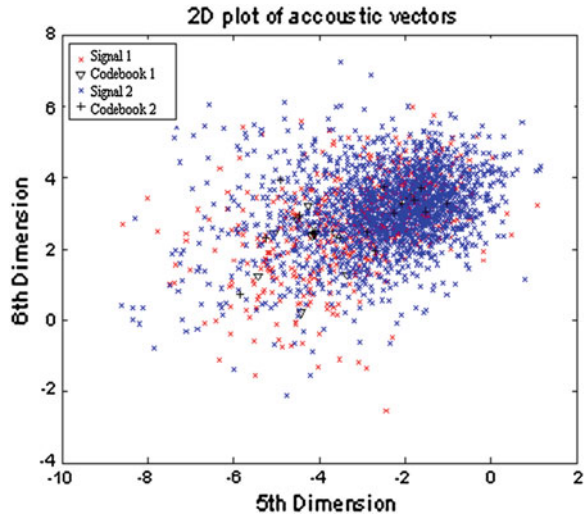
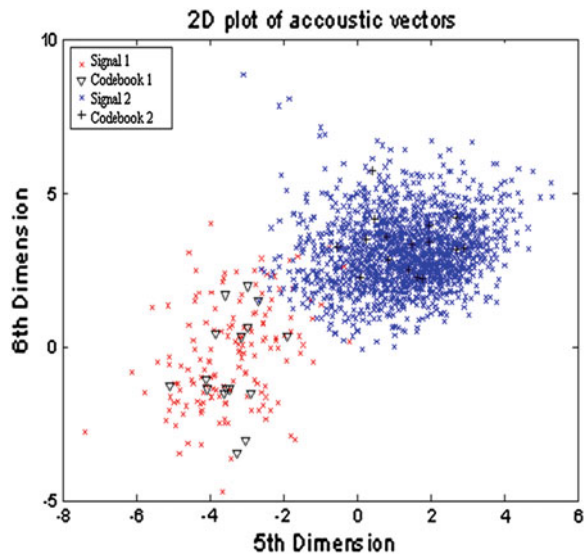
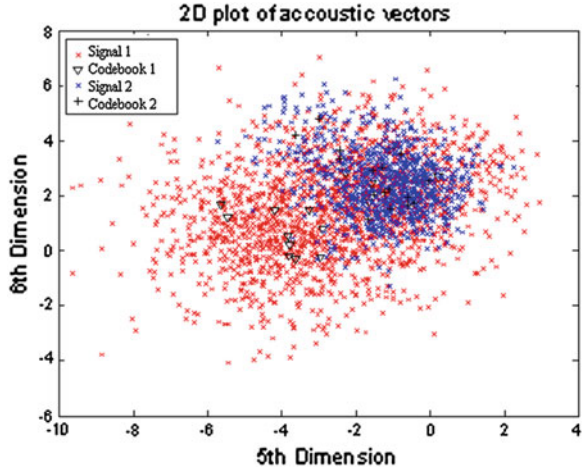


Fig. 6 Codebook of acoustic vectors for RPW with RB signal (∇ symbol-RPW and + symbol—RB)



signal in red color, and signal 2 stands for the RB signal in blue color. Figure 6 shows the acoustic space used in 2D by the two signals is spread out without any overlap. The space used by signal 1 is not used by the other signals. The code words of each signal are separated, and there is no overlap between the two code words. The Euclidean distance between the two signals is calculated as 8.5067 which is greater than the threshold. The verdict simply announces the absence of RPW.

Fig. 7 Codebook of acoustic vectors for RPW with moving vehicle signal (∇ symbol—RPW and + symbol—moving vehicle)



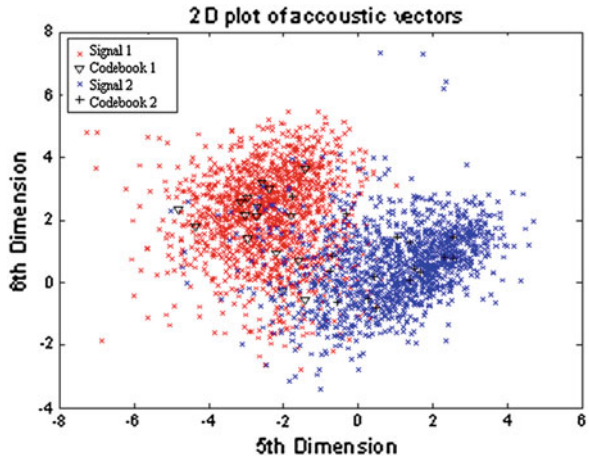
2.4 Comparison of RPW Signal with Moving Vehicle Signal

The two signals considered in the test are RPW and moving vehicle signal. Signal 1 represents the RPW signal in red color, and signal 2 stands for a moving vehicle signal in blue color. Figure 7 shows that the acoustic space used in 2D by the two signals is spread out without overlapping. The space used by signal 1 is partially used by the other signal. The code words of each signal are separated. The Euclidean distance between the two signals is calculated as 9.239 which is greater than the threshold. The verdict simply announces the absence of RPW.

2.5 Comparison of RPW Signal with Wind

The two different signals taken for test are RPW and wind signal. Signal 1 represents the RPW signal in red color, and signal 2 stands for a windy signal in blue color. Figure 8 shows that the acoustic space used in 2D by the two signals is spread out without any overlap. The space used by signal 1 is not used by the other signals. The code words of each signal are separated, and there is no overlap between the two code words which are shown by inverted triangles and plus signs. This confirms that the signal 2 used in this test does not contain the RPW signal. The Euclidean distance between the two signals is calculated as 11.10 which is greater than the threshold. The verdict simply announces the absence of RPW.

Fig. 8 Codebook of acoustic vectors for RPW with wind signal (∇ symbol-RPW and + symbol—wind)



2.6 Comparison of RPW Signal with Human Whisper

In this test, the two different signals considered are RPW and human whisper. Signal 1 represents a RPW signal in red color, and signal 2 stands for a human whisper signal in blue color. Figure 9 shows that the acoustic space used in 2D by the two signals is spread out without any overlap. The space used by signal 1 is not used by the other signals. The code words of each signal are separated, and there is no overlap between the two code words which are shown by inverted triangles and plus signs. This confirms that the signal 2 used in this test does not contain the RPW signal. The Euclidean distance between the two signals is calculated as 8.165 which is greater than the threshold, and the verdict simply announces the absence of RPW.

Fig. 9 Codebook of acoustic vectors for RPW with human whisper signal (∇ symbol-RPW and + symbol—whisper)

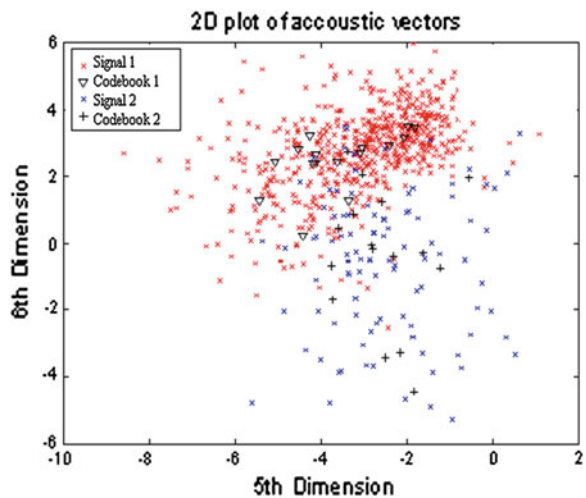


Table 1 Results of vector quantization $SP = 8$

Reference RPW signal against	Euclidean distance	Correlation
Noninfested	15.5	–
RPW	4.4597	0.55
RB	8.5067	0.28
Human whisper	8.165	0.4
Wind	11.10	<0
Moving vehicle	9.239	<0

Table 1 summarizes the outputs obtained for the different sources and that the least Euclidean distance is obtained by the unknown RPW signal against the reference signal. Hence, it is easy to identify the presence of RPW signal from other external sources.

3 Results and Discussion

The results presented above from the tests prove that the existence of the RPW pests can be identified in a nondestructive manner. However, it is necessary to confirm the detection of presence of the pest by means of a method achieved by the binary classification. Using this binary classifier, a 2×2 matrix containing true positive, true negative, false positive, and false negative can be created. This may be defined in terms of correctly identified, correctly rejected, incorrectly identified, and incorrectly rejected, respectively. Based on the analysis done on 42 samples, the 2×2 matrix created is furnished below: true positive (22), false positive (2), true negative (15), and false negative (3). From the matrix, in a collection of 42 samples, 22 acoustic activities of RPW were correctly identified and 2 samples were incorrectly identified. There were 15 activities correctly rejected and 3 incorrectly rejected. The performance index with values obtained from above definitions is listed in the Table 2.

Table 2 Performance index

Predictive parameters	Percentage
Positive predictive value (PPV)	91.6
Negative predictive value (NPV)	83.3
Sensitivity (SEN)	88
Specificity (SPE)	88.2

4 Conclusion

Analysis on recorded data of the beetle along with the external sources is performed. Sound of RPW could be recognized by mel-frequency cepstral coefficient and by vector quantization. Using this study, a simple, economical noninvasive system can be implemented to sense the beetle in situ. The result was validated by dissection of tree. Future scope can be to increase the number of samples to increase the accuracy. Another method is to try other methods of feature extraction and matching.

References

1. B. Martin, V. Juliet, A novel approach to identify red palm weevil on palms, in *The Proceedings of 2nd International Conference on Chemical, Material and Metallurgical Engineering ICCMME* (2012)
2. B. Martin, V. Juliet, Distinguishing features of RPW from RB present in palms by signal processing, in *The Proceedings of International Conference on Trends in Industrial measurements and Automation—TIMA2011* (2011), pp. 668–671
3. B. Martin, V. Juliet, Extraction of features from acoustic activity of RPW using MFCC, in *The Proceedings of International Conference on Recent Advances in Space Tech Services & Climate Change—RSTS&CC* (2010), pp. 194–197
4. Q. Li, Y. Huang, Robust speaker identification using an auditory based feature, in *ICASSP* (2002), pp. 4514–4517
5. S. Singh, E.G. Rajan, Vector quantization approach for speaker recognition using MFCC and inverted MFCC. *Int. J. Comput. Sci. Secur.* **1**(3) (2011)
6. S. Chin et al., A speaker verification system. ELEC 499A Final Report (2002), pp. 1–34
7. L. Tan, M. Kamjanadecha, Modified mel frequency cepstral coefficient, in *Proceedings of the Information Engineering Post Graduate Workshop* (2003), pp. 127–130
8. T. El Bachir, Design of an automatic speaker recognition system based on adapted MFCC and GMM methods for Arabic speech. *Int. J. Comput. Sci. Netw. Secur.* **10**(10), 45–50 (2010)
9. K.K. Paliwal, Chapter 7 Book on MFCC quantization in distributed speech recognition (Springer, Berlin), pp. 295–350
10. V. Tiwari, MFCC and its application in speaker recognition. *Int. J. Emerg. Technol.* **1**(1), 19–22 (2010)
11. W. Yutai, et al., Speaker recognition based on dynamic MFCC parameters, in *International Conference on Image Analysis and Signal Processing* (2009), pp. 406–409
12. P. Tellez, J. Savage, Isolated sentences recognition using vector quantization and neural network, in *SPECOM'2006* (2006)
13. H.B. Kekre, V. Kulkarni, Speaker identification by using vector quantization. *Int. J. Eng. Sci. Technol.* **2**(5), 1325–1331 (2010)

VNS-Based Heuristic for Identical Parallel Machine Scheduling Problem

S. Bathrinath, S. Saravana Sankar, S.G. Ponnambalam
and I. Jerin Leno

Abstract Minimization of make span and minimization of number of tardy jobs in identical parallel machine scheduling problems are proved to be NP-hard problems. Many researchers have attempted to solve these combinatorial optimization problems by employing different heuristic algorithms. While providing a satisfactory solution to the production environment for each of the above-said objectives, still remains as a challenge, most of the time, the need has been to have satisfactory solutions optimizing simultaneously the above-said two objectives. In this research work, an attempt is made to address this issue and heuristic algorithms using simulated annealing algorithm (SA) and variable neighborhood search algorithm (VNS) have been developed to provide near-optimal solutions. The developed heuristics are tested for their efficiency on a very large data sets generated as per the prescribed procedure found in the literature. Based on the results of experiments, it is inferred that the VNS-based heuristics outperforms the SA-based heuristics consistently both in terms of solution quality and consistency.

Keywords Identical parallel machine scheduling · Variable neighborhood search algorithm · Simulated annealing algorithm · Make span · Number of tardy jobs

1 Introduction

In our daily life, scheduling plays an important role for performing all activities in time. When considering a manufacturing system, scheduling deals with the allocation of resources to tasks over given time periods and its goal is to optimize one

S. Bathrinath (✉) · S. Saravana Sankar
Kalasalingam University, Krishnankoil, Virudhunagar, Tamil Nadu, India
e-mail: bathri@gmail.com

S.G. Ponnambalam
Monash University Malaysia, Sunway Campus, Bandar Sunway 46150
Selangor, Malaysia

I. Jerin Leno
National College of Engineering, Maruthakulam, Tirunelveli, Tamil Nadu, India

or more objectives [1]. In general, minimizing the make span and minimizing the number of tardy jobs also minimize the time the shop is operating which also implies to minimize the support cost as well as maximize the use of resources. In this work, we are focusing on identical parallel machine scheduling with the objective of minimizing make span and number of tardy jobs simultaneously.

Earlier Graham [2] proposed LPT; Garey and Johnson [3] proposed MULTIFIT algorithm; and Lee and Massey [4] presented the COMBINE algorithm in which the MULTIFIT has obtained the initial solution by LPT, and these were used to minimize make span. Later, Gupta and Ruiz-Torres [5] proposed LISTFIT algorithm, Lee et al. [6] presented simulated annealing approach (SA), and Liang et al. [7] proposed variable neighborhood search algorithm (VNS) for minimizing make span in single objective identical parallel machine scheduling problems. Geiger [8] presented a randomized variable neighborhood search to solve multi-objective flow shop scheduling problem. Briand and Ourari [9] presented the computation of good quality lower bound and upper bound, and Yin et al. [10] proposed branch and bound procedure with two agents for minimizing number of tardy jobs in the production scheduling.

In this work, an attempt has been made to develop a methodology to solve two different objectives by considering them as a single objective using combined objective function (COF). The proposed algorithm is experimented with benchmark problems available in the literatures and found consistent.

2 Problem Description

A set $N = \{J_1, J_2, \dots, J_i, \dots, J_n\}$ of n jobs are to be scheduled on m identical parallel machines indexed by a set $M = \{M_1, M_2, \dots, M_j, \dots, M_m\}$. Each job $J_i \in N$ has only one operation and a deterministic processing time (p_i) for the operation which includes any setup time required. Also, each job J_i has a fixed due date d_i , before which the job is expected to be completed. The objective of this problem is to find the optimal schedule $S = \{S_1, S_2, \dots, S_j, \dots, S_m\}$ where S_j is a subset of jobs assigned to machine M_j , such that $\max \{C_1(S), C_2(S), \dots, C_j(S), \dots, C_m(S)\} = C_{\max}(S)$ is minimum where

$$C_j(S) = \sum_{p_i \text{ of } J_i S_j} p_i \quad (1)$$

and also to minimize the number of tardy jobs

$$N_t = \sum_{j=1}^n U_j \quad (2)$$

where U_i will assume the value 1 for the job J_i if the job J_i is completed in time, that is, if the due date d_i for the job J_i is smaller than completion time C_i , otherwise 0.

Hence, the COF is formulated as below:

$$\text{Minimize COF} = \delta \times C_{\max}(S) + (1 - \delta) \times N_T \quad (3)$$

where δ and $(1 - \delta)$ are weightage factors assigned to the make span and number of tardy jobs, respectively, $0 < \delta < 1$.

Assumption All jobs are independent and available at time zero, and each machine can process only one job at time and preemption of job is not permitted.

3 Proposed Methodology

In this work, VNS has been implemented to obtain optimal solution using COF presented in Eq. 3.

3.1 Proposed Variable Neighborhood Search Algorithm (VNS)

VNS algorithm is a local search algorithm which employs a systematic change of neighborhood where it is not constrained with local optimum when attempting to reach global optimum. To implement VNS, several perceptions have to be considered. For instance, (1) a local minimum with respect to one neighborhood structure is not necessary to be a local minimum for another neighborhood structure, (2) a global minimum is considered to be a local minimum with respect to all possible neighborhood structures, and (3) local minima with respect to one or several neighborhoods are relatively close to each other. This paper is mainly based on the concept of Geiger's algorithm [10], and the detailed study is designed according to the characteristics of bi-criteria identical parallel machine scheduling problem.

3.1.1 Initialization

A set of neighborhood structures (N_k) and stopping criterion are first determined during the initialization phase. In this process, two different neighborhood structures ($k = 1, 2$) are defined and the stopping criterion depends on the predetermined maximum number of evaluations. The initial sequence of jobs is obtained from Grahams (1969) longest processing time (LPT) algorithm. The assignment of jobs to the machine is performed by first available machine (FAM) rule where it is used to assign the unscheduled jobs to the available machine at earliest time among all others.

3.1.2 Neighborhood Structure

In this section, two different types of neighborhood structure are illustrated. In the first neighborhood structure $N_k = 1$, any two jobs in the sequence are selected randomly and both the jobs are swapped to get the new sequence. For instance, there are six numbers of jobs, given the current job sequence as 3-4-2-6-5-1 as well jobs 3 and 1 swapped; the resulting sequence will become 1-4-2-6-5-3. In the second neighborhood structure $N_k = 2$, any one of the job is selected randomly and is inserted in different position to get the new sequence. For instance, given the current job sequence as 3-4-2-6-5-1, job 1 is selected and inserted to the first position of the sequence which results in the new job order as 1-3-4-2-6-5.

3.1.3 Pseudocode for the Proposed Variable Neighborhood Search Algorithm

- Step 1: Initialization of the parameters: number of jobs, number of machines, weightage values for make span and number of tardy jobs, stopping condition.
- Step 1.1: Select the two different sets of neighborhood structures N_k , where $k = 1$ and 2 which are used in the search process.
- Step 1.2: Find the initial job sequence by LPT. The initial sequence of jobs is allocated to the machines by means of FAM rule.
- Step 1.3: Choose the stopping condition as the maximum number of times the search process has to be executed, i.e., 1,000 times.
- Step 2: Based on the initial sequence generated by step 1.2, find the combined objective fitness function COF1 using the relation given in Eq. (3) and set that sequence as current.
- Step 3: Set $x = 1$;
- Step 4: Generate the new sequence of jobs by choosing any one of the following neighborhood structures randomly.
- Step 5: For the newly generated sequence of jobs, apply the FAM rule to obtain the combined objective fitness function COF2 using the Eq. (3) and go to step 6.
- Step 6: If $\text{COF2} < \text{COF1}$, then set $\text{COF1} = \text{COF2}$ and set newly generated sequence as current sequence.
- Step 7: Increment the value of x as one and check $x > \text{Stopping Condition}$. If yes, go to step 8, otherwise go to step 4.
- Step 8: Obtain the optimal sequence from the current sequence.

Table 1 Summary of computational experiments

Experiment names	<i>m</i>	<i>n</i>	<i>p</i>
E1	3, 4, 5	2 <i>m</i> , 3 <i>m</i> , 5 <i>m</i>	U(1, 20), U(20, 50)

4 Computational Experiments

Gupta and Ruiz-Torres [5] proposed several uniform distribution generating schemes for generating problem sets to conduct experiments as shown in Table 1. Several computational experiments have been carried out using *m*, *n*, and range of uniform distribution used to generate processing times. For generating the due dates, it has to be lie between the processing time as well as the average value of the processing time with the tightness. The tightness is found to be 1.10 from [11] which means the due date is at most as bigger as the 10 % of the average processing time. The due dates are calculated using the relation taken from the literature as given by

$$d_j = \text{random} \left(p_j, \left(\sum_{j=1}^n p_j / m \right) \times \text{TG} \right) \tag{4}$$

where TG is the tightness.

The algorithms are coded in MATLAB R2010a and are executed in Intel® Core™ i5 CPU M430 @ 2.27 GHz 2.27 GHz with 4 GB RAM.

5 Results and Discussion

The relative performance of the one algorithm with respect to another algorithm is calculated for the proposed meta-heuristics. For example, a value of *c/d* in VNS/SA means that among the one hundred problems, there are *c* number of problem instances for which VNS yields a better solution than SA, *d* problems for which SA performs better than VNS, and 100-*c*-*d* problems for which both VNS and SA yield the same solution. Table 2 shows the relative performance of the meta-heuristics for the experiment level E1, and its weight value has been taken as 0.5. This table also shows the computation time in seconds for the proposed VNS and SA. From this table, column 12 shows the relative performance measures of VNS over SA.

From these results, it is clear that VNS is superior to SA for the identical parallel machine scheduling problem with bi-criteria optimization of minimizing make span and minimizing number of tardy jobs simultaneously.

Table 2 Results for experiment E1 when $\delta = 0.5$

m	n	P_{gen}	SA			VNS			$\delta = 0.5$		
			Min	Mean	Max	CPU time	Min	Mean	Max	CPU time	VNS/SA
3	6	U(1, 20)	5.1	8.71	11.7	0.15	3.5	6.82	10.7	102.2	82/5
	9		8	13.06	18.8	0.17	7.1	11.10	15.2	130.7	94/2
	15		17.7	21.98	26.7	0.25	12.5	18.25	22.9	169.8	99/0
4	6	U(20, 50)	18.4	24.21	30.3	0.16	17.2	22.58	25.4	119.3	62/10
	9		29.4	35.96	42.4	0.18	26.5	33.8	40.3	141.06	79/20
	15		49.3	59.17	66.1	0.22	44.3	57.4	64.5	175.6	74/24
5	8	U(1, 20)	6.2	9.55	13.9	0.19	4.5	7.81	11.5	150.81	81/11
	12		9	14.29	19.7	0.17	8.1	11.5	17.0	166.41	97/2
	20		18.1	23.96	28.8	0.3	15.0	20.06	25.2	210.97	97/0
6	8	U(20, 50)	19.4	24.7	29.5	0.19	17.9	24.1	28.1	135.63	64/25
	12		30.3	36.6	44.5	0.21	28.1	31.1	40.1	170.58	57/40
	20		49.2	59.91	68.8	0.25	50.6	58.3	66.2	180.4	65/34
7	10	U(1, 20)	5.4	10.18	13.9	0.21	4.4	9.0	12.6	162.71	80/14
	15		10.6	15.34	20.6	0.24	8.5	12.4	18.2	190.13	94/5
	25		19.3	26.03	31.4	0.36	15.0	20.6	29.5	285.1	94/2
8	10	U(20, 50)	21.3	25.92	31.7	0.2	19.8	24.8	30.2	173.54	50/40
	15		31	37.85	45.4	0.25	30.1	35.8	44.4	164.4	55/44
	25		53	62.01	70.2	0.36	53.0	60.7	70.4	265.9	68/30

6 Conclusion and Scope for Future Research

The effectiveness of the proposed meta-heuristics is analyzed by the number of test problems taken from the literatures, and their weight values have been considered. The relative performance of the each meta-heuristics over other has been tested. Some of the major findings from the present work have been stated as follows.

COF proposed for the identical parallel machine scheduling problem confirms to be an effective and comprehensive measure, as multiple decisions are frequently involved in this dynamic and competitive environment. The optimal solutions can be obtained by changing the weight values.

For considering the bi-objective identical parallel machine scheduling problem, the meta-heuristics has been proposed and analyzed the effectiveness of the proposed algorithm and found that the VNS gives the better schedule with minimum performance measures considered.

The outcome of this paper leaves scope for further research toward employing a local search mechanism to further optimize the optimal solution. The proposed algorithm can be extended toward non-identical parallel machines. Multi-objective optimization algorithms such as NSGA-II and MACO can also be considered to produce a pareto-optimal front.

References

1. M.L. Pinedo, *Scheduling: Theory, Algorithms and Systems* (Springer, Berlin, 2012)
2. R.L. Graham, Bounds on multiprocessor timing anomalies. *SIAM J. Appl. Math.* **17**, 416–429 (1969)
3. M.R. Garey, D.S. Johnson, *Computers and intractability: a Guide to the Theory of NP Completeness* (1979)
4. C.Y. Lee, J.D. Massey, Multiprocessor scheduling: combining LPT and MULTIFIT. *Discrete Appl. Math.* **20**, 233–242 (1988)
5. J.N.D. Gupta, A.J. Ruiz-Torres, A LISTFIT heuristic for minimizing makes span on identical parallel machines. *Prod. Plannand Control* **12**, 28–36 (2001)
6. W.C. Lee, C.C. Wu, P. Chen, A simulated Annealing approach to make span minimization on identical parallel machines. *Int. J. Adv. Manuf. Technol.* **31**, 328–334 (2006)
7. Y.C. Liang, Y.M. Hsiao, C.Y. Tien, Metaheuristics for drilling operation scheduling in Taiwan PCB industries. *Int. J. Prod. Econ.* **141**(1), 189–198 (2013)
8. M.J. Geiger, Randomised variable neighbourhood search for multi objective optimisation. In *Proceedings of EU/ME Workshop: Design and Evaluation of Advanced Hybrid Meta-Heuristics*, pp. 34–42 (2004)
9. C. Briand, S. Ourari, Minimizing the number of tardy jobs for the single machine scheduling problem: MIP-based lower and upper bounds. *RAIRO-Oper. Res.* **47**(1), 33–46 (2013)
10. Y. Yin, C.C. Wu, W.H. Wu, C.J. Hsu, W.H. Wu, A branch-and-bound procedure for a single-machine earliness scheduling problem with two agents. *Appl. Soft Comput.* **13**(2), 1042–1054 (2013)
11. S. Bathrinath, S. Saravanasankar, S.G. Ponnambalam, B.K.V. Kannan, Bi-objective optimization in identical parallel machine scheduling problem. In *SEMCCO 2013*, vol. 8297, pp. 377–388 (2013)

Green Algorithm for Virtualized Cloud Systems to Optimize the Energy Consumption

P. Prakash, G. Kousalya, Shriram K. Vasudevan and K.S. Sangeetha

Abstract In recent days, most of the cloud users request data center in the cloud environment by applying an exhaustive data-centric workflows which leads to the major energy consumption. The major energy breaks out from the data center and makes way to CO₂ emission which impacts the global warming. In this paper, we introduce optimized energy utilization in deployment and forecast (OEUDF) for data-intensive workflows in virtualized cloud systems which help to reduce the energy in the cloud workflow environment. In this approach, initially, we compute the optimal data-accessing energy path (ODEP) which helps us to deploy and configure the virtual machines; secondly, it computes the rank, according to that it will schedule the workflow activities in the cloud environment. If any unscheduled activities are in the submission pool, then OEUDF finds the suitable virtual machine and reconfigures the data center by minimizing the energy utilization. The experiment result indicates that the proposed algorithm gradually reduces the energy consumption.

Keywords Energy utilization · Virtual machine · Workflow · Data-centric and cloud environment

P. Prakash (✉) · S.K. Vasudevan

Department of Computer Science and Engineering, Amrita University, Coimbatore, Tamilnadu, India

e-mail: npprakash@gmail.com

S.K. Vasudevan

e-mail: shriramkv@gmail.com

G. Kousalya

Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore, Tamilnadu, India

e-mail: kousir@gmail.com

K.S. Sangeetha

Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, Tamilnadu, India

e-mail: sangee4u@gmail.com

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_75

701

1 Introduction

Cloud computing is upcoming paradigm providing a highly obtainable, resilient, and elastic platform for various distributed computing applications [1, 2]. In cloud environment, user's quality of service (QoS) was affected by data centers performance. Data centers play vital role in management of resources and service stipulation [3]. To keep the agreeable QoS, many data centers up their machines throughout the day and night. It impacts the energy consumption of data centers rapidly. Furthermore, it increases the running cost of the cloud provider. In the darker side, this may lead into CO₂ emanation harms [4]. As a result, the modern-day cloud provider needs a revamp of their data center architecture. There are different types of workflow applications such as computation and data-intensive workflows. These applications attract ample amount of research in the field of energy aware scheduling. Cloud computing has several advantages over other performance computing such as implementation of virtual machines, on-demand resource management, and expandable service capability. On the other hand, it may raise some challenges when the cloud environment tries to implement the energy aware algorithm for data-intensive workflows. Improper data transfer between the nodes may lead to the degradation of the performance with high energy expenditure [5]. Reducing the energy of these two kinds at the same time is very complex [6]. To address the above issues, we proposed a new scheduling algorithm, OEUDF for data-intensive workflows in virtualized cloud systems. The proposed scheduling algorithm consists of two phases: deployment of VM machines and scheduling of workflows. The rest of this paper is organized as follows. In Sect. 2, we summarize the related work. In Sect. 3, we present the formal definitions on workflow scheduling and the problem description. In Sect. 4, the proposed algorithm is presented. In Sect. 5, we analyze the scheduling model and present the result details. Finally, Sect. 6 concludes the paper with a brief discussion of our future work.

2 Related Work

Initial work related to energy aware management is dedicated to mobile devices with the aim of improving the lifetime of a battery [7]. Subsequently, the focuses have been moved to data centers [8] and virtualization of machines in cloud environment. The authors Nathuji and Schwan [9] have explained the architecture of energy management system for virtualized data centers where resources are alienated into local and global polices. On the local level, the system takes the major advantages of guest operating system's power executive approaches. Applying a live migration of VMs, the VMs are consolidated that will be handled by global polices. Nevertheless, the global policies are not addressed the requirements of QoS. In contrast, our work spotlights the VM allocation by considering energy level of each VMs.

The scheduling of workflow is going to be considered as one of the hard problems, but few of them can be reduced in polynomial time under some assured

conditions. For example, Benoit et al. explained an algorithm for scheduling pipeline workflow by considering some constraints related to energy [10] which runs polynomial time complexity.

Dynamic voltage and frequency (DVFS) [11] based algorithms are mostly used in energy aware scheduling. The reason for that is processor contributes major portion of the total energy consumption. For example, the customary Max–Min algorithm is united with DVFS mechanism and formed a new method called MMF-DVFS heuristic by Rizvandi et al. [12].

Yu and Shi [13] explained a stratagem for multiple workflows. The strategy reveals that it collects all the tasks and checks whether it is ready or not. If the tasks are ready, then rank the task which helps to decide the order in which tasks are going to be executed. On the other hand, if any task which comes into the workflow has very low rank, then the higher rank task needs to wait for so long time, and this becomes true if the cloud is dynamically scalable. The author considered only the running time of a tasks without worrying other factors such as cost and QoS.

In recent times, many researchers put their efforts into understanding the mutual relation between the characteristics of application and the total energy consumption. The interchange between the energy utilization and the structure of the application [14] was clearly expressed by Cho and Melhem. Kang proposed an I/O group scheduler [15], which delegates the I/O requests to the dedicated VM instance, which helps to reduce the cost spending in MapReduce applications.

In conclusion, the above-mentioned related works have their own benefits within their limitations, but none of them are specifically developed for multiple workflows. Based on the survey, our approach takes both the structural feature of the application and the energy aware approaches.

3 Problem Description

There are so many applications, and middleware is mixed up in the cloud environment as depicted in Fig. 1. At first, users can submit their description of the workflow and it moved into the portal where the pools of workflow description are placed. After that, workflow management machine is responsible for transferring the workflow from user interface into cloud environment. Then, the scheduling of the VMs is taken care by the hypervisor which is responsible for resource management.

The directed acyclic graph is used to represent the workflow. While running a data-intensive workflow, the data which is presented in the activity node is transferred from a storage node into running node. The resultant data can be placed back into the respective storage node. So there is need for mapping the computing activities and the resources with respect to the storage nodes. Scheduling data-accessing tasks and computing task are going to be tough in the data-centric workflows. So we explained VM power model initially and then energy consumption model for the same by using DAG scheme of scheduling.

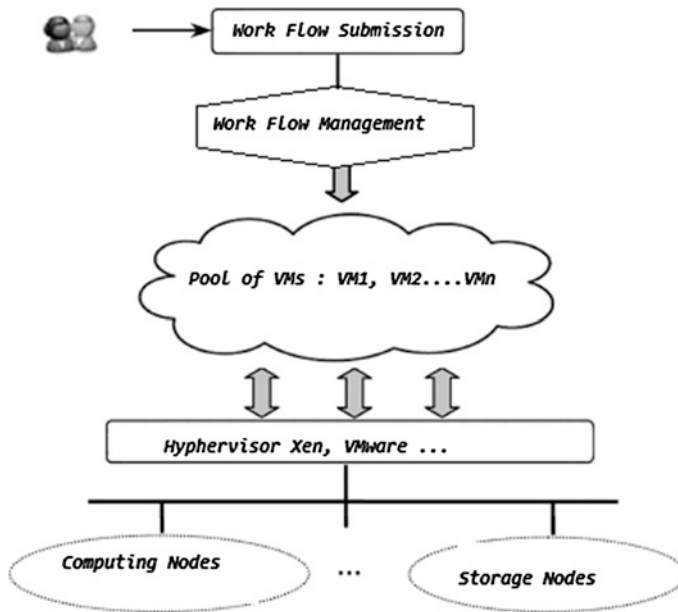


Fig. 1 Cloud workflow environment

4 Energy Utilization Replica of Data-Intensive Workflows

Based on the problem description, there are two different phases which include deployment of VMs and scheduling with respect to DAG. The numbers of VM instances are mapped into physical resources. So the power model can be related to virtual machine not with respect to the physical machine. The scheduling phase consists of assigning a workflow into the active VMs and updating the total energy consumption due to the running time variant of each VM. For executing these two phases, we explained VM power model initially and then energy consumption model for the same by using DAG scheme of scheduling.

4.1 Energy Consumption Model

The power expenditure of machine consists of stagnant part P_s and vibrant part P_v . P_s is the permanent power consumption for keeping the machine in running state even there are no jobs that are executing on it, while P_v is related to the dynamic utilization of power-consuming components. Typically, the power model of a physical machine is formulated as

$$P(t) = P_s + \sum_{k \in \Omega} P_i(t) \quad (1)$$

where $P_i(t)$, the vibrant power utilization of module i is the set of power-consuming components, which often consists of central processing unit, graphical processing unit, memory, disk, etc.

The total energy utilization of a given scheduling scheme shown spending under a given scheduling scheme is shown as

$$\text{Energy}(G, K) = \sum_{i=1}^n [\text{Energy}_c(n_i, K_{i,i',i''}) + \text{Energy}_d(n_i, K_{i,i',i''})] \quad (2)$$

where $n_i \in N$ and K can be denoted as $\{K_{i,i',i''} \mid i \in n, i' \in m, \text{ and } i'' \in k\}$.

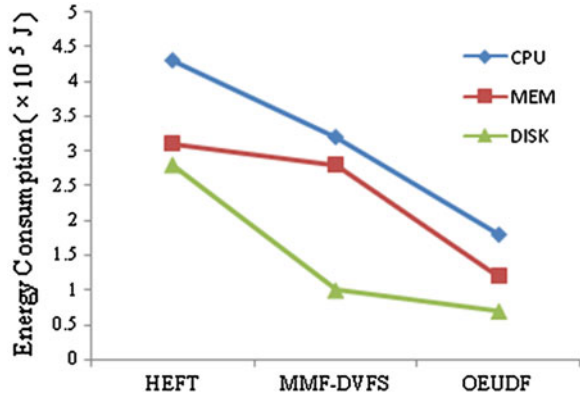
4.2 Algorithm

Input: original DAG: $G = (N, E)$, bandwidth matrix: BM , power model of all physical nodes. G is a directed graph where N represents set of activities representing with respect to the computing tasks and E is the set of edges and each $e_{i,j}$ indicates that e_j depends on e_i .

Start

1. $nset := \{\}$;
2. $index := 0$;
3. **for** each n_i do
4. Find an s_j that satisfies optimized OEDP(n_i);
5. **If** exists a VM that uses s_j then
Continue;
6. **end if**
7. $index := index + 1$;
8. Create VM_{index} and add it to $nset$;
9. **end for**
10. Sort the VMs in the $nset$ by arranging ascending order.
11. Compute $rank(n_i)$ for all the activities by traversing G from n_{end} to n_{start} .
12. Sort the activities according to the rank.
13. **while** any unscheduled activities do
From the list select the task n_1 (First Task).
for each VM_k do
Compute $EBT(K_{i,i',i''}) + T_{exec}(n_i)$;
end for
After inserting the task assign n_1 to $VM_{i'}$ to meet the requirement of n_1 which minimizes $EBT(K_{i,i',i''}) + T_{exec}(n_i)$;
if $s_v < d_i^{out}$ then
Reconfigure the virtual disk to meet the requirements.
end if
14. **end while**

Fig. 2 Results by comparing HEFT, MMF-DVFS, and OEUDF



EBT is earliest Begin Time is main idea behind for most scheduling algorithm, it is defined as earliest beginning time of an activity n_i . ODEP is used to calculate the minimal energy utilization from n_{start} to the current activity.

5 Experiment Setup and Result

We set up an experiment environment with the help of CloudSim. The proposed algorithm was compared with different other algorithm with respect to the energy consumption. We compared the results which are obtained from OEUDF algorithm with HEFT and MMF-DVFS. The following Fig. 2 illustrates how the energy was utilized in different components of the cloud environment. This graph clearly shows that OEUDF yields better result compared to some of the existing algorithms.

6 Conclusion

In this paper, we have addressed the energy aware scheduling of data-intensive workflow in virtualized cloud systems using OEUDF. We defined the problems which optimize the energy consumption by meeting the QoS requirements. The experimental results show that the proposed algorithm outperforms the entire existing algorithm. In the future, we plan to integrate some workload aware mechanism, VM migration, and storage of VMs across the cloud systems with efficient manner. Furthermore, we are planning to build an efficient resource scheduler for dynamic configuration of the VM at different levels of the cloud environment.

References

1. D.W. Sun, G.R. Chang, S. Gao, L.Z. Jin, X.W. Wang, Modeling a dynamic data replication strategy to increase system availability in cloud computing environments. *J. Comput. Sci. Technol.* **27**(2), 256–272 (2012)
2. M. Sedaghat, F. Hernandez, E. Elmroth, Unifying cloud management: Towards overall governance of business level objectives, in *Proceedings of the 11th IEEE/ACM International Symposium Cluster, Cloud and Grid Computing* (2011), pp. 591–597
3. A. Iosup, N. Yigitbasi, D. Epema, On the performance variability of production cloud services, in *Proceedings of the 11th IEEE/ACM International Symposium Cluster, Cloud and Grid Computing* (2011) pp. 104–113
4. S.K. Garg, C.S. Yeob, A. Anandasivamc, R. Buyyaa, Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers. *J. Parallel Distrib. Comput.* **71**(6), 732–749 (2011)
5. G. Juve, E. Deelman, G.B. Berriman, B.P. Berman, P. Maechling, An evaluation of the cost and performance of scientific workflows on Amazon EC2. *J. Grid Comput.* **10**(1), 5–21 (2012)
6. W. Fang, X. Liang, Y. Sun, A.V. Vasilakos, Network element scheduling for achieving energy-aware data center networks. *Int. J. Comput. Commun. Control* **7**(2), 241–251 (2012)
7. R. Neugebauer, D. McAuley, Energy is just another resource: Energy accounting and energy pricing in the nemesis OS, in *Proceedings of the 8th IEEE Workshop on Hot Topics in Operating Systems* (2001), pp. 59–64
8. E. Pinheiro, R. Bianchini, E.V. Carrera, T. Heath, Load balancing and unbalancing for power and performance in cluster-based systems, in *Workshop on Compilers and Operating Systems for Low Power* (2001), pp. 182–195
9. R. Nathuji, K. Schwan, Virtualpower: Coordinated power management in virtualized enterprise systems. *ACM SIGOPS Operating Syst.* **41**(6), 265–278 (2007)
10. A. Benoit, P.R. Goud, Y. Robert, Performance and energy optimization of concurrent pipelined applications, in *Proceedings of the 24th IEEE International Symposium Parallel and Distributed Processing* (2010), pp. 1–12
11. D. Zhu, R. Melhem, B.R. Childers, Scheduling with dynamic voltage/speed adjustment using slack reclamation in multi processor real-time systems. *IEEE Trans. Parallel Distrib. Syst.* **14** (7), 686–700 (2003)
12. N.B. Rizvandi, J. Taheri, A.Y. Zomaya, Y.C. Lee, Linear combinations of DVFs-enabled processor frequencies to modify the energy-aware scheduling algorithms, in *Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing* (2010), pp. 388–397
13. Z. Yu, W. Shi, A planner-guided scheduling strategy for multiple workflow applications, in *International Conference on Parallel Processing—Workshops* (2008) pp. 1–8
14. S. Cho, R.G. Melhem, On the interplay of parallelization, program performance, and energy consumption. *IEEE Trans. Parallel Distrib. Syst.* **21**(3), 342–353 (2010)
15. H. Kang, Y. Chen, J.L. Wong, S. Radu, J. Wu, Enhancement of Xen’s scheduler for MapReduce workloads, in *Proceedings of the 20th International Symposium High Performance Distributed Computing*

Lecture Notes in Computer Science: S Transform for the Analysis of Impulse Faults in Transformer

N. Vanamadevi, S. Santhi and R. Saranya

Abstract The tremendous and rapid growth of digital signal processing has motivated researchers to apply improved signal analysis techniques for the fault diagnosis of transformers. This paper demonstrates the application of Stockwell transform for the detection and analysis of impulse faults in transformers. Stockwell transform (S transform) is a time–frequency transformation method of signal analysis that conveys information directly in terms of time and frequency, and hence, interpretation of result becomes easier. Further, it produces the progressive resolution of the wavelet transform (WT) and maintains a direct link to the Fourier transform. The proposed method is validated through simulation of faults in a lumped parameter model of a layer winding transformer. The results are encouraging and pave way to develop an automated impulse fault classification system.

Keywords Stockwell transform · Frequency domain analysis · Short-time Fourier transform · Wavelet transform

1 Introduction

Power transformer is highly expensive equipment in an electrical power system, and any fault during the operation may cause unwanted interruption of power supply resulting in enormous economic loss. Hence, it is required to take precaution right from the design and testing stage. All transformers have to undergo a routine

N. Vanamadevi (✉) · S. Santhi · R. Saranya
Department of Instrumentation Engineering, Annamalai University,
Annamalai Nagar, Chidambaram, Tamil Nadu, India
e-mail: vanamadevinraju@gmail.com

S. Santhi
e-mail: santhi.sathyamurthy@gmail.com

R. Saranya
e-mail: rakshana.saran@gmail.com

test namely impulse test as explained in standards such as IEC-60076, Part IV, 2002 [1]. Manufacturing defects or inadequacy of insulation may lead to failure against impulse voltage stresses. The detection and location of the type of fault are highly essential for taking proper remedial measures. Thus, it is necessary to untank the winding in order to carry out a destructive one-minute power frequency test for visually locating the fault. However, a priori knowledge can enhance the process of identification and serve to locate the faults in transformers.

In case of occurrence of any fault during the test, the neutral current and/or winding current contain typical signatures based on the nature and location of the fault [2]. Since these current signals are non-stationary in nature, faults can be classified from such a transformation technique that has the time–frequency localization property [3].

Earlier methods for fault diagnosis during impulse test have been related to visual comparison of the oscilloscopic recording of the neutral currents at reduced- and full-voltage levels. The time domain analysis is not possible when the excitation waveforms at reduced and full voltage are not identical. The transfer function method has been acclaimed as an improved assessment technique that involves frequency domain approach [4, 5]. Model-based approaches for the impulse fault identification have also been attempted [6, 7]. Application of wavelet transform for impulse fault detection has also been demonstrated [8]. A number of techniques combining wavelet transform and fuzzy logic have been successfully applied to power system fault identification problems where the techniques revolve on utilizing the low-frequency components generated during fault conditions [9]. An effort also has been laid out to facilitate the feature extraction strategy using wavelet transform and artificial neural network principle for the impulse fault classification [10, 11].

In the present work, we propose a method for transformer impulse fault detection and analysis based on S transform with good time–frequency localization of various impulse fault types. The method is validated through the simulation of the lumped parameter model of a transformer winding. In order to represent dielectric faults that can occur within windings, we consider fault models for breakdown fault, partial discharge, and presence of non-linear element.

2 Motivation

Signal analysis techniques on transformation principle have long since been applied to extract salient features describing a fault event. A method already exists for the detection of various types of impulse faults through triple filter approach and subsequent frequency domain analysis [12]. However, transformer impulse fault detection through frequency domain analysis (FRA) has the limitation of loss of time information. Short-time Fourier transform (STFT) has the drawback of limited time resolution, and wavelet transform (WT) is only timescale representation that does not provide direct time–frequency record of a signal [13]. Gabor transform applies Gaussian window to compute the time–frequency representation of a signal

and provides better record than STFT. As S transform is an extension of the Gabor transform and employs frequency-dependent window for computation of Fourier transform, we are motivated to formulate the objective to apply it for the detection of impulse faults to enable automated classification upon extraction of relevant discriminating features from the time–frequency distribution of the signal.

3 Signal Analysis Based on Time–Frequency Transformation Principle

Most of the real-life signals are non-stationary in nature and require non-stationary signal analysis for process evaluation. The standard Fourier analysis plays an important role in signal processing, because of the fact that it allows the decomposition of a signal into individual frequency components and establishes the relative intensity of each component. The Fourier transform has been the most commonly used tool for analyzing frequency properties of a given signal, while after transformation, the information about time is lost and it is hard to identify where a certain frequency occurs.

The time–frequency analysis is concerned with the study of non-stationary signals and has many important applications. In this analysis, the main purpose is to have a good representation of a signal and the distribution of its energy both in time and frequency. Because of the uncertainty relation, it is not possible to have a rigorous representation of the time–frequency distribution of the energy of a signal at any scale of resolution in time and frequency [14]. STFT and Gabor transform are widely used time–frequency transformations, while Wigner Ville distribution is suitable for nonlinear and non-stationary signal analysis.

4 Theory of S Transform

S transform is a variable-window STFT or an extension of wavelet transform. It is based on scalable localizing Gaussian window and supplies the frequency-dependent resolution. The advantage of S transform is that it provides multiresolution analysis while retaining the absolute phase of each component of the signal. Further, it conveys information directly in terms of time and frequency, and hence, interpretation of the result becomes easier. Recently, researchers focus their attention toward S transform as it produces the progressive resolution of the wavelet transforms and maintains a direct link to the Fourier transform. It not only estimates the local power spectrum but also calculates the local phase spectrum.

S transform of a non-stationary signal $x(t)$ is given as

$$S(\tau, f) = \frac{|f|}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x(t) e^{-0.5f^2(\tau-t)^2} e^{-i2\pi ft} dt. \tag{1}$$

Here, τ is time of spectral localization and f is Fourier frequency.

5 Impulse Faults in a Transformer

All transformers are demonstrated for their dielectric integrity through impulse test which is a routine test. There is a possibility that a transformer may fail during the impulse test. In order to enable failure identification during the test, relevant test records are necessary. Transformer manufacturers place lightning arresters (non-linear device) in winding section as a protecting device. Presence of a nonlinear device may also be treated as undesirable. The existing transfer function measurement approach for the impulse fault detection does not guarantee the distinction between breakdown and partial discharge. Hence, a signal analysis method for the analysis of impulse faults is essential and is treated in this paper through a simulation work. The above-said faults are represented using models as shown in Fig. 1. Partial discharge is represented by the well-known ‘abc’ model, and the presence of partial discharge or nonlinear device faults does not disqualify the device [6]. Partial discharge is an incipient fault that may develop as a major fault at a later stage. The nonlinear devices are modeled using zener diodes as shown in Fig. 1. Breakdown fault is simulated using a resistor in series with a switch that is closed at a chosen time and remains closed for the entire duration of the response. This is reasonable since several studies have treated it in this manner [6].

Furthermore, impulse faults in a transformer winding that are mostly considered are series faults and shunt faults in the winding sections. Series fault is due to the short between turns and shunt fault in view of the short that occurs between a turn and ground.

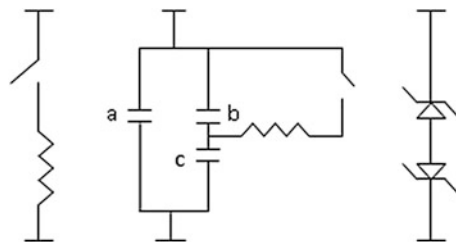


Fig. 1 Schematic of breakdown, partial discharge, and nonlinear element faults models

Table 1 List of acronyms of impulse faults

Fault No.	Fault	Acronym
0	No fault	NF
1	Series fault in line end	SFL
2	Series fault in mid end	SFM
3	Series fault in neutral end	SFN
4	Shunt fault in line end	SHFL
5	Shunt fault in mid end	SHFM
6	Shunt fault in neutral end	SHFN
7	Breakdown fault	BD
8	Partial discharge	PD
9	Presence of a nonlinear device	NL

6 Lumped Parameter Model

In order to simulate the detection of winding faults that may occur due to impulse test, a ten-section lumped parameter model of a specially designed transformer winding is considered. The winding length is divided into three regions, namely line end (winding-sections 1 up to 3), mid-winding (winding-sections 4 up to 7), and neutral end (winding-sections 8 up to 10). The faults considered for the analysis are tabulated in Table 1. The lumped parameter model of the transformer winding considered for simulation is shown in Fig. 2. A single-layer winding is initially considered to study the detection of impulse faults of a transformer. The values of the model parameters as given below correspond to a typical layer winding as mentioned in [15]. A ten-section model is treated in this work with the assumption that the mutual inductance effect is negligible.

Depending on C_g , C_s , and L values, each of the transformer winding is found to have its own natural frequencies of oscillation that largely characterizes its impulse responses. The model is subjected to excitation similar to a standard lightning impulse (LI) of 1 V amplitude, 1.2/50 μ s front time, and fall time, respectively, and the current through a current viewing resistor R is recorded under no fault and various fault simulated conditions as indicated in Table 1. To simulate a series fault, a short is placed between sections through a switch that is closed at a known time and removed at a known time.

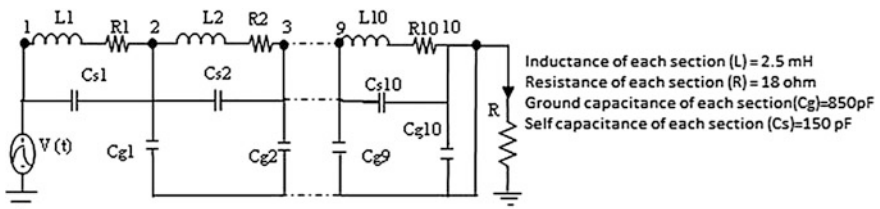


Fig. 2 Lumped parameter model of transformer winding

7 Validation of the S Transform Algorithm Through Simulation

In order to study the efficiency of S transform for time–frequency analysis of a non-stationary signal, the S transform algorithm is implemented through MATLAB coding. A non-stationary chirp signal is simulated as shown in Fig. 3. This chirp signal is a linear chirp with frequency components increase from 10 to 40 Hz over the time range 0–1 s. Figure 4 shows the frequency spectrum of the chirp 1 signal. It is evident from the spectrum that the signal contains frequency components in the range 10 to 40 Hz with constant magnitude, but the time information is lost due to frequency transformation.

Hence, the same signal is analyzed using S transform and the time–frequency distribution is shown in Fig. 5 as an image plot. The intensity of the image is indicated in the right side as a color bar and indicates the region of energy concentration of the signal. Further, it is obvious that the chirp signal has frequency components increasing from 10 to 40 Hz over the time range 0–1 s, and certainly, this time–frequency transformation provides a better representation of the chirp

Fig. 3 Simulated chirp signal

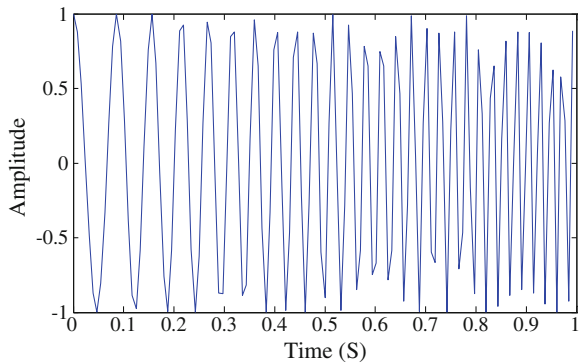
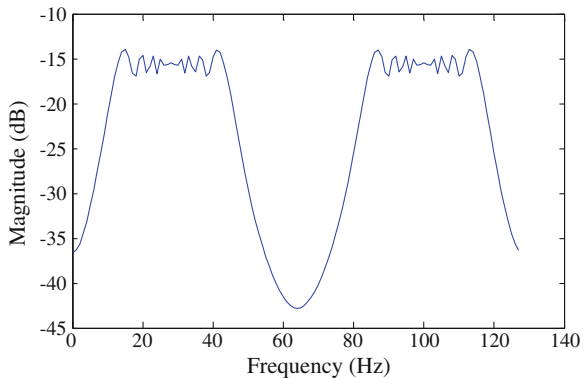


Fig. 4 Spectrum of simulated chirp signal



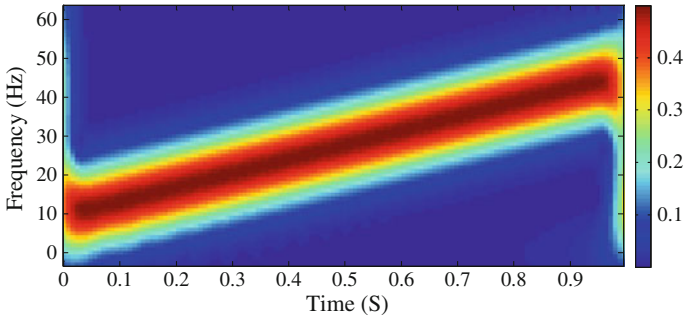


Fig. 5 S-transformed chirp signal

signal than the Fourier transformation. In addition, it also indicates the location of concentration of highest energy of the signal. Further, the linear increasing frequency trend is also observable in time–frequency representation of the signal. The color bar indicates the energy concentration in the signal in terms of S transform coefficients. It is evident that S transform has the ability to retain time information of the signal.

8 Analysis of Impulse Faults in a Transformer Using S Transform

This section presents the simulation results of impulse fault analysis with a lumped parameter model of transformer winding using S transform. The various faults tabulated in Table 1 have been simulated in the line end, middle, and neutral end of the lumped parameter model of the winding. The respective winding currents under no fault and various fault simulated conditions have been subjected to analysis based on Fourier transform and S transform method. The frequency domain constituent of winding current reflects changes in resonant frequencies, but the time information is lost due to Fourier transformation. To analyze the frequency contents of the winding current along with time information preservation, the winding current is transformed to time–frequency plane by S transform algorithm implemented through MATLAB coding. Figures 6, 7, and 8 show the time–frequency distribution of winding current under NF, SFM, and SHM condition, respectively. It is evident that compared to Fourier transform, S transform provides a better visual interpretation for the analysis of impulse faults.

A close observation of time–frequency distribution of winding current under NF, SFM, and SHM reveals that it is possible to identify the presence of a fault, but mere visual inspection could not assist in distinction between SFM and SHM. Hence, it is desirable to extract features that could serve better for distinguishing between faults that are difficult to classify. In this regard, for automated analysis,

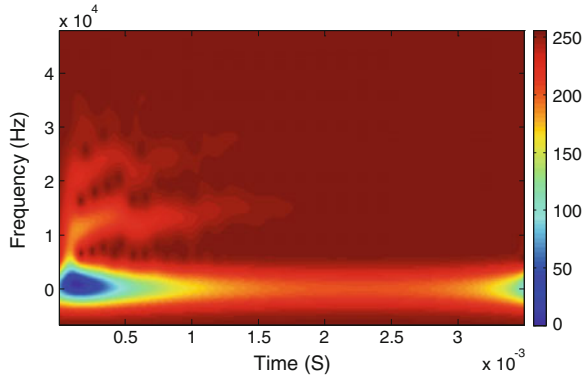


Fig. 6 S-transformed winding current under NF

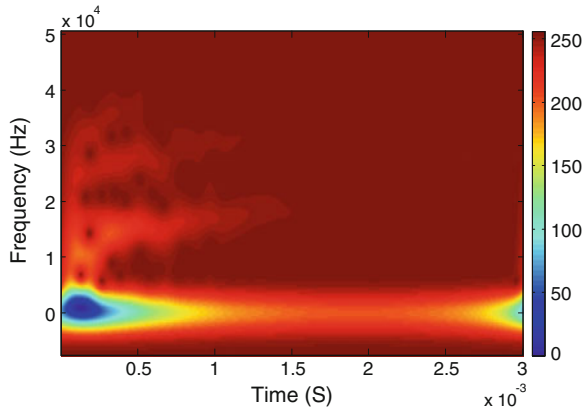


Fig. 7 S-transformed winding current under SFM fault

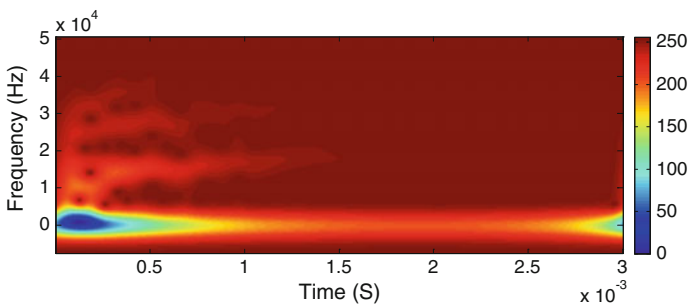


Fig. 8 S-transformed winding current under SHM fault

Table 2 Feature extracted from S-transformed signal under no fault and various faults

Fault types	Resonant frequency (KHz)	TOO (ms)	Percentage change in signal energy in time–frequency domain	Percentage change in S transform coefficient
NF	12.87	0.234	Base	Base
BD	10.45	0.234	3.68	26.69
PD	12.24	0.234	2.99	26.53
NL	7.13	0.234	7.15	12.04
SFL	12.50	0.101	10.65	16.75
SFM	12.30	0.102	19.23	7.77
SFN	13.83	0.103	6.49	15.17
SHL	15.33	0.102	7.75	29.21
SHM	15.67	0.102	6.42	30.99
SHN	16.10	0.102	7.82	34.41

features such as coefficients corresponding to dominant resonant frequency of the winding and energy concentration of the signal in the time–frequency plane have been extracted for no fault and various fault simulated conditions. Table 2 shows the result of feature extraction. Extraction of features has also been made for simulated winding currents with other fault models as shown in Fig. 1 to analyze the breakdown fault, partial discharge phenomena, and the presence of a nonlinear device in the winding section. The S-transformed winding current corresponding to BD and PD and presence of nonlinear element fault are shown in Figs. 9, 10, and 11, respectively. The results of feature extraction show an evidence for better distinction between BD and PD.

It is observed that the changes are clear from 0.1 to 0.2 ms since the fault is introduced at 0.1 ms and persists for a total duration of 0.1 ms. The percentage change in S transform coefficient corresponding to dominant resonant frequency between no fault and faulty conditions is more than the percentage change in energy concentration. The results of feature extraction for winding faults have been shown

Fig. 9 S-transformed winding current under BD fault

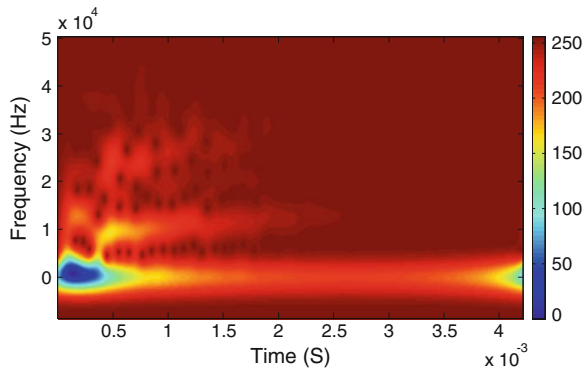


Fig. 10 S-transformed winding current under PD fault

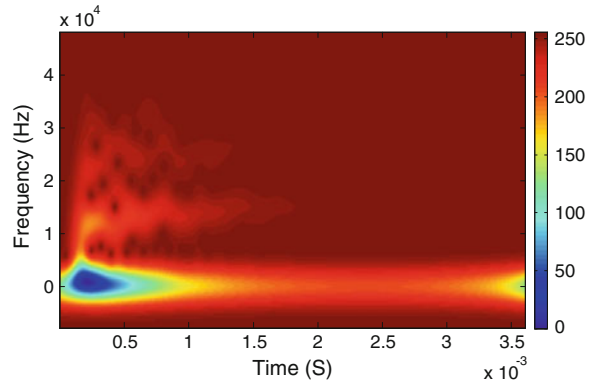
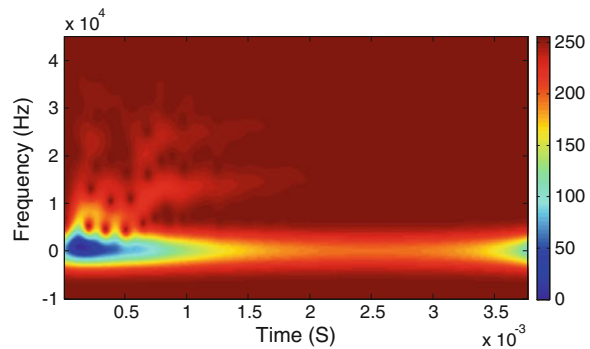


Fig. 11 S-transformed winding current under NL fault



only for faults simulated at one section corresponding to line end, middle, and neutral end. Simulation can be extended by considering faults in other sections and also simultaneous occurrence of series and shunt fault in a section. This provides considerable fault data to design an automated fault analysis system.

9 Conclusion

The application of S transform for the impulse faults analysis has been validated through simulation work on a lumped parameter model of a transformer winding. The algorithm is initially validated for simulated non-stationary signal and extended to impulse fault analysis. It has been observed that frequency transformation does not indicate changes clearly if the resonant frequencies lie in noise-prone regions due to monotonic decreasing trend of the amplitude in the LI excitation. The results of S transform-based analysis of impulse faults for faults at one section corresponding to line end, middle, and neutral end alone are provided even though satisfactory results have been obtained for other fault types and combination of faults to page limits the presentation.

References

1. IEC 60076—Part IV, Guide to the lightning impulse and switching impulse testing—Power transformers and Reactors. IEC, Geneva, Switzerland (2002)
2. R. Vanaja, K. Udayakumar, Fault location in power transformers during impulse test. *IEEE Power Eng. Soc. winter meet.* (2000) (pp. 2199–2204)
3. L. Satish, Short-time Fourier and wavelet transforms for fault detection in power transformers during impulse tests. *IEEE Proc. Sci. Meas. Technol.* **145**, 77–84 (1998)
4. R. Malewski, B. Poulin, Impulse testing of power transformer using the transfer function method. *IEEE Trans. Power Delivery* **3**, 476–489 (1988)
5. T. Leibfried, K. Feser, Monitoring of power transformers using the transfer function method. *IEEE Trans. Power Delivery* **14**, 1333–1341 (1999)
6. S. Arun kumar, V. Sandeep, S. Shankar, M. Gopalakrishnan, K. Udayakumar, V. Jayashankar, Impulse testing of power transformers—a model reference approach. *IEEE Proc. Sci. Meas. Technol.* **151** (2004) (pp. 25–30)
7. M. Arivamudhan, S. Santhi, Model based approach for fault detection in power transformer using Particle swarm intelligence. *Recent Advancements in System Modelling Application* **188**, 287–300 (2013). Springer
8. P. Purkait, S. Chakravorti, Pattern classification of impulse faults in transformers by wavelet analysis. *IEEE Trans. Dielectr. Electr. Insul.* **9**, 555–561 (2002)
9. N. Vanamadevi, S. Santhi, Impulse Fault Detection and Classification in power transformers with wavelet and fuzzy based technique. *Recent advancements in Syst. Model. Appl.* **188**, 261–273 (2013). Springer
10. N. Vanamadevi, M. Arivamudhan, S. Santhi, Detection and classification of impulse faults in transformer using wavelet transform and artificial neural network. *IEEE Int. conf. Sustain. Energ. Technol.* (2008) (pp. 72–76)
11. L.P. Mao, R.K. Aggarwal, A novel approach to the classification of the transient phenomena in power transformers using combined wavelet transform and neural network. *IEEE Trans. Power Delivery* **16**, 654–660 (2001)
12. S.S. Sahu, G. Panda, N.V. George, An improved S-transform for time frequency analysis. *IEEE Int. conf., IACC.* (2009) (pp. 315–319)
13. S. Santhi, V. Jayashankar, V. Jagadeesh Kumar, Time frequency analysis of method for the detection of winding deformation in transformers during short circuit test. *IEEE Int. Instrum. Meas. Technol. Conf.* (2008) (pp. 1–5)
14. R.L. Allen, D.W. Mills. *Signal analysis time, frequency, Scale and structure.* IEEE Press, Wiley, Newyork (2004)
15. S. Jayalalitha, V. Jayashankar, Fuzzy logic based impulse test analysis. *IEEE Mid-Summer Workshop Soft Comput. Ind Appl* (2005)

Defensive Mechanism to Guard Against Packet Droppers in Mobile Ad Hoc Network

S. Madhurikkha and R. Sabitha

Abstract Mobile ad hoc network (MANET) is a group of mobile (or temporarily stationary) nodes which provides the ability to stream voice, data, and video between arbitrary pairs of devices utilizing the others as relays to avoid the need for infrastructure. There are many techniques which are employed in order to provide robust MANET capability like self-forming, link adaptation, transparent IP Networking, and multicast support. Its rapidly changing network topology, which is unpredictable over time, makes MANET vulnerable to wide range of attacks. It is very tedious to detect some attacks when it becomes a part of network. Ad hoc on-demand distance vector (AODV) is a popular reactive routing protocol but exposed to well-known packet dropping attack, where a malicious node intentionally drops packets without forwarding them to destination. In this paper, we discuss the security mechanisms, namely data routing information (DRI), cross-checking, and retard-mode operations, to defend against packet dropping attack in MANET with results simulated in ns-2 to show the improvement in packet delivery ratio.

Keywords Ad hoc networks · Routing protocols · AODV · Packet dropping attack · MANET

1 Introduction

Mobile ad hoc network (MANET) is a system of wireless mobile nodes that self-organize dynamically and frequently change network topologies. It is insecure by nature due to unreliability of wireless links between nodes, and lack of security

S. Madhurikkha (✉)
Sathyabama University, Chennai, India
e-mail: madhurikkha@gmail.com

R. Sabitha
Jeppiaar Engineering College, Chennai, India
e-mail: sabitha_ramados@yahoo.com

features is incorporated in ad hoc environment. They have special features like wireless links, high mobility, multiple hops, dynamic topology, and decentralized control which make them vulnerable to various attacks. Most of research work focuses on prevention and detection of malicious nodes from network. Nodes flood other nodes with routing traffic, advertise nonexistent links, drop packets, and change the contents of packets, thus inflicting failure in network. One of the most popular routing protocols, ad hoc on-demand distance vector (AODV), is used in MANET. However, AODV is vulnerable to packet dropping attack. A malicious node stealthily drops some or all data packets or control packets without forwarding them to destination. A group of nodes can drop packets in collaboration in network at such a rate that message communication in network may get degraded or even disrupted. Due to lack of physical protection and reliable mechanisms, packet dropping attack posts a serious threat to routing in MANETs. In [1], authors have shown that black hole nodes (malicious node falsely advertise good route to destination on route discovery process) cooperate and work in groups in MANET and have proposed a solution to identify black and cooperative black hole attack. In this paper, the mechanisms proposed for cooperative black hole attack [1], i.e., data routing information table, cross-checking mechanism, and retard-mode operation (Sect. 5), are adapted to defend against packet dropping attack.

2 Related Works

A number of works have been done to enforce the security problem on the area of ad hoc network community. This section lists some of these works. Zeshan et al. [2] proposed a twofold solution to detect and identify malicious nodes in network by setting T_{max} (maximum threshold) and monitoring nodes to declare misbehaving nodes. Sen et al. [3] proposed a cooperative scheme used to detect malicious node, as every node in network monitors the behavior of its neighbors upon abnormal action. Distributed algorithm is used to confirm attack in network. Since only trusted nodes are used for securing routing, it is an overhead and malicious nodes are not isolated in this method. Medidi et al. [4] proposed a method where a detection manager locates malicious nodes that drop packets in MANET by setting rules for nodes with low false-positive rate. Marti et al. [5] proposed a mechanism watchdog and path rater to detect malicious node in MANETs. Bhalaji et al. [6] proposed an association-based routing using DSR protocol to enhance security against selective packet drop attack which is based on trust value and threshold parameters between nodes. Shandilya and Sahu [7] proposed a distributed approach to detect and prevent the flooding attack. The technique depends on selection of threshold values, and the concept of delayed queue reduces the accidental black listing of node, but data flooding is not prevented. Djahel et al. [8] proposed a state of art on securing MANET from packet dropping attack, but it is not valid always due to dynamic nature of MANET and mainly based on assumption that the majority of nodes are misbehaving.

3 Overview of AODV Routing Protocol

Ad hoc on-demand vector (AODV) is a descendent of destination-sequenced distance vector (DSDV) routing protocol. It is a reactive routing protocol which establishes route to destination only on demand. Whenever source node wishes to route packets to destination, it first checks its own routing table to determine whether a route to destination is already available. If so, it routes the packets to destination. If not, the source node initiates a route discovery process where it broadcasts a route request (RREQ) message to its neighbors which is further propagated until it reaches an intermediate node with fresh route to destination node or destination node itself. The intermediate nodes on receiving RREQ make an entry in their routing table for the node (which forwarded RREQ message) and source node. If the destination sequence number present in routing table is lesser than or equal to number present in RREQ packet, the node relays a further request to its neighbors. If the sequence number is higher, it denotes a 'fresh route' and packets can be sent through this route. All neighboring nodes on the reverse path make entry of the nodes from which it received RREP. The source node on receiving the updated route to destination node starts routing data packets through neighboring nodes that responded to it first with RREP. However, AODV protocol achieves limited security and lacks scalability and latency time.

4 Packet Dropping Attack

In a packet dropping attack, a malicious node intentionally drops the packets they receive. To conduct its attack, the malicious node must initially belong to the route, and then, it starts the action which is the data dropping [6]. The manner with which the malicious node fits in the data route differs. The packet dropping attack can be of any one way given below.

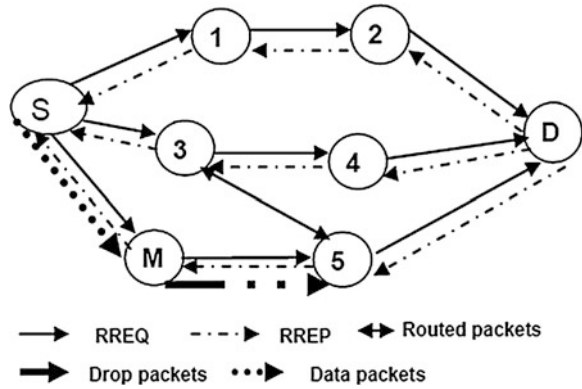
- (i) Dropping control packets
- (ii) Selectively dropping packets
- (iii) Group of nodes collaboratively drop packets

We shall discuss the above-mentioned ways of dropping packets using AODV protocol in MANET.

4.1 Control Packet Dropping

A malicious node drops control packets like RREQ, RREP, or RERR packets to keep use of failed routes, which results in a denial of service [8]. Dropping RREQ packets will exclude the malicious node from routes and avoid receiving data

Fig. 1 Selective packet dropping



packets to forward from it. So, malicious node mainly drops RERR (route error) packets to gain routes. The source node is not updated with the RERR message, and it sends packets to destination which is accepted by the malicious node. Now, the malicious node gains the route and initiates a denial of service.

4.2 Selective Data Packet Dropping

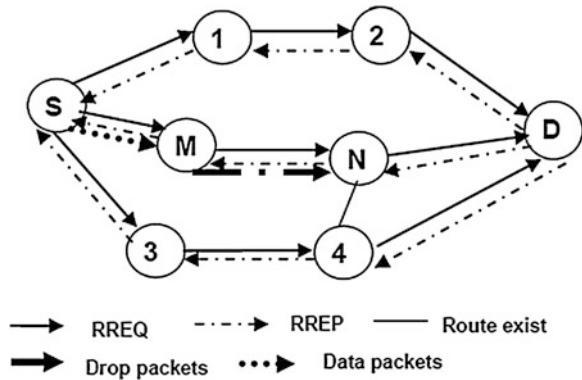
A malicious node becomes a part of routing data to destination node and starts receiving data packets and drops them selectively and forwards the rest of packets to its neighbors [6]. It is difficult to isolate such nodes in the network. In Fig. 1, S is the source node and D is the destination node. Nodes 1–5 act as intermediate nodes. Node M is malicious. The source node (S) starts a route discovery process by transmitting RREQ packets to neighboring nodes. The malicious node is a part of network (M) and receives RREQ. The source node starts forwarding the data packets after receiving the RREP message from the destination node (D). Malicious node M being part of routing data packets starts dropping some data packets and forwards others to next hop node (3). The packet dropping is shown with lines in the Fig. 1. This is difficult to detect since it drops the packets selectively.

4.3 Packet Dropping in Groups

Groups of malicious node collaboratively drop the packets without forwarding it to the destination. This activity makes the network to break from message communication between nodes and even disrupt the whole topology. In Fig. 2, node S is source node and D is the destination node. Nodes 1–4 act as intermediate nodes.

The malicious nodes M and N being part of RREQ respond with an RREP to source node. Now, on receiving the RREP from M, source node transmits the data

Fig. 2 Collaborative packet dropping



packets. On the receipt of data packets, malicious node M simply drops them or forwards all data packets to malicious node M which in turn drops all packets without forwarding to intended destination. The destination node does not know any message on what is happening in the path as the malicious nodes completely block it.

5 Methodologies Adapted

In this section, the proposed methodologies for defending against packet dropping attack are discussed

1. Data routing information (DRI) table with status bit
2. Cross-checking
3. Retard-mode operation.

5.1 Data Routing Information with Status Bit (DRI)

In this method during route discovery process, the nodes which respond to RREQ message of source node must send two bits of additional information. Each node must maintain an additional DRI table in which bit ‘1’ stands for ‘true’ and bit ‘0’ stands for ‘false.’ The first bit ‘from’ denotes whether any data packets routed from the nodes in node field. The second bit ‘through’ stands for routing data packets through the node in the node field. The last bit called as status bit is updated based on the two-bit entries in the DRI table. For example from Fig. 3, a sample database maintained by node 5 is shown in Table 1. The status bit is updated as ‘T’ (True) for node 4 since the node has routed at least once. The entry 1 1 for node D implies that node 5 has routed data packets from and through node D and the status bit is ‘T’.

Fig. 3 Sample network for DRI table entry

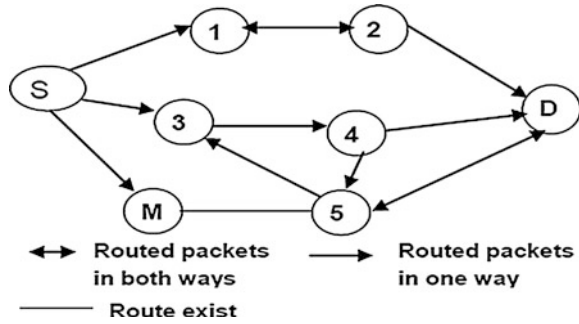


Table 1 DRI table of node 5

Node #	DRI		
	From	Through	Status bit
4	1	0	T
M	0	0	F
D	1	1	T
3	0	1	W

The 0 0 entries for node M denotes that node 5 has not routed from and through node M. The status bit is updated as ‘F’ (False) for the node M. The entry 0 1 for node 3 implies that node 5 has not routed data packets from 3 but has routed data packets through 3. The status bit is updated as ‘W’ (Wait).

The entry for the node 3 is given as ‘W’ because it may have not got any chance to route data packets from it or it may be new mobile node to the network. So the node 3 is put in retard-mode operation which will be discussed in the later section.

5.2 Cross-Checking

The proposed scheme relies on reliable nodes; that is, nodes through which data packets are routed previously by source node are known to be trustworthy. The proposed model is depicted in Fig. 4. The intermediate node (IN) on its RREP provides details regarding its next hop node (NHN) and its DRI table entry for that NHN. On receiving the RREP from IN, source node checks its own DRI table for the status bit to verify whether IN is a reliable node. If status bit of NHN is ‘T’, it is reliable; otherwise, NHN is unreliable. If NHN is reliable, the source node checks whether IN is a malicious node; that is, if the status bit of DRI table entry for IN in NHN table is ‘F’ (NHN has not routed data from and through IN), then IN is a malicious node. If the entry is ‘W’, then that route is not selected for routing but sent to retard mode for further checking. If IN is malicious node, then source node

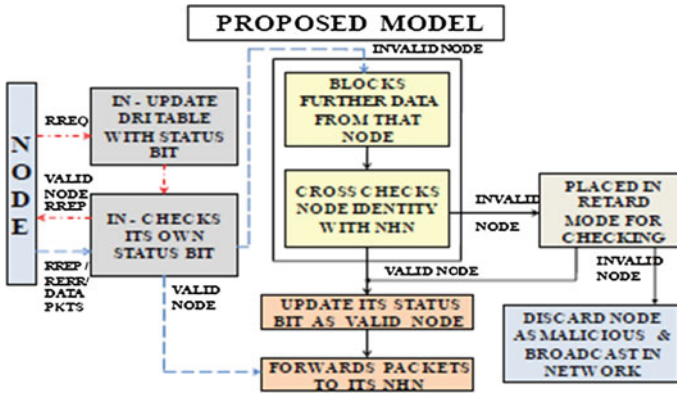


Fig. 4 Proposed model for packet dropping attack in MANET

identifies all nodes in reverse path from IN to the node which has generated RREP as malicious nodes.

//Algorithm for Packet dropping attack

Notifications:

SN: Source node IN: Intermediate node FP: Further packet FR_q : Further request FR_p : Further reply NH: Next hop

1. SN sends control or data packets
2. SN receives packets
3. **If**(IN’s DRI information for SN is valid) {
Route data packets from the source node.}
4. **Else** {
5. **Do** {
6. SN blocks FP from IN
7. Send FR_q and ID of IN to NHN
8. Receive FR_p , NHN of current NHN-DRI entry with Status bit for NHN’s next hop, DRI entry with status bit for current NHN
9. **If** (NHN is a reliable node) {
10. Check IN for Packet dropping attack using status bit of DRI entry & Retard-mode operation.
11. **If** (IN is not a malicious node) **then**
12. Accept data packets from the source node
13. **Else** {
14. IN is a malicious node
15. All nodes in the network broadcast the node as malicious}}
16. **Else**
17. Current IN = NHN
18. **} While** (IN is not a Valid node)}

Hence, nodes M and N are marked as malicious nodes and this information is propagated through network.

5.3 Retard-Mode Operation

In this method, the nodes which have the status bit entry with 'w' (Wait) are placed in a queue and further checked for its validity. The retard mode is used when a new node joins the network due to mobility but may not be a malicious node. Since the source node checks the DRI table entry with status bit before routing the data packets, the status bit entry of these new nodes, nodes resumed after ideal state, and nodes moved within network due to mobility will have the status bit as 'W', but they may be a good node.

These nodes cannot participate in routing and neglected as malicious node due to their status bit entry as they have not routed any packets. To avoid losing a good node in participating for routing, retard mode is used where a queue is maintained for the nodes which have the status bit entry as 'W'. These nodes' activities are checked for some time, and then, their status bit will be changed from 'W' to 'T' or 'F' based on the nodes' routing information.

6 Simulations

The experiments for the evaluation of proposed scheme have been carried out using the network simulator ns-2. The performance of AODV under attack and the improved AODV under attack is shown in Table 2.

The metric packet delivery ratio is tabulated against number of mobile nodes in the network, i.e., 30. The maximum delivery ratio is achieved between the connections 10 and 20. However, the delivery ratio falls slightly when maximum

Table 2 No. of connections versus packet delivery ratio

No. of connections	Packet delivery ratio AODV with attack-modified AODV with attack	
5	40	40
10	41	60
15	41	61
20	42	62
25	35	50
30	30	48

connection is established. The packet dropping attack has a severe impact on the normal AODV under attack than on the modified AODV algorithm. The proposed algorithm increases the packet delivery ratio to 59% from 57 % resulting in improvement from AODV.

7 Conclusion and Future Work

In this paper, one of the attacks, i.e., packet dropping attack, is studied in MANET. A security protocol has been proposed to identify the ways of packet dropping nodes in MANET and thereby diverting a secure routing path from source node to the destination node avoiding the malicious nodes. The results have been simulated with ns-2 and compared with the AODV and modified AODV security mechanism. As a future scope of work, packet dropping attack can be discussed with different routing protocols in MANET and simulated using ns-2. The simulations can be done for the throughput and packet overhead.

References

1. J. Sen, S. Koilakonda, A. Ukil, A mechanism for detection of cooperative black hole attack in mobile ad hoc networks, in *Proceedings of IEEE International Conference on Intelligent systems, Modeling and Simulation* (2011)
2. M. Zeshan, S.A. Khan, et al., Adding security against packet dropping attack in mobile ad hoc networks, in *Proceedings of ACM International Seminar on Future Information Technology and Management Engineering* (FITME, 2008)
3. J. Sen, G. Chandra, P. Balamuralidhar, et al., A distributed protocol for detection of packet dropping attack in mobile ad hoc networks, in *Proceedings of IEEE International conference on Telecommunication* (2007)
4. S.R. Medidi, M. Medidi, S. Gavini, Detecting packet-dropping faults in mobile ad-hoc networks. (IEEE, 2003)
5. S. Marti, T. Giuli, K. Lai, M. Baker, Mitigating routing misbehaviour in mobile ad hoc networks. in *Proceedings of International Conference on Mobile Computing and Networking* (2000)
6. N. Bhalaji, A. Shanmugam, Reliable routing against selective packet drop attack in DSR based MANET. *J. Softw.* **4**(6), 536–543 (2009)
7. S.K. Shandilya, S. Sahu, A trust based security scheme for RREQ flooding attack in MANET. *Int. J. Comput. Appl.* **5**(12) (2010)
8. S. Djahel, F. Nait-abdesselam, Z. Zhan, Mitigating packet dropping problem in mobile ad hoc networks: proposals and challenges, in *IEEE Communications Survey and Tutorials* (IEEE, 2010)

Real-Time Intrusion Prediction Using Hidden Markov Model with Genetic Algorithm

T. Divya and Kandasamy Muniasamy

Abstract As the use of Internet increases, cyber attacks and their severity also increase. Since it is not possible to compromise on security, intrusion detection systems (IDSs) become critical component in a secure organization. IDSs detect an attack only after it has occurred. When use in a high-traffic network, IDSs produce a large number of alerts. The false-positive (FP) rate increases with this. In this paper, we propose a framework for predicting future attacks by combining two machine-learning methods: genetic algorithm (GA) and hidden Markov model (HMM). It has two major components in which the first component makes use of GA to derive efficient intrusion detection rules and thereafter a precise detection of attacks. The second component uses HMM to predict the next attack class of the attacker. So combining these together is a good idea and gives a good intrusion prediction capability with reduced FP rate.

Keywords Intrusion prediction · False positive · Genetic algorithm · Hidden markov model

1 Introduction

It is true that in this information era, the Internet and local area networks give us a convenient way of communication, information, storage, etc. But, there is a bad side that we are always threatening by the interference of intruders in our communication. With the increased usage of Internet, the number and the severity of the new attack types also increased. So proper defense mechanism is needed to ensure that the communication is safe.

T. Divya (✉) · K. Muniasamy
TIFAC CORE in Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: divyatharissil@gmail.com

K. Muniasamy
e-mail: kandamuniasamy@yahoo.com

IDSs are the tools widely used to monitor and analyze the network traffic to detect intrusions. There are two types of intrusion detection mechanism—misuse detection and anomaly detection [1]. In misuse detection, the predefined attack patterns are comparing with the current traffic to detect if any. Anomaly detection checks the behavior change in the traffic from one that is defined as normal. If the network is a large one with a large traffic, these IDSs produce large number of alerts that make it very tough to manage. This also leads to a large false-positive (FP) rate. If the attacker is performing a multistep attack, these IDSs are not capable of detecting the aimed attack even though it produces alerts for individual steps. If it is possible to predict the chance of an upcoming attack, we can save our network by taking proper response mechanism.

In this paper, we are proposing a framework to deal with two important things—to reduce the FP rate and to predict the chance of a future attack. To achieve these two, we used two machine-learning techniques. Genetic algorithm (GA) is used to reduce the FP rate by deriving strong rules for attack detection. To facilitate real-time intrusion prediction, we applied the prediction component, which uses hidden Markov model (HMM) on the output of detection component.

This paper is organized as follows: Next, we will discuss the background and related works, and in Sect. 3, we will discuss our proposed real-time intrusion prediction system. Section 4 is the conclusion.

2 Background and Related Works

As the use of Internet increases, cyber attacks and their severity also increase. Since it is not possible to compromise on security, IDSs become critical component in a secure organization. Intrusion detection includes the monitoring and analyzing of a computer and/or network to detect security breaches. IDSs are the tools used for this purpose. One of the major drawbacks of IDSs is that they can only detect the attack after it has occurred. At that time, the network and systems might have been compromised. This is not a good way of cyber defense. The other problem is that when used in a high-traffic network, these IDSs produce large number of alerts. The FP rate also increases with this. So it will become a good defense technique if it is possible to predict the probability of an upcoming attack as well as reducing the FP rate of alerts.

Two categories of related works are there: works for reducing false positives in intrusion detection system and the works related to intrusion prediction. Gong et al. [2] propose a GA-based approach to intrusion detection. As the intrusions become known, it is able to generate and upload new. GA is used to derive a set of classification rules from network audit data. Seven network features including both categorical and quantitative data fields were used when encoding and deriving the rules. A simple but efficient and flexible fitness function, i.e., the support-confidence framework, is used to select the appropriate rules. Although it is cost-effective and adaptive, it cannot deal with very large dynamic network traffic in real time.

Zhu and Ghorbani [3] utilized neural networks for extracting attack strategies without using prior knowledge. They used a multilayer perceptron which is used to extract the causal correlation between a pair of alerts. An Alert Correlation Matrix (ACM) is used to store correlation strengths of any two types of alerts. Yu and Frincke [4] propose Hidden Colored Petri-Net (HCPN) to predict intruder's next goal. HCPN can describe the relationship between different steps carried out by intruders, model observations (alerts), and transitions (actions) separately, and associate each token element (system state) with a probability (or confidence). Haslum et al. [5] propose a model based on HMM to predict the next step of an anomaly. In this model, distributed system attacks are simulated in four steps. Based on observations from all IDSs in the network, the system mode can be moved among states. Thus, each time, prediction of the next goal can be estimated by the probability of each state. It will be better if it is possible to generate new classification rules for the network traffic and update them automatically in the IDSs.

3 Real-Time Intrusion Prediction System

The architecture of the proposed real-time intrusion detection system is given in Fig. 1. It has two major components: (1) rule generation for intrusion detection system using GA and (2) intrusion prediction using HMM.

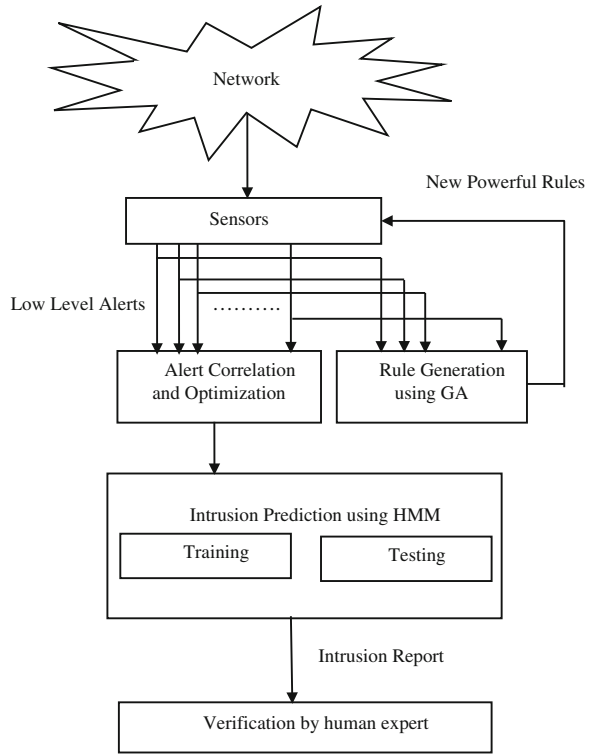
Over all idea: We assume that there is more than one sensor in our network. A sensor can be intrusion detection system, intrusion prediction system, or firewall. We are simulating the attacks, and the alerts are generated by the sensors. Logs of these are collected for further processing. After preprocessing the data, GA is used to generate new powerful rules. This is performed in offline. Then prediction is performed in a real-time manner using HMM. We are explaining the framework in detail now.

3.1 Rule Generation for Intrusion Detection System Using Genetic Algorithm

The sensors in the network produce alerts corresponding to the simulated attacks. If they are combined properly, new powerful rules can be generated. These rules are updated in the sensors and it reduces the FP rate. To achieve this, we are using GA with the same idea proposed by Gong et al. [2].

Several network features have higher possibilities to be involved in network intrusions [6, 7]. Here, seven network features are taken as input to the GA. They are duration, protocol, source port, destination port, source IP, destination IP, and attack. Here, duration (H: M: S format) is represented using three genes, source IP

Fig. 1 Architecture of the proposed real-time intrusion prediction system



and destination IP (a.b.c.d format) are represented using four genes. All other features are represented using single gene. So a vector containing 15 points is obtained. These 15 values are obtained through the feature extraction from the alerts generated by the sensors.

Rules are represented in the IF {Condition} THEN {Act} form (A → B form). The fitness function used here is as follows:

$$\text{Fitness} = (w1 * \text{Support}) + (w2 * \text{Confidence}) \tag{1}$$

where

$$\text{Support} = |A \text{ and } B| / N \tag{2}$$

$$\text{Confidence} = |A \text{ and } B| / |A| \tag{3}$$

- N total number of network connections
- $|A \text{ and } B|$ number of network connections that matches the rule
- $|A|$ number of network connections matching A
- $w1, w2$ weights used to control the balance between the two terms

The above fitness function is based on the support-confidence framework [8]. The chromosomes which cross a threshold with their fitness value are taken for mutation and crossover. New rules are updated in the knowledge base of the sensors, which are then used in generating alerts.

3.2 Alert Correlation and Optimization

Before moving to prediction component, alert optimization is performed. It is done to enhance the performance of prediction component. It includes two subparts: alert correlation and alert optimization. Alert correlation is done with the help of ACM proposed by Zhu and Ghorbani [3]. Optimization is used to modulate the severity of alert. Severity of an alert is calculated as follows:

$$\text{Alert Severity} = \text{Alert Severity} * e^{[(F*N)/(K*A)]} \quad (4)$$

where

N Alert frequency

F Extracted effect of alert from ACM

K Constant which controls the correlation effect

A Predefined acceptable alert per day.

3.3 Intrusion Prediction Using Hidden Markov Model

Next, we are explaining how the attack prediction is performed using HMM. An HMM is specified using the following parameters:

- (1) $N \rightarrow$ Number of states in the model.
- (2) $M \rightarrow$ Number of distinct observation symbols per state.
- (3) $A \rightarrow$ State transition probability distribution.
- (4) $B \rightarrow$ Observation symbol probability distribution.
- (5) $\Pi \rightarrow$ Initial state distribution.

In our case, we have four states: normal, attempt, progress, and compromise [5]. While performing an attack, a system is assumed to be move among these states. The observations are the alerts from the detection components. They cause the system to move among states. We took increased severity of alerts as our observation as given in [4]. For each observation, the HMM will update the state probability distribution and calculate the intrusion probability value. Based on this value, the probability of occurrence of an attack in the future will be predicted.

Initially, the four states are distributed as $\{1, 0, 0, 0\}$. It means, initially, the system is fully normal. As the attack progresses, the probability of the normal state will be reduced and the probability of other states will be increased accordingly. The prediction of an attack is performed using the probability value of the Compromise state. If this probability is above a predefined threshold value, let it be some 90 or 95 % (0.9 or 0.95), then the chance of a future attack is predicted.

4 Conclusion

In this paper, we discussed a method to reduce false positives in IDSs and to utilize the output of those IDSs for the prediction of a future attack. GA is used to produce new powerful rules that are then updated in IDSs to reduce FP rate. HMM is used for intrusion prediction. Combining GA and HMM gives a good prediction on future multistep attack.

References

1. T. Xia, G. Qu, S. Hariri, M. Yousif, in *An Efficient Network Intrusion Detection Method Based on Information Theory And Genetic Algorithm*. Performance, Computing, and Communications Conference, 2005. IPCCC 2005. (2005), pp. 11–17
2. R.H. Gong, M. Zulkernine, P. Abolmaesumi, A software implementation of a genetic algorithm based approach to network intrusion detection, in *Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (2005)
3. B. Zhu, A.A. Ghorbani, Alert correlation for extracting attack strategies. *Int. J. Netw. Secur.* **3**, 244–258 (2006)
4. D. Yu, D.A. Frincke, Improving the quality of alerts and predicting intruder's next goal with hidden colored Petri-Net. *Comput. Netw.* **51**, 632–654 (2007)
5. K. Haslum, A. Abraham, S. Knapkog, A, Dips framework for distributed intrusion prediction and prevention using hidden markov models and online fuzzy risk assessment, in *3rd International Symposium on Information Assurance and Security*, pp. 183–188 (2007)
6. W. Li, Using genetic algorithm for network intrusion detection, in *Proceedings of the United States Department of Energy Cyber Security Group*, pp. 1–8 (2004)
7. M. Middlemiss, G. Dick, Feature selection of intrusion detection data using a hybrid genetic algorithm/KNN approach, in *Design and Application of Hybrid Intelligent Systems*, pp 519–527 (2003)
8. Lu Wei, A.I. Traore, Detecting new forms of network intrusion using genetic programming. *Comput. Intell.* **20**(3), 475–494 (2004)

Detection of Power Quality Disturbances Based on Adaptive Neural Net and Shannon Entropy Method

D. Kavitha, P. Renuga and M. Seetha Lakshmi

Abstract Detection of power quality (PQ) events is a vital task for the power system monitoring and control. This paper presents a new scheme for the revealing of PQ disturbances using adaptive neural net (ANN) and information theory which employs neural net as a harmonics extracting unit and the difference entropy as a feature extracting unit. Simulations on six signals, such as ideal sine wave, interruption, voltage sag, voltage swell, impulse, and oscillation transient, are done with and without the presence of harmonics and the begin and end instants of disturbances are accurately tracked. The robust nature of the algorithm allows accurate estimation in the presence of noises about 10 db and the results of detection show that the proposed method has good compliance on determination of attributes of the signals.

Keywords Neural network · Difference entropy · Power quality · Detection · Harmonics

1 Introduction

Power quality (PQ) has become a momentous problem for both utilities and customers, for its adverse effects on equipments. Electric loads have become more vulnerable to power quality. Load equipments with microprocessor-based controls and power electronics devices even though worsen the PQ are more sensitive to power variations. They may suffer failure, malfunction, or hardware damage during PQ

D. Kavitha (✉) · P. Renuga · M.S. Lakshmi

Department of Electrical and Electronics Engineering, Thiagarajar College of Engineering, Madurai, Tamilnadu, India
e-mail: dkavitha@tce.edu

P. Renuga
e-mail: preee@tce.edu

M.S. Lakshmi
e-mail: mseetha@tce.edu

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_79

737

events [1, 2]. When characterizing system disturbances, it is sufficient to monitor only the voltage signals. The process of converting raw measurement data into suitable PQ detection data involves data selection, information extraction, and assimilation and report presentation. These steps require advanced technology. Hence, the analysis of PQ disturbances, especially detection and location, needs to be attaching importance to. Nowadays, the scholars have carried through abundant researches on PQ detection and location, using wavelet transform [3], short-time Fourier transform, HHT transform [4], and mathematical morphology [5], et al., and have gained plentiful productions. Practically, the PQ signals are often captured with noise; therefore, the effective methods able to process such signals are indispensable. In the previous similar works [1–5], the ideal sine wave and PQ disturbances such as interruption, voltage sag, voltage swell, impulse, and oscillations transients are analyzed using different methods including Shannon entropy and the starting and ending instants of the disturbances are determined. The drawback in those works is that the authors have assumed a pure sinusoidal condition which cannot be a practical one.

Adaptive neural net (ANN) is able to handle the signals with noise and can extract the harmonics from fundamental effectively [6–8]. Entropy which can open out the developing direction of the things and measure the uncertainty and disorder of material systems has widely applied in power system [9, 10]. This paper presents a new simple PQ detection method on basis of adaptive neural network and difference entropy. The adaptive neural network has been applied in the area of PQ for the estimation of harmonics, since the network convergence is fast and the results are accurate. In this paper, a combined method using ANN and Shannon entropy is used which employs neural net as a harmonics extracting unit and the difference entropy as a feature extracting unit. The organization of the paper is as follows: In Sect. 2, the basic theories and proposed methodology including ANN and Shannon entropy are explained. Next, Sect. 3 describes the results and discussions on proposed detection method. Conclusions are finally shown in Sect. 4.

2 Proposed Methodology

a. Adaptive neural networks

In this paper, ANN is implemented to eradicate the monitoring problems occur due to the presence of harmonics. Neural networks have self-adapting and super-fast computing features that make them well suited to handle nonlinearities, uncertainties, and parameter variations that can occur in the system. The error minimization between actual and estimated signals is actually done. The general form of a harmonic signal is given by Eq. 1,

$$f(t) = \sum_{k=1}^N A_k \sin(\omega_k t + \varphi_k) \quad (1)$$

where A_k is the amplitude, and φ_k is the phase angle of the k th harmonic signal. N is the total harmonics to be estimated. ω_k is the k th harmonic frequency. The Eq. 1 can be rewritten as Eq. 2 to model the adaptive neurons effectively.

$$f(t) = \sum_{n=1}^N [X_n \cos(n\omega t) + Y_n \sin(n\omega t)] \tag{2}$$

$$X_n = A_n \sin \varphi_n \tag{3}$$

$$Y_n = A_n \cos \varphi_n \tag{4}$$

where X_n and Y_n are the amplitude of the cosine and sine components of order- n harmonic. The vectorial notation of Eq. 2 is given in Eq. 5.

$$f(t) = W^T \cdot x(t) \tag{5}$$

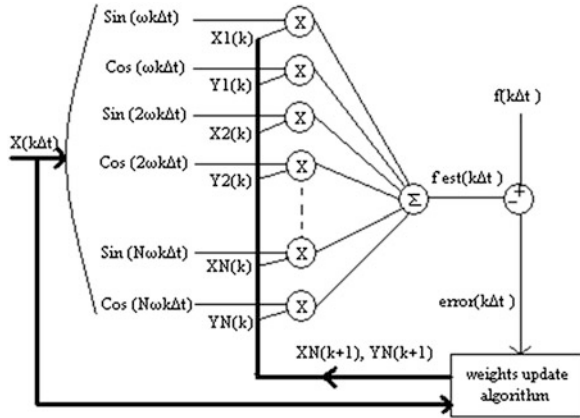
where

$$W^T = [X_1 Y_1 X_2 Y_2 \dots X_N Y_N] \tag{6}$$

$$X(t) = \begin{pmatrix} \cos(w_1 t) \\ \sin(w_1 t) \\ \cos(w_2 t) \\ \sin(w_2 t) \\ \dots \\ \dots \\ \cos(w_N t) \\ \sin(w_N t) \end{pmatrix} \tag{7}$$

The signals are sampled at a uniform rate Δt , so time values are discrete, $k\Delta t$ with $k = 0, 1, 2, \dots$. The dot product presented in Eq. 20 is carried out by one Adaline neuron, where W is the network weights vector. After the initial estimation, an adaptive algorithm updates the weights. Thus, the estimated signal converges to the actual one. Figure 1 shows the network topology and the weights update algorithm. At time k , $x(k)$ is the proposed signal model and $f(k)$ is the actual signal. The neurons, taking into account their weights $W(k)$, carry out an estimation $f_{est}(k)$. The error $e(k)$ is the difference between the actual signal and its estimation. An Adaline algorithm allows the weights to be used in the next iteration $W(k + 1)$ to be obtained, which minimizes that error. After this iterative process, the estimated signals adapt to the actual signals. In adaptive neural network technique, number of inputs for the neural net depends on the number of harmonics to be estimated. Total number of inputs required for the estimation of 7 harmonic terms is 14 since each harmonic term requires two inputs, one for sine value and other for cosine value. The learning rate is taken as 0.1. The waveform to be estimated is sampled with the frequency of 3,200 Hz and hence the total number of samples given is 64.

Fig. 1 Adaptive network topology



b. Shannon Entropy

Entropy in information theory is a measurement unit of information based on probability-statistic model, which can show the complexity and disorder degree of information. When a disturbance occurs, the signals' magnitude and frequency will change and its entropy will change correspondingly. In this way, we can detect disturbance by entropy variety. Equations 8–10 explain the concept of difference entropy.

$$D(n) = f'(n + 1) - f'(n), n = 0, 1, \dots, N - 2 \tag{8}$$

$$P_n = P_{[D(n)]} = \frac{\text{abs}[n]}{\sum \text{abs}[D(n)]} \tag{9}$$

$$H(n) = - \sum_{n=1}^L P_n \log_2(p_n) \tag{10}$$

The six different signals considered for detection are ideal sine wave, interruption in the signal for few milliseconds, voltage sag, voltage swell, impulse in the signal for few milliseconds, and oscillations transients. These six signals will possess the characteristics of the following equations

Ideal sine waveform

$$f(t) = A \sin(\omega t) \tag{11}$$

Interruption

$$f(t) = [1 - \alpha(f(t_2 - t_1))]\sin(\omega t) \quad (12)$$

where $\alpha > 0.9$ is amplitude change factor; $0.5T < t_2 - t_1 < 3$ s is durative time of disturbance.

Voltage sag

$$f(t) = [1 - \alpha(f(t_2 - t_1))]\sin(\omega t) \quad (13)$$

where $\alpha = 0.1 \sim 0.9$ is amplitude change factor; $0.5T < t_2 - t_1 < 30T$ is durative time of disturbance.

Voltage swell

$$f(t) = [1 + \alpha(f(t_2 - t_1))]\sin(\omega t) \quad (14)$$

where $\alpha = 0.1 \sim 0.8$ is amplitude change factor; $0.5T < t_2 - t_1 < 30T$ is durative time of disturbance.

Impulse

$$f(t) = \sin(\omega t) + \alpha [f(t_2) - f(t_1)] \quad (15)$$

where $\alpha = 1 \sim 3$ is amplitude of impulse; $1 \text{ ms} < t_2 - t_1 < 3 \text{ ms}$ is durative time of disturbance.

Oscillation transient

$$f(t) = \sin(\omega t) + \alpha e^{-c(t-t_1)} \sin(\beta \omega t) [f(t_2) - f(t_1)] \quad (16)$$

where $\alpha = 0.1 \sim 0.8$ is amplitude of oscillation; $\beta = 0.1 \sim 0.5$ is fluctuation frequency relative coefficient; c is damped oscillations coefficient; $0 < t_2 - t_1 < 2T$ is durative time of disturbance.

3 Results and Discussions

Test signals are generated using MATLAB software using Eqs. 11–16. The sampling frequency is assumed 4 kHz and three full wave is considered for analysis. PQ disturbances are created between 25 and 37.5 ms.

The proposed algorithm is applied to determine the starting and ending instants of the disturbances present in the practical voltage signal. The consumer units in distribution systems have inductive and nonlinear characteristics. These nonlinear

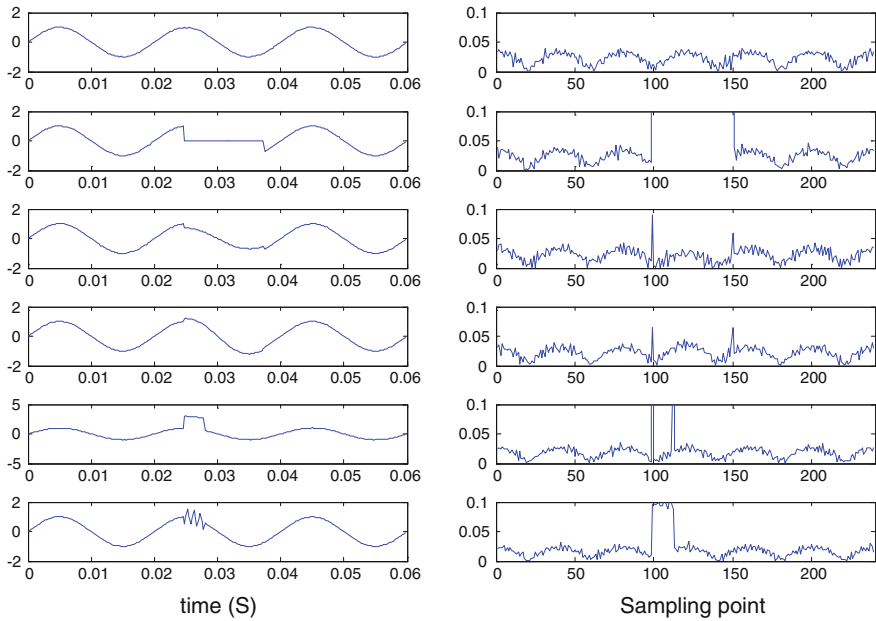


Fig. 2 Test signals and their entropy without harmonics

loads inject harmonic currents of several orders into utility electric system leading to non-sinusoidal situations. Hence, it becomes necessary to analyze the signal including harmonics. If the Shannon entropy alone is applied for the signals in the presence of harmonics, simulations show that the time of transient disturbances cannot be obtained. Hence, adaptive neural network-based harmonic detection algorithm is initially used to split up the harmonics present throughout the signal and Shannon entropy is applied for the signal after removing the harmonics.

Figure 2 shows the entropy for the six signals considered in the absence of harmonic components. The sampling frequency is fixed as 4 kHz. Figure shows the signal with harmonics and the results of ANN algorithm (Fig. 3 and Table 1).

The signal is distorted with a THD of 11.5 %. After the estimation of harmonics, the signal is split up into individual harmonics. The harmonics are summed up and the value of harmonics at each sampling point is determined. The harmonic components are determined using the following equation.

$$HC = \sum_{k=1}^N A_k \sin(w_k t + \varphi_k) - A_1 \sin(w_1 t + \varphi_1). \tag{17}$$

where HC is harmonic component. HC is an array of values at each sampling point. These components are fixed constant and will be extracted from the signal to remove the harmonics for entropy estimation. Difference entropy for ideal sine wave is distributing average implying that there is no disturbance in it. The

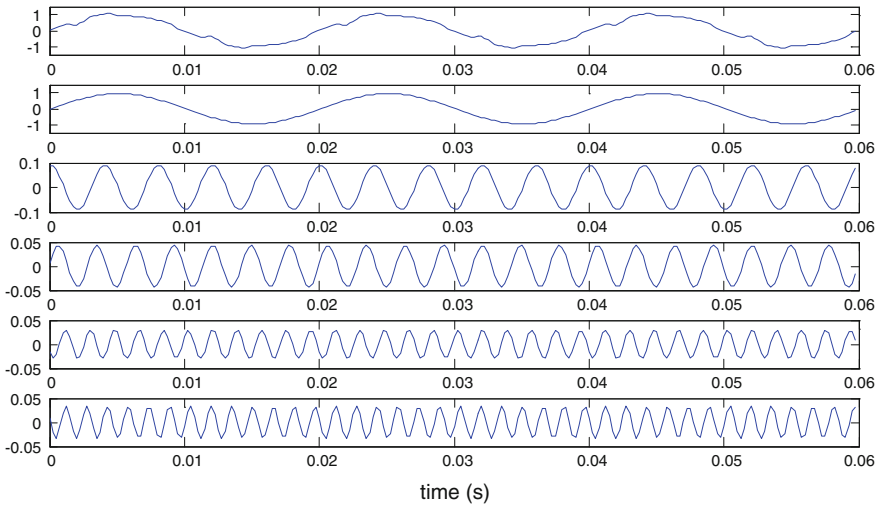


Fig. 3 Output waveforms from ANN module

Table 1 Results from ANN

Harmonics (h)	1	3	5	7	9	11	13
Magnitude (p.u)	0.950	0.000	0.090	0.043	0.000	0.030	0.033
Phase angle (rad)	-0.0353	0.0000	1.4329	0.1379	0.0000	-2.5674	2.837

difference entropy value increases drastically in the instant of disturbance. Figures 5 and 6 show the variation in entropy for different noise values. The starting instant and ending instants of a disturbance occurring in a signal can be recorded accurately and effectively using the proposed method (Fig. 4).

a. Effect of noise in detection

Random noise is added to the signal with various signal-to-noise ratio ranges from 40db to 8db. The simulation results are given for noises of 40 and 10 db. From the results, it is obvious that the noise plays important role in the performance of ANNDE method in Fig. 5. The stronger noise heavily affects the performance of the method.

b. Effect of amplitude in detection

From the figure, it is seen that the sharp thresholds are obtained using ANNDE method even in the presence of noise. The method failed to obtain the disturbing instants in some cases in the presence of noise. When voltage sag magnitude is around 0.9, the proposed method is not able to feature out the sag. Also for voltage swell, the magnitude around 1.1 is undiscoverable. All values except the stated are discoverable. Table shows a summary of exhaustive simulations done (Table 2).

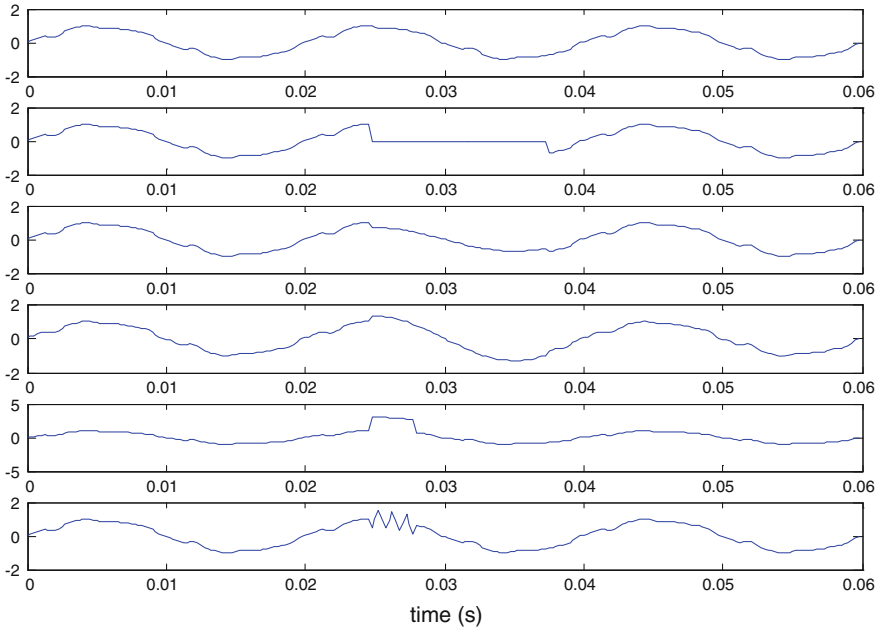


Fig. 4 Test signals with harmonics

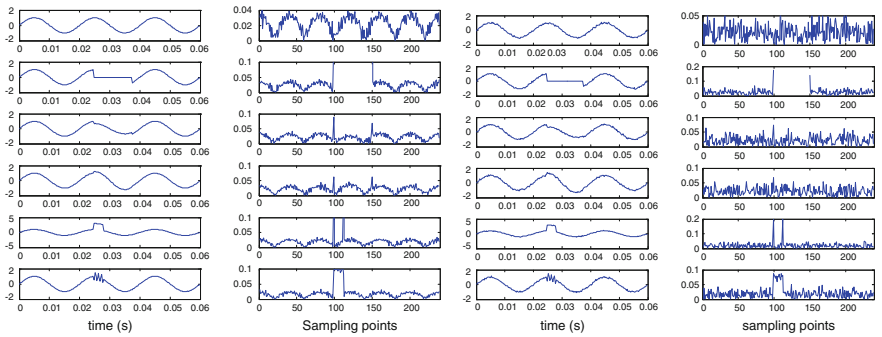


Fig. 5 Test signals with harmonics and their entropy at 40 and 10 db

Table 2 Summary of exhaustive simulations

S. No	Disturbances	Noise	Amplitude	Begin instant detection	End instant detection
1	Interruption	db > 10	0	√	√
2	Voltage sag	db > 15	a > 0.85	x	x
			0.1 ≤ a ≤ 0.85	√	√
3	Voltage swell	db > 15	1.1 ≤ a ≤ 1.8	√	√
4	Impulse	db > 10	1 ≤ a ≤ 3	√	√
5	Oscillation transient	db > 10	0.1 ≤ a ≤ 0.8	√	√

4 Conclusion

Power quality disturbances detection methodology based on ANN and Shannon difference entropy algorithm is proposed in this paper. Vast simulations are done to validate the method adaptableness and found that the performance of the algorithm is better than the conventional algorithms. ANN extracts the harmonics present in the signal and difference entropy algorithm can extract characteristic hiding in the signals. The proposed method has good noise resistance ability. The algorithm is simple and easy to implement in real time.

References

1. Z. Jing, H. Zhengyou, Q. Qingquan, Detection of power quality disturbances based on generalized morphological filter and information theory. *IEEE*. (2009)
2. Yop Chung, Dong-Jun Won, Joong-Moon Kim et al., Development of a network-based power quality diagnosis system. *Electr. Power Syst. Res.* **77**, 1086–1094 (2007)
3. A.S. Yilmaz, A. Subasi, M. Bayrak et al., Application of lifting based wavelet transforms to characterize power quality events. *Energy Convers. Manage.* **48**(1), 112–123 (2007)
4. F. Jurado, J.R. Saenz, Comparison between discrete STFT and wavelets for the analysis of power quality events. *Electr. Power Syst. Res.* **62**(3), 183–190 (2002)
5. S. Ouyang, J. Wang, A new morphology method for enhancing power quality monitoring system. *Int. J. Electr. Power Energy Syst.* **29**(2), 121–128 (2007)
6. V. Suresh Kumar, D. Kavitha, K. Kalaiselvi, P.S. Kannan, Harmonic Mitigation and Power Factor Improvement using Fuzzy Logic and Neural Network Controlled Active Power Filter. *Journal of Electrical Engineering & Technology.* **3**(3), 520–527 (2008)
7. D. Kavitha, A.F. Zobaa, P. Renuga, V. Suresh Kumar, NSGA-II optimized neural network controlled active power line conditioner under non-sinusoidal conditions. *Int. Rev. Electr. Eng. (IREE).* **6**(2011)
8. V. Suresh Kumar, D. Kavitha, K. Kalaiselvi, P.S. Kannan, Optimal estimation of harmonics in power system using intelligent computing techniques. *Proceedings of IEEE Neural Networks conference.* (2007)
9. Catarina Moreira, A. Wichert, Finding academic experts on a multisensor approach using Shannon's entropy. *Expert Syst. Appl.* **40**(14), 5740–5754 (2013)
10. Yuequan Bao HuiLi, J. Ou, Structural damage identification based on integration of information fusion and shannon entropy. *Mech. Syst. Signal Process.* **22**, 1427–1440 (2008)

Texture Feature Extraction Using MGRLBP Method for Medical Image Classification

Suganya Ramamoorthy, R. Kirubakaran
and Rajaram Siva Subramanian

Abstract Texture is an important significant property of medical images based on which images can be characterized and classified in a content-based image retrieval and classification system. This paper examines the feature extraction methods to ameliorate texture recognition accuracy by extracting the rotation-invariant texture feature from liver images by the individual Gabor filter method and by multi-scale Gabor rotation-invariant LBP (MGRLBP) method. The features extracted from both the approaches are tested on a set of 60 liver images of four different classes. The classification algorithms such as support vector machine (SVM) and k-nearest neighbor (KNN) were used to evaluate the extracted features from both methods, showing advancing improvements with the MGRLBP method over the individual method in the classification task.

Keywords Texture · Feature extraction · Texture analysis · Rotation invariant

1 Introduction

In addition to color and shape, texture is another prominent and essential feature for medical image classification process. The classification activity can be done through similarity matching with the help of the extracted texture features. Texture of an image can be defined by its distinct properties such as coarseness, inherent direction, and pattern complexity which are usually sensed from the variations in

S. Ramamoorthy (✉) · R. Kirubakaran
Department of CSE, Thiagarajar College of Engineering, Madurai, India
e-mail: rsuganya@tce.edu

R. Kirubakaran
e-mail: kirubakaran@tce.edu

R.S. Subramanian
Department of ECE, Thiagarajar College of Engineering, Madurai, India
e-mail: rajaram_siva@tce.edu

the scale and orientations in the texture pattern. Texture analysis has become an active research area in the past decade, and researchers have proposed numerous methodologies to automatically analyze and recognize texture which helps the medical professionals to carry out the diagnosis. The basic research on texture analysis starts with the derivation of texture energy measures and visual, textural features such as coarseness, contrast, directionality, entropy, and homogeneity. However, most of the existing texture analysis methods were carried out with the assumption that texture images are acquired from the same viewpoint, i.e., same scale and orientation. The multiview representation of the image can enhance texture analysis process, and it is proved to be much effective in qualifying texture features of a medical image.

2 Related Work

Visual features of liver tissues can be explained numerically in a number of ways [1]. The simple gray-level distributions are frequently used, because the intensities describe physical properties of liver tissues effectively. Second-order statistics such as the gray-level co-occurrence matrices (GLCM) and run length (RLE) [2] have also been widely incorporated for additional feature information. Another type of popular feature extraction techniques is based on filters and is also used for certain applications in liver imaging.

Tesar et al. [3] examined the specific 3D extent of Haralick features for effective segmentation of three-dimensional CT scans of the abdominal area. This approach is not suitable for complex textures. Bharathi et al. [4] used a new statistical approach for the discrimination of normal liver and abnormal liver using orthogonal moments as features due to the extraneous nature, but computation time is high. Huang et al. [5] extracted texture features from the ultrasonic liver images using three different spatial techniques such as gray-level histogram, gray-level difference statistics, and gray-level co-occurrence matrix which significantly discriminates the normal liver and fatty liver images. The size of the dataset used is too small in this approach. Riaz et al. [6] proposed a novel descriptor, autocorrelation Gabor features (AGF) for the classification of gastroenterology images using invariant Gabor descriptors. Farhan used autocorrelation homogeneous texture (AHT) as a region-based descriptor, but the author did not use color descriptors for classification.

Mittal et al. [2] proposed a system and use four different feature extraction techniques such as first-order statistics (FOS), spatial gray-level dependency matrix (SGLDM), gray-level run-length matrix (GLRLM), and texture energy measures (TEM) to identify focal liver lesions in B-mode ultrasound images, but the classification accuracy is low. Parekh [7] used gray-level co-occurrence matrix and wavelet decomposition matrix methods for feature extraction process to recognize human skin diseases. The author considered only 8-bit images for analysis. Rajendra Acharya et al. [8] proposed a descriptive method for glaucoma detection using a combination of texture and higher-order spectra features (HOS) in which

the feature extraction is performed using GLCM and run-length matrix methods and the classification is performed with SVM, naive Bayesian, and random-forest classifiers. The limitation with this approach was that the size of the diverse medical images used for classification is too small.

Song et al. [9] proposed an automatic classification method for different categories of HRCT lung images in which higher descriptiveness was obtained by combining multiple methods such as local binary pattern, Gabor filters, and histogram of oriented gradients method. The limitation with this approach is that the classification accuracy of complex lung tissue category is low. Kayaalti et al. [10] proposed a noninvasive and fast approach to specify fibrosis using texture properties. Omer used gray-level co-occurrence matrix, discrete wavelet transform, and discrete Fourier transforms for feature extraction. The success rate of this approach is very low and not sufficient for clinical diagnosis. Virmani et al. [11] differentiate normal and cirrhotic liver based on Laws' masks analysis. Jitendra used a methodology of designing computer-aided diagnostic system with optimal Laws' texture features and a neural network classifier for effective discrimination between normal and cirrhotic-segmented regions, but the dataset used for the classification is too small.

3 Research Contribution

A medical image classification system is a system that classifies the multiple categories of medical images from the available large digital medical database, through feature extraction for encoding the image features as feature descriptors. This feature extraction process extracts the most promising and effective feature as descriptors from the query images. In the proposed system, a rotation-invariant texture feature descriptor is extracted using individual rotation-invariant Gabor filter method and by MGRLBP method, which combines the rotation-invariant property of LBP and multi-scale property of Gabor filters. Based on the extracted rich-texture feature, the liver images are classified using machine learning algorithms such as support vector machine (SVM) classifier and k-nearest neighbor (KNN) classifier. The performance of the texture classification can be measured by calculating the accuracy of the classifiers.

4 Proposed MGRLBP Method

A texture feature is a characteristic that can capture a certain visual property of an image either globally for the whole image or locally for objects or regions. Texture is the variation of data at different scales and at different rotations. The proposed MGRLBP method incorporates two feature extraction methods such as local binary pattern (LBP) and Gabor filter to extract the rotation-invariant texture feature which discriminates the normal liver from the affected liver effectively. The LBP feature

describes the spatial structure of local image texture in which the rotation invariance can be achieved. On the other hand, the multi-scale and multi-orientation representation of Gabor filters is often demonstrated as a highly effective texture descriptor, and its multi-scale nature is quite useful for computing multi-resolution LBP features. Therefore, to incorporate rich-texture information while attempting to minimize intracategory variations, the proposed MGRLBP feature extraction method combines the multi-scale property of Gabor filters and the rotation-invariant property of LBP features. The proposed system is shown in Fig. 1.

The rotation-invariant texture feature is extracted in the system using Gabor filters in which all Gabor functions of certain scale with different orientation angles as 30°, 45°, 60°, and 90° are summed together. Then, for each Gabor-filtered image, a rotation-invariant LBP feature is computed for each pixel in the image as follows:

$$I^s(x, y) = \sum_{r=0}^{R-1} I^{s,r}(x, y) \tag{1}$$

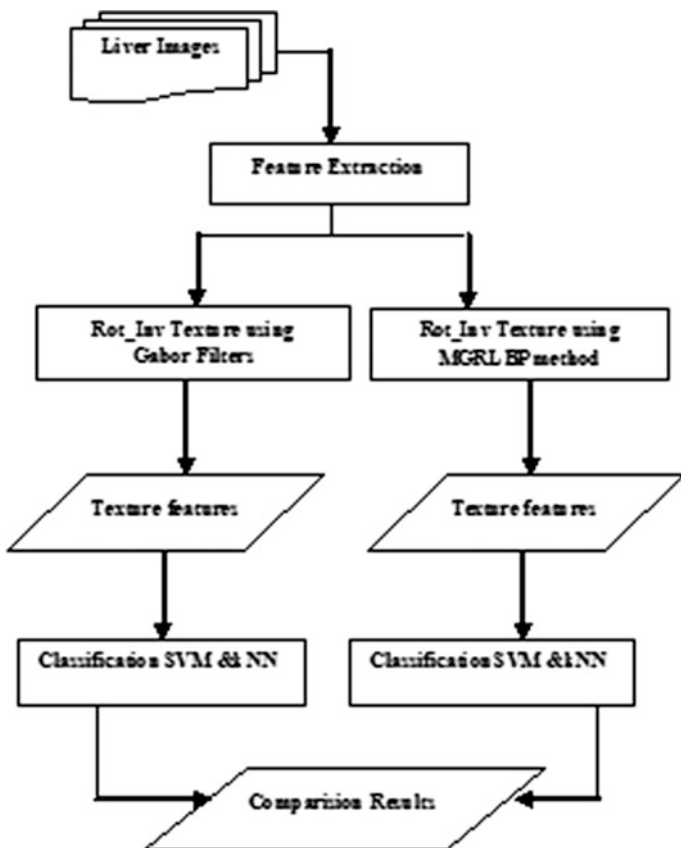


Fig. 1 MGRLBP method for texture feature extraction and classification

where r and s denote the orientations and scale, respectively, and $I^{s,r}$ denotes the Gabor-filtered images, obtained by convolving image I with Gabor functions of R orientations. On the other hand, frequency and orientation representations of Gabor filters resemble human visual system, and they have been found to be very effective for representing texture. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The rotation-invariant texture feature can be extracted by modulating the conventional Gabor filter method with respect to a constant scale, i.e., by summing all the orientations of 30° , 45° , 60° , and 90° at each scale level, which extracts features from a specific scale band covering all the orientations of the specified liver image.

5 Dataset Evaluation and Metrics

For the evaluation purpose, ultrasound liver images of four different categories such as hemangioma, fatty liver, cyst liver, and normal liver are collected from various hospitals and scan centers along with the guidance of the medical professionals. There are totally 56 images of which 15 images are normal, 15 images are hemangioma, 13 images are fatty liver, and 13 images are cyst liver.

To evaluate the proposed method, invariant texture features are extracted from all the four different categories of liver tissues using both the conventional Gabor filter and the MGRLBP method. The rich invariant texture is extracted under different orientations as 30° , 45° , 60° , and 90° with constant scale and describes the unique variations in each tissue category. Then, the medical image classification task is performed based on the extracted invariant texture features. Two standard classifiers, namely SVM and k-Nearest Neighbor (kNN), were used. Support vector machine is supervised learning model that takes a set of input data, which of two classes predicts the output, making it a non-probabilistic binary linear classifier. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, used for classification, and the hyperplane that has the largest distance to the nearest training data point of any class reduces the generalization error of the classifier. The distance between the image features is calculated using the Euclidean distance formula:

$$E(I, J) = \sqrt{\sum |f_i(I) - f_i(J)|^2} \quad (2)$$

where $f_i(I)$ and $f_i(J)$ denote the feature values of images I and J . Similarly, kNN is a nonparametric learning algorithm. k-NN assumes that the data are in a feature space and operate on the premises that classification can be done by relating the unknown to the known based on certain distance or similarity function. Each of the training data consists of a set of vectors and class label associated with each vector. It may be either + or - (for positive or negative classes). But k-NN works equally well with arbitrary number of classes.

6 Results and Discussion

Performance of the classification is measured by calculating retrieval accuracy using the following relation:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positive} + \text{True Negatives} + \text{False Positives} + \text{False Negatives}}$$

where true positive refers to the proportion of the positive images that are correctly identified and true negative refers to the proportion of the negative images that are correctly identified from the total positive and negative images. Performance of the proposed MGRLBP feature extraction method is compared with conventional Gabor method. MGRLBP method gives more satisfied classification results (88 % accuracy) than individual feature extraction method (79 % accuracy). Figure 2 shows the implementation of MGRLBP method on Hemangioma liver category in which the Gabor filter with multiple orientations is summed together. Then, the local binary pattern is applied for the Gabor-filtered image and the changes in intensities are observed with the histogram. The classification accuracy of texture features for different liver tissues is shown in Table 1, and Table 2 shows the performance comparison between the feature extraction methods. From Table 2, it can be inferred that performance of the classification process is more enhanced with the compound method than the conventional method.

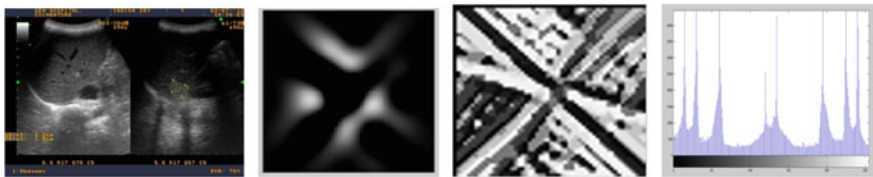


Fig. 2 Implementation of MGRLBP method on hemangioma liver category

Table 1 Classification accuracy of texture features for various liver tissue categories

Liver tissue categories	MGRLBP	Gabor
Hemangioma	82.52	77.87
Fatty liver	80.70	71.26
Cyst liver	79.88	73.34
Normal liver	88.31	79.65

Table 2 Performance comparison between feature extraction methods

S. No.	Classification techniques	Gabor filters	MGRLBP method
1	k-nearest neighbor	73.33	86.65
2	Support vector machine	79.86	88.23

7 Conclusion and Future Work

A combined feature extraction method is proposed which increases the texture discriminative ability on different classes of liver images, and it reduces the intraclass variations within the medical image. The accuracy of the classification system is also improved to (88 %) in MGRLBP method when compared with the individual feature extraction methods (79 %). In future, texture feature can be combined with other features such as gradient and intensity to calculate the feature vector, and also, the classification is performed by other classification schemes such as sparse-based approximation method and dictionary construction method.

Acknowledgments The authors convey their heartfelt thanks to Dr. R. Sambath, Radiologist, and Dr. P.S. Rajan, MS of GEM Hospital, Coimbatore, for providing the medical image dataset used in this paper, and also Dr. Kasthurimohan, MD of Malar Hospital, Dindigul, and Dr. Mahalakshmi, DGO of Meenakshi Mission Hospital, Madurai, for their motivation and support for conducting this work and valuable suggestions at different stages of the work.

References

1. I. Sluimer, A. Schilham, M. Prokop, B. van Ginneken, Computer analysis of computed tomography scans of the lung: a survey. *IEEE Trans. Med. Imaging* **25**(4), 385–405 (2006)
2. D. Mittal, V. Kumar, N. Khandelwal, N. Kalra, Neural network based focal liver lesion diagnosis using ultrasound images. *Comput. Med. Imaging Graph.* **35**(4), 315–323 (2011)
3. L. Tesar, A. Shimizu, D. Smutek, H. Kobatake, S. Nawano, Medical image analysis of 3D CT images based on extension of Haralick texture features. *Comput. Med. Imaging Graph.* **32**(6), 513–520 (2008)
4. V.S. Bharathi, V.S. Raghavan, L. Ganesan, Texture classification using Zernike moments, in *Proceedings of 2nd FAE International Symposium* (2002), pp. 292–294
5. Y. Huang, X. Han, X. Tian, Z. Zhao, J. Zhao, D. Hao, Texture analysis of ultrasonic liver images based on spatial domain methods, in *3rd International Congress on Image and Signal Processing (CISP)* (2010)
6. F. Riaz, F.B. Silva, M.D. Ribeiro, M.T. Coimbra, Invariant gabor texture descriptors for classification of gastroenterology images. *IEEE Trans. Biomed. Eng.* **59**(10), 2893–2904 (2012)
7. R. Parekh, Using texture analysis for medical diagnosis. *IEEE Computer Society* (2012)
8. U. Rajendra Acharya, S. Dua, X. Du, S. Vinitha Sree, C.K. Chua, Automated diagnosis of glaucoma using texture and higher order spectra features. *IEEE Trans. Inf. Tech. Biomed.* **15** (3), 449–455 (2011)
9. Y. Song, W. Cai, Y. Zhou, D.D. Feng, Feature-based image patch approximation for lung tissue classification. *IEEE Trans. Med. Imaging* **32**(4), 797–808 (2013)
10. O. Kayaalti, B.H. Aksebzeci, K. Deniz, M. Ozturk, S. Kara, Staging of the liver fibrosis from CT images using texture features, in *International Conference on Health Informatics and Bioinformatics (HIBIT)* (2012)
11. J. Virmani, V. Kumar, N. Kalra, N. Khandelwal, Prediction of cirrhosis from liver ultrasound B-mode images based on laws' masks analysis, in *International Conference on Image Information Processing (ICIIP)* (2011)
12. KH. Hwang, H. Lee, D. Choi, Medical image retrieval: past and present. *Healthc. Res. Inf.* **18** (1), 3–9 (2012)

A Novel Lightweight Protocol for Address Assignment in Ad Hoc Networks Based on Filters

M. Anusuya Shyamala and R. Velayutham

Abstract Ad hoc networks play a vital role in latest technologies, but assigning address to the nodes in ad hoc network is a major challenge due to its lack of infrastructure. A lightweight protocol is proposed in this work which helps to configure the mobile ad hoc nodes which are based on a distributed address database that is stored in filters which reduces the control load and also makes the proposal strong in partition of networks and to loss of packets. The performance of the protocol is evaluated by considering the joining nodes, partition merging events, and initialization of networks. The result of control message reduction and address collisions for the proposed protocol is shown in simulation results.

Keywords Ad hoc networks · FAP · DAD · MANETconf

1 Introduction

Mobile ad hoc networks are a special category of networks exclusively classified on the basis of attributes like dynamic topology and infrastructurelessness [1]. Due to the dynamic nature and inherent infrastructureless architecture, solutions developed for configuration and deployment of infrastructure-oriented networks cannot be directly applied in MANETs [2]. Assigning address in ad hoc networks is a challenging one due to its self-organized environment. Network address translation (NAT) or dynamic host configuration protocol (DHCP) [3] feels difficult with the distributed feature of ad hoc networks and also does not address with the network partition and node merging.

M. Anusuya Shyamala (✉) · R. Velayutham
CSE Department, Einstein College of Engineering, Tirunelveli, Tamil Nadu, India
e-mail: mshyashya@gmail.com

R. Velayutham
e-mail: rsvel_kumar@yahoo.co.uk

An efficacious approach called as filter-based addressing protocol (FAP) is proposed in this work [4]. A distributed database allocated address is maintained by FAP, thus database is stored in filters. These addresses are stored in a tamped manner. Both the bloom filter and the sequence filter are considered to design a filter-based protocol that ensures both the univocal address configuration of the nodes joining the network and detection of address collisions because each node can easily identify whether the address has been already allotted or not. Hash of this filter is used here as the partition identifier in order to find the network partitions. The hash of the filters is exchanged with the neighbors so that address collisions could be found by detecting the small control overhead where neighbors using different filters and then the filters are maintained and distributed with the nodes in the network .

This paper is structured as follows. Section-I contains the introduction. Overview of the related work is given in Sect. 2. Section 3 contains the proposed system. System design is given in Sect. 4. Simulation results are given in Sect. 5. Finally, conclusion of the paper is given in Sect. 6.

2 Existing System

2.1 Duplicate Address Detection (DAD)

Duplicate address detection (DAD) protocol is used as an addressing protocol where all the joining nodes selects an address in a random manner and informs the network with an Address REQuest message (AREQ). If the randomly chosen address is already allocated to another node, the already existing node announces the duplication to the joining node by replying with an Address REPLY message (AREP). When the joining node receives an AREP, the joining node again selects another address and repeats the flooding process or else it allocates the chosen address. This protocol does not mind about network partitions and is not appropriate for ad hoc networks. The WDAD proposed aims at extending the DAD mechanism [5]. The idea behind WDAD is that duplicate addresses may be tolerated as long as packets reach the destination node intended by the sender .

The main drawback of WDAD is its dependency on the routing protocol. It requires some changes to the routing layer to support the introduction of the key identity. Each node will be identified at the routing layer by a kind of virtual address consisting in the combination of the IP address and the key value. In addition, WDAD detects address duplication based on local routing information; thus, it is totally adapted to proactive routing where each node maintains a complete routing table. For reactive routing, it is not the case where the nodes cache partial routing information for only ongoing and relayed connections, which reduces the possibility of detecting in moderate delays address duplication.

2.2 MANETconf

MANETconf is based on a “common distributed address table” where each node is able to assign IP addresses and maintains an allocation table that contains already allocated addresses and pending allocations [6]. Thus, the synchronization of these distributed tables constitutes the most critical and complex task of this protocol.

The advantages of this protocol are that it guarantees address uniqueness and it is totally distributed in terms that each node has the possibility to assign new addresses. In addition, it generates no unnecessary address changes when networks merge because only nodes involved in duplication release their IP addresses.

The problems of this protocol are its high complexity in terms of communication, table maintenance, and synchronization. The mechanism for assigning new addresses is bandwidth consuming; it consists of a network flood and a large number of unicast. All nodes should give their permission to the initiator to assign a new by another node. That’s why each node generates large delays. Finally, this protocol is very sensible to network losses because of its dependency on unicast communication.

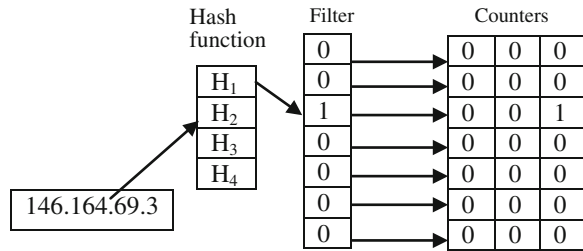
3 Proposed System

The proposed system is mainly used to auto-configure the network addresses dynamically, solving collisions with a low control load and in joining or merging events. This work uses FAP to achieve the addressing problems. FAP uses a distributed compact filter to represent the current set of allocated addresses [4]. Frequent node joining events can be simplified by the filters that are present at every node and also reduce the control overhead that is required to solve address collisions implicit in random address assignments. The filter signature is the hash of the address filter. It is considered as a partition identifier. Detection of network merging events can be easily found by filter signature. Two different types of filters are used in this work. They are bloom filter and sequence filter. The bloom filter is purely based on hash functions, and the sequence filter is used for compressed data based on the address sequence.

3.1 Bloom Filters

The bloom filter is a compact data structure which is most often used in distributed applications. It consists of an m -bit vector that represents a set $A = \{a_1, a_2, \dots, a_n\}$ through a set of independent hash functions h_1, h_2, \dots, h_k , and the elements are inserted into the filter. Initially, all the bits in the filter are set to zero, and after that, all the elements are hashed where the output is set to 1 [7]. To check whether an

Fig. 1 Bloom filter



element is present in A, check whether the bits corresponding to $h_1(a_j), h_2(a_j), \dots, h_k(a_j)$ are all set to 1, and if any one bit is 0, then a_j is not present in A. This shows the false-positive probability that an element $a_j \notin A$ be recognized as being in A. Such cases may occur when the bits at the positions $h_1(a_j), h_2(a_j), \dots, h_k(a_j)$ are all set by previously inserted elements (Fig. 1).

3.2 Sequence Filters

Sequence filter stores and compacts addresses based on the sequence of addresses [7]. By concatenating the initial element that is the first element of the address sequence with an n-bit vector address range, the filter is created. Here, each address suffix is represented by one bit, indexed by Δ , which gives the distance between the initial element a_{suffix} and the current element a_{suffix} . The address with the given suffix is considered as inserted into the filter if the bit is in 1, or else if the bit is in 0, it says that the address is not in the filter. Since the available address is represented by its respective bit, neither false positives nor false negatives are present in sequence filter.

3.3 Selection of Filter

Network characters like number of nodes in the network and number of available addresses make to select the best filter for FAP. False-positive and false-negative rates of the filter are also considered. False negatives are not presented by the bloom filters, which show a membership test for an element that was placed in the filter is always positive. Though filters present a false-positive probability, a membership test of an element that was not placed into the bloom filter may be positive, whereas the sequence filter size is constant for the number of elements and the address range size increases. As a result, the bloom filter is more suitable for a large address range, whereas the sequence filter is more enough to a large number of elements. Thus, the filter can be selected based on our need.

4 System Design

4.1 Network Initialization

Auto-configuration of initial set of nodes is handled in network initialization. Two different events can occur here: gradual initialization, where the node arrives at enough time gaps between them, and abrupt initialization, where all nodes arrive at the same time. The first node chooses a partition identifier, and joining nodes are handled through the joining node procedure by the first node [3]. High control load can occur when partition merging events are triggered; such case may happen when all the nodes join the network at the same time and choose different partition identifier. It may also lead to address collisions by causing inconsistencies in the address allocation. FAP is suitable for both gradual and abrupt initialization using Hello message, and AREQ message is used to advertise that the previously available address is now allocated. AREQ messages have a unique identifier number which can be used to distinct the AREQ messages of the nodes.

4.2 Node Ingress and Network Merging Events

After the initialization phase, every node broadcasts periodic Hello message containing its address filter signature. After receiving a Hello message, the neighbor node detects merging events by evaluating the signature and the signature in the message by comparing. The nodes which have already joined the network alone can send Hello messages, can receive the request from a node, and can detect merging events. A node turns ON and listens to a medium for a time period, and if it receives a Hello message, the node assumes that it is a joining node and not an initiator node [3]. Again the joining node asks the host for the source node to send the address filter network using Address Filter (AF) message. The host node checks the bit after receiving the AF if the message is used in a node joining procedure or in a partition merging procedure (Fig. 2).

4.3 Node Departure

The address of the node leaving from the network should be available for other nodes, and when a node departing it floods the network with a notification to remove the address, the available address may scarce with time which can be identified in the address filter by the fraction of bits set to 1 in the bloom filter and in the sequence filter and also by the fraction of counters greater than one in the

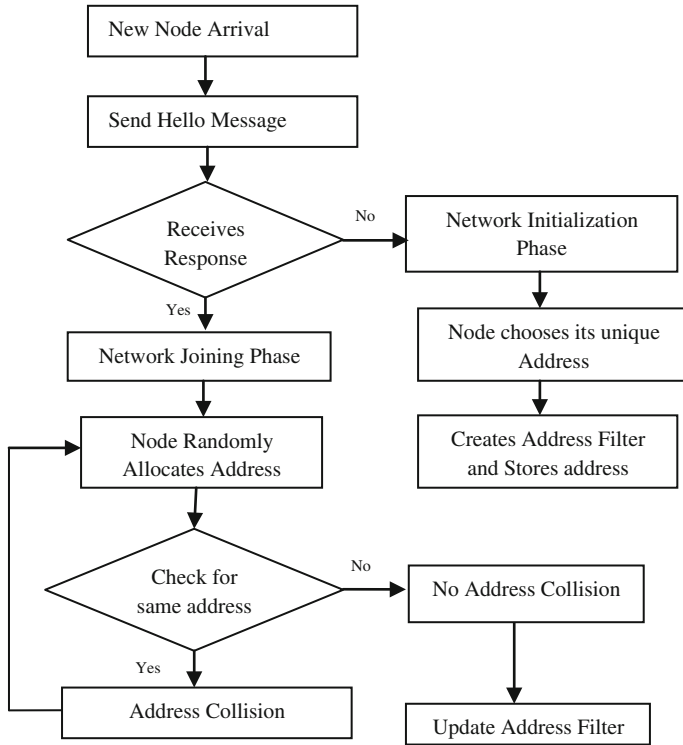


Fig. 2 Workflow architecture

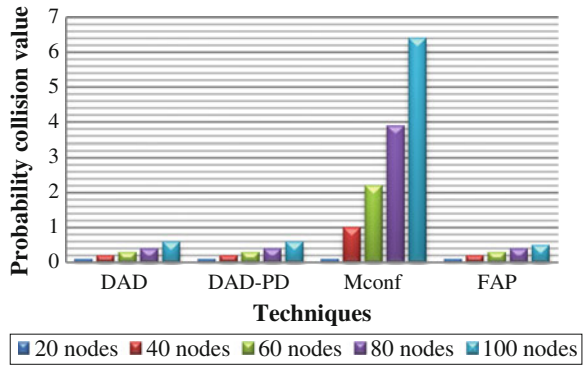
counter bloom filter. Hence, whenever the filter is updated, the nodes verify this fraction in the address filter every time. If the fraction value reaches the threshold, address filter is reseted (i.e.) filter returns to initial as it is full.

5 Result Analysis

5.1 Probability of Collisions in FAP

A collision normally occurs in two different cases. First case is when any two joining nodes produce AREQs of same address and also with the partition identifier, and here, joining nodes do not note that their addresses are same because the message from other node looks like a retransmission of its message to the first node. Second case is when two disjoint partitions own exactly the same filters, and here, signatures of the Hello messages are same for both the partitions and as a result, the network would have an address collision; thus, partition merging procedure is not started in this case.

Fig. 3 Collision graph



The graph (Fig. 3) shows that MANETconf is vulnerable to address collisions on increasing the number of nodes. In this protocol, the abrupt initialization occurs through parallel partition merging procedures, which are not robust and cause collisions in the addresses.

5.2 Control Overhead Estimation

The procedures in addressing protocol like network initialization, node joining/leaving, and merging reduce the available bandwidth by generating the overhead. FAP performs effective when compared to DAD-PD when one or more address collisions occur because this increases the number of floods only in DAD-PD and floods are the most costly operation. The ratio of the number of nodes in the network and the number of available addresses determines the value. MANETconf has a larger overhead compared to FAP because MANETconf is based on the assumption that all nodes must agree before allocating an address which demands many control messages. The initiator floods the network asking whether all nodes agree with the availability of chosen address (Fig. 4).

If all the other nodes agree with the chosen address, then the initiator floods the network again to allocate the chosen address. Besides the overhead caused by all the flood events, depending on the routing protocol, each unicasted message flow can imply in a flood to search for a route between the source and the destination nodes. Hence, an intensive use of uncast to different destinations, such as in MANETconf, can generate a high control overhead compared to FAP. Thus, the proposed protocol produces low control overhead.

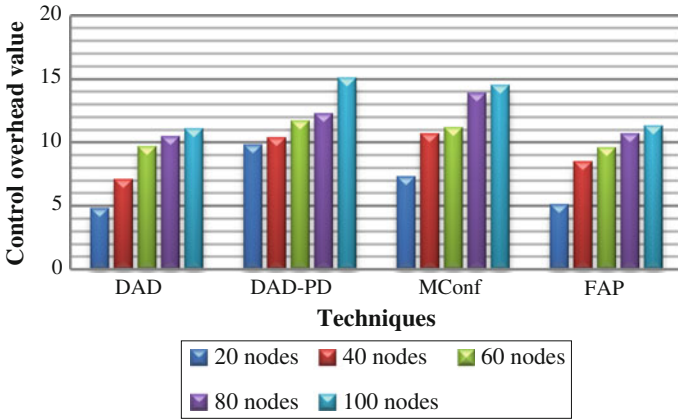


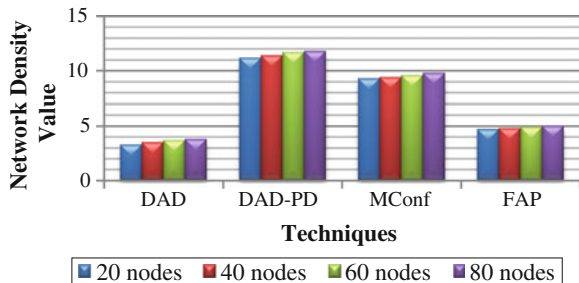
Fig. 4 Control overhead graph

5.3 Impact of Network Density

The impact of the network density is evaluated and the number of transmissions of flooding messages in abrupt network initialization. DAD-PD suffers a greater influence on the control load compared to FAP, MANETconf, and DAD because it floods the network many times due to the false positives in the partition merging detection. Compared to MANETconf and DAD-PD, FAP has a lower control load (Fig. 5).

The delay of FAP is lesser than the delay of DAD and MANETconf for a high number of nodes. DAD-PD has the greatest overlap delay and becomes unstable in networks with more number of nodes.

Fig. 5 Network density graph



6 Conclusion

Address assignment in ad hoc networks should be automatic, fast, and without collisions. A filter-based addressing protocol (FAP) is proposed, which uses address filters to reduce the control load and the delay to allocate addresses. Filters allow an accurate partition merging detection and increase the protocol robustness. Simulation results show that FAP resolves all the address collisions during partition merging. In the initialization of the network, the control load of FAP is more or less same as the control load of DAD which is a simple protocol that does not handle partition. Therefore, FAP is an efficient proposal to configure addresses automatically in the network. FAP withstands to message losses, which is considered as an important issue for ad hoc networks, which has fading channels and high bit error rates. Thus, the proposed protocol efficiently solves all the address collisions. FAP initialization process is also considered as the simplest one compared to other proposals.

References

1. S. Khalid, A. Mahboob, Design and implementation of id based MANET auto-configuration protocol. *Int. J. Commun. Netw. Inf. Secur. (IJCNIS)* **5**(3), 141–151 (2013)
2. S. Rafiul Hussain, S. Saha, A. Rahman, SAAMAN: scalable address auto configuration in mobile ad hoc networks. *J. Netw. Syst. Manage.* **19**, 394–426 (2010)
3. Z. Fan, S. Subramani, An address auto configuration protocol for IPv6 hosts in a mobile ad hoc network. *Comput. Commun.* **28**(4), 339–350 (2005)
4. N.C. Fernandes, M.D.D. Moreira, O.C.M.B. Duarte, An efficient and robust addressing protocol for node auto configuration in ad hoc networks. *IEEE Trans. Networking* **21**(3), 845–856 (2013)
5. N.H. Vaidya, Weak duplicate address detection in mobile ad hoc networks, in *Proceedings of the 3rd ACM International Symposium On Mobile Ad Hoc Networking & Computing*, pp 206–216 (2002)
6. S. Nesargi, R. Prakash, MANETconf: configuration of hosts in a mobile Ad Hoc network, in *Proceedings of 21st Annual IEEE INFOCOM*, vol 2, pp 1059–1068 (2002)
7. N.C. Fernandes, M.D. Moreira, O.C.M.B. Duarte, An efficient filter-based addressing protocol for autoconfiguration of mobile ad hoc networks, in *Proceedings of 28th IEEE INFOCOM*, pp 2464–2472 (2009)

Structural Refinement: An Effective OCL-Based Testing Approach

A. Jalila and D. Jeya Mala

Abstract Formal software development begins with formal specification of the user requirements or design of a system. Hence, formal specification languages are used to resolve ambiguities in user requirements or detect design errors at the early software life cycle. Furthermore, specification-based functional testing derives test inputs from its formal specification. However, formal specifications are expressed using set theory or predicate logics which are non-executable. Thus, functional test execution over abstract expressions would be impossible. Therefore, there is a need to refine abstract specification into a form that can be executable. In this paper, an automatic functional testing framework using object constraint language (OCL) formal specification has been proposed. The major objective of this paper is to describe how the refinement processes are integrated into the specification-based testing framework.

Keywords OCL · Structural refinement · Fitness function · Specification testing

1 Introduction

Formal methods are used to avoid specification errors or resolve ambiguities in the user requirements at the early phases of software life cycle. Moreover, formal specifications act as the formal proof of correctness of a program. Therefore, systems are formally specified using any one of the specification languages, namely object constraint language (OCL), Z, B, VBM, alloy, etc. However, specification-based testing is necessary to improve confidence in the system being developed. Hence, the delayed fault and failure detection increases the cost of error correction

A. Jalila (✉) · D. Jeya Mala
Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
e-mail: mejalila@gmail.com

D. Jeya Mala
e-mail: djmce@tce.edu

and maintenance [1]. Fault-based testing is the most prominent form of specification-based testing [2]. In the proposed approach, OCL predicate-based mutation test is performed. OCL-based testing is preferable due to the following reasons:

- OCL is platform independent; hence, it is applicable to all type of systems.
- Test coverage is improved; hence, it is measured from formal specification.
- OCL describes the precise model of a system; hence, functional test results will be accurate.
- OCL specification refinement process requires less effort.

The proposed approach endeavors that the most of abstract OCL specifications are complex predicates and not suitable for automated testing. Thus, performing testing directly from abstract OCL specification would be ineffective. Therefore, it is necessary to refine the abstract specifications into executable form. Thus, in our approach, abstract OCL predicates are refined into equivalent simple Boolean expressions for effective testing. Then, mutants are automatically generated for all methods of the system under test (SUT) by applying predicate-based [3] fault classes. In this paper, test adequacy criterion based on branch coverage and method-wise mutation scores are used to evaluate the fitness function. For the purpose of this work, the terms specification, condition, constraints, and expressions are used interchangeably. Furthermore, SUT would mean the OCL specification of the system which is derived from the user requirement specification and component would mean class of the SUT.

The remainder of this paper is organized into the following sections. Section 2 starts with related earlier works. The background of the proposed OCL specification-based testing is elaborated in Sect. 3. Section 4 discusses the OCL-based structural refinement process. Section 5 deals the experimentation of the proposed algorithm with various system specifications, and conclusions are discussed in Sect. 6.

2 Related Works

Significant work has been available to generate test cases from OCL specification using various mapping techniques.

Brucker et al. [4] proposed a theorem-prover-based test data generation from OCL specifications. In their research, they included language features such as recursive query operations, object graph, and equivalent relation. They used the concept called alias closure to derive an object graph. In their approach, object graphs are implicitly represented based on the equivalent relation. They have translated OCL invariants into recursive HOL predicates.

Mohammed et al. [5] described the partition analysis technique to identify the domains of the operation of a specification. In their study, they used mathematics

inherent which is derived from various constraints of the OCL specification to generate test data. They reduced the mathematical expression of the OCL specification by transforming it into disjunctive normal form (DNF). Their study focused on three concepts, namely formal definition of the notion of whole-part, behavioral inheritance and active objects for distributed testing.

Ali et al. [6] depicted a methodology to generate test data from OCL constraints that are derived from UML state and class diagrams. Their study used search-based algorithms to derive test data from OCL constraints. They defined a set of heuristics based on OCL constraints and branch distance functions for various types of OCL expressions. Empirical evaluation was done using two statistical tests, namely Fisher's exact test and a paired Mann Whitney U-test).

There are many other works which have refined OCL constraints into various intermediate formats such as SQL queries [7], constraint programming expressions [8], equation logic [9], and first-order logic [10] to ensure correctness of a program. However, in the proposed approach, OCL constraints are translated into equivalent simple Boolean expression predicates for improved reasoning.

From the existing research works, it has been inferred that OCL expressions are the important source for functional testing. However, majority of the prior works discussed the concept of the test case generation from OCL specification using simple expressions. However, OCL predicates include collection operations which are excluded in their work or the implementation details are not clear. Hence, the proposed approach deals the issues related to automated OCL specification-based testing.

3 Backgrounds

This section describes the quick overview on OCL and specification-based testing.

3.1 *Object Constraint Language (OCL)*

OCL standard stands for object constraint language [11]. It was proposed by the Object Management Group (OMG). OCL defines the system in the form of constraints or predicates. It is based on the simple mathematical notations or predicate logic that enables rigorous analysis and reasoning about the specifications. In general, OCL is used for precise model specification. OCL expressions can be derived both from the user requirements specification (during the requirement specification phase) or design documents (during the design phase). In the proposed approach, OCL constraints have been derived from user requirement specification of the system.

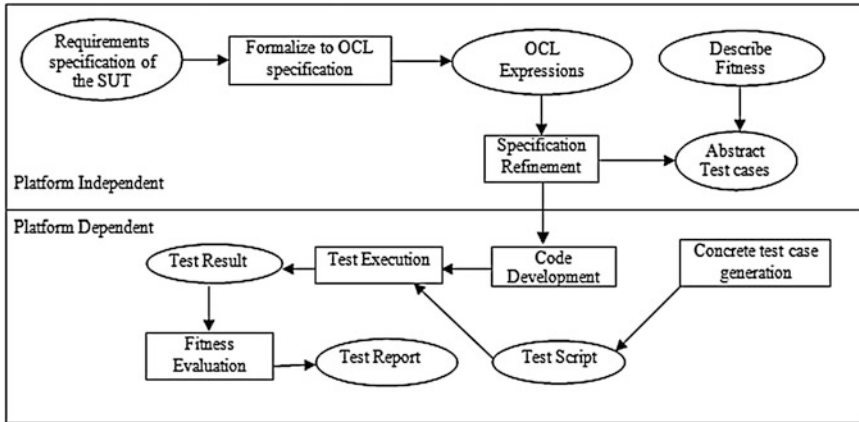


Fig. 1 Specification-based testing framework

3.2 Automatic OCL Specification-Based Testing

In general, OCL specification predicates are written in textual format. Hence, test can be generated and executed manually. However, automatic test generation and execution using OCL predicate guarantees the reliability of a system. Hence, the proposed work deals the issues related to automated OCL specification-based testing. A typical automatic OCL specification-based testing framework as shown in Fig. 1 requires six major modules, namely formalization, specification refinement, code development, test case generation, test execution, and evaluation. The details of the framework are explained below.

3.3 Fault-Based Testing

There are different types of faults that can occur in the specification of the system. Hence, the fault-based testing approach hypothesizes certain types of faults which frequently occur in the specification or program of a system. Those hypothesized faults are named as fault classes. Each fault class is designed to detect a particular type of fault. Moreover, the cost of testing can be affected by the choice of fault class. Hence, there are five predicate-based fault classes as proposed by Tai et al. [3] which have been used in the proposed approach. These fault classes include Boolean operator fault (BOOF), incorrect relational operator fault (IROF), incorrect parenthesis fault (ICPF), incorrect Boolean variable fault (IBVF), and incorrect arithmetic expression fault (IAEF). Table 1 furnished the details of the fault classes which have been used in the proposed approach to generate mutants.

Table 1 Tai's predicate-based fault classes

Fault class	Mutant generation	
	Original predicate	Mutant
BOOF	$a \neq b$	$a = b$
IROF	$a > b$	$a < b$
ILPF	$a \rightarrow \text{size}() = 0$	$a \rightarrow \text{size}(\neq 0)$
IBVF	$x = \text{True}$	$y = \text{True}$
IAEF	$x = a + b$	$x = a - b$

3.4 Fitness Function

In the proposed approach, method-wise mutation score and branch coverage values are calculated to evaluate the test adequacy criteria.

Method-wise mutation score: If a class 'C' contains methods $m = \{1 \dots n\}$, F_m is the number of methods exercised by the test case T and C_m is the total number of methods defined for the component 'C'. Thus, the method-wise mutation score (MS) against test case (T) is given by

$$MS_m(T) = \frac{|F_m|}{|C_m|} * 100 \quad (1)$$

Branch coverage: Most of the executable form of OCL preconditions, post-conditions, and invariants are branch statements. If executable expression of a class 'C' contains branch statements $bs = \{1 \dots n\}$, then branch coverage value (BCV) against test case (T) is given by

$$BCV_{bs}(T) = \frac{|N_{bs}|}{|T_{bs}|} * 100 \quad (2)$$

4 OCL-Based Structural Refinement

OCL uses first-order predicate logic to capture the software specifications precisely [7]. However, specifications written in OCL are essentially in textual form which is non-executable. Thus, performing testing directly from abstract OCL specification would be ineffective. Therefore, it is necessary to refine the abstract OCL specifications into executable form. The structural mapping or refinement produces executable expressions with the structure corresponding to that of the abstract OCL expressions, which are more suited to code. Figure 2 illustrates the structural mapping function.

Structural Mapping $[\]^E$: Structural mapping is defined as a function $[\]^E$ which translates each part of abstract OCL expression (S_A) into its counterpart

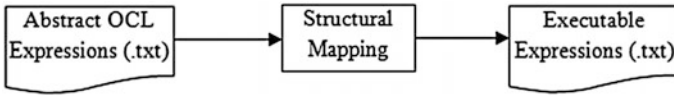


Fig. 2 Structural mapping function

Table 2 Structural mapping for the relational and Boolean operations

Relational operations (S_A)	Structural mapping (S_E)	Boolean operations (S_A)	Structural mapping (S_E)
$a \lt \! \! \! \triangleright b$	$[[\text{if } a \neq b, \text{ then return True}]]^E$	a	$[[\text{if } a = \text{True}, \text{ then return flag}]]^E$
$a > b$	$[[\text{if } a > b, \text{ then return True}]]^E$	$a \text{ and } b$	$[[\text{if } a \text{ and } b = \text{True}, \text{ then return flag}]]^E$
$a \geq b$	$[[\text{if } a \geq b, \text{ then return True}]]^E$	$a \text{ or } b$	$[[\text{if } a \text{ or } b = \text{True}, \text{ then return flag}]]^E$
$a \leq b$	$[[\text{if } a \leq b, \text{ then return True}]]^E$	$a \text{ implies } b$	$[[\text{If } a = \text{True}, \text{ then } b = \text{True}]]]^E$
$a < b$	$[[\text{if } a < b, \text{ then return True}]]^E$	$a \text{ xor } b$	$[[\text{If } a = \text{True}, \text{ then } b = \text{False}]]^E$
$a = b$	$[[\text{if } a = b, \text{ then return True}]]^E$	$\text{If } a \text{ then } b \text{ else } c$	$[[\text{if } (\text{exp } a) = \text{True}, \text{ then return } b \text{ else return } c]]^E$
$a \lt \! \! \! \triangleright b$	$[[\text{if } a \neq b, \text{ then return True}]]^E$		

executable expression (S_E). Hence, the behaviors of S_A include all those of S_E . Therefore, S_A is equivalent to S_E . It is formally represented as

$$[[\]]^E: S_A \sqsubseteq S_E \quad \text{or} \quad [[\]]^E: S_A \rightarrow S_E$$

The proposed work describes the structural mapping functions for the following OCL elements, namely relational, Boolean type-based OCL operations, standard and iterative OCL collection operations. Table 2 describes the structural mapping function for the relational and Boolean operators.

4.1 Standard OCL Collection Operations

Table 3 depicts the structural mapping for the standard OCL collection operations.

4.2 Iterative OCL Collection Operations

Table 4 depicts the structural mapping function for the iterative OCL collection operations.

Table 3 Structural mapping of the standard OCL collection operations

Standard OCL operations (S_A)	Structural mapping (S_E)
$a \rightarrow \text{includes}(b)$: Boolean	$[[\text{loop: } i = 1 \text{ to } b.\text{size()} \text{ if } b[i] \neq \text{null then return True}]]^E$
$a \rightarrow \text{excludes}(b)$: Boolean	$[[\text{loop: } i = 1 \text{ to } b.\text{size()} \text{ if } b[i] == \text{null then return True}]]^E$
$a \rightarrow \text{includesAll}(b:T)$: Boolean ($b:\text{Collection}(T)$): Boolean	$[[\text{loop: } i = 1 \text{ to } a.\text{size()}, \text{ loop: } i = 1 \text{ to } b.\text{size()} \text{ if } a[i] == b[i]]^E$ $\text{then return True}]]^E$
$a \rightarrow \text{excludesAll}(b:T)$: Boolean	$[[\text{loop: } i = 1 \text{ to } a.\text{size()}, \text{ loop: } i = 1 \text{ to } b.\text{size()} \text{ if } a[i] \neq b[i]]^E$ $\text{then return True}]]^E$
$\text{isEmpty}()$: Boolean	$[[\text{if } a.\text{size()} \neq 0 \text{ then return True}]]^E$
$\text{notEmpty}()$: Boolean	$[[\text{if } a.\text{size()} \neq 0 \text{ then return True}]]^E$

Table 4 Structural mapping of iterative OCL collection operations

Collection operations (S_A)	Structural mapping (S_E)
$\text{isUnique}(b:T)$: Boolean	$[[\text{loop } a = 1 \text{ to } a.\text{size}()-1 \text{ if } a[i] \neq a[i] + 1 \text{ then return True}]]^E$
$c \rightarrow \text{exists}(b P)$: Boolean, where p is the predicate	$[[\text{loop } i = 1 \text{ to } a.\text{size}(), \text{ if } a[i] == b \text{ then return True}]]^E$
$a \rightarrow \text{forAll}(v1, v2 \dots vn: T p)$: Boolean, where p is the predicate	$[[\text{loop } i = 1 \text{ to } a.\text{size}(), p[a[i]]]]^E$
$a \rightarrow \text{select}(b:T)$: Boolean, where b is any object	$[[\text{loop } i = 1 \text{ to } a.\text{size}(), \text{ if } \text{exp}[a[i]] == \text{True then Collection}[i] = a[i] \text{ return Collection}]]$
$a \rightarrow \text{reject}(b:T)$: Boolean, where b is any object	$[[\text{loop } i = 1 \text{ to } a.\text{size}(), \text{ if } \text{exp}[a[i]] \neq \text{True then Collection}[i] = a[i] \text{ return Collection}]]$

5 Experimentation and Result Analysis

For the experimental purpose of this research work, the OCL specification of six real-time applications has been selected. Table 5 describes the brief summary of these case study applications. Figure 3 presents the sample OCL expression for method ‘calatt’ of the class ‘admin’, and its equivalent structural mapping is given in left-hand side of the Table 6. The mutants are generated for the method ‘calatt’

Table 5 Case studies

Application name	App-ID	No. of classes
Patient monitoring	PMS	22
Blood bank	BBMS	19
Library management	LMS	15
Payroll	PAYROLL	6
E-commerce	E-com	32
Banking	BMS	19

```

context admin::calatt a:attendance)
pre:a.workday>1
pre:a.noofhol>=0
post: self.noofpre>=self.workday
post:a.noofpre=(a.noofpre-
a.workday) + a.noofhol
    
```

Fig. 3 Sample abstract OCL expression

based on the fault classes [3] as discussed in Sect. 4. After the mutant generation, the test cases are generated and executed against mutants. Then, the test adequacy criteria are analyzed based on the mutation score and branch coverage values as shown in Table 6.

Table 7 depicts the method-wise mutation score, branch coverage, test size, and execution time taken for the various case studies when applying random search.

Table 6 Sample for the test case generation using OCL specification

OCL expressions after structural refinement	Equivalent Java program with its mutant	
class admin {	public class admin{	
calatt(Integer wd, Integer hol, Integer pre) {	public calatt(Integer wd, Integer hol, Integer pre) {	
attendance a = new attendance ()	attendance a = new attendance ()	
a.wd = workday;	a.wd = workday;	
a.hol = noofhol;	a.hol = noofhol;	
a.pre = noofpre;	a.pre = noofpre;	
If (workday > 1&&noofhol > = 0) {	If (workday > 1 && noofhol <=0) {	
if (noofpre >=workday)	if (noofpre >=workday)	
noofpre = workday + noofhol	noofpre = workday + noofhol else	
else	noofpre = (workday-noofpre) + noofhol	
noofpre = (workday-noofpre) + noofhol	noofpre = (workday-noofpre) + noofhol	
} return a;}}	} return a;}}	
Fitness evaluation		
<i>Test cases</i>	<i>Mutation score</i>	<i>Branch coverage</i>
Testcase1: {18, 2,18}	Method: 100 %	B1-33 %
Testcase2: {18,0,16}	Method: 100 %	B1&B3-67 %

Table 7 Result analysis

Application name	Criteria			
	Mutation score (%)	Branch coverage (%)	Test size	Execution time (s)
PMS	50	74	1126	18.232
BBMS	56	68	982	32.63
LMS	58	62	612	45.678
PAYROLL	67	63	250	31.567
E-com	62	56	1549	34.3
BMS	59	54	677	12.34

6 Conclusion

The major objective of the proposed work is to demonstrate that the implementation conforms to the specification. In this research work, a novel approach has been adopted to automate refinement of the OCL predicates. The proposed framework is used to generate efficient test cases at the specification level. We have applied our approach to many real-time applications and observed the effectiveness of specification-based testing using OCL predicates. From the above study, it has been inferred that OCL predicates support effective testing with more test coverage and thereby improve the overall software quality. As a future work, it has been proposed to adopt other optimization techniques to minimize test requirements with less test runs.

Acknowledgment This paper is a part of the UGC major research project supported by University Grants Commission (UGC), New Delhi, India.

References

1. F. Elberzhager, F.L. Alla Rosbach, J. Münch, R. Eschbach, Reducing test effort: a systematic mapping study on existing approaches. *Inf. Softw. Technol.* **54**, 1092–1106 (2012)
2. D.R. Kuhn, Fault classes and error detection capability of specification based testing. *ACM Trans. Softw. Eng. Methodol.* **8**, 411–424 (1999)
3. K.C. Tai, M.A. Vouk, A.M. Paradkar, P. Lu, Evaluation of a predicate-based software testing strategy. *IBM Syst. J.* **33**, 445–457 (1994)
4. A.D. Brucker, M.P. Krieger, D. Longuet, B. Wolff, A specification-based test case generation method for UML/OCL, in *International Conference on Models in Software Engineering* (2011)
5. M. Benattou, J.M. Bruel, N. Hameurlain, Generating test data from OCL specification, in *Proceedings of the ECOOP'2002 Work-Shop on Integration and Transformation of UML Models* (2002)
6. S. Ali, M.Z. Iqbal, A. Arcuri, L. Briand, Generating test data from OCL constraints with search techniques. *IEEE Trans. Softw. Eng.* **39**, 1376–1402 (2013)
7. F. Heidenreich, OCL-Codegenerierung für Deklarative Sprachen, Master's thesis (2006)

8. J. Cabot, R. Clariso, D. Riera (2007), UMLtoCSP: a tool for the formal verification of 'UML/OCL models using constraint programming, in *Proceedings of the 22nd IEEE/ACM International Conference on Automated Software Engineering, ASE '07*
9. J. Carsí, I. Ramos, A. Boronat, A. Gomez, The MOMENT: MOdelManagement 'framework project (2008)
10. B. Beckert, R. Hahnle, P.H. Schmitt, *Verification of object-oriented software: the KeY approach* (Springer, Berlin, 2007), p. 4334

Dynamic Architecture and Performance Analysis of Secure and Efficient Key Management Scheme in Multicast Network

N.M. Saravanakumar, R. Keerthana and G.M. Mythili

Abstract Group key management in multicast networks plays a crucial role in data communication environment. Also, key management deals with distribution of keys among group members and maintaining the keys. The lack of security services, communication overhead, computation overhead, etc. Enables us to concentrate on creating new innovative ideas. A key distribution algorithm in the reviewed protocols does not provide much security in group communication networks. The contribution of this research is to investigate all the available key management schemes and to design the secure and efficient key management scheme in multicast network for achieving a secure communication between the group members. The proposed scheme, dynamic architecture and performance analysis of secure and efficient key management scheme, provides a secure variable UID-based key management scheme which protects non-group members from the access of data and generation of static group key which reduces computation cost at any change in the multicast network. The periodic renewal of group key using proactive secret sharing scheme provides greater security during communication. The proposed authentication process ensures data integrity and confidentiality during communication and the encryption mechanism used in dynamic architecture and performance analysis of secure and efficient key management scheme completely eliminates the communication, computation overhead, and network traffics. The analysis shows that the proposed key management scheme comprises of the most reliable methods for key generations and key distributions and provides better performance in terms of security requirements and other services.

N.M. Saravanakumar (✉) · R. Keerthana · G.M. Mythili
Department of Computer Science and Engineering, Bannari Amman
Institute of Technology, Sathyamangalam, India
e-mail: saravanakumaar2008@gmail.com

R. Keerthana
e-mail: keerthi5111990@gmail.com

G.M. Mythili
e-mail: mythilebe@gmail.com

Keywords Multicast network • Key management • Rekeying • Invertible matrix

1 Introduction

Multicast has not yet been extensively used since it is an extension to IPv4 standard because craving for on-demand and group conferencing-type services is not very large. It also plays a vital role in computer networks and communication systems which include many emerging applications based on multicast group communications. The extensive applications of group communication include file/software updating, video–audio transmission, multiparty video games, news, chat, video-conferencing, military application, simulation services. The efficiency and scalability of such group communications can be increased by using multicast approach especially in case of multimedia services such as video on demand and video-conferencing. The members in multicast networks are organized as groups.

2 Dynamic Architecture and Performance Analysis of Secure and Efficient Key Management Scheme

The multicast network in SEKMS has time-based cluster structure. Initially, Key Generation Center/Group Controller (KGC/GC) assigns the number of sub-groups (cluster) to be constructed and their respective subscription span values based on which the members are grouped. For instance, we consider the number of sub-groups under KGC/GC to be 3 whose subscription spans are 30 days, 6 months, and 1 year, respectively, as shown in Fig. 1 where SGC1, SGC2, and SGC3 are sub-group controllers. Let there be 8 members with subscription span less than or equal to 30 days who are to be grouped under SGC1, 2 members with subscription span more

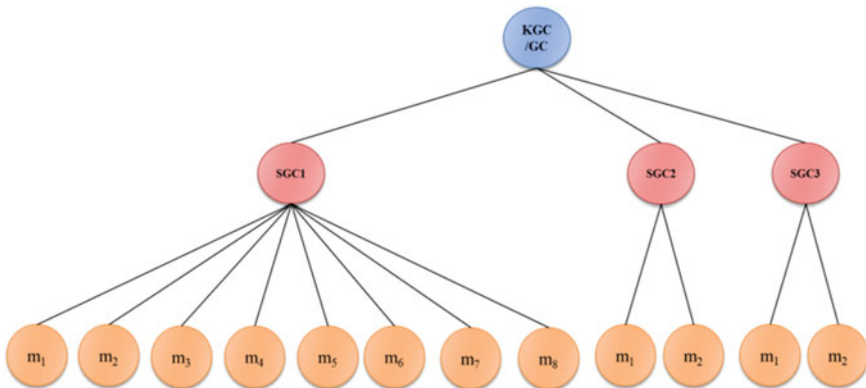


Fig. 1 SEKMS sample network

than 1 month and less than or equal to 6 months under SGC2, and 2 members with subscription span more than 6 months and less than or equal to 1 year under SGC3.

There are two databases maintained by the KGC/GC.

- (a) Existing members' database that stores the data about members under each SGC. These data include member UIDs, subscription spans of each member.
- (b) Leaving members' database that contains the data of the members which left previously [1].

2.1 Generation of UID

In this scheme, UID for each member in the sub-group is generated using modified Huffman coding technique as shown in Fig. 2. The term Huffman coding refers to the use of a variable length code for encoding a source symbol where the variable length code has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol.

Initially, KGC/GC will randomly generate unique binary values to SGC1, SGC2, and SGC3 [2] (Table 1).

2.2 Sub-Group Key and Group Key Generation

The group key computation method is used in multiparty Diffie–Hellman and TGDH protocol to generate group keys. In SEKMS, we modify this idea by making the group key GK independent of sub-group key SGK.

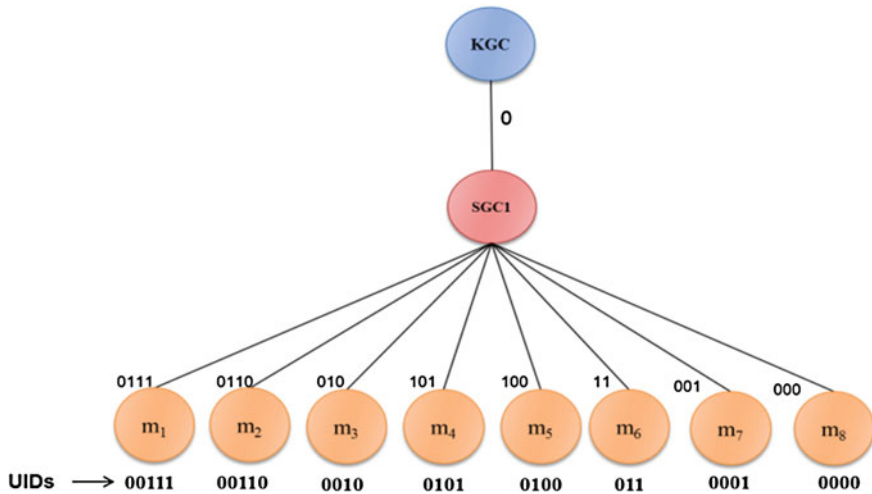


Fig. 2 SEKMS UID generation

Table 1 SEKMS—UID generated for each member

SGC	m (members)	UID
SGC1 (0)	m ₁ (0111)	00111 (UID _{1,1})
	m ₂ (0110)	00110 (UID _{2,1})
	m ₃ (010)	0010 (UID _{3,1})
	m ₄ (101)	0101 (UID _{4,1})
	m ₅ (100)	0100 (UID _{5,1})
	m ₆ (11)	011 (UID _{6,1})
	m ₇ (001)	0001 (UID _{7,1})
	m ₈ (000)	0000 (UID _{8,1})
SGC2 (1)	m ₁ (0)	10 (UID _{1,2})
	m ₂ (1)	11 (UID _{2,2})
SGC3 (00)	m ₁ (0)	000 (UID _{1,3})
	m ₂ (1)	001 (UID _{2,3})

- (a) *Generate sub-group Keys using Partial Keys from Members*: Each member under a sub-group sends the partial key, $L_{i,j}$ to SGC, where $i = 1, 2, 3, 4$ and $j = 1, 2, 3, \dots$. The SGC then uses these partial keys to compute the sub-group keys (SGKs). Here, Z_N^* which is the set $1, 2, N - 1, N$ is the prime, and L is a randomly chosen prime number for respective member. For example, from Fig. 2, the resulting sub-group key of SGC_1 is given by Eq. 1.

$$SGK_1 = H(L_{1,1} || L_{2,1} || L_{3,1} || L_{4,1} \dots L_{8,1} || K_1) \quad (1)$$

The resulting SGK is sent to each member and is used for encryption and decryption of the message exchange among the members within the sub-group [4].

- (b) *Generate Group Keys using the Partial Keys from SGCs*: The KGC/GC collects the partial key of each sub-group. Consider Fig. 2. Let partial keys of SGCs be $K_1, K_2,$ and $K_3,$ respectively. The KGC/GC receives all these keys and generates its own partial group key, and computes the common group key as shown in Eq. 2.

$$GK = H(K_1 || K_2 || K_3 || K_{GC}) \quad (2)$$

Further, this group key is broadcasted to each sub-group which is used for decryption or encryption during the communication between different sub-groups under KGC/GC. The SGKs and GK are distributed in this network using proactive secret sharing scheme [12]. For each GK and SGK to be distributed to the sub-groups and the members, time periods, T_{GK} and T_{SGK} , are set and divided into periods of time [6].

2.3 Public–Private Keys and Signature Generation

Each user is given long-term public and private keys. The *KGC/GC* randomly chooses a secret key and computes and publishes the corresponding public key. *SEKMS* uses the idea of RSA to construct a private–public key pair, where the *KGC/GC* calculates (1) public key (M, E) , where M is the product of any two large prime numbers, a and b , and E is the number prime with respect to M and (2) private key $(a, b, d, \phi(M))$, where d is the part of private key of *KGC/GC* and is equal to $e^{-1} \text{ mod } \phi(M)$. The *KGC/GC* determines a primitive element α in $GF(a)$ and $GF(b)$. Then, it chooses a one-way hash function. Here, $(\alpha, h())$ is a public information, where $h()$ gives unique output for different inputs [7].

Each *SGC* provides *UIDs* of the member under it to the *KGC/GC* to obtain the signature $S_{i,j}$ for each $UID_{i,j}$ of a member $m_{i,j}$, where $i = 1, 2, 3, \dots$, represents each member and $j = 1, 2, \text{ or } 3$ represents the *SGC*. If *KGC/GC* confirms the correctness and the relationship between $m_{i,j}$ and $UID_{i,j}$, then it calculates $S_{i,j}$ using Eq. (3) and distributes $S_{i,j}$ to each *SGC* where each *SGC* distributes them to the respective members as shown in the Fig. 11.

$$S_{i,j} = \text{UID}_i^d \text{ mod } M \tag{3}$$

Both public–private keys pair and signatures are distributed using proactive secret sharing scheme.

2.4 Generation Self-invertible Matrix

As decryption requires inverse of the any matrix used for encryption, so while decryption one problem arises that is, inverse of the matrix does not always exist. If the matrix is not invertible, then encrypted text cannot be decrypted. In order to overcome this problem, we suggest the use of self-invertible matrix generation method. Moreover, this method eliminates the computational complexity involved in finding inverse of the matrix while decryption. A is called self-invertible matrix if $A = A^{-1}$. The analyses presented here for generation of self-invertible matrix are valid for matrix of +ve integers that are the residues of modulo arithmetic on a prime number [16].

2.4.1 Generation of Self-invertible 3×3 Matrix

$$\text{Let } A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{22} & A_{21} \end{bmatrix} \tag{4}$$

where A_{11} is a 1×1 matrix = $[a_{11}]$, A_{12} is a 1×2 matrix = $[a_{12} \ a_{13}]$, A_{21} is a 2×1 matrix = $\begin{bmatrix} a_{21} \\ a_{31} \end{bmatrix}$, and A_{22} is a 2×2 matrix = $\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$. If A is self-invertible then $A_{11}^2 + A_{12}A_{21} = I$, $A_{11}A_{12} + A_{12}A_{22} = 0$, $A_{21}A_{11} + A_{22}A_{21} = 0$ and $A_{21}A_{12} + A_{22}^2 = I$. Since A_{11} is a 1×1 matrix = $[a_{11}]$ and $A_{21}(a_{11}I + A_{22}) = 0$. For non-trivial solution, it is necessary that $a_{11}I + A_{22}$. That is $a_{11} = -$ (one of the eigen values of A_{22}). $A_{21}A_{12}$ can also be written as

$$A_{21}A_{12} = \begin{bmatrix} a_{21} & 0 \\ a_{31} & 0 \end{bmatrix} \begin{bmatrix} a_{12} & a_{13} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a_{21}a_{12} & a_{21}a_{13} \\ a_{31}a_{12} & a_{31}a_{13} \end{bmatrix}$$

So $A_{21}A_{12}$ is singular and

$$A_{21}A_{12} = I - A_{22}^2 \tag{5}$$

Hence, A_{22} must have an eigen value ± 1 . It can be shown that $\text{Trace}[A_{21}A_{12}] = A_{12}A_{21}$.

Since it can be proved that if $A_{11} = a_{11} = -$ (one of the eigen values of A_{22}), then any non-trivial solution of the Eq. (5) will also satisfy $A_{12}A_{21} = I - a_{11}^2$

2.4.2 A General Method for Generating Self-invertible Matrix

Let $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$ be an $n \times n$ self-invertible matrix partitioned to

$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{22} & A_{21} \end{bmatrix}$. A_{11} is a 1×1 matrix = $[a_{11}]$, A_{12} is a $1 \times (n - 1)$ matrix = $[a_{12}$

$a_{13} \dots a_{1n}]$, A_{21} is a $(n - 1) \times 1$ matrix = $\begin{bmatrix} a_{21} \\ a_{31} \\ \dots \\ a_{n1} \end{bmatrix}$, A_{22} is a $(n - 1) \times (n - 1)$

matrix = $\begin{bmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}$

$$\text{So, } A_{12}A_{21} = I - A_{11}^2 = 1 - a_{11}^2 \tag{6}$$

and $A_{12}(a_{11}I + A_{22}) = 0$. Also, $a_{11} = -$ (one of the eigen values of A_{22} other than 1) since $A_{21}A_{12}$ is a singular matrix having the rank 1 and

$$A_{21}A_{12} = I - A_{22}^2 \quad (7)$$

So, A_{22}^2 must have rank of $(n - 2)$ with eigen values $+1$ of $(n - 2)$ multiplicity. Therefore, A_{22} must have eigen values ± 1 . It can also be proved that the consistent solution obtained for elements A_{21} and A_{12} by solving the Eq. (7) term by term will also satisfy the Eq. (6).

Algorithm

1. Select A_{22} , a non-singular $(n - 1) \times (n - 1)$ matrix which has $(n - 2)$ number of eigen values of either $+1$ or -1 or both.
2. Determine the other eigen value 1 of A_{22} .
3. Set $a_{11} = -\lambda$
4. Obtain the consistent solution of all elements of A_{21} and A_{12} by using the Eq. (7).
5. Formulate the matrix.

3 Rekeying

Any member may leave or join the sub-group at any time. Whenever there is any change in the number of members in a sub-group, rekeying is done. In this section, the rekeying is discussed with respect to single leave, single join, multiple leaves, and multiple joins situations in the group. In the database of *KGC/GC*, the data of the member are deleted and put in leaving member database as soon as the subscription span is finished [13].

3.1 Single-Member Leave Event

When a member's subscription span is completed, the data about this member in the database of *KGC/GC* are removed and inserted in the leaving member database. The signatures, public-private keys, and the group key of the other members remain same. Only the sub-group key is changed as explained in the previous sections [14].

3.2 Single-Member Join Event

- Multiple leave events
- Multiple leave events from the same sub-group
- Multiple leave events from different sub-groups

3.3 Multiple Join Events

- Multiple join events in the same sub-group
- Multiple join events in different sub-groups

4 Result and Discussions

SEKMS is compared to LKH and OFT and the following results were obtained.

4.1 Computation Cost

Like all other approaches, there is a rekeying process in SEKMS, except that the group key remains same for any change in the number of members along with ensuring securities necessary. The rekeying is done only in the sub-group where the change takes place. Table 3 gives a brief comparison of computation costs for group key generation of OFT, LKH, and SEKMS for a single-member join and leave. Figures 3 and 4 gives the statistical comparisons.

Here, m is the sub-group size and n is the group size. When a member joins the group, the sub-group key is regenerated along with the UID, public–privates keys pair and signature. For each newly joining member, three new keys and one UID are generated. If there are n members joining at the same time, then $4n$ computations are done [15].

Fig. 3 Computation cost at joins

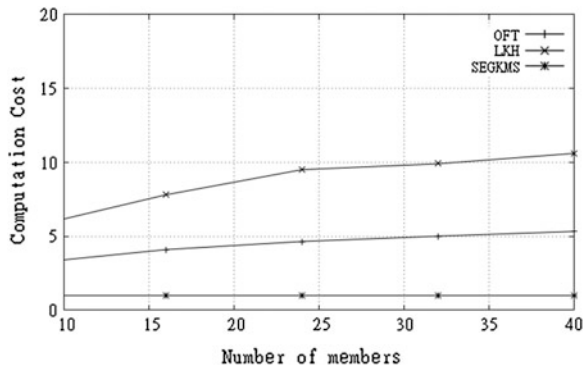
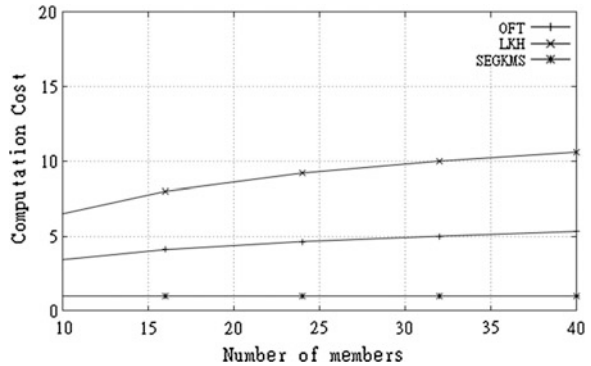


Fig. 4 Computation cost at leaves



5 Conclusion

This paper has proposed a scalable and efficient key management scheme for multicast networks, which results in security of the data exchanges. SEKMS achieves flexibility in the size of each sub-group, i.e., any number of members can be joined in SEKMS. Since the joining of a member is based on the subscription span and whenever subscription span is completed, the member leaves the sub-group. The communications and key distributions in SEKMS ensure the security of the message exchanged between the members. Use of database for the current members in group and the members left previously helps in achieving forward and backward secrecy. It generates variable length *UIDs* and hence variable length signatures and session keys. When an intruder tries to access the information being exchanged, it is difficult to obtain without the use of the required keys. In SEKMS, only the communicating members know the sub-group keys, group keys, and other necessary keys required for the decryption of the data received. Even if the group key is static, the data cannot be accessed without the present sub-group key and other keys as the session keys change for each communication. If the intruder was the member under the group in past, it cannot access the data without its information being present in the database. Whenever a past member or a newly joined member tries to access any information, it is possible only if the session falls under their subscription span.

References

1. J. Bibo, H. Xiulin, A survey of group key management, in *International Conference on Computer Science and Software Engineering* (2008)
2. A.T. Sherman, D.A. McGrew, Key establishment in large dynamic groups using one-way function trees. *IEEE Trans. Software Eng.* **29**(5), 444–458 (2003)
3. L. Dondeti, S. Mukherjee, A. Samal, Scalable secure one-to-many group communication using dual encryption. *Comput. Commun. ACM* **23**, 1681–1701 (1999)

4. E. Munivel, J. Lokesh, Design of secure group key management scheme for multicast networks using number theory, in *CIMCA, IAWTIC, and ISE* (2008)
5. C. Isabella, E. Robert, in *Proceedings of IEEE INFOCOM'99 Key Management for Secure Internet Multicast using Boolean Function Minimization Techniques. 18th Annual Joint Conference of the IEEE Computer and Communications Societies* (1999)
6. J. Lakshmanaperumal, K. Thanushkodi, N.M. Saravana Kumar, K. Saravanan, T. Purusothaman, Efficient key management scheme for secure multicast in MANET. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **10**(11) (2010)
7. W. Liming, W. Chuan, Efficient key agreement for large and dynamic multicast groups. *Int. J. Netw. Secur.* **3**(1), 8–17 (2006)
8. S. Vijaya, B.H. Ravi Krishna, A group key management approach for multicast cryptosystems. *J. Theor. Appl. Infor. Technol.* (2009)
9. M.M. Nasreldin Rasslan, Y.H. Dakroury, H.K. Aslan, A new secure multicast key distribution protocol using combinatorial boolean approach. *Int. J. Netw. Secur.* **8**(1), 75–89 (2009)
10. M.S. Bouassida, I. Chrismet, O. Festor, Group key management in MANETs. *Int. J. Netw. Secur.* **6**(1), 67–79 (2008)
11. L. Pavithira, T. Purushothaman, An energy efficient topology aware key management scheme for multicasting in ad-hoc networks. *Int. J. Wisdom Based Comput.* **1**(3) (2011)
12. S. Pitipatana, A. Nirwan, Elliptic curve cryptosystem-based group key management for secure group communications, in *IEEE Military Communications Conference* (2007)
13. R. Srinivasan, V. Vaidehi, R. Rajaraman, S. Kanagaraj, Secure group key management scheme for multicast networks. *Int. J. Netw. Secur.* **11**(1), 33–38 (2010)
14. P. Roberto, V. Luigi, M. Mancini, Alessandro. Key management for high bandwidth secure multicast. *J. Comput. Secur.* 693–709 (2004)
15. J. Jabeen, T. Purusothaman, A new scalable and reliable cost effective key agreement protocol for secure group communication. *J. Comput. Sci.* **7**(3), 328–340 (2011)
16. D. Luciano, G. Prichett, From caesar ciphers to public-key cryptosystem. *Coll. Math. J.* **12**(1), 2–17 (1987)

Mining Undemanding and Intricate Patterns with Periodicity in Time Series Databases

S. Sridevi, P. Saranya and S. Rajaram

Abstract Existing periodic pattern mining algorithms detect only specific type of periodicity, i.e., either symbol, sequence, or segment periodicity. Thus, the user is not able to find the combination of any two periodicities using existing algorithms. To overcome the above problem, this paper describes an approach called STNR (suffix tree-based noise resilient) algorithm to detect undemanding and intricate patterns. Undemanding covers symbol and sequence periodicity, and intricate covers segment periodicity. The benefit of STNR algorithm is to detect all the three types of periodicity in a single run. The main contribution in this work is to compare the performance of this algorithm using suffix tree and suffix cactus which act as an underlying data structure. The result shows that the outcome of the STNR algorithm using suffix cactus was more efficient in terms of space and time complexity.

Keywords Time series · Periodicity detection · Symbol periodicity · Sequence periodicity · Segment periodicity

1 Introduction

Pattern mining plays a most important role in data mining. It includes periodic pattern mining, sequence pattern mining, frequent pattern mining, and episode mining. This paper is concerned with periodic pattern mining. Periodic patterns can be extracted from time series databases. In general, time series is represented as 5 tuples [1]. They are starting position (stPos), pattern (X), period value (p),

S. Sridevi (✉) · P. Saranya
Department of CSE, Thiagarajar College of Engineering, Madurai 600015, Tamil Nadu, India
e-mail: sridevi@tce.edu

P. Saranya
e-mail: saranyasivan@tce.edu

S. Rajaram
Department of ECE, Thiagarajar College of Engineering, Madurai 600015, Tamil Nadu, India
e-mail: rajaram_siva@tce.edu

confidence (conf), and ending position (endPos). The analysis of time series data is done in order to predict future trends and behavior of the system.

2 Related Work

Elfeky et al. [2] proposed two separate algorithms for periodicity detection in time series databases. The first algorithm namely CONV is based on convolution technique and detects only symbol periodicity. It fails to perform well when the time series database contains noise. The second algorithm [3] namely WARP detects only segment periodicity. It works well in the presence of noise.

Huang and Chang [4] proposed algorithms for finding asynchronous periodic patterns; here, the periodic occurrences are shifted in an allowable range along the time axis.

Dong [5] proposed data cube-based mining approach to mine segment-wise periodic patterns in time-related databases. They integrated data cube and Apriori data mining techniques for mining segment-wise periodicity. It is related to a fixed-length period and showed that data cube provides an efficient structure and a convenient way for interactive mining of multiple level periodicity. It seems difficult for the user to find segment-wise periodicity for varying period lengths.

Han et al. [6] proposed max-subpattern hit-set algorithm to detect all the partial periodic patterns for a given period ' p ' in time series based on the max-subpattern hit set, for a given min_conf threshold. Partial periodicity mining requires only two scans over the time series database, even for mining multiple periods. The algorithm requires the end user to specify the period value in prior. It seems difficult for the user to provide period value in advance.

Yang and Wang [7] proposed an algorithm called InfoMiner+ to simultaneously mine significant patterns. They used generalized information gain measure to include a penalty for gaps between pattern occurrences.

Rasheed et al. [1] proposed suffix tree-based noise resilient algorithm (STNR) for periodicity detection in time series databases. It can detect symbol, sequence, and segment periodicities in time series. But the limitation here is it is not possible to skip intermediate events in time series data.

Nishi et al. [8] proposed Apriori-based level-by-level sequential pattern mining approach to mine a specific pattern. The proposed periodicity detection algorithm can detect all the three types of periodicity in a single run. It generates all possible interesting patterns by allowing event skipping among intermediate events.

3 Design Methodology

The goal of analyzing the time series database is to find whether and how frequent a pattern is repeated within the time series. The main contribution in this work is to compare the performance of this algorithm using suffix tree and suffix cactus as an

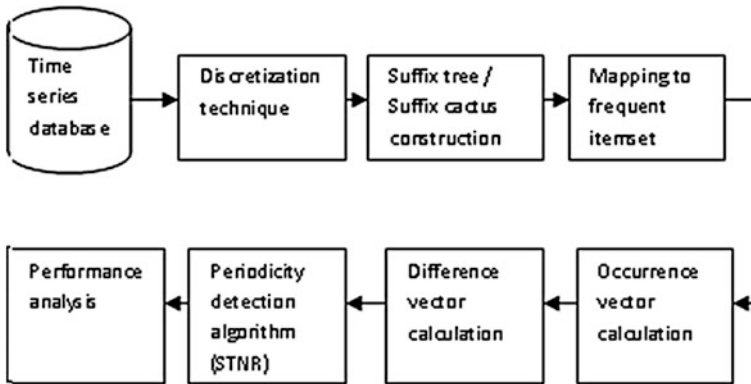


Fig. 1 Block diagram

underlying data structure. The block diagram shown in Fig. 1 explains various modules and the algorithms used in each module.

The time series database is taken as an input as shown in [9]. With the help of discretization technique, data are preprocessed and then, suffix cactus was constructed. Mapping to frequent item is done in order to calculate occurrence vector. The occurrence vector was used to calculate the index positions in the original string. The difference between two occurrence positions is calculated using difference vector. Finally, the performance analysis based on STNR—suffix tree-based noise resilient—algorithm was done using suffix tree and suffix cactus.

3.1 Discretization Technique

Discretization technique can be thought as a mapping among the range of values of an entity and an ASCII character which represents a specific event. Here, time series was discretized into symbols. If the ranges of transactions are within {1–2,000}, then it is denoted as ‘A’; if there are {2,000–4,000} transactions, then it is denoted as ‘B’, likewise {4,001–6000} transactions: ‘C’, {6,001–8,000} transactions: ‘D’, {8,001–10,000} transactions: ‘E’, {10,001–12,000} transactions: ‘F’, {12,001–14,000} transactions: ‘G’, and {>14,000} transactions: ‘H’.

3.2 Comparison of Suffix Tree with Suffix Cactus

The proposed STNR—suffix tree-based noise resilient—algorithm uses suffix tree as an underlying data structure and finds out periodicities within time series databases. Suffix tree has number of applications such as string processing, used to find

a substring in the original string. It also finds frequent substring and other string matching problems. Each substring is terminated with special character \$, where \$ denotes the end marker for the string.

The suffix cactus is an intermediate approach for storing and searching of values. The performance and size of the suffix cactus are between the suffix tree and suffix array. The suffix tree is an index-based data structure where the queries related to strings can be executed fast. The searching time and size depend on the length of the string. The suffix array also allows fast searching in string and text. But for better performance of the suffix array implementation, the size taken is 6 bytes per symbol. The suffix array is slower in cases on problems involving multiple string expressions. Thus, there is a need to go for suffix cactus. The suffix cactus has a size of 10 bytes per symbol that is in between the suffix tree and suffix array. It has similarities with both the earlier methods and can perform well in any problem that can be solved using the above two structures.

3.3 Periodicity Detection Algorithm

STNR detects all the three types of periodicity in single run. They are symbol, sequence, and segment periodicity. They are described as follows: A time series 'T' is said to have symbol periodicity if at least one symbol is repeated periodically. A time series 'T' is said to have sequence periodicity if more than one symbol is repeated periodically. A time series 'T' is said to have segment periodicity if a whole pattern or segment is repeated periodically.

This STNR algorithm detects whether the generated pattern is periodic or not using the formula's 1 and 2 given below.

$$\text{Mean } p = \text{sumPer} - \frac{P}{\text{count}(p) - 1}; \quad (1)$$

$$\text{Conf}(p) = \frac{\text{count}(p)}{\text{perfect}_{\text{periodicity}(p, \text{stPos}, X)}}; \quad (2)$$

where 'p' is the periodicity, count (p) is total number of periodicities, and sumPer is average value of periodicities. Confidence of a periodic pattern is the ratio of its actual periodicity to its expected perfect periodicity. The pseudocode for finding whether the pattern is periodic or not is as follows:

Algorithm

```

Input: The time series array T[ ] and the pattern P
Output: The type of periodicity detected with the period
value 'p'

Length = T.length;
a = 0; b = 0;
while(pattern!=NULL)
  for(i = 0, i < Length, i++)
    if(T[i] == P) then
      occ_vec[a] = i;
      a = a + 1;
    end if
  end for
  Length_2 = occ_vec.length;
  for(j = 0, j < Length_2, j++)
    diff_vec[b] = occ_vec[j+1] - occ_vec[j];
    b = b + 1;
  end for
  if(all values in diff_vec array are equal) then
    P to periodicity with period 'p' = diff_vec;
    stPos = diff_vec[0];
  end if
  count(p) = total number of periodicities available;
  sumPer = average periodicities;
  Mean p =  $\frac{sumPer}{count(p)-1}$ ;
  Conf(p) =  $\frac{count(p)}{perfect_{periodicity}(p,stPos,X)}$ ;
  if(Conf(p) >= threshold) then
    Add 'p' to period list;
  end if
end while

```

4 Experimental Results and Discussions**4.1 Dataset Selection**

The Tafeng dataset is all about a collection of grocery items transactions based on products and customers. This dataset [9] holds transactions data for 4 years.

The time series dataset was preprocessed in order to get accurate results as shown in Table 1.

By analyzing this dataset, the user can be able to predict whether the number of transactions are at peak or down on particular dates.

In this section, an experimental result on Tafeng grocery dataset was made. Then, comparison between our proposed approach and other related approaches were made as shown in Table 2. The proposed STNR algorithm using suffix cactus

Table 1 Sample preprocessed dataset

Date	Unique_Products	Unique_Customers	No. of transactions	Transaction class
01/11/2000	781	193	1,229	A
02/11/2000	3391	1029	7,667	D
03/11/2000	3595	1069	7,628	D
04/11/2000	4530	1221	10,032	F
05/11/2000	5616	1563	6,706	H

Table 2 Comparison of various existing algorithms

Features	STNR suffix cactus	STNR suffix tree	WARP	CONV
Periodicity detection	All type	All type	Segment	Symbol
Time complexity	$O(n \log n)$	$O(n^2)$	$O(n^2)$	$O(n \log n)$
Noise resilient	Good	Good	Good	Worst

is an improvement over suffix tree which follows Apriori-based level-by-level sequential pattern mining approach to mine a specific pattern. It allows the user to skip intermediate events.

4.2 Performance Analysis

The three periodicities are analyzed, and the performance is evaluated by considering datasets with varying size. The size of the dataset is increased with 1 kB each time, and the execution time to detect the symbol, sequence, and segment periodicities is obtained in the performance analysis. Figure 2a shows the time taken for detecting symbol periodicities for input symbols 'a' and 'b'. The time for string 'g' is less since the occurrence of 'g' is less and so the time taken for identifying the periodicity is also less. The time values are plotted against the dataset size.

The same is done for sequence and segment periodicities also with the input strings as 'ab', 'bb' and 'cd' for sequence periodicities and 'abb' and 'cde' for segment periodicities. Here, also the time taken for detecting periodicities for 'cd' and 'cde' is less since they have less number of occurrence vectors. The calculated time values are plotted in graphs in Figs. 2b and 3.

The performance analysis shows that the STNR using suffix cactus was more time efficient. Furthermore, the time increases for increase in the input string size since it needs more processing for identifying periodicities of long strings. The average values of the time for all three periodicities are plotted in graph in Fig. 4.

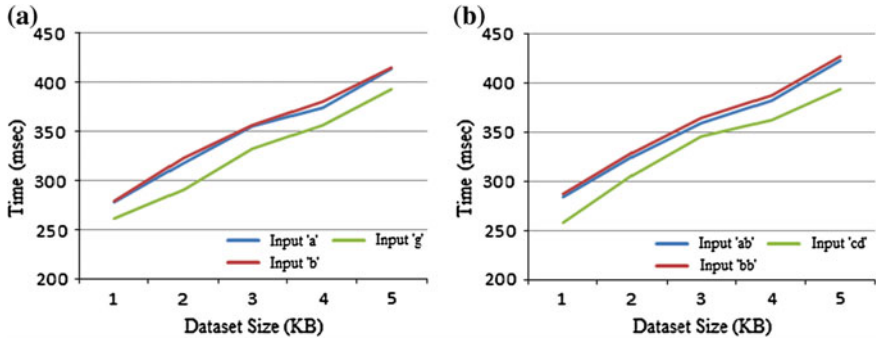


Fig. 2 a Symbol periodicity analysis with time and b sequence periodicity analysis with time

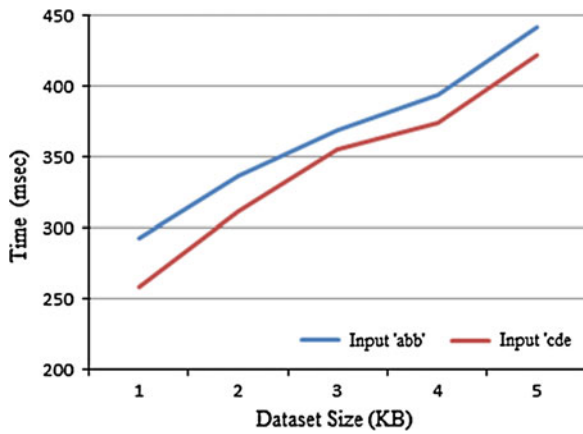


Fig. 3 Segment periodicity analysis with time

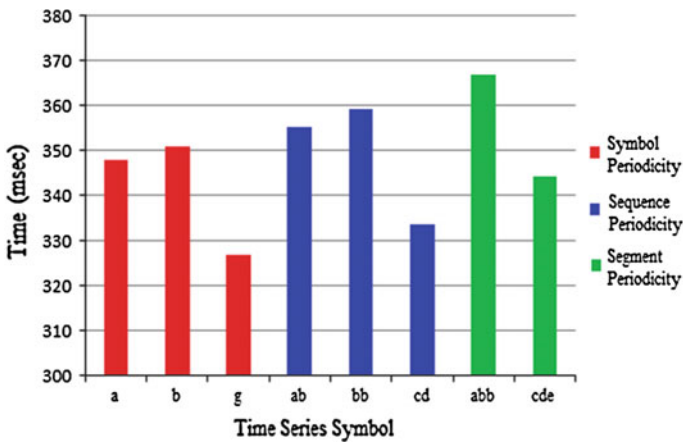


Fig. 4 Overall periodicity detection performance with time

5 Conclusion and Future Work

To mine periodicity in time series database, the existing algorithms face difficulty or time consuming to make decision whether there exists symbol, sequence, or segment periodicity. Thus, the proposed STNR algorithm uses both suffix tree and suffix cactus as an underlying data structure. The suffix cactus is an intermediate approach for storing and searching of patterns. It can be seen as a compact variation of suffix tree. The STNR algorithm can detect all the three types of periodicities such as symbol, sequence, and segment periodicity in a single run. The result shows that the suffix cactus is more space efficient than the best-known suffix tree. It is not only space efficient but also runs faster because it is a simpler structure. Our future work will focus on trend analysis, by which we can analyze the peak sales of the unique products and vice versa in order to make appropriate decisions and corrective measures to boost the optimum profit.

References

1. F. Rasheed, M. Alshalalfa, R. Alhadjj, Efficient periodicity mining in time series databases using suffix trees. *IEEE Trans Knowl Data Eng.* **23**(1) (2011)
2. M.G. Elfeky, W.G. Aref, A.K. Elmagarmid, Periodicity detection in time series databases. *IEEE Trans. Knowl. Data Eng.* **17**, 875–887 (2005)
3. M.G. Elfeky, W.G. Aref, A.K. Elmagarmid, Warp: time warping for periodicity detection, in *Fifth IEEE international conference on Data Mining* (2005), pp. 8–12
4. K.-Y. Haung, C.H. Chang., SMCA: a general model for mining asynchronous periodic patterns in temporal databases. *IEEE Trans. Knowl. Data Eng.* **17**, 774–785 (2005)
5. J. Han, G. Dong, Y. Yin, Mining segment-wise periodic patterns in time related databases, in *Proceedings of ACM International Conference on Knowledge Discovery and Data Mining* (1998), pp. 214–218
6. J. Han, G. Dong, Y., Yin, Efficient mining of partial periodic patterns in time series databases, in *Proceedings of the 15th International Conference on Data Engineering* (1999), pp. 106–115
7. J. Yang., W. Wang., P.S. Yu, InfoMiner+: mining partial periodic patterns with gap penalties, in *Proceedings of the Second IEEE International Conference on Data Mining* (2002), pp. 725–728
8. M.A. Nishi, C.F. Ahmed, M.D. Samiullah, B.-S. Jeong, Effective periodic pattern mining in time series databases. *J Expert Syst Appl* **40**, 3015–3027 (2013)

Group-Based Access Technique for Effective Resource Utilization and Access Control Mechanism in Cloud

Lavanya Selvaraj and Saravana Kumar

Abstract Due to the demand on large volume of data access, the cloud computing is currently a popular model in computing world to process large volumetric data using clusters of commodity systems. It delivers computing resources as a service in which data security and access control is one of the most challenging ongoing research works in cloud computing, because of users outsourcing their sensitive data to cloud providers. Existing solutions that use pure cryptographic techniques to mitigate these security and access control problems suffer from heavy computational overhead on the data owner as well as the cloud service provider for key distribution and management. Protecting data and handling resources for solving problem is a difficult task. In order to overcome this problem, the recent trend is to use cloud computing, supporting resource sharing. The objective of this project is to solve the above challenging problem using group-based instance access control technique that ensures effective utilization of resources within a particular time slot and the verification scheme for secure data access in cloud environment. This work also suggests a proactive secret sharing scheme between data owner and cloud service provider, cloud service provider and the user for secure data access that alleviates the problem of key distribution and management at cloud service provider. In this context, the above approach provides effective utilization of resources and security as well.

L. Selvaraj (✉)

Department of Information Technology, Sri Krishna College
of Engineering and Technology, Coimbatore, India
e-mail: lavanyas@skcet.ac.in

S. Kumar

Department of Computer Science, Bannari Amman Institute
of Technology, Erode, India

© Springer India 2015

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324, DOI 10.1007/978-81-322-2126-5_85

793

1 Introduction

With the advancement in science and technology, the cloud computing has become a necessity nowadays when an enterprise plans to increase its' capacity or capabilities on the fly without investing on new infrastructure, training new personnel, buying new software licenses, etc. It includes any subscription-based or pay-per-use service that extends the enterprise's existing IT capabilities, over the Internet in real time.

It is significantly necessary to utilize security controls and policies that protect sensitive data no matter where it exists, as point solutions by their very nature provide only limited visibility. We are investigating on secure cloud computing that due to the extensive complexity of the cloud, it will be difficult to provide a holistic solution to securing the cloud, at present. Therefore, our goal is to make increment enhancements to securing the cloud that will ultimately result in a secure cloud. The major security challenge with clouds is that the owner of the data may not have control of where the data are placed. This is because if one wants to exploit the benefits of using cloud computing, one must also utilize the resource allocation and scheduling provided by clouds. Therefore, we need to safeguard the data in the midst of untrusted processes.

The effective cloud security solution should incorporate three key capabilities:

1. Data lockdown
 2. Access policies
 3. Security intelligence
- First, make sure that data are not readable to the outside world and that the solution offers strong key management.
 - Second, implement access policies that ensure only authorized users can gain access to sensitive information
 - Third, incorporate security intelligence that generates log information, which can be used for behavioral analysis to provide alerts that trigger when users are performing actions outside of the norm.

Next, there are lot many techniques available that gives details about how the resources are effectively utilized. But there is still an issue that ensures a problem of resource utilization in cloud environment. To overcome the above problems, we suggested a new technique that solves the problem of resource utilization and enhances the security with new key management approach.

2 Reviewed Schemes

Due to the development of cloud technology, the market research and analysis firm IDC suggest that the market for cloud computing services was \$16bn in the year 2008 and will rise to \$42bn/year by 2012 [1]. Cloud computing can also be defined as "a type of parallel and distributed system consisting of a collection of

interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers.” In recent past, various commercial models are developed that are described by “X as a Service (XaaS)” where X could be hardware, software, or storage, etc. [3].

Successful examples of emerging cloud computing infrastructures are Microsoft Azure, Amazon’s EC2 and S3, and Google App Engine, etc. Cloud computing also faces the data security challenges as that of any other communication models. As data owners store their data on external servers, there have been increasing demands and concerns for data confidentiality, authentication, and access control [4].

Work done in [4] proposes cryptographic access control model as shown in Fig. 1 which we have also considered as the system model in our work. The model depicted in Fig. 2 has three participants: data owner (DO), cloud service provider (CSP), and the user. The DO places the outsourced data on the CSP which the user wants to access. As the CSP is untrusted, DO places encrypted data on CSP. Upon receiving a data access request from the user, DO sends required keys and a certificate to the user.

User then presents the certificate to CSP and gets the encrypted data upon successful verification by CSP. The model described guarantees confidentiality, integrity, and authentication, but the problem with this model is that the owner should be always online when the user wants to access the data. The key management between all the communicating parties is also cumbersome.

The general principle of cryptography has also been used with access control lists (ACLs) for ensuring access control and confidentiality to the data storage on

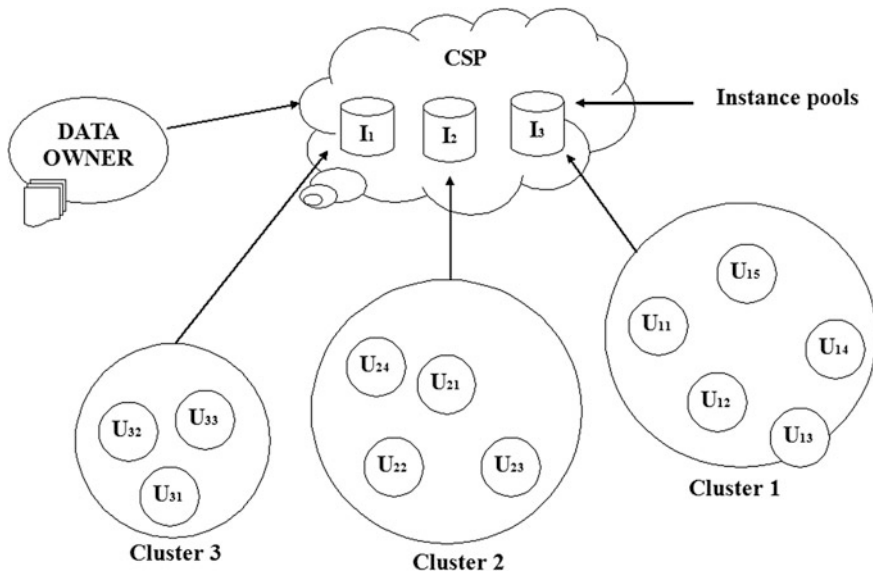


Fig. 1 Grouping of instances makes resource utilization efficient

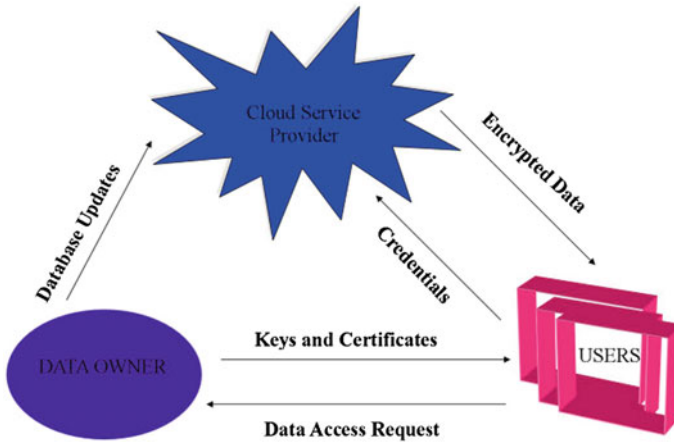


Fig. 2 Secure data access

untrusted servers. Use of ACLs or file groups reduces the complexity of data encryption and key management. However, ACLs or file groups still lack scalability and fine-grainedness for confidentiality and access control in cloud computing [4]. Access control policies based on data attributes and encryption as suggested in [4] also become cumbersome as it is computationally challenging to derive a unique logical expression for every user in the cloud.

In this paper, we address the issue of access control and propose a scalable and efficient data access control mechanism using group-based access control technique and provide security over cloud environment. Data owner encrypts the outsourced data with a symmetric key which is shared only with the user. The CSP and user generate a symmetric key using RSA protocol for the purpose of secure communication between them. The remainder of the paper is organized as follows. Sect. 3 discusses our proposed scheme. In Sect. 4, we analyze our proposed scheme in terms of security. Finally, Sect. 4 concludes the paper and presents future research directions.

3 Proposed Scheme

3.1 Effective Resource Utilization and Access Control Mechanism

The DO is the one who is providing the data in cloud environment. The CSP gets the resources from the data owner and allows access to consumer with enhanced security policy. Here, the proposed work intimates that CSP should group the instance family such as general purpose, compute optimized, GPU instances, storage optimized, memory optimized, and micro-instances according to services

utilized by the consumer. CSP groups the users as clusters based on instance family requested by the consumer. The instance requested by the consumers is high; those customers are grouped under one cluster. The customer who needs less instance utilization is grouped as another cluster and so on. Likewise, the CSP should be able to assign the instance to the consumers according to the requirements with the help of jumper-firefly algorithm.

3.2 Group-Based Access Technique

Figure 1 shows the grouping of users based on the resource usage or the resource utilization based on timings. Here, the CSP groups the instances (In Fig. 1, I_1, I_2, I_3, \dots , etc., are the instance pools where instance families are grouped based on the services such as memory usage, I/O operations needed, processor, storage area of the consumer, etc.) and assigns the valid consumer who wants to use the particular resources.

Consider, the instance pool I_1 contains general purpose instance family with instance type, processor architecture, instance storage, physical processor, etc., and I_2 contains compute-optimized instance family with instance type, processor architecture, instance storage, physical processor, and so on. If new consumer requesting the CSP since the instance family has been grouped, the CSP can easily identify the instance for the new consumer and assign it to particular cloud where its instance services match.

By grouping instances by the above criteria, the maintenance as well as managing the instances and users in a cloud environment is an easy task. Also, search time and service time are reduced by doing grouping. The data integrity and access control mechanism are defined in further sections.

3.3 Data Security and Access Control Mechanism

This mechanism is used to protect the user from unwanted usage of resources and enables access control toward the cloud infrastructure. It is mainly structured for providing security toward outsourced data. The following steps are provided as service and these reside in CSP, and the consumer who needs high security over data can access this **security as a service**.

User identity, public–private key generation, and signature generation: User identity (UID) is a nonnegative integer assigned uniquely to the user by the CSP. Each user is given long-term public and private keys. The DO randomly chooses a secret key and then computes and publishes the corresponding public key. The system uses the idea of RSA to construct a private–public key pair, where they DO calculates

1. Public key (M, e) , where M is the product of any two large prime numbers, a and b , and e is the number prime with respect to M and
2. Private key $(a, b, d, \varphi(M))$, where d is the part of private key of DO and is equal to $e^{-1} \bmod \varphi(M)$.

Each CSP provides $UIDs$ of the user under it and obtains the signature S_{ij} for each UID_{ij} of a user U_{ij} , where $i = 1, 2, 3 \dots$, represents cluster and $j = 1, 2, 3$ represents the user. If DO confirms the correctness and the relationship between U_{ij} and CSP_i , then it calculates S_{ij} using Eq. (1) and distributes S_{ij} to each CSP where each CSP distributes them to the respective users.

$$S_{ij} = UID_i^d \bmod M \quad (1)$$

Secure and scalable data access: To greatly achieve a secure data access, if any user wants to access the data from the CSP, then they will be given a cluster key (CK) which is provided by CSP and the master key (MK) which is provided by DO. The process is as follows.

Assume U_{11}, \dots, U_{ij} are the consumers, C_1, \dots, C_i are the number of clusters generated by CSP, CK_{11}, \dots, CK_{ij} are the cluster keys, and I_{11}, \dots, I_{ij} are the instances (i.e., resources) used by the consumers in the cloud environment. Here, i is the number of cluster where $i = 1 \dots n$ and j is the number of users where $j = 1 \dots n$. If U_{ij} wants to access the resource I_{ij} in a particular cluster, then the corresponding CSP should assign the CK_{ij} to the user U_{ij} to use the resource I_{ij} . The CK_{ij} is the cluster key for accessing the resource in a particular cluster which is dynamic.

Initially, CSP assigns a random prime number called cluster key CK_{11} to U_{11} in cluster C_1 . Initially, CK_{11} (generated by CSP) is XORed with user key (UK_{11}) which is generated by the user while joining.

$$CK_{11} \oplus UK_{11} = CK_{\text{new}} \quad (2)$$

Now CK_{new} is taken as a new cluster key to access the data (i.e., the above is satisfied if only one user U_{11} is present in cluster₁). This key will be deleted as soon as the user stop accessing the data or their time expires. (Note: The CK_{ij} is different for different cluster, and it is known only to the CSP not even to the user so that the spoofing is impossible.) If more users are added to the cluster, then the steps are proceeding as follows.

Consider U_1 to U_4 are in cluster C_i and its cluster key is CK_{ij} which is some K 's where K is an integer provided by CSP. U_{15} is a newly joined user wants to access the data/resources. Then, the CK_{new} for user U_{15} is calculated as follows.

$$CK_{ij} \oplus UK_{15} = CK_{\text{new}} \quad (3)$$

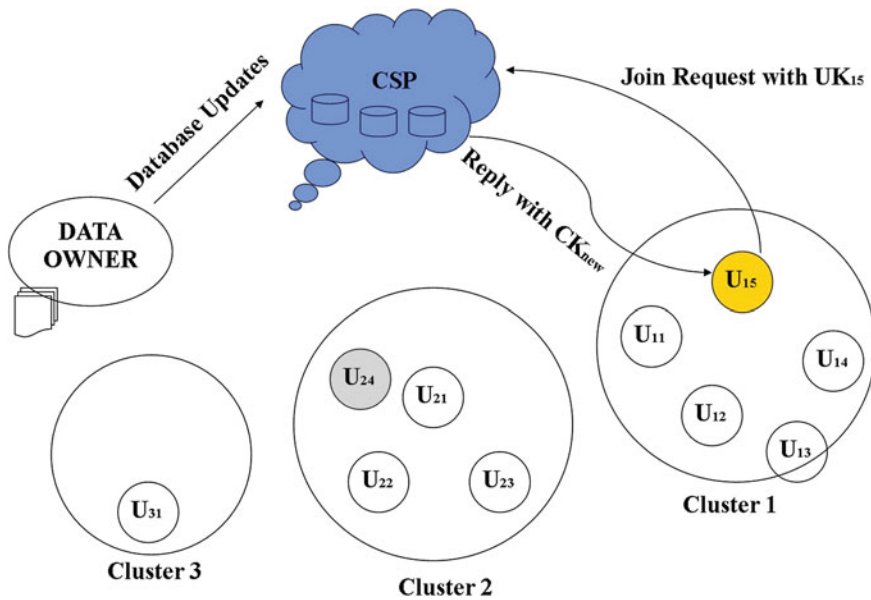


Fig. 3 Generation of keys

Now, the CK_{new} should be used by the user U_{15} to access the data/resources present in a cloud. Similarly, the above procedure is followed for all the clusters. The general formula is (Fig. 3)

$$CK_{ij} \oplus UK_{ij} = CK_{new} \tag{4}$$

Two access control lists, namely existing list and leaving list, are maintained to keep track of user’s log. These two lists are maintained by DO, and it ensures access control in cloud environment (Fig. 4).

The existing list consists of a different parameter such as UID_{ij} , CK_{ij} , and I_{ij} used by the particular user. The leaving list contains the UID_{ij} of the user to ensure authenticity. The purpose of maintaining UID in leaving list is when the user left the cluster once will have the previous keys for accessing the resource. If the user tries to access the resource with old keys whose UID is in leaving list will be considered as attacker/hacker.

Communication between CSP and consumer, CSP and DO in the cloud Environment: Let us know how to provide security during communication between CSP and the users (i.e., CSP and U_{ij}). Here, the UID_{ij} are the user identities assigned to each user in the cluster C_i , public key (d) and private key (e), and signatures (S_{ij}) of each U_{ij} are generated by CSP using RSA.

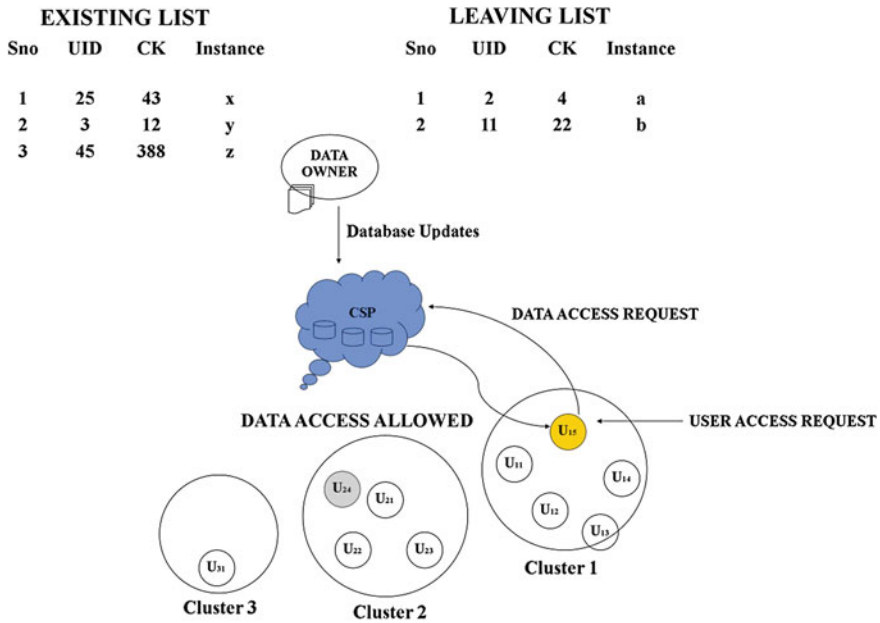


Fig. 4 An example of data access scenario with lists

1. U_{ij} selects a random number R_{ij} and computes two public keys x_{ij} and y_{ij} as follows:

$$x_{ij} = S_{ij} \cdot \alpha^{R_{ij}} \text{ mod } M \tag{5}$$

$$y_{ij} = R_{ij}^e \text{ mod } M \tag{6}$$

2. U_{ij} uses the identification number ID_i of CSP and finds P_{ij} by doing the operation one-way function of $h(x_{ij}, y_{ij}, ID_i)$, then computes

$$P_{ij} = S_{ij} \cdot R_{ij}^{h(x_{ij}, y_{ij}, ID_i)} \text{ mod } M \tag{7}$$

3. U_{ij} sends (ID_i, x_{ij}, y_{ij}) to CSP.

Similarly, member CSP selects the random number R_{kj} , then computes x_{kj} , y_{kj} , and P_{kj} , and sends $(UID_{ij}, x_{kj}, y_{kj})$ to U_{ij} . U_{ij} and CSP have to verify whether (ID_i, x_{ij}, y_{ij}) and $(UID_{ij}, x_{kj}, y_{kj})$ are sent from the user U_{ij} and the CSP, respectively. It is done by checking

$$P_{kj}^e = \text{ID}_i \cdot y_{kj}^{h(x_{kj}, y_{kj}, \text{ID}_{ij})} \pmod{M} \quad (8)$$

Consider P_{kj} from Eq. 9. From Eq. 8,

$$P_{kj}^e = (\text{ID}_i^d \pmod{M})^e \cdot \left(R_{kj}^{h(x_{kj}, y_{kj}, \text{ID}_i)} \pmod{M} \right)^e \quad (9)$$

Mathematically, $(G^x \pmod{n})^y = (G^y \pmod{n})^x = G^{xy} \pmod{n}$ and $(G^x \pmod{n}) \pmod{n} = (G^x \pmod{n})$ because n is a very large number. According to RSA, $d = e^{-1} \pmod{\varphi(n)}$ and $d * e = 1 \pmod{\varphi(n)} = 1$. Similarly, member verifies at his end. The communicating members compute a secret session key (SK). The computation of SK s is as follows: consider the communication between user U_{ij} and CSP. They compute the secret SK_{ij} (user's session key) and SK (CSP's session key), respectively, as follows:

$$SK_{ij} = \left(\frac{x_{kj}^e}{\text{ID}_i} \right)_{kj}^R \pmod{M} \quad (10)$$

$$SK = \left(\frac{x_{ij}^e}{\text{UID}_{ij}} \right)_{ij}^R \pmod{M} \quad (11)$$

Using these session keys, they can communicate successfully with each other. If the session key is same for both (i.e., CSP and the consumer), then the consumer and the CSP are authenticated and ensure data integrity. Session key is used to encrypt/decrypt the data between the CSP and the consumer.

3.4 Benefits of the Proposed Scheme

Whenever a consumer requests or relieves, is notified to CSP once for each change. Hence, only one message is sufficient for any change in the environment. The number of messages exchanged at any change in a infrastructure is 1. The cluster key is dynamic for any change in the number of users along with ensuring the security is necessary. Here, i is the cluster size and n is the group size. When a consumer joins the group, the cluster key is regenerated along with the UID, public–privates keys pair, and signature. For each newly joining user, three new keys and one UID are generated. If there are n members joining at the same time, then $4n$ computations are done. Whenever a user leaves the cluster or if n members are leaving the cluster at the same time, no keys are required to generate. This technique avoids denial of service attack because the trusted users with valid user identities are allowed to use the cloud infrastructure, avoid side channel attacks

since it uses strong cryptographic technique called hash function which is unbreakable, and avoid phishing attack by the usage of digital signature in this system.

4 Conclusion

As we suggested that a set of security protocols to secure the data of a data owner in the cloud infrastructure, the combined cryptographic approach is used to protect the outsourced data. The group-based model for access control mechanism along with public key encryption provides efficient utilization of resources and assures secure data access. The RSA key management technique is efficient for large number of cloud user scenario and ensures security. The public key, signature, and private key, cluster keys, ciphers that are proposed between cloud service provider, data owner, and consumer ensures a secure cloud infrastructure has been assured in the last section. Future extensions will include enhancement of key distribution and resource scheduling techniques.

References

1. E. Gleeson, Computing industry set for a shocking change. MoneyWeek (2009)
2. S.D.C. Di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, P. Samarati, A data outsourcing architecture combining cryptography and access control, in *Proceedings of ACM Workshop on Computer Security Architecture (CSAW'07)* (2007)
3. S. Yu, C. Wang, K. Ren, W. Lou, Achieving secure, scalable, and fine-grained data access control in cloud computing, in *Proceedings of IEEE INFOCOM* (2010), pp. 1–9
4. W. Wang, Z. Li, R. Owens, B. Bhargava, Secure and efficient access to outsourced data, in *Proceedings of ACM Cloud Computing Security Workshop* (2009), pp. 55–65

Design of Fuzzy Logic-Based pH Controller for High-Pressure-Rated Modified CSTR System

Jithin Kannangot, Ponnusamy Lakshmi and Keppayan Thiruppathi

Abstract In this paper, control of pH in the high-pressure-rated environmental continuous stirred tank reactor (CSTR) where deep sea conditions are mimicked is considered. Since the pressure in deep sea is higher, the pressure inside the environmental CSTR is also kept at an elevated value. Temperature inside the system is considered to be constant. Control of pH inside such a high-pressure-rated environmental CSTR system is a tedious task. First-order plus dead time (FOPDT) model is derived from the real-time system open-loop pH curve. Conventional PI controller and fuzzy logic controller are developed for the control of pH and their performance indices also compared. MATLAB Simulink block is used for controller simulation and result comparison. PI controller is tuned using minimum ISE Zhuang and Atherton method tuning protocols. Simulation result of PI controller and fuzzy logic controller are presented for comparing their efficiency; also various performance indices are calculated, and the best control action is identified. Fuzzy logic controller gives better servo and regulatory response than PI controller.

Keywords CSTR · First-order plus dead time · Conventional PI controller · Fuzzy logic controller · Ziegler Nichols and Cohen Coon methods

J. Kannangot (✉) · P. Lakshmi
College of Engineering Guindy, Chennai, Tamil Nadu, India
e-mail: jithinkannangot@gmail.com

P. Lakshmi
e-mail: lakshmi_p_2000@yahoo.com

K. Thiruppathi
National Institute of Ocean Technology, Chennai, Tamil Nadu, India
e-mail: thiru@niot.res.in

1 Introduction

pH is the degree of acidity or alkalinity of a solution. It is measured as $-\log[H^+]$, and the range of pH value is between 0 and 14. Due to the logarithmic relationship between hydrogen ion concentration and pH value, pH process is highly nonlinear [1, 2]. Control of pH is a difficult task due to the nonlinearity and higher sensitivity to disturbances of the process [3, 4]. In this work, deep sea condition is mimicked using high-pressure-rated environmental CSTR system to study the growth of microorganisms. For that, pH inside setup has to be maintained at a predetermined level at elevated pressure. Due to various conditions like elevated pressure and enzymes produced by microbes inside the environmental CSTR, the pH value is going to vary frequently [5]. For the survival of microbes, this variation should be controlled. Around the neutral point (pH value 7), the pH curve has a huge gain; that is, even a small change in acid/base flow is going to vary the pH value drastically [3]. The control of pH around this region is really challenging.

The system has to be modeled accurately for the design of controller. In this work, it is modeled as a first-order plus dead time (FOPDT) system from the real-time open-loop response. The system is having significant delay due to various reasons like the delay for uniform mixing of acid/base in the five-liter environmental CSTR.

Generally, all systems are initially checked with conventional controllers including P, PI, and PID [6] since it is easy to develop and implement. If the response is not satisfactory, advanced controllers are considered. When the system is nonlinear with significant delay, conventional controllers cannot give satisfactory result [7–9]. Fuzzy logic controller is a suitable alternative in such case. It can deal with nonlinear systems efficiently. Mamdani-type fuzzy controller is more suitable for nonlinear systems [9]. Membership functions and rule base are the main part of fuzzy logic controller. These can be developed from experience about the process. The range of membership function depends upon the operating region of the process [10]. By trial and error, the range, gain values, as well as rule set can be modified and FLC can be implemented. In this work, a conventional PI controller and fuzzy logic controller are developed for the system and their performance is compared.

In Sect. 2, a detailed description about the experimental setup and system designing is given. Section 3 describes about the designing of PI controller and fuzzy logic controller for the system. In Sect. 4, the simulation results are compared, and Sect. 5 contains conclusion.

2 Experimental Setup Description

The physical system used to create deep sea conditions in the environmental CSTR system includes a high-pressure-rated double-jacketed reactor vessel, multipoint serial server, digitally controlled heater/chiller system, modem, temperature sensor,

pH sensor, and PC. The reactor vessel used here has a volume of 5L and can be emptied via a drain pipe controlled by valve. The working pressure and temperature, at which any reactor can be used, will entirely depend upon the design, size, and nature of the material used for construction [11]. Since all materials tend to vary their strength according to change in temperature, any pressure rating must be stated in terms of the temperature at which it is applicable. We have selected 350-bar pressure-rated reactor vessel, 140-bar pressure-rated special pH probe, and 400-bar pressure-rated special positive displacement dosing pumps for real-time experiments. Communication between them is established using Delphi 6 software running on a PC. The Multiport Industrial Serial Server SE5008/SE5016 in the system is equipped with serial communication ports RS232, RS485, and RS422. It is a gateway between ethernet (TCP/IP) and serial communication ports. It allows most of the serial devices to be connected to a new or existing ethernet network. The simulation model of the system is shown in Fig. 1.

The pressure inside the vessel is kept at an elevated value by using pneumatic pressure building technique in the environmental CSTR system through a solenoid valve. The CSTR is kept closed to maintain the pressure. Acid/base is injected to the vessel by using a special high-pressure pump called 400-bar pressure-rated positive displacement pump at the rate of 1 ml/min. Added acid/base is well mixed using a hermetically sealed magnetic stirrer which rotates at a speed of 300 rpm. Special type of pH probe is used to measure the pH inside the setup since it should be able to withstand in the elevated pressure. In this experiment, strong acid (HCL) and strong base (NaOH) of one molarity is prepared and used to conduct real-time experiments. The model parameters of the system and pH probe details are given in Tables 1 and 2.

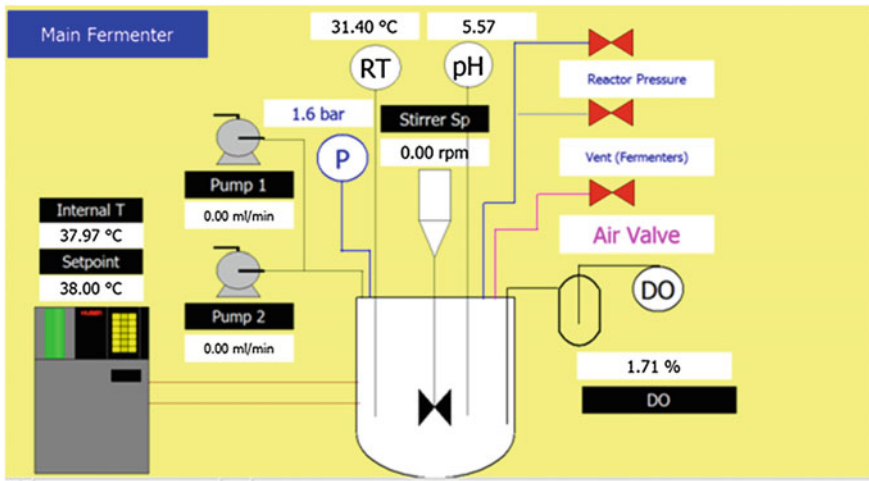


Fig. 1 Diagram of simulation model for the system

Table 1 Model parameters of the system

Name	Value
Number of fermenter vessels	1
Fermenter vessel volume	5 L
Working pressure	Up to 350 bar
Working temperature	30 °C
Reactor material construction:	Corrosion 316 stainless steel with PTFE inner coating
Sterilization	There is a steam sterilization system to sterilize all the reactor vessels before and after the process

Table 2 pH probe details

Name	Value
Sensitivity of probe in voltage	mV
Resolution of the probe	0.001
Range	0–14
Pressure rating	Maximum 140 bar

Considering the delay factors, system is modeled as FOPDT system whose general transfer function model is given below,

$$G(s) = \frac{Ke^{-\tau_d s}}{\tau s + 1} \quad (1)$$

The real-time open-loop response of the system is shown in Fig. 2.

The transfer function model of the system obtained from the open-loop response is

$$G(s) = \frac{0.276e^{-5.005s}}{3.2s + 1} \quad (2)$$

The simplified model of the system is shown in Fig. 3.

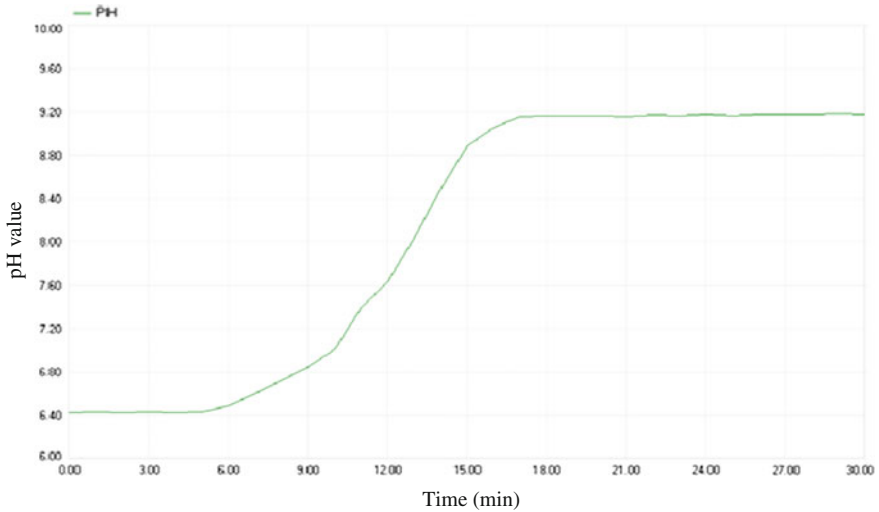


Fig. 2 Real-time open-loop response of the system

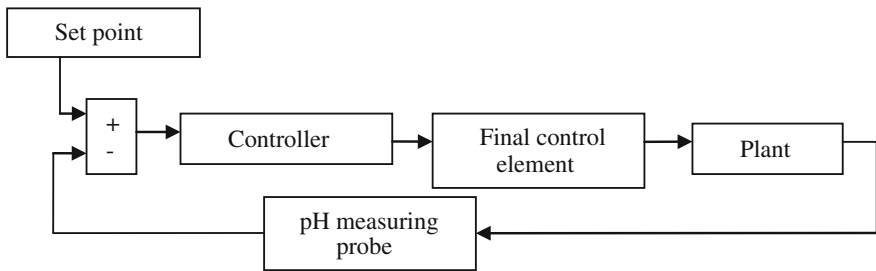


Fig. 3 Simplified model of the system

3 Design and Performance Comparison of PI and FUZZY Controllers

In this section, the conventional PI controller and fuzzy controller design is discussed; also the responses of the system with these two controllers are compared.

3.1 PI Controller Design

Initially, the conventional PI controller is designed for the system. Various tuning methods are available for calculating K_p and T_i values of PI controllers. The Ziegler Nichols and Cohen Coon methods give satisfactory values for K_p and T_i if the

quarter decay ratio, that is, the ratio between the delay time and the time constant of the system, is ≤ 1 ($\frac{\tau_d}{\tau} \leq 1$). But for this system, the ratio is $\frac{\tau_d}{\tau} = 1.56$. So, it is clear that these tuning methods are not going to give satisfactory values for k_p and T_i . Minimum ISE Zhuang and Atherton (Zhuang and Atherton, 1993) method is used to find controller parameters in this case.

$$K_p = \frac{1.072}{K} \left(\frac{\tau}{\tau_d} \right)^{0.560} \quad T_i = \frac{\tau}{0.690 - 0.155 \frac{\tau_d}{\tau}} \quad (3)$$

Using the above equations, obtained parameter values are $K_p = 3.6$ and $T_i = 7.05$

3.2 Fuzzy Logic Controller

In this section, a fuzzy logic controller is developed for the system based on Mamdani’s methodology [12]. Membership functions for the fuzzy controller can be generally defined using human knowledge of the process [13, 14]. The range of membership functions depends upon the operating range of neutralization process [15]. For this controller, two input variables, error (e) and change in error (Δe), as well as one output variable (u) is used.

The membership functions for the variables are shown in Figs. 4, 5 and 6.

The corresponding rule table is given in Table 3.

Fig. 4 Membership function for e

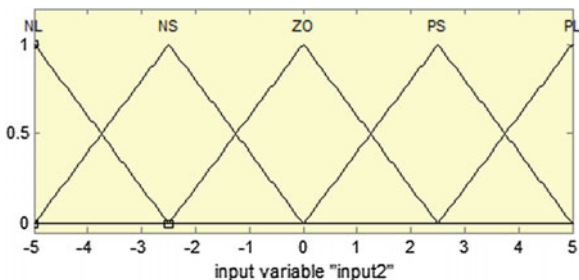


Fig. 5 Membership function for Δe

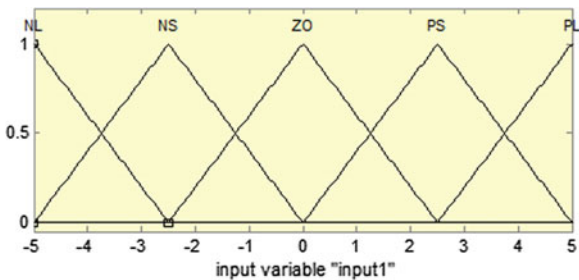


Fig. 6 Membership function for u

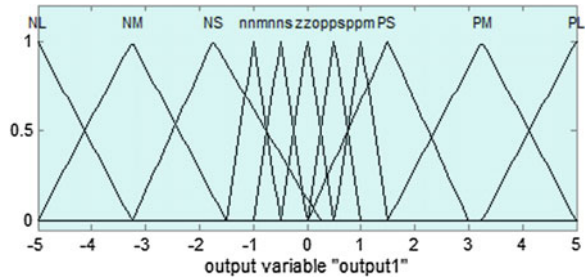


Table 3 Rule table of FLC

e	Δe				
	NL	NS	ZO	PS	PL
NL	PM	PS	ppm	pps	zzo
NS	ppm	pps	pps	zzo	nns
ZO	pps	pps	zzo	nns	nns
PS	pps	zzo	nns	nns	nnm
PL	zzo	nns	nnm	NS	NM

4 Results and Discussion

The PI and fuzzy logic controller performance for the system is simulated and compared using MATLAB Simulink. The PI controller is tuned with minimum ISE Zhuang and Atherton tuning rule [3]. Membership function and rule base for fuzzy logic controller is developed from the experience about the process, and it is fine-tuned by trial and error method. Both the servo and regulatory response of the system with the two controllers are compared. It is observed that overshoot is minimized with fuzzy controller compared to PI controller, and the output settles faster in case of fuzzy logic controller.

4.1 Servo Response

The servo response with two step changes are shown in the below Fig. 7. A positive step change of 23.07 % of initial value is applied at 150th minute. Again a negative step change of 38.46 % of initial value is applied at 300th minute. The response with PI and fuzzy controller is compared.

Fuzzy logic controller gives a better response because it is able to adapt the nonlinear variation in pH quickly compared to PI controller. For every condition in error and change in error, a corresponding output is defined for FLC which makes it better than conventional controller with fixed parameters.

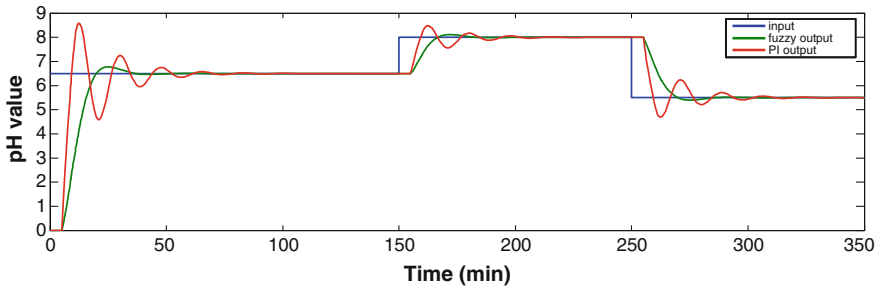


Fig. 7 Servo response of PI controller and FLC

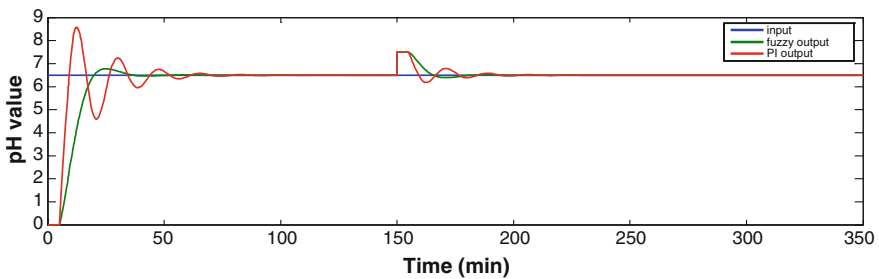


Fig. 8 Regulatory response of PI controller and FLC

4.2 Regulatory Response

After the response is settled, a step disturbance of 15.38 % of the initial value is introduced at 150th minute. The regulatory response with PI controller and fuzzy controller is compared (Fig. 8).

The maximum peak overshoot (M_p) for PI controller is 30 %, and for FLC, it is 3 %, also the settling time (T_s) for PI controller is 80 min, and for FLC, it is 50 min.

The efficiency of both the controllers is compared using the performance indices integral absolute error (IAE) and integral time absolute error (ITAE) which are tabulated in Table 4.

Table 4 Performance indices for PI and fuzzy controller

Response	Performance indices	PI	Fuzzy
Servo response	IAE	125.1	120.3
	ITAE	11,700	10,310
Regulatory response	IAE	89.41	87.06
	ITAE	2,916	2,300

5 Conclusion

The paper describes about the implementation of PI and fuzzy logic controller for a high-pressure-rated environmental CSTR system. The results for both controllers are compared from the simulation results, and it is observed that fuzzy logic controller shows better performance in terms of Mp (%), Ts, IAE, and ITAE in servo response and regulatory response.

References

1. R. Ibrahim, *Practical Modelling and Control Implementation Studies on a pH Neutralization Pilot Plant* (2008), pp. 1–198
2. K.S. Saji, M. Sasi Kumar, Fuzzy sliding mode control of a pH process. IEEE international conference on communication control and computing technologies (2010), pp. 276–281
3. K. Jiayu, W. Mengxiao, X. Zhongjun, Z. Yan, Fuzzy PID control of the pH in an anaerobic wastewater treatment process (2009), pp. 978–982
4. O.L.R. Jacobs, P.F. Hewkin, C. While, Online computer control of pH in an industrial process. IEEPROC **127**, 161–168 (1980)
5. C. Culberson, R.M. Pytkowicz, Effect of pressure on carbonic acid, boric acid and the pH in seawater. *Limnol. Oceanog.* XIII, 403–417 (1968)
6. M. Zhuang, D.P. Atherton, Automatic tuning of optimum PID controllers. IEEE Proc. **140**, 216–224 (1993)
7. R. Muthu, E. El Kanzi, Fuzzy logic control of pH neutralization process. IEEE Proc. ICECS 1066–1069 (2003)
8. S.R.S. Abdullah, M.M. Mustafa, R.A. Rahman, T.O.S. Imm, H.A. Hassan, A fuzzy logic controller of two position pump with time-delay in heavy metal precipitation process. IEEE international conference on pattern analysis and intelligent robotics (2011), pp. 171–176
9. I. Muhaini, M. Noor, Control of pH level using fuzzy controller (2009)
10. M.J. Fuente, C. Robles, O. Casado, F. Tadeo, Fuzzy control of neutralization process. IEEE Proc. (2002)
11. K.P. Thiruppathi, R.L. Ponnusamy, R. Kirubakaran, M.A. Atmanand, Design and optimization of temperature controller for high pressure rated modified CSTR system. IEEE Proc. ISPPCC 1–6 (2013)
12. H.M. Genc, E. Yesil, I. Eksin, M. Guzelkaya, O. Aydin Tekin, A rule base modification scheme in fuzzy controller for time-delay systems. *Expert Syst. Appl.* **36**, 8476–8486 (2009)
13. M. Parekh, M. Desai, H. Li, In line control of pH neutralization based on fuzzy logic. IEEE Trans. Compon. Packing Manuf. Technol. **17**(2), 192–201 (1994)
14. P. Ramanathan, K.C. Sukanya, S. Mishra, S. Ramasamy, Study on fuzzy logic and PID controller for temperature regulation of a system with time delay. IEEE international conference on energy efficient technologies for sustainability (2013), pp. 274–277

Level Control of Quadruple Tank Process with Finite-Time Convergence Using Integral Terminal Sliding Mode Controller

Sekaran Sankaranarayanan, Lakshmi Ponnusamy
and Sangapillai Sutha

Abstract This paper addresses the level control of quadruple tank process (QTP) operating in minimum phase mode. To control the level in bottom tanks, a non-linear controller standard sliding mode controller (SMC) is designed initially. To improve the controlling effort and finite-time convergence of process variable, terminal sliding mode controller (TSMC) is designed. With same operating conditions, integrality is added to TSMC results in integral terminal sliding mode controller (ITSMC) for QTP with relative degree two, which results in better asymptotic error convergence. Undesirable chattering effect in final control element is reduced by introducing Exponential Multilevel Switching Variable Gain is added to the designed controllers. Finally, the efficacy of the proposed scheme is demonstrated by conducting simulation studies using MATLAB on QTP to prove ITSMC is significantly superior to others in terms of robustness, better tracking, rejecting the disturbances with better quality, and performance indices.

Keywords Quadruple tank process · Sliding mode controller · Terminal sliding mode controller · Integral terminal sliding mode controller · Non minimum phase

1 Introduction

A complex and challenging multivariable quadruple tank process (QTP) is proposed in [1], and it also enumerates the various working configurations such as minimum phase (MP) and non minimum phase (NMP) conditions. Different

S. Sankaranarayanan (✉) · L. Ponnusamy · S. Sutha
Department of EEE, College of Engineering Guindy, Anna University, Chennai, India
e-mail: kokilamsankar@gmail.com

L. Ponnusamy
e-mail: lakshmi_p_2000@yahoo.co.in

S. Sutha
e-mail: suthaa_s@yahoo.co.in

analysis techniques are given in elaborate with numerical examples in [2]. Various controllers are implemented in case of QTP from the day of its proposal, and in such a case, application of sliding mode controller (SMC) is made first for QTP that is explained with detailed modeling and control law is given in [3]. In order to understand the concept of SMC to highly complex and nonlinear systems, various numerical solutions are given with detailed explanation in [4, 5]. A similar experiment to QTP is a hybrid tank system where bottom two tanks are interconnected, which is controlled with a SMC in [6], and it makes a useful introduction to SMC applied QTP. To make finite-time convergence in the output response, concept of terminality is used along with SMC to give an improved version called as terminal sliding mode control (TSMC). With TSMC, various processes are controlled [7, 8] using fuzzy concept. The TSMC is designed for SISO system in [9] and extends many MIMO processes in an elaborate manner along with robustness concept in [10–13]. Further advancement is made in TSMC to get better tracking performance and improved working conditions for some parameters and control effort. This leads to the concept of integral terminal sliding mode controller (ITSMC). These concepts are explained with numerical solution for another process in case of relative degree of the system equal to one [14–16].

This paper is structured as follows. In Sect. 2, a description and a dynamic model of the QTP are provided. The design of PI controller, standard SMC, TSMC, and ITSMC are developed in Sect. 3. In Sect. 4, the simulation results are presented and the performances are analyzed to show the efficacy of proposed controller. Finally, Sect. 5 contains conclusions.

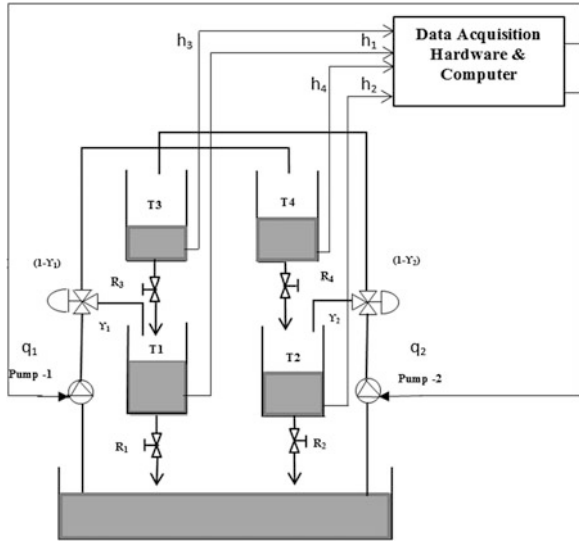
2 Model of Quadruple Tank Process

The structure of laboratory QTP alone is taken as in [1]. The control voltage limit is (0–5) V converted to actuator valve opening from (0 to 100) %. The schematic diagram of the process is given in Fig. 1.

2.1 *Dynamic Model of the Quadruple Tank Process*

A nonlinear process model could be achieved by using mass balance and Bernoulli's law,

Fig. 1 Schematic diagram of quadruple tank process



$$\begin{aligned} \frac{dh_1(t)}{dt} &= -\frac{\beta_1 a_1}{A_1} \sqrt{2gh_1(t)} + \frac{\beta_3 a_3}{A_1} \sqrt{2gh_3(t)} + \frac{\gamma_1 K p_1}{A_1} V_1(t) \\ \frac{dh_2(t)}{dt} &= -\frac{\beta_2 a_2}{A_2} \sqrt{2gh_2(t)} + \frac{\beta_4 a_4}{A_2} \sqrt{2gh_4(t)} + \frac{\gamma_2 K p_2}{A_2} V_2(t) \\ \frac{dh_3(t)}{dt} &= -\frac{\beta_3 a_3}{A_3} \sqrt{2gh_3(t)} + \frac{(1-\gamma_2) K p_2}{A_3} V_2(t) \\ \frac{dh_4(t)}{dt} &= -\frac{\beta_4 a_4}{A_4} \sqrt{2gh_4(t)} + \frac{(1-\gamma_1) K p_1}{A_4} V_1(t) \end{aligned}$$

A_i cross-sectional area (cm^2), a_i cross section of the outlet hole (cm^2), h_i water level (cm) and β_i outlet valve ratio of tanks ($i = 1, 2, 3$) and 4, respectively. γ_i fraction of water flow from pump ($i = 1$ and 2), v_i voltage input (volts) to the control valve ($i = 1$ and 2), and g gravitation due to gravity (cm/s). Based on the condition that a sum of flow ratios lies between $1 < (\gamma_1 + \gamma_2) \leq 2$; then, the system works in MP mode.

With process parameters given in Table 1, transfer function matrix is calculated as given in [1].

$$G(s) = \begin{bmatrix} \frac{6.9605}{(29.2896s+1)} & \frac{4.6403}{(24.9535s+1)(29.2896s+1)} \\ \frac{3.0617}{(29.4612s+1)(21.1366s+1)} & \frac{4.5925}{(21.1366s+1)} \end{bmatrix}$$

Table 1 Process parameters

Process parameters	MP operating point
$A_i(\text{cm}^2)$	176.7146
$a_i(\text{cm}^2)$	2.01062
Y_1, Y_2	(0.6, 0.6)
$(h_1^0, h_2^0, h_3^0, h_4^0)$ in cm	(19.61, 13.9, 6.326, 8.169)
K_{P1}, K_{P2}	(70, 64)
(V_1^0, V_2^0) , in (V)	(3.5, 3.5)
$(\beta_1, \beta_2, \beta_3, \beta_4)$	(0.6, 0.7, 0.4, 0.385)

Based on steady state process gain matrix by $G(s = 0)$, and is represented by

$$\Lambda = G(0)G^{-T}(0) = \begin{bmatrix} 1.800 & -0.800 \\ -0.800 & 1.800 \end{bmatrix}$$

Relative gain analysis (RGA) indicates the output–input pairings should be $y_1 - u_1$ and $y_2 - u_2$. It is shown that for this process, y_1 is paired with u_1 and y_2 paired with u_2 .

3 Various Sliding Mode Control Strategies

3.1 Conversion to Standard Form

Assumption 1 The bottom two tanks are operated under standard or constant voltages, so that derivative of control law becomes zero and the relative degree of the system becomes 2.

Remark 1 The assumption is valid under the open loop condition that a constant voltage is applied to the tanks. On converting the system to companion form, the derivative of the control input becomes zero.

With the above-mentioned assumptions, the equations are rewritten as

$$\begin{aligned} \dot{h}_1 &= -K_{11}\sqrt{h_1} + K_{31}\sqrt{h_3} + \frac{\gamma_1 k_{p1}}{A} V_{10}; & \dot{h}_2 &= -K_{22}\sqrt{h_2} + K_{42}\sqrt{h_4} + \frac{\gamma_2 k_{p2}}{A} V_{20} \\ \dot{h}_3 &= -K_{33}\sqrt{h_3} + \frac{(1 - \gamma_2)k_{p2}}{A} V_2; & \dot{h}_4 &= -K_{44}\sqrt{h_4} + \frac{(1 - \gamma_1)k_{p1}}{A} V_1 \end{aligned} \tag{1}$$

where

$$K_{ij} = \frac{\beta_{ij} a}{A} \sqrt{2g}.$$

- V_{i0} constant input voltage
- i index of the tank, for which the dynamics are explained
- j index of the tank written with respective i th tank

Following equations are derived with the help of assumptions and equations mentioned above which lead to the standard form of the equation.

$$x_1 = h_1 - h_{ref1}; \quad x_3 = h_2 - h_{ref2}.$$

where

- x_1, x_3 error generated in tanks 1 and 2, respectively, system or error variable
- h_{ref1}, h_{ref2} reference height of tanks 1 and 2

Thus, for obtaining the phase, variable format or simply a standard form of equations following procedure is followed.

$$\dot{x}_1 = x_2 = \dot{h}_1; \dot{x}_2 = \ddot{h}_1; \quad \dot{x}_3 = x_4 = \dot{h}_2; \dot{x}_4 = \ddot{h}_2$$

The complete phase variable equation format is

$$\begin{aligned} \dot{x}_1 = x_2; \quad \dot{x}_2 &= -\left(\frac{K_{11}}{2\sqrt{h_1}}\right)x_2 - \left(\frac{K_{21}K_{33}}{2}\right) + \left(\frac{K_{31}k_{p2}(1-\gamma_2)}{2\sqrt{h_3}A}\right)V_2 \\ \dot{x}_3 = x_4; \quad \dot{x}_4 &= -\left(\frac{K_{22}}{2\sqrt{h_2}}\right)x_4 - \left(\frac{K_{42}K_{44}}{2}\right) + \left(\frac{K_{42}k_{p1}(1-\gamma_1)}{2\sqrt{h_4}A}\right)V_1 \end{aligned} \tag{2}$$

3.2 Sliding Mode Control

Generally, sliding mode control is used for tracking the desired trajectory even any functional parameter or input parameter changes.

$$S_1 = x_2 + m_1x_1; \quad \dot{S}_1 = \dot{x}_2 + m_1x_2 \tag{3}$$

where $m_1 > 0$ is a positive constant. The constant determines the slope of the sliding surface. For checking the stability of the sliding surface, a positive definite Lyapunov can be taken and the differentiation of the function as

$$V_{n2}(S_1) = \frac{1}{2}S_1^2 > 0; \quad \dot{V}_{n2}(S_1) = S_1\dot{S}_1 < 0$$

For making $\dot{V}_{n2}(S_1) < 0$

$$V_2 = - \left(\frac{K_{31}k_{p2}(1-\gamma_2)}{2\sqrt{h_3}A} \right)^{-1} \left[\left(m_1 - \frac{K_{11}}{2\sqrt{h_1}} \right) x_2 - \left(\frac{K_{31}K_{33}}{2} \right) + \alpha * \text{sgn}(S_2) \right] \quad (4)$$

Similarly, for control law, V_1 is derived in a similar way

$$V_1 = - \left(\frac{K_{42}k_{p1}(1-\gamma_1)}{2\sqrt{h_4}A} \right)^{-1} \left[\left(m_2 - \frac{K_{22}}{2\sqrt{h_2}} \right) x_4 - \left(\frac{K_{42}K_{44}}{2} \right) + \alpha * \text{sgn}(S_1) \right] \quad (5)$$

3.3 Terminal Sliding Mode Control

Terminality is introduced to bring a finite settling time in the sliding mode controller output response. This can be achieved by introducing v fractional power.

$$S_1 = x_2 + m_1 x_1^v; \quad \dot{S}_1 = \dot{x}_2 + m_1 v x_1^{(v-1)} \dot{x}_1$$

where

v Fractional power ranges from 0 to 1

$$V_2 = - \left(\frac{K_{31}k_{p2}(1-\gamma_2)}{2\sqrt{h_3}A} \right)^{-1} \left[\left(m_1 v x_1^{(v-1)} - \frac{K_{11}}{2\sqrt{h_1}} \right) x_2 - \left(\frac{K_{31}K_{33}}{2} \right) + \alpha * \text{sgn}(S_2) \right] \quad (6)$$

Similarly, for first tank, control law is derived

$$V_1 = - \left(\frac{K_{42}k_{p1}(1-\gamma_1)}{2\sqrt{h_4}A} \right)^{-1} \left[\left(m_2 v x_3^{(v-1)} - \frac{K_{22}}{2\sqrt{h_2}} \right) x_4 - \left(\frac{K_{42}K_{44}}{2} \right) + \alpha * \text{sgn}(S_1) \right] \quad (7)$$

Expression for finite-time convergence is obtained by substituting $S_1 = 0$ in the sliding surface equation, $t = \frac{|x_0|^{(1-v)}}{m_1^{(1-v)}}$ seconds.

3.4 Integral Terminal Sliding Mode Controller

Slight changes are made in the sliding manifold, where a power is added in the equation for bringing terminality concept.

$$S_1 = x_2 + 2m_1x_1^v + m_1^2 \int_{t_0}^{t_f} x_1^v dt; \quad \dot{S}_1 = \dot{x}_2 + 2m_1x_2vx_1^{(v-1)} + m_1^2x_1^v$$

where

- t_f Total time of running;
- t_0 Initial time of running the system

$$V_2 = -\left(\frac{K_{31}k_{p2}(1-\gamma_2)}{2\sqrt{h_3A}}\right)^{-1} \left[\left(2m_1vx_1^{(v-1)} - \frac{K_{11}}{2\sqrt{h_1}}\right)x_2 - \left(\frac{K_{31}K_{33}}{2}\right) + m_1^2x_1^v + \alpha * \text{sgn}(S_2) \right] \tag{8}$$

In similar way, control voltage V_1 is obtained,

$$V_1 = -\left(\frac{K_{42}k_{p1}(1-\gamma_1)}{2\sqrt{h_4A}}\right)^{-1} \left[\left(2m_2vx_3^{(v-1)} - \frac{K_{22}}{2\sqrt{h_2}}\right)x_4 - \left(\frac{K_{42}K_{44}}{2}\right) + m_2^2x_3^v + \alpha * \text{sgn}(S_1) \right] \tag{9}$$

$$t = \frac{2m_1x_1^{(v+1)}}{(v+1) - m_1^2x_1^{(v+1)}} \text{ (s)}$$

The above expression gives the finite time of convergence.

3.5 Multilevel Switching Control

To reduce the effect of chattering, $\text{sgn}(S)$ is replaced by exponential multilevel switching variable gain; thus, the undesirable chattering is reduced. From various computational studies, the system is classified into three different LS¹ with different error occurrence condition and operating region given below.

Different operating regions	Different error occurring regions	Different switching levels
LO—lower level OR ¹	ZR—around zero EC ¹	L—low LS ¹
MO—middle level OR ¹	MR—medium EC ¹	M—middle LS ¹
HO—higher level OR ¹	HR—high EC ¹	H—high LS ¹

Exponential multilevel switching variable gain is given by

$$g_0 = \frac{\text{(Variable switching level(upper--lower)limits)}}{\text{(full level switching (upper--lower)limits)}} \tag{10}$$

Table 2 Combinational logic

Set point	Error range	Switching level
LO	ZR	L
MO	ZR	L
HO	ZR	M
LO	MR	M
LO	HR	H
MO	MR	H

Then, final form of the smooth transition-based multilevel switching variable gain is incorporated by taking the inverse Laplace transform of $G_F(s)$. Selection of g_0 depends upon the combinational logic given in Table 2. The values of $L = 0.35$, $M = 0.22$, $H = 0.27$.

$$L^{-1}((G_F(s)) * |S|) = L^{-1} \left(\left(\frac{1}{\left(\frac{1}{g_0}\right)s + 1} \right) * |S| \right) = (g_0)e^{-(|S|*g_0)} \tag{11}$$

4 Simulation Results and Discussion

4.1 Servo Response

To test the ability of proposed ITSMC, multiple step changes in the process variables are given (initially 9.523 % of positive step change at 600 s and 44.828 % of negative step change at 1,200 s) for tank 1 and 2. Among all controllers, ITSMC gives better performance and the chattering reduced in the control input by the exponential multilevel switching variable gain. Figure 2 gives the output and control input response.

4.2 Regulatory Response

Figure 3 depicts the performance of the three SMC controllers for disturbance rejection. After reaching the steady state, sudden disturbance d_1 (drain 11.11 % of water in tank 1 and 2) is introduced at 250 s and d_2 (22.22 % increase in the flow rate of tank 1 and 2) at 500 s, respectively. From the response, ITSMC is superior to other SMCs.

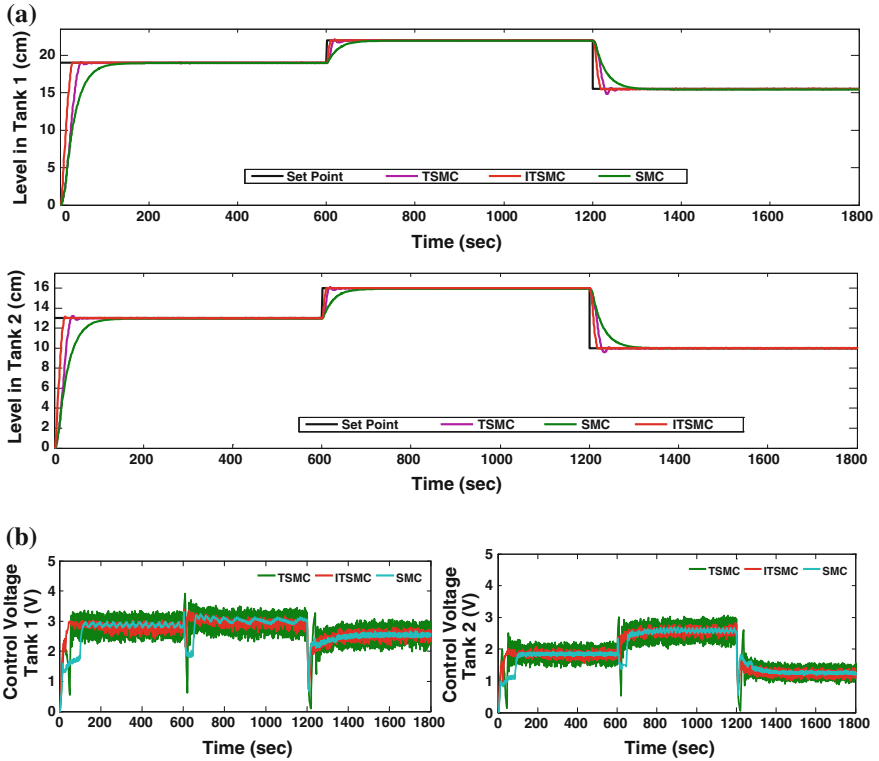


Fig. 2 a Servo response and b control input of tank 1 and 2

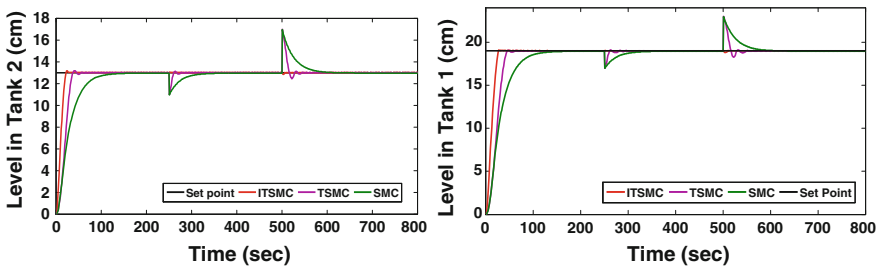


Fig. 3 Regulatory response for tank 1 and 2

4.3 Robustness Response

The robustness of the SMCs is tested by varying the valve coefficient β_1 and β_2 from its nominal value of 60 % and 70 % (0.6, 0.7). Disturbances are created at 600 and 1,200 s by adjusting the position of outlet valves to positive 20 % (0.4, 0.5) and

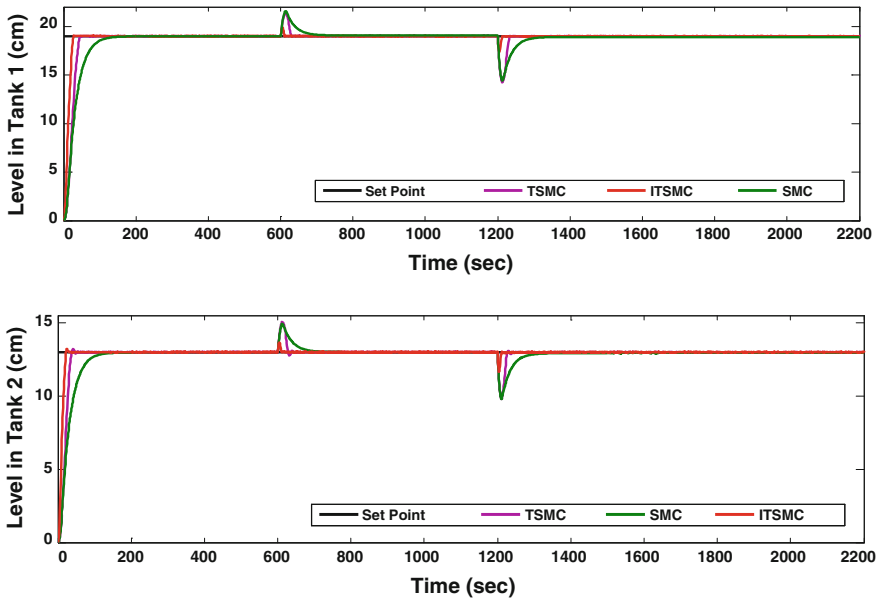


Fig. 4 Robustness of level in tank 1 and 2 by adjusting β_1, β_2 position to +20 % and -20 % change

in negative 20 % (0.8, 0.9), respectively, from its nominal position. From the simulated results, the proposed ITSMC brings the system back to operating conditions. Figure 4 illustrates the robustness performance of ITSMC in maintaining the tank levels 1 and 2.

Table 3 Comparison of performance indices (where E error indices; Q quality indices)

Performance indices			Controllers		
			<i>SSMC</i>	<i>TSMC</i>	<i>ITSMC</i>
E	<i>OS (%)</i>	<i>h₁ loop</i>	–	–	–
		<i>h₂ loop</i>	–	–	–
	<i>T_s (s)</i>	<i>h₁ loop</i>	243	57	36
		<i>h₂ loop</i>	221	64	31
Q	<i>IAE</i>	<i>h₁ loop</i>	946	593.7	316.6
		<i>h₂ loop</i>	678.8	399.4	210.6
	<i>ISE</i>	<i>h₁ loop</i>	7,843	6,634	3,310
		<i>h₂ loop</i>	3,719	2,943	1,395
	<i>IAE</i>	<i>h₁ loop</i>	781.3	492.8	236.7
		<i>h₂ loop</i>	546.8	303.9	138.7
	<i>ISE</i>	<i>h₁ loop</i>	7,280	6,168	3,013
		<i>h₂ loop</i>	3,260	2,532	1,158

4.4 Comparison of Performance Indices

Table 3 shows the effectiveness of proposed ITSMC is compared in terms settling time and peak overshoot and error indices such as ISE, IAE, and ITAE for h_1 loop and h_2 loop for set point tracking (servo) and disturbance rejection (regulatory).

5 Conclusion

In this paper, standard SMC, TSMC, and ITSMC controllers have been presented for nonlinear minimum phase QTP. From the simulation results, proposed scheme gives superior performance in terms of response with minimum error values, better tracking, robustness, and fast disturbance rejection capabilities with reduced chattering effect.

References

1. K.H. Johansson, The quadruple—tank process: a multivariable laboratory process with an adjustable zero. *IEEE Trans. Control Syst. Technol.* **8**, 456–465 (2000)
2. B.W. Bequette, *Process Control Modeling, Design and Simulation* (Prentice Hall of India, India, 2004)
3. P.P. Biswas, R. Srivastava, S. Ray, A.N. Samanta, Sliding mode control of quadruple tank process. *Mechatronics* **19**, 548–561 (2009)
4. E.M.M. Farrokhi, Sliding-mode state-feedback control of non-minimum phase quadruple tank system using fuzzy logic, in *18th IFAC World Congress* (2011)
5. J.J. Slotine, W. Li, *Applied Nonlinear Control* (Prentice Hall, New Jersey, 1991)
6. N.B. Almutairi, M. Zribi, Sliding mode control of coupled tanks. *Mechatronics* **16**, 427–441 (2006)
7. C.-K. Lin, Non-singular terminal sliding mode control of robot manipulators using fuzzy wavelet networks fuzzy systems. *IEEE* **14**, 849–859 (2006)
8. M. Zhihong, X.H. Yu, Terminal sliding mode control of MIMO linear systems. *IEEE* **44**, 1065–1070 (1997)
9. S. Ding, C. Zhang, X. Li, Terminal sliding mode control of second-order systems with bounded input, in *Control and Decision Conference (CCDC)* (2013), pp. 265–269
10. F.-J. Chang, E.-C. Chang, T.-J. Liang, Digital—signal-processor-based DC/AC inverter with integral compensation terminal sliding mode. *Control Power Electron. IET* **4**, 159–167 (2011)
11. M. Chen, Q.-X. Wu, R.-X. Cui, Terminal sliding mode control for a class of SISO uncertain nonlinear systems. *ISA Trans.* **520**, 186–206 (2013)
12. S.-Y. Chen, F.-J. Lin, Robust non-singular terminal sliding mode control for non-linear magnetic bearing systems. *IEEE* **19**, 636–643 (2011)
13. Y. Wu, X. Yu, Z. Man, Terminal sliding mode control design for uncertain dynamic systems. *Syst. Control* **34**, 281–287 (1998)

14. C.-S. Chiu, Derivative and integral terminal sliding mode control for a class of MIMO nonlinear systems. *Automatica* **48**, 316–326 (2012)
15. J. Jayaprakash, D. Davidson, P.S.H. Jose, Comparison of controller performance for MIMO process. *Int. J. Emerg. Technol. Adv. Eng.* **3**, 51–59 (2013)
16. N. Kitdormat, S. Khoo, L. Xie, Integral terminal sliding mode control approach for multi-robot formation. *Control Inf.* 98–103 (2009)

An Enhanced Security Framework for a Cloud Application

B. Balamurugan and P. Venkata Krishna

Abstract Cloud computing has gained momentum over the past few years where numerous business domains have shown a radical shift from their traditional software. Cloud apps can run weeks with minimum cost and user maintenance, which overcomes the drawback of a traditional software approach. However, cloud security is still in its infancy stage; hence, there are a numerous security issues possible. The purpose of this paper is to devise a mechanism to store and retrieve data securely from cloud, overpowering as many threats as possible. The paper starts with a general introduction about cloud security and its advantages, which leads way to the major security challenges known about cloud and a framework to overcome it. We specifically concentrate on secure access of data over cloud, sharing data between multiple cloud users and the cloud data center. A simulator “*Cloudsim*” has been used.

Keywords Access control · Cloud computing · Security · Cloudsim · Encryption/decryption · Digital signature

1 Introduction

The cloud computing has shown tremendous growth potential being in its early years. Cloud model is adapted for achieving economies of scale; it has enhanced information sharing speed and has helped in developing new services. In 2012–2013, the cloud industry is expected to generate £5.79 billion in revenue [1]. This will be a 6.5% increase than the previous year, and this evolution of cloud will lead to exponential increase in the storage of data than previous years [2]. But there

B. Balamurugan (✉) · P. Venkata Krishna
VIT University, Vellore, India
e-mail: balamuruganb@vit.ac.in

P. Venkata Krishna
e-mail: pvenkatakrishna@vit.ac.in

are two major factors such as the concern over security and the limited capital expenditure that are expected to limit growth over the coming years [32]. Management of private data by a trust of the third party is also a major security concern [1]. Cloud computing is a collection or cluster resources, which include hardware and software that are provided as service over the Internet. Resources that are provided on cloud must be “on-demand,” and “at scale,” in a multitasking environment [3, 4]. “On-demand” means that cloud services must be available based on the customers requirement. Cloud gives us an illusion of huge amount of resource being available at all time, which implies that services are “at scale” [3, 5]. Cloud computing adds capabilities to the small-, medium-sized, and large-scale businesses by providing them with a scalable, low-cost, and flexible services. Some of the cloud service providers and the types of service they provide are stated in Table 1.

Breaches in cloud had affected major cloud provider including Apple Inc. The issue of data security in the cloud has come under increased scrutiny after several cases of data wipe from a number of interconnected devices [6]. Security breaching incidents in Apple’s cloud environment *icloud* had forced Apple Inc to temporarily disable the ability of *icloud* account holders to change any details over the phone, and customers who need to modify any details will need to use the company’s *iforgot* service [40]. The high-profile breach in cloud security has forced Apple Inc to look at its security protocols. The case clearly states the outburst of the cloud had made the cloud a very attractive and negligible domain; on the contrary, there are still several avenues of security in cloud, where tremendous improvement is in need [41]. The paper suggests security enhancement [43] in terms of authorization, authentication and integrity, and the ways of achieving it through encryption/decryption, access control, and digital signature.

Table 1 Cloud service providers and their provided service

Service provider	Service	Types of service
Amazon	Elastic compute cloud (EC2) [20]	Iaas
	Amazon simple storage (s3) [21]	Iaas
	Amazon Elastic Beanstalk [22]	Paas
Oracle	Sun Cloud [23]	Iaas
	Force.com [24]	Saas
	Force.com platform [25]	Paas
Microsoft	Windows Azure [26]	Iaas
IBM	Blue cloud [27]	Iaas
VMware	CloudFoundary.com [29]	Paas
	CloudFoundary.org [29]	Saas
Google	Google App Engine [28]	Paas

2 Related Works

Cloud security is a top priority concern for cloud users [7]. Certain issues are responsible for this growing concern such as system downtime or business interruption, exposure or loss of data during file transfers to the cloud, concerns over encryption of data, physical security of cloud service provider data center, shared technology vulnerabilities (e.g., multitenant environment), malicious activity from insiders or privileged administrators at cloud providers, and identifying and authenticating users. Various methods of encryption have been provided for the security of files on cloud; the choice of encryption methodology is provided with the user, and also data segregation and privacy are provided [32]. Work has also been done on improving the ISO 27001:2005 standard, and results show that this certificate does not provide enough security either to cloud service provider or to cloud customers [33]. A method has also been proposed using .NET Framework to improve the security of a cloud platform [34]. Research has been done for retention and detection of HTTP-DoS [8] and XML-DoS attacks [9] using cloud trace back methods and cloud protector, respectively [35]. In addition to this method, a hidden Markov model based on clustering which uses data mining techniques is used as an intrusion detection system [36]. To provide security to data storage on cloud, Sobol sequence is used to maintain integrity of the data [37], and this method is claimed to be better than pseudo random method used by homo-morphic verification scheme to assure the availability, reliability, and integrity of data on a distributed system [38]. Filho has also proposed a method which uses hash function to stop malpractice in a P2P system, but the problem arises when the size of the file is too large which makes the system un-usable [39]. Security at different levels such as network level, host level, and application level is important to keep the cloud secure and provide efficient cloud service. The survey results in finding several attacks and security threats such as SQL injection attack, cross-site scripting attack, man-in-the-middle attacks, DNA attacks, CAPTCHA attacks, cookie poisoning, and denial of service [10]. These attacks have been done with major cloud providers.

3 A Cloud Application: Dropbox

3.1 Overview

In order to examine and testify the cloud application security, a prominently used cloud application Dropbox [11] is considered as a case study. Dropbox is a famous file sharing and storage tool on cloud said to be vulnerable to many attacks [42]. Dropbox has 25 million users currently, spread over 175 countries; it has 200 million files saved on cloud [12]. Dropbox offers 2 GB of free space and upto 10 GB on subscription. Files are available in both off-line and online mode. It works with all platforms such as Windows, Linux, iPhone, android, and Blackberry. Dropbox

stores all the vital information about a user in a config.db file; this file is stored on client system. Taking this file and putting it into another system starts Dropbox immediately and joins the system to a synchronized group. One way of protecting the authentication file is to set more restrictive permission on the folder where the authentication file is stored. Another way adopted by Dropbox is encrypting the config.db file, which would make the file more secure [12]. All files are synced, encrypted, and stored securely on Amazon's (S3) Simple Storage Service over several data center [11]. Dropbox uses AES256 to encrypt all the files stored on Dropbox. In addition to this, the files are also protected by an individual's passwords, thus providing two levels of security [12] Dropbox has privacy policy and technical control, which prohibits its employee from viewing any user file [13]. The only information available to the employee is in the form of a metadata file, i.e., name and location of a file [12].

3.2 Attacks on Dropbox

There is an extensive use of passwords in the online services provided by companies such as Amazon [14], Google, *icloud*, etc. An attacker can capture the password from one such site and reuse it on another site to gather a person's information [15]. One such problem came to light when an attacker was able to log into a Dropbox employee's account by reusing the employee's password from another compromised site and had access to the encrypted customer information. The attacker used the email-id of these users and flooded them with spam mails [15–17]. The enterprises should have a watch on the file sharing activities of their employee because the more we transfer onto the cloud, the less control we have on our data [15]. Thus, a monitored use of Dropbox and related application of cloud are necessary. From a recent study, it is known that the users do not know what their service providers are doing to keep their data safe on cloud [18]. One must verify the security features provided by any cloud application based on the level of security one needs. As a solution to the Dropbox password breach [15], the company has reset the password of all the users whose credentials were uploaded by the hackers, similar to what LinkedIn did when their user accounts were hacked. But according to security experts, these measures are not secure enough. Dropbox now is using a two factor authentication, which detects odd user behavior and audits logs of user account.

Dropbox recommends a third party encryption for its advanced users who are capable of managing their own encryption key. The most famous third party encryption is provided by TrueCrypt. Other provides include EncFS, SecretSync, and BoxCrypto [12].

3.3 Mitigation Technique

Dropbox has added this feature in its beta version for some users. The two level of authentication is for providing more security at login. The security code can either be given via a text message or by using any Time-based One-Time Password (TOTP) [13]. This feature is displayed on the security tab of the account page. User have also criticized the new method which would fail in a scenario where the cell phone to which secure code is sent gets lost, which would make it impossible for user to login.

Besides security Dropbox must also concentrate on the persistence of the data that is stored. Consider a situation where a user has not accessed his account for a very long time. Dropbox has no way to keep the user posted about the status of the file and last access time. We propose a method in which Dropbox sends periodic mails to the users, intimating them about the status of their account [13].

4 Introduction to *Cloudsim*

Cloudsim is a simulation tool that was developed as a built in project, to enable researchers, industrialists to explore specific design issues in cloud without concerning themselves with the low lying infrastructure. *Cloudsim* allows users to have seamless simulation and experimentation on any cloud infrastructure or services. In each of our proposed method, we have used *Cloudsim* as the simulator [19].

4.1 Framework

As a first step, every user creates a cloudlet through which the request is sent. The cloudlet is configured by assigning the required parameter. The second step is to initialize the virtual machine by supplying in unique ID, owner ID, the MIPS, the CPUs, RAM, bandwidth, amount of storage, virtual machine monitor, cloudlet-Scheduler policy for cloudlets. All the cloudlets are added to a list, and similarly all the Vm's are added to another list. The cloudlet and the Vm's are bounded together by a broker who maintains a list of cloudlets and their corresponding Vm's. The Vm's list is submitted to the Host via VmProvisioner, CPU allocation, memory allocation, storage allocation, bandwidth allocation units, which take care of allocating the Vm's to the host based on the resource available with the host. The host on behalf of the data center receives the request and forward it to the data center broker, which stores the data and retrieves data from the data center. All responses from the data center are sent through the host to the Vm's and cloudlet. Here, the user, host and the data center are singular processes and Cloudlet and Vm's are threads as shown in Fig. 1.

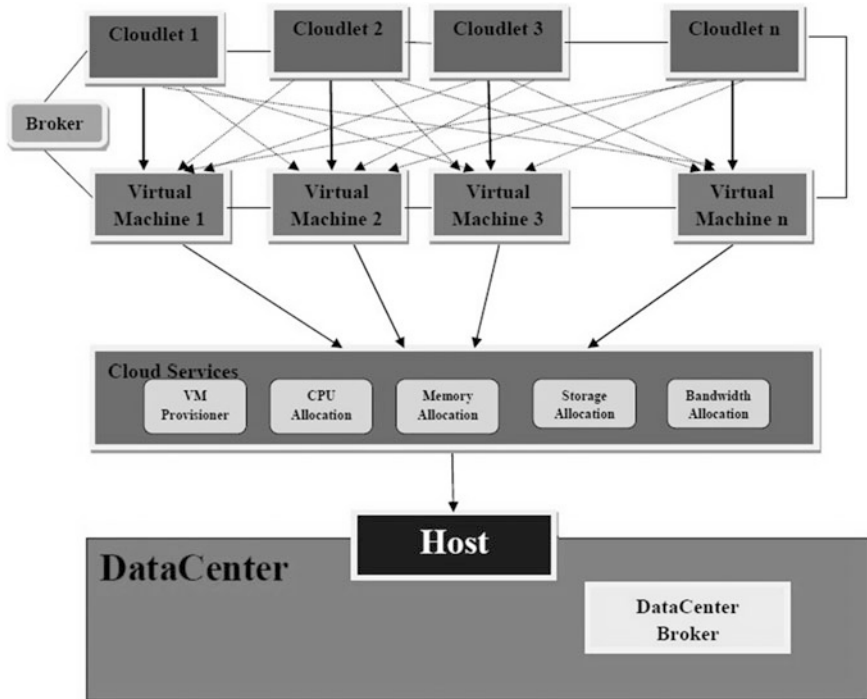


Fig. 1 Framework for implementation of a cloud application on *Cloudsim*

5 Proposed Method

In order to increase the security of a file stored in cloud, we propose three different methods: access control, encryption and decryption, and digital signature. In the *Cloudsim* framework, when a file needs to be added into the data center, it is initially added to cloudlet via the cloudlet constructor, the cloudlet along with virtual machine are bound to broker, this broker in turn contacts the host and creates an instance in host. A data center in *Cloudsim* consists of multiple hosts who provide cloud services; hence, a file added to host is automatically part of the data center. In each of the proposed methods, the secure file is stored by the above steps to the data center.

5.1 Access Control

A common challenge faced by cloud developers is controlling access to the resource provided across cloud. Access should be provided to clients who are authorized, at any time as they demand, from any location while maintaining the basic security concerns over integrity, confidentiality, and availability. In a cloud environment, this

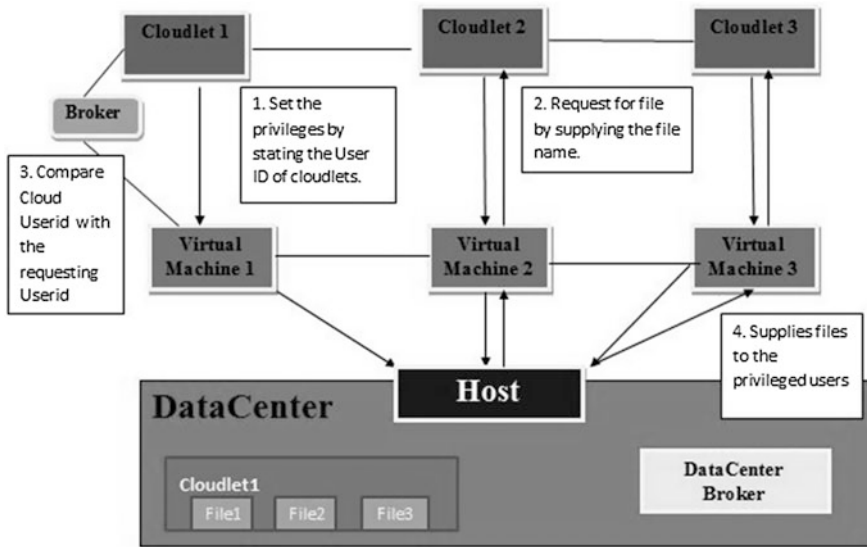


Fig. 2 Access control in cloud

becomes a major challenge since resources are spread over an uncontrolled environment, where each data can mingle with data owned by another client. A mitigation technique for this problem is allowing access only to plaintext file.

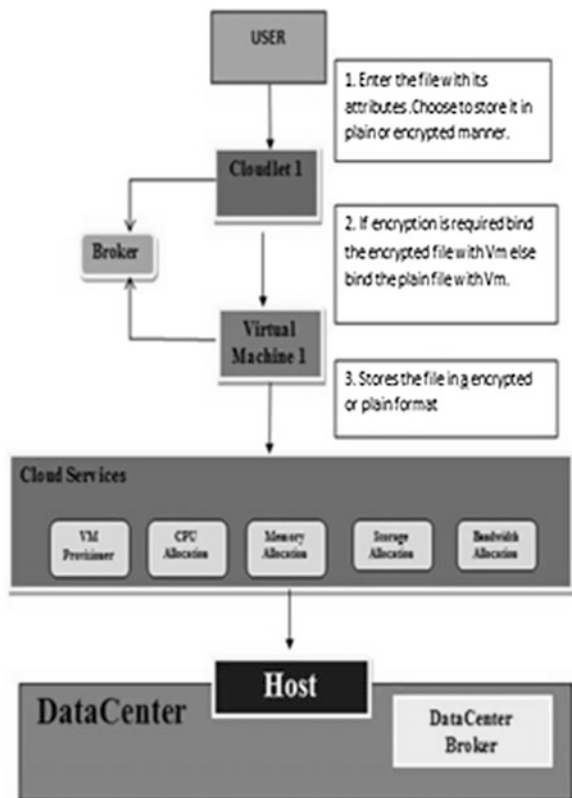
On creating a cloudlet, a list of user ids of privileged cloudlets are provided, to indicate which cloudlets can share data. When a cloudlet wants to access a file stored by another cloudlet, the broker checks the user id's in the list. If a suitable match is found, i.e., if the cloudlet is privileged, it is given access. In addition, a cloudlet that stores encrypted files does not want them to be accessed by any other cloudlet. Applying this in access control scenario, we check whether the file desired by any cloudlet is stored in encrypted format. In case the file is encrypted, no access is provided to it. Only a plaintext file is allowed to be accessed across multiple cloudlets. Hence, we control access of a file over cloud, by restricting access only to cloudlets which are privileged and also to a plaintext file and preventing access to encrypted file. *Cloudsim* method used for this purpose is `cloudletObject.getRequiredFiles()`, which returns a list of files stored in the cloudlet Fig. 2 depicts pictorially access control in *cloudsim*. The execution time is 6,300 ms.

5.2 Encryption and Decryption

Encryption is a process or technique by which we transform the plaintext message to an encoded message, which is also called ciphertext, so that an illegitimate user cannot read it. Decryption, on the other hand, is just the reverse of encryption where the ciphertext is converted back to plaintext message [30]. There are a numerous

algorithms that are available for encryption such as RSA (public key encryption), DES (private key encryption), and AES (private key encryption). We are using Advanced Encryption Standard (AES), which is a symmetric key encryption technique. In AES, a single key (private) is used both to encrypt as well as to decrypt. This key is shared between communicating parties over a secure channel so that one can easily decrypt the file encrypted by another. It is a 128-bit block cipher technique, which is more advanced than its counterpart DES. Comparing with RSA, AES is considerably better since it takes less execution time [31]. Due to these advantages, AES is used frequently and has become a globally accepted standard of encryption. Depending on the user's preference, a file can be encrypted using AES and then stored into the data center. Similarly, the file can be decrypted by the cloudlet. Here, predefined methods of Java have been used for encryption and decryption, while creating the cloudlet, the encrypted or plaintext file is added as a parameter in the cloudlet constructor. The cloudlet and virtual machine are together grouped in a broker. In case an external file needs to be added, cloudlet-object.addRequiredFile (File filename) can be used. Similarly, the encrypted file will be decrypted before presenting to the cloudlet. The execution time is 8,305 ms. Figure 3 depicts encryption and decryption pictorially over cloud.

Fig. 3 Encryption and decryption



5.3 Hash Code

Hash code is a one-way security technique in which we can create a hash code for a string using the hash function (e.g., md5) but cannot get back the original string from a given hash code, using the same hash function or any other means. The hash code for a given string is unique when passed through a given hash function. Therefore, comparison of two hash codes can be used as a technique for validation. If the hash codes for the string at both ends match, then it is considered as genuine, else rejected. Here, we are applying the same concept with cloudlet id's. A hash code of the cloudlet id is sent to the host. The data center is contacted to get the host list using `DataCenterObject.getHostList()`, from this list the list of VM object using `HostObject.getVmList()`. Cloudlet id's are retrieved from the VM list by contacting the broker. The hash code of the retrieved cloudlet id's is compared to match the hash code received, from the requesting cloudlet. If a match is found, then it can be assumed that cloudlet is genuine otherwise we regret the cloudlet sending request. For generating a hash code of each cloudlet, we use the function `Object.hashCode()`, which returns a hash code. The execution time is 7,308 ms. Figure 4 depicts hash code in cloud framework.

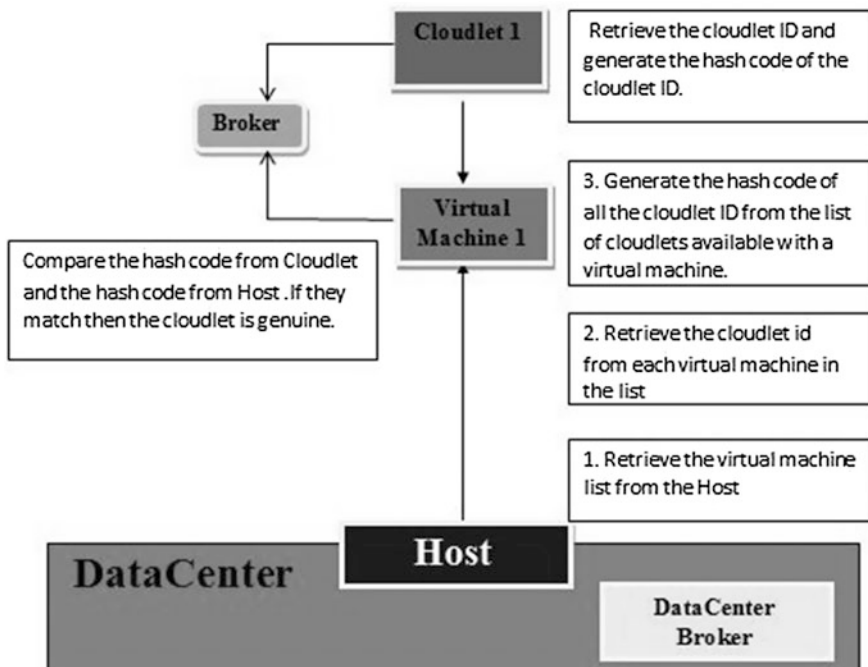


Fig. 4 Message digest in cloud

5.4 Analysis of the Efficiency of a Cloud Application

Figure 5 depicts a graph which shows that when we increase the number of cloudlets, keeping the VM constant at 1, the time taken for creating each cloudlet and binding with an available broker increases exponentially. In case we want to decrease the time, we can associate every cloudlet with an individual VM, thereby reducing its load and computation time. The overall execution time of a cloud application increases with the increase in number of tasks.

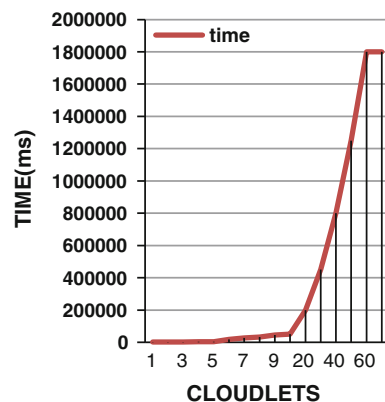
In regard to the comparative study of our framework with Dropbox cloud application, we can conclude on certain issues which have shown that the proposed method has significantly covered many security issues at an initial level.

Dropbox mainly has two folders personal and shared; personal folder holds data that the user does not wish to share with others. Whereas in shared folder, owner can add users to access the files and information stored in that folder. Owner adds other users by sharing a link of the shared folder. Both personal and shared folders are stored on the cloud based on certain subscription plan, i.e., as free user or premium user, based on the subscription Dropbox provides security on files. Dropbox provides encryption of data to ensure security for premium account holders.

In our proposed method, user decides whether a file needs to be encrypted or not before putting it on cloud; the encrypted data therefore cannot be accessed by illegitimate user. On creation of cloudlet it can also store a list of user id whose hold the privilege to access the encrypted files. In Dropbox, access is provided when user gives the request to the owner for the file who assesses the user and decides whether access must be granted or not.

The proposed method provides authentication at infrastructure level as compared to Dropbox that provides authentication at basic application level.

Fig. 5 Efficiency of the cloud



6 Conclusion

Cloud computing has brought a technical revolution today. It is now in a position to dominate the IT market. Companies such as IBM, Microsoft, Google, and Amazon have major shares on cloud and have successfully transformed from traditional grid computing to a diverse cloud. One key factor that is a threat to this transformation is security aspects of cloud. Many companies are concerned that cloud is not yet a completely safe framework for storing proprietary information and provide access to most of the services in a feasible manner. Hence, we must focus on solidifying the security aspects of cloud in a broader dimension. The proposed methods project improving security in three concrete dimensions—access control, where a file stored by one can be read by others depending on the permission allowed, encryption and decryption, where the file being stored is encrypted first using AES algorithm to maintain its confidentiality, message digest, this primarily ensures that a cloudlet trying to access a data center host is genuine by comparing the hash code's of the virtual machine's id and the id stored by the host. All the methods significantly improve security over a cloud and in turn enhance user experience.

References

1. M.N. Islam, M. Mia, M. Chowdhury, M.A. Matin, Effect of security increment to symmetric data encryption through AES methodology, in *Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, SNPD'08*, (2008), pp. 291–294
2. B.P. Rimal, E. Choi, I. Lumb, A taxonomy and survey of cloud computing systems, in *Fifth International Joint Conference on INC, IMS and IDC 2009, NCM'09*, (2009), pp. 44–51
3. T. Olavsrud, D. Muse, *How secure is the cloud?*, *IT Pros Speak Up* (2009)
4. A. Kesarwani, C. Gupta, M.M. Tripathi, V. Gupta, R. Gupta, V.K. Chaurasiya, in Implementation of Chinese wall model in cloud computing for enhanced security. *International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, vol 22(24), (2011), pp. 411–413
5. D.L. Gazzoni Filho, P.S.L.M. Barreto. Demonstrating data possession and uncheatable data transfer. *IACR Cryptology ePrint Arch.* (2006), p. 150
6. Y. Fu, S. Luo, J. Shu. Survey of secure cloud storage system and key technologies. *Jisuanji Yanjiu Yu Fazhan/Comput. Res. Dev.* **50**(1), pp. 136–145 (2013)
7. R. Bhadauria, R. Chaki, N. Chaki, S. Sanyal. A survey on security issues in cloud computing. (2011)
8. Y. Chen, V. Paxson, R.H. Katz. What's new about cloud computing security? (2010)
9. M. Kirkpatrick, IBM Unveils Blue Cloud—What Data Would You Like to Crunch? (2007)
10. M. Jensen, J. Schwenk, N. Gruschka, L.L. Iacono, On Technical Security Issues in Cloud Computing, *IEEE CLOUD*. pp. 109–116, 2009
11. A. Chonka, Y. Xiang, W. Zhou, A. Bonti, Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks. *J. Network. Comput. Appl* **34**(4), 1097–1107 (2011)
12. M. Almorsy, J. Grundy, A.S. Ibrahim, Collaboration-based cloud computing security management framework. *Cloud computing (CLOUD)*, *IEEE International Conference.* (2011), pp. 364–371

13. P. Kumar, N. Nitin, V. Sehgal, K. Shah, S.S.P. Shukla, D.S. Chauhan. A novel approach for security in cloud computing using hidden markov model and clustering. *Inf. Commun. Technol (WICT)*. pp. 810–815 (2011)
14. R.N. Calheiros, R. Ranjan, A Beloglazov, C.A. De Rose, R. Buyya. Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Softw: Pract. Exper.* **41**(1), pp. 23–50 (2011)
15. S. Ristov, M. Gusev, M. Kostoska, A new methodology for security evaluation in cloud computing, MIPRO, in *Proceedings of the 35th International Convention*. pp. 1484–1489, 2012
16. A. Chonka, Y. Xiang, W. Zhou, and A. Bonti. Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks. *J. Netw. Comput. Appl.* **34**(4), pp. 1097–1107, 2011
17. M. Firdhous, O. Ghazali, S. Hassan, A memoryless trust computing mechanism for cloud computing. *NDT* **1**, 174–185 (2012)
18. P. Syam Kumar, R. Subramanian, D. ThamizhSelvam, Ensuring data storage security in cloud computing using Sobol Sequence. *Parallel distributed and Grid Computing (PDGC)*, 1st International Conference. pp. 217–222, 2010
19. C. Wang, Q. Wang, K. Ren, W. Lou. Ensuring data storage security in cloud computing. *IACR Cryptology ePrint Archive*. (2009)
20. M. Mulazzani, S. Schrittwieser, M. Leithner, M. Huber, E. Weippl. Dark clouds on the horizon: using cloud storage as attack vector and online slack space. *USENIX Security Symposium*. (2011)
21. P.V. Krishna, S. Misra, D. Joshi, A. Gupta, M.S. Obaidat, Secure Socket Layer (SSL) Certificate verification using learning automata. *Secur. Commun. Netw.* (2013)

Enhanced Role-Based Access Control for Cloud Security

B. Balamurugan and P. Venkata Krishna

Abstract There has been significant work done on access control previously. The proposed access control system is aiming at providing more security to enterprise resource on cloud by limiting the access of resources. Each enterprise has a different hierarchical structure based on their organizational policy. This organizational hierarchy is built using various roles that are interdependent on each other. Therefore, instead of defining the access policy for each user, the system defines access for each role. There are many standards which are available for access control such as XACML and SAML. The proposed system uses the concept of XACML for designing the policies. In addition to this feature, the system has a role request module that enables the user to hold other roles as well. This project provides a security solution via RBAC system for cloud security (RBACCs) to improve the security of data on cloud. This paper shows how the existing security standards such as XACML can be used to create and manage the policies of RBACCs. These policies are defined and designed according to current business scenario. The paper also studies the existing work done on cloud security and gives a comparative study between the proposed and the existing systems. In the end, we also proposed a few extensions to the existing system that can be made based on the security needs.

Keywords Access control · Cloud security · Cyber security · XACML · Access policies

B. Balamurugan (✉) · P. Venkata Krishna
VIT University, Vellore, India
e-mail: balamuruganb@vit.ac.in

P. Venkata Krishna
e-mail: pvenkatakrishna@vit.ac.in

1 Introduction

Cloud computing is an emerging trend in the IT industry and many enterprises are adopting this technology. The key features of cloud computing such as multitenancy, scalability, and elasticity, Pay as you need has lead to its adoption by enterprises of all scales. Cloud computing is increasing fame because it provides IT resources on demand, which means if an enterprise get a new business requirement, it can get the resources just by paying for what it wants immediately on demand. Interest in cloud computing is growing because it provides high computing resources in no time at a genuine cost. The users of cloud services can buy what they need. Cloud computing is identified as a scalable IT enabled capability provided to users as services using the Internet technologies [1]. Cloud computing has gained momentum in recent times as big organization is seeing cloud and cloud-related technologies as a profitable domain to invest in. IT giants are spending a lot behind the study and research of cloud computing, to explore cloud computing. In the coming years, the spending by both IT organizations as well as non IT organization would go up by large amount. In fact, cloud computing is expected to have a compounded growth in the years to come [1]. Although we have many benefits of using cloud technologies, these are important area which one must consider. These are security and privacy aspects of cloud. As we know that cloud services are provided over the Internet, thus posing a threat to the data being stored on cloud. To keep the resources safe on cloud is a big challenge faced by cloud service provider [1]. Because the data on cloud are also on Internet, therefore, they are prone to all sorts of threat. We mainly concentrate on security, dealing with how safe the data are stored, and privacy, dealing with providing privacy to the user [1, 17]. Cloud service providers are collectively working toward providing more privacy in all its models. Organizations today face numerous different requirement which attempts to provide adequate protection to such information [1]. Issues such as connectivity, reliability, and interoperability also are focused.

Security in cloud computing has become a major concern in the IT world recently. Cloud computing is vulnerable to many types to attack that we see on the Internet. Few of the cloud vulnerabilities are listed below:

1. Data breaches on cloud had created an environment of distrust among the cloud users who put important data on cloud.
2. The second greatest threat in cloud computing is the loss of data by theft. As cloud is available on the Internet, it is vulnerable to hacking and other cyber crimes. If the data on cloud goes into wrong hands, it will create huge loss for the cloud users.
3. If the bandwidth of the network is not good, a cloud user may not be able to realize all the features that a cloud service provider wants to give him [2].
4. If an attacker gains access to a cloud user credentials, he or she can eavesdrop on the transaction and activities, manipulate data, return false information, and redirect clients to illegitimate sites.

5. With the increase in traffic on cloud, the services become unavailable to a large number of users. This creates a denial of service attack where the users are unable to use the cloud services.

2 Literature Review

Role-based access control system was initially proposed by Sadhu [3] in which permissions are associated with roles, and users are made member of appropriate role. Roles are closely related to the concept of user groups in access control. The basic concept of role-based access control (RBAC) originated when security demands increased. In the flat RBAC model [3] proposed by Sandhu et al., the basic RBAC has been described. The main feature of this model is many roles can be assigned on permission to many users. This model also talks about the session and role mapping where the role is valid for particular sessions. Flat RBAC gets its features from traditional group-based access control. This model expects each role to be assigned at least one permission and each user to be assigned at least one role and does not require any administrative permission. Another model that we are going to see is role hierarchy model [4] proposed by Sandhu et al. which is a slight variation of the model proposed in [3]. The model [4] proposes that a user can have one or more roles in a particular session based on the permission provided to him. These models propose the initial access control and can be used to secure data. The ARBAC model [5] proposed by Mon and Naing aims at providing privacy to private clouds and defines permission based on just three types of attributes where each subject is associated with attributes which define the identity and characteristics of the subjects. These attributes can be subject's name, organization, etc. Second attribute proposed in [5] is the resource attributes which is defined as an entity that is acted upon by a subject. Resources have attributes that can be used to make access control decisions. And the last is environment attributes which describe the operational, technical, and even situational environment or context in which the information access occurs. This model states that data access is granted only if there exists an authorization policy which matches the above-specified attributes. In addition to that, this model also checks the security level of the user to match the security level of the resource that he wants to access. Mon and Naing [5] proposes an algorithm and checks if a policy is available for the user in policy set, if yes then check the constraints of user policy to match the user credentials. coRBAC model [6] is an optimized RBAC model for cloud computing environment. This model uses dRBAC model. dRBAC uses PKI and domain information in the certificate to identify the different enterprises and organizations. The enterprise or organization builds an internal RBAC system and uses its own certificate to sign and issue certificates for their internal users and users of other companies for authentication and permission assignments [7]. This model tries to integrate RBAC along with key encryption techniques to provide security to enterprise data. Next is

refined RBAC model [8] which extends the features of a simple RBAC model so that it can be used at cloud layer. Cloud computing has security issues; therefore, a RBAC model can be used to provide security at cloud level. Model proposed in [8] defines cloud resources as protected objects and maintains session and permission for the access of cloud resources. Users and roles both are mapped to session and can be altered during a session. This model mainly differs from the other models because the cloud resources are treated as protected objects where they can be accessed during a session. A RBAC model for high-performance computing system proposed by Pereira [9] suggests the use of RBAC system along with XACML [10] protocol. In XACML, the XML files are designed for access control. Policies are created against which the requested are checked. If the constraints in policy match the constraints in the request given by the user, the user is granted permission. This model [9] does not maintain any session, but the access period can be defined in the XACML policies. There are other types of access control also available such as the discretionary access control (DAC) [11] and mandatory access control (MAC). These two techniques can be applied for individual user. RBAC is a modified form of DAC and MAC. ORBAC [12] model proposes various ontology for access control keeping the basic definition same. This model is not concerned with cloud resources.

3 Challenges

Access control is an important information security mechanism [13], according to user identity and the attribution of a predefined group of users to restrict access to certain information items and limit the use of certain function. Only a user who is privileged to access a resource can view and operate on the resource. Security data are a big challenge that enterprises are facing. With more enterprises adopting cloud technologies, the risk to data has increased in folds [17]. Enterprises are making huge investment to keep their data safe on cloud. There are various solutions that have been proposed, and still, a lot of research is being done to improve the security on cloud [18]. RBAC system for cloud security (RBACcs) aims at providing more security and control over the resources on cloud.

Access control is a very old technique to provide security in enterprise. To apply this access control technique [14] to provide security of cloud resources is a challenging task. Cloud computing when adopted by enterprise creates a threat to the potentially sensitive data on cloud. The proposed RBAC technique tries to provide access control to enterprise data on cloud. This is role-based because access privileges are defined based on the role. This system decides whether to permit or deny the access of resources to a user.

RBAC system [15] aims at providing additional security to enterprises that are using cloud technology for their business requirement. Data security is a very essential aspect of cloud security as everything is on Internet. All the security threats faced on Internet usage are also faced when we adopt cloud technology.

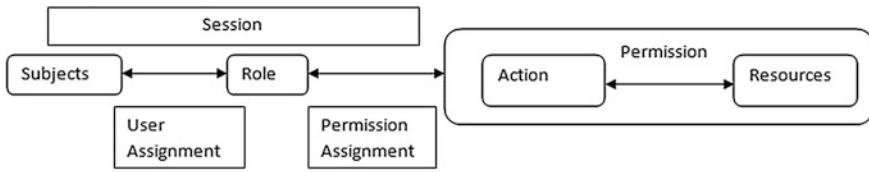


Fig. 1 Basic RBAC model

In the recent year, cloud computing has gained popularity because of its features discussed in the section above.

The proposed system can be used by any enterprise for providing additional security to enterprise data. The high-level architecture of the basic RBAC model is illustrated in the Fig. 1. The proposed RBAC system plans to limit the access of resource on cloud. This is achieved by categorizing the user based on their role in an organization. For every role, job functionality is defined, according to which a user is allowed to access the resources. This system has three main components: “default access module,” “user request module,” and “role inheritance module.” These modules are discussed in detail in the following section. The user is divided into two categories: the general user and an administrator who monitors the entire user and their actions.

The access to cloud resources can be achieved in a controlled manner on using the proposed system. The system also provides a solution to upload large data as blob files in containers on cloud.

4 Statement of Assumption

In the proposed system, we are considering a scenario where we are using a RBAC system in an enterprise. Here, we are considering our enterprise to be an IT organization where role-based work is assigned to employees. An organization may use cloud resources for its various resource requirement. This is where access control mechanism control comes into play. An organization may have various resources which should be access in a controlled manner. For this, we have proposed a role-based approach to access control where instead of defining access limit for a vast number of employees, we define access for a set of defined roles of an organization. Each role has a set of defined access privileges which enable them to access cloud resources. In the proposed mechanism, we also consider special scenario where under special situation, one may have to inherit another defined role for a certain period of time. In such situations, an administrator will be governing the role inheritance process. Another scenario is where the user wants to explicitly access a set of resources with his current role in such a situation the user may have to put a request. This request is checked with the parameters supplied in the request, if the request parameters match any of the available policy, the users’ request is

granted. With the change in scenarios, policies will keep changing; therefore, an administrator is given the privilege to create policy based on the changing business needs. There is a local database that maintains user information along with the information about his operation on cloud. The database also makes a note of the resources on cloud so that a reference is available at local level.

5 Proposed System

The proposed RBAC for cloud security aims at providing more security to resource available on cloud by allowing its users to access the resources in a controlled manner. The proposed system has categorized the users into two segments. One is the generic users who are registered to the system, and other is the administrator who monitors the activity on the software. Administrator has the responsibility permitting and denying any request for role inheritance.

Figure 2 depicts the entire system architecture of the proposed system which defines access rights and privileges based on job role assigned to a user when he joins an enterprise. If we define access privileges for each user, it will be difficult to maintain as the project he is a part of. On the cloud layers, we are using blob services to store enterprise resources. Therefore, we create container for every project that comes and put all resources related to that project into their respective containers.

Where a user logs into the system, he can view the files that he has default access to. Say a user A is assigned a group X. Therefore, all the resources in group X container are available to user A as default access resource. For any other resource that fall out of the default access scope, a request needs to be created. In a “create request” scenario, the user needs to provide a request ID, resource for which access is requested and the type of action the user wants to perform on the resource. In our application, we have limited the action to read and commit.

In the file read scenario, a users can download the file from cloud directly if he/she is permitted. In the commit scenario encompasses editing an existing resource and uploading or uploading new resource to cloud.

A request is handled using XACML module. An administrator creates policies based on the current business needs. A policy file is an XML file which has been designed according to Sun’s XACML standards [16]. A policy contains subjects, actions, resources, and rules. Here, the subject is the role for which the policy has been designed. Action is the operation that the above-mentioned role can perform on the resource specified in the resource tag. Followed by the resource tag is the rule tag. This is the most important part of any policy file as this is used for decision making. The decision on whether to permit or deny a particular request is made after checking the rules available. A rule tag has two main attributes: one is the rule Id and other is the effect. The rule Id attribute can be given any value according to user choice, mainly used for reference purpose. The effect attribute usually has permit or deny value.

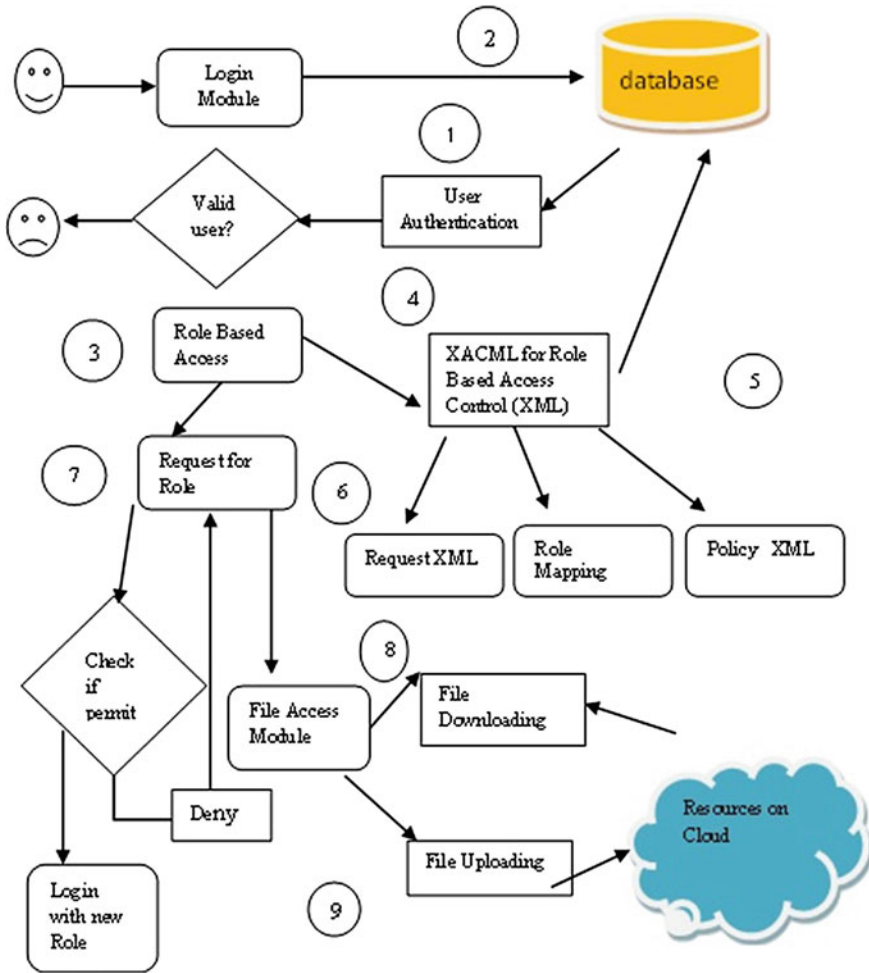


Fig. 2 Proposed system architecture

If requesting attributes match the permit rule, then the user requested operation is permitted else the requested is denied. For every role and for every action, there is a policy designed. This is depicted in Formula 1 given below.

Formula 1: P-Policy R-Role A-Action => P X R XA (policy per role per action).

Role acquisition is also a very important module of this project. Here, we constraint the user to hold only one role at a time. This role can be a primary role or a secondary role. A primary role is defined as the basic role that is assigned to the user when he registers with the system. Secondary role is a new role which he acquires on approval from the administrator. For inheriting a particular role, a user needs to put a request to the administrator. In this scenario, a user needs to wait until the approval comes from administrator. When an administrator logs in, he can

view the entire request coming from different users. These requests can be approved by checking the check box placed along with each request. If the administrator leaves the box unchecked, the user request does not get approved. Once administrator approves the user to have the new role, the user requested role goes to secondary role. This secondary role becomes active immediately and remains active for next 24 h. After the end of this set time period, the user's role is set back to the primary role. In our application, we are limiting the user to request for only one role at a time. If a user already holds a secondary role, then the user is not allowed any role until his current secondary role expires.

This role acquisition plays a very important role in scenario where say in a team a developer had to leave due to unwanted circumstances, then any other job role can take up developer's role as their secondary role and manage the situation.

Apart from these features, there is the file upload component which allows user to upload data to cloud. The user needs to mention the container reference and also the file path in the local directory of the system. This application after uploading the data on cloud keeps reference of the currently available resources on cloud. This is done with the help of a scheduler. Scheduler concept is implemented in two scenarios. Firstly, it is used to get the current data from the cloud. The scheduler is set to trigger every half an hour so that at equal intervals, the local database get refreshed with latest data from cloud. Second scenario where scheduler is used is when the secondary role expires. The time at which the secondary role is granted is set, and the scheduler keep on checking if its validity has expired or not.

In addition to this, administrator can manage all users by adding new user and deleting the users, but he cannot add or remove content from cloud as it is not in his scope of responsibility. The local database stores all user-related information locally. This database contains all user profile-related information such as user ID, role, and group, and it also contains policy-related information.

6 Case Study

Modules of the proposed system case study are role-based login module, role inheritance module, user request module, default file access module, and scheduler module (Table 1).

6.1 Role-Based Login Module

This module allows the user with the right credential to login to the system. On login, the users' role is verified and according his role definition his account his customized giving him the access functionalities that he is entitles to perform. This module takes into account that a user is prevented from resources and functionalities that he is not allowed to access. This system defines two major functionalities:

6.2 Role Inheritance Module

Role inheritance module allows the user to inherit roles. To use this module, user has to first create a request for the role he wants to have. This request is visible in administrator account from where he can permit or deny a request based on the changing business scenario. The algorithm for this module is described below.

6.3 User Request Module

This module is an additional feature of the proposed system. This module allows a user to explicitly access a file which he is not permitted to. This module uses XACML component which has policies defined for every business scenarios. Each user request is checked with the policy available to find a match for the constraints. If a right match is found, the users' request is granted else it is denied.

6.4 Default File Access Module

This module is present in every users' account. When a user registers into the system, he is assigned a group. All the files in that group is accessible to the members of that group. Therefore, this module lists the file that a user can use by default.

6.5 Schedule Module

The proposed system uses the concept of scheduler to get the latest data from cloud. This is achieved using the Quartz Scheduler. The system has a servlet file that loads automatically when the server starts. This scheduler triggers another class which connects with cloud to get all the resource reference to local database. This module helps to get the cloud resource references which can be viewed the administrator or users.

1. User logs in using the user authentication module.
2. If the user credentials match with the data from the database, the user is allowed to log into the system
3. User on entering the system has two options either he can go for role inheritance module to request for a resource.
4. Request for resource by creating the request and mentioning the request attributes. If the request parameter matches the policy parameter that is available, then the request for resource is granted, go to step 8 or 9.

5. Create request for new role.
6. Request for resource by creating the request and mentioning the request attributes. If the request parameter matches the policy parameter that is available, then the request for resource is granted, go to step 8 or 9.
7. Check if the role has been permitted or not, if not then hold the current role else get the new role.
8. Download the file
9. Upload the file

7 Result

The proposed system tries to integrate the RBAC with XACML standard to provide a secure system for the access of cloud resources. This system successfully achieves to provide a secure access to cloud resources. This system has tried integrating the positive features of the below-mentioned algorithms. Comparative studies of all the systems are given in Table 2.

Table 2 Comparative studies of all the systems”Please suggest whether the phrase ‘No implemented for cloud system’ can be changed to ‘No cloud implementation’ in Table 2.” ->

S. no.	Model name	Advantages	Disadvantages
1.	Flat RBAC	Defines role, user, permission, and sessions	No role inheritance feature and no cloud implementation
2.	Role hierarchy model	Defines roles, user, permission, and session and role inheritance	No implemented for cloud system
3.	Attribute RBAC	Subject, resource, and environment Attributes are defined and use security-level checks	No role inheritance and no cloud implementation
4.	coRBAC	Uses dRBAC and digital signature certificate for authentication and authorization	No role inheritance and no cloud implementation
5.	Refined role-based access control model	Defines cloud resources as protected objects and maintains session and permission for the access of cloud resources	No role inheritance
6.	RBAC model for HPCS	Uses standard access control implementation such as XACML	No role inheritance
7.	O-RBAC	Defines role, user, permission, and sessions, ontology concept implemented	No role inheritance and no cloud implementation

7.1 Efficiency Analysis of Proposed System

The proposed system is more efficient than the already existing systems because the proposed system covers all the functionalities of the existing system and at the same time try to integrate it with the cloud environment. The proposed system aims as providing features of role inheritance in cloud along with the use of industry standards such as XACML. The graph below shows the various parameters against which models' functionality can be measured. The graph below shows comparative study of all the models versus the level of protection provided by them: Graph: 6.1.1.1 (Fig. 3 and Table 3).

Fig. 3 Comparative study of access control models

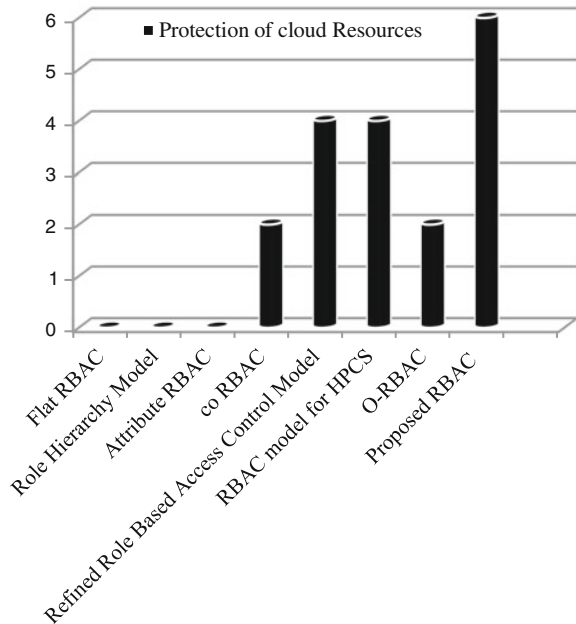


Table 3 Models and their protection attribute

S. no.	Model name	Protection to cloud*
1	Flat RBAC	0
2	Role hierarchy model	0
3	Attribute RBAC	0
4	coRBAC	2
5	Refined role-based access control model	4
6	RBAC model for HPCS	4
7	O-RBAC	2
8	Proposed RBAC	6

Table 4 shows the scale in which the user session and roles are mapped in various models (Fig. 4).

Table 4 User session mapping rate

Sno	Model name	User session and role mapping ^a
1	Flat RBAC	2
2	Role hierarchy model	2
3	Attribute RBAC	2
4	coRBAC	4
5	Refined role-based access control model	2
6	RBAC model for HPCS	4
7	O-RBAC	2
8	Proposed RBAC	4

^a 4—high, 2—average, 0—no protection

Fig. 4 Comparative study of all the models versus the level of user session

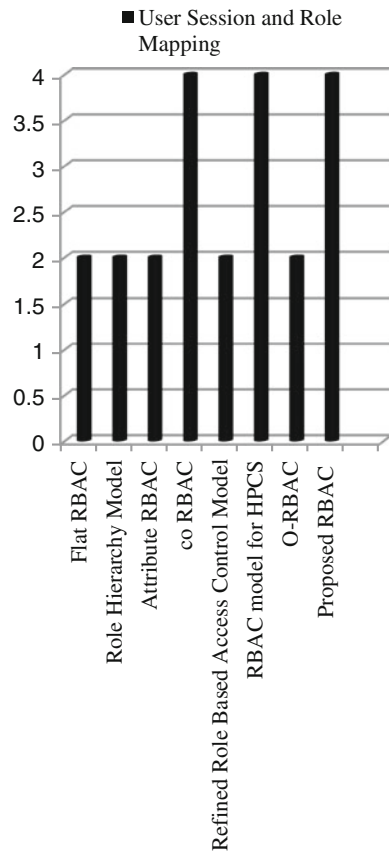


Fig. 5 Comparative study of all the models versus the level of role inheritance

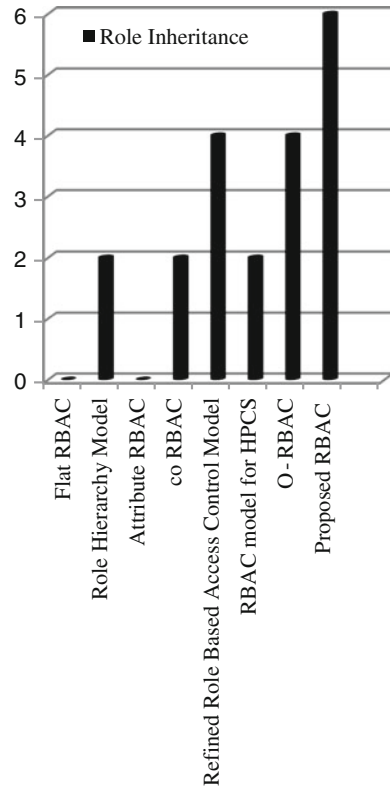


Table 5 Model name and role inheritance

Sno	Model name	Role inheritance ^a
1	Flat RBAC	0
2	Role hierarchy model	2
3	Attribute RBAC	0
4	coRBAC	2
5	Refined role-based access control model	4
6	RBAC model for HPCS	2
7	O-RBAC	4
8	Proposed RBAC	6

^a 6—high, 4—average, 2—low, 0—no protection

The graph (Fig. 5) shows comparative study of all the models versus the level of role inheritance by them.

The Table 5 shows the scale in which the role inheritance is measured in various models.

8 Conclusion

The proposed system tries to provide security solution for cloud resources using RBAC for enterprise data. The system tries to include the basic definition of RBAC along with feature of role inheritance and user request modules. The proposed system uses XACML for developing policies for access control. In general, we can say that RBAC models can be used to implement three important security principles: least privilege, separation of duties, and data abstraction. RBAC has great potentials to provide security and can be successfully implemented in cloud computing. As there are no particular standards or best practices that are available and the unexplored vulnerability of cloud security, there are a lot of solutions that can be provided.

Security through access control on cloud has great scope for improvement in future.

9 Future Work

The proposed system can be modified to add new time constraints like if an administrator wants to customize the time for which role inheritance is valid. Moreover, the system can be made fully automated by removing the administrator role completely. The use of XACML can be explored more and can be modified to provide better security via policy.

Future research should include more formal identification process different entities in RBAC in relation to cloud computing, industry standards, and best practices for using RBAC in cloud computing and large-scale experiments to show that RBAC is a viable solution for cloud security. Over the years as the maturity of cloud computing will increase, we can improve the standards of proposed system so that it covers all aspects of cloud computing.

References

1. [Paperback] T. Mather, S. Kumaraswamy, S. Latif, Cloud security and privacy: an enterprise perspective on risks and compliance (Theory in Practice)
2. S. Misra, P. Venkata Krishna, V. Saritha, H. Agarwal, L. Shu, M.S. Obaidat, Efficient medium access control for cyber physical systems with heterogeneous networks. *IEEE Syst. J.* (2013)
3. R. Sandhu, V. Bhamidipati, E. Coyne, S. Ganta, C. Youman, The ARBAC97 model for role-based administration of roles: preliminary description and outline. In *ACM Workshop on Role-Based Access Control* (1997), pp. 41–50
4. R. Sandhu, D. Ferraiolo, D. Richard Kuhn, The NIST model for role-based access control: towards a unified standard. In *ACM Workshop on Role-Based Access Control*, (2000), pp. 47–63

5. EE. Mon, TT. Naing, The privacy-aware access control system using attribute-and role-based access control in private cloud. In *2011 4th IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, vol. 28(30) (2011), pp. 447–451
6. Z. Tianyi, L. Weidong, S. Jiaying, An efficient role based access control system for cloud computing. In *CIT* (2011), pp. 97–102
7. M. Raykova, H. Zhao, S.M. Bellovin, Privacy enhanced access control for outsourced data sharing. In *Financial Cryptography* (2012), pp. 223–238
8. W. Li, H. Wan, X. Ren, S. Li, A refined RBAC model for cloud computing. In *ACIS-ICIS* (2012), pp. 43–48
9. A.L. Pereira, RBAC for high performance computing systems integration in grid computing and cloud computing. In *IPDPS Workshops* (2011), pp. 914–921
10. S. Godik, T. Moses, Oasis extensible access control markup language (xacml) version 1.1. Oasis committee specification (2003)
11. S.T. Vinter, Extended discretionary access controls. In *IEEE Symposium on Security and Privacy*, (1988), pp. 39–49
12. W.T. Tsai, Q. Shao, Role-based access-control using reference ontology in clouds. In *ISADS* (2011), pp. 121–128
13. M.R. Sadasivan, M.K. Sangeetha, S. Karthik, A survey on access control of cloud data. *IJAR CET* **1**(8) (2012)
14. V. Suhendra, A survey on access control deployment. In *FGIT-SecTech* (2011), pp. 11–20
15. S.L. Reeja, Role based access control mechanism in cloud computing using co-operative secondary authorization recycling method (2012)

Model Predictive Controllers for Nonminimum-phase Quadruple-tank Process

Keerthi Chacko, Lakshmi Ponnusamy and Sangapillai Sutha

Abstract Predictive controllers are used in many in many industries recently. In this paper, a comparison between conventional Proportional plus integral (PI) controller using Ziegler–Nichols tuning method (Z–N) and predictive controllers such as dynamic matrix controller (DMC) and receding horizon control (RHC) are designed. Above-mentioned controllers are applied to a bench mark process quadruple-tank process (QTP) which is operated in nonminimum phase. The simulation for the process and above-mentioned controllers are obtained from MATLAB. Proposed controllers’ performance is supported with better tracking, disturbance rejection, and performance indices in terms of integral time absolute error (ITAE), integral absolute error (IAE) integral square error (ISE).

Keywords Ziegler–Nichols tuning method · Quadruple-tank process · Dynamic matrix controller · Receding horizon control · Relative gain array · Nonminimum phase

1 Introduction

Industrial control problems are difficult to control since they depend on various parameters and there may be interactions among the loops [1, 2]. The multivariable QTP process [3] is explained with the design of various controllers for it. Design of PID controller for a two-by-two system is discussed in [4].

K. Chacko (✉) · L. Ponnusamy · S. Sutha
Department of EEE, College of Engineering Guindy, Chennai, India
e-mail: keerthichacko@gmail.com

L. Ponnusamy
e-mail: p_lakshmi@annauniv.edu

S. Sutha
e-mail: suthaa_s@yahoo.com

The presence of right half plane zero comes along with serious limitations on the performance of the controller when it is in nonminimum-phase mode [5, 6]. The traditional use of transfer function models and state-space models for predictive controllers are explained in [7]. The technique of using step response models and finite impulse response models for the creation of predictive controllers is elaborated in [2]. Explanation about discrete-type model predictive controllers is given in [8, 9]. Various tuning rules for PI is discussed in [10] based on the system characteristics. Similarly, [11] suggests tuning rules for PI in reduced-order model. The theory and design of PI tuning rules is given in [12]. The use of decentralized type of controllers for a multivariable system is discussed in [13]. The receding horizon-type control for the use of multivariable systems is discussed in [5, 14]. Some of the advanced controllers like nonlinear model predictive control are discussed in [15–17].

The rest of the paper is structured as follows. In Sect. 2, a description and a model of the four-tank plant are provided and the benchmark control problem is presented. In Sect. 3, the controller design, PI, and predictive controllers were designed. The simulation results are compared in Sect. 4. Finally, Sect. 5 contains conclusions.

2 Quadruple Tank Process

The laboratory QTP consists of four similar cylindrical tanks of same dimensions connected in interacting fashion along with two fixed-speed pumps and two control valves. It is shown schematically in Fig. 1. The tanks 2 and 3 and 1 and 4 are fed from the same pumps and the outputs of tanks 3 and 4 pour in tanks 1 and 2, respectively. This setup is interfaced with personal computer (PC) via interfacing modules and USB ports. Process signals from the four tank-level transmitters are interfaced with the PC, and it sends outputs to the individual control valves through interfacing units using LabVIEW software.

2.1 Model of Quadruple Tank Process

A nonlinear process model using mass balance and Bernoulli's law is given as

$$\begin{aligned}
 \frac{dh_1}{dt} &= -\frac{a_1}{A_1} \sqrt{2gh_1} + \frac{a_3}{A_1} \sqrt{2gh_3} + \frac{\gamma_1}{A_1} q_1 \\
 \frac{dh_2}{dt} &= -\frac{a_2}{A_2} \sqrt{2gh_2} + \frac{a_4}{A_2} \sqrt{2gh_4} + \frac{\gamma_2}{A_2} q_2 \\
 \frac{dh_3}{dt} &= -\frac{a_3 \sqrt{2gh_3}}{A_3} + \frac{(1-\gamma_2)}{A_3} q_2 \\
 \frac{dh_4}{dt} &= -\frac{a_4 \sqrt{2gh_4}}{A_4} + \frac{(1-\gamma_2)}{A_4} q_1
 \end{aligned} \tag{1}$$

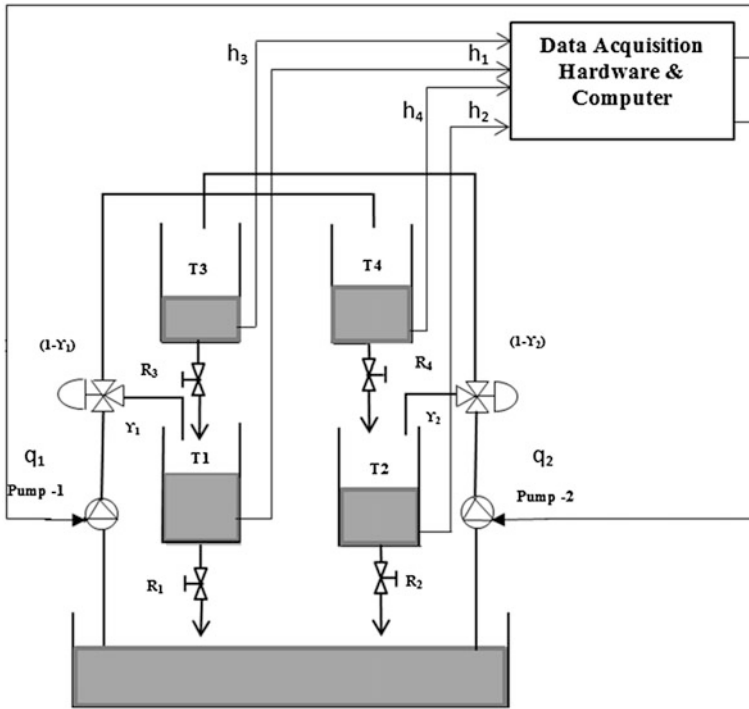


Fig. 1 Schematic diagram of QTP

where

- A_i Cross section of tank (cm^2); $i = 1, 2, 3, 4$
- a_i Cross section of the outlet hole (cm^2)
- h_i Water level in tank (cm)
- γ_i Fraction of water flow from pump i ; $i = 1, 2$
- q_i Inflow rate (cm^3/s)
- g Gravitation due to gravity (cm/s^2)

The linear model is obtained by linearizing the above nonlinear model at an operating point given by the equilibrium levels and flows,

$$\frac{dx}{dt} = \begin{bmatrix} -\frac{1}{T_1} & 0 & \frac{1}{T_3} & 0 \\ 0 & -\frac{1}{T_2} & 0 & \frac{1}{T_4} \\ 0 & 0 & -\frac{1}{T_3} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_4} \end{bmatrix} x + \begin{bmatrix} \frac{\gamma_1}{A_1} & 0 \\ 0 & \frac{\gamma_2}{A_2} \\ 0 & \frac{(1-\gamma_2)}{A_2} \\ \frac{(1-\gamma_1)}{A_4} & 0 \end{bmatrix} u \tag{2}$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x$$

Table 1 Parameters of process

Process parameters	NMP
A_i (cm ²)	176.7146
a_i (cm ²)	2.01062
Υ_1, Υ_2	0.4, 0.42857
h_1, h_2, h_3, h_4 in cm	(14.65, 16.42, 5.07, 5.6)
(q_1^0, q_2^0) in cm ³ /s	(117, 117)

where the time constants are

$$T_i = \frac{A_i}{a_i} \sqrt{2h_i^0/g}$$

To design the controllers, the relation between manipulated variables and control input is required. The corresponding transfer function matrix is

$$G(s) = \begin{bmatrix} \frac{\gamma_1 c_1}{1+sT_1} & \frac{(1-\gamma_2)c_1}{(1+sT_3)(1+sT_1)} \\ \frac{(1-\gamma_1)c_2}{(1+sT_4)(1+sT_2)} & \frac{\gamma_2 c_2}{1+sT_2} \end{bmatrix} \tag{3}$$

where $c_1 = T_1 k_1 k_{p1}/A_1$ and $c_2 = T_2 k_2 k_{p2}/A_2$ if the sum of flow ratios lies between $0 < (\Upsilon_1 + \Upsilon_2) \leq 1$, then the system works in NMP mode. By substituting all parameters in Eq. (3), the open-loop transfer function matrix in NMP case is found. The process parameters are given in Table 1.

The open-loop transfer function matrix is calculated and given in Eq. (4).

$$G(S) = \begin{bmatrix} \frac{0.25}{1+15.18s} & \frac{0.36}{(1+8.94s)(1+15.18s)} \\ \frac{0.4}{(1+9.39s)(1+16.08s)} & \frac{0.28}{1+16.08s} \end{bmatrix} \tag{4}$$

2.2 Analysis of System

Relative gain array (RGA) and singular value decomposition (SVD) are useful to obtain intensity of process interaction based on the open-loop gains. RGA elements should be positive and nearly equal to unity for best pairing and for a lesser interaction. Thus, the relative gain array for a 2×2 system can be expressed as Λ , which is calculated as $G(0) * G(0)^{-T}$. RGA matrix indicates that the output–input pairings should be $y_1 - u_2$ and $y_2 - u_1$.

$$\Lambda = \begin{bmatrix} -0.9459 & 1.9459 \\ 1.9459 & -0.9459 \end{bmatrix}$$

The Niederlinski Index (NI), is $NI = \det[G(0)] / \prod_{i=1}^n G_{ii}(0)$. The system is stable ($NI = 1.9459$) if $NI > 0$. Multivariable zeros can be found either by calculating the $\det[G(s = 0)] = 0$. The zeros in which one lies in LHS of S-plane and another in the RHS which shows the system are in NMP. The two zeros are located in 0.04528 and -0.2640 .

3 Controller Design

3.1 PI Controller Design

The decentralized controller structure [3] and the decentralized control law $u = \text{off-diag} \{C_1, C_2\} (r - y)$, where r is reference input and y is process output. The decentralized control strategy can be represented as $C(s) = \begin{bmatrix} 0 & C_1(s) \\ C_2(s) & 0 \end{bmatrix}$, where each one of the controllers follows the equation, $G_{Cl}(s) = K_{cl} \left(1 + \frac{1}{T_{il}s} \right)$, $l = 1, 2$.

The off-diagonal elements of process can be converted into first-order plus dead-time (FOPDT) model by using Skogestad half-rule method [7]. The following Z-N rule is used to find PI controller settings.

$$k_{cl} = \frac{0.9\tau}{k_p\theta_d}; \quad \tau_{il} = 3.33\theta_d \tag{5}$$

3.2 Predictive Controller

3.2.1 Dynamic Matrix Controller

An objective function based on output predictions over a prediction horizon of P time steps is minimized by selection of manipulated variables over control horizon of M control moves. The model used here is finite step response model.

$$S^T = [S_1 S_2 \dots S_N]$$

where $S_1 S_2 \dots S_N$ are step response coefficients for model length N . The predicted output is given by,

$$Y_{k+j} = S_1 \Delta U_{k+j-1} + S_2 \Delta U_{k+j-2} + \dots + S_j \Delta U_k + S_N U_{k-N+j} + S_{j+1} \Delta U_{k-1} + S_{j+1} \Delta U_{k-1} + \dots + S_{N-1} \Delta U_{k-N+j+1} \tag{6}$$

This can be written as

$$Y = S_f \Delta U_f + S_{\text{past}} \Delta U_{\text{past}} + S_N U_p \quad (7)$$

The control move is given by the matrix

$$\Delta U_f = \left(S_f^T S_f + W \right)^{-1} S_f^T E \quad (8)$$

where E is unforced error.

3.2.2 Receding Horizon Control

The augmented model [2] will be used in the design of predictive control. The objective is to minimize the errors between the predicted output and the set point signal.

Steps involved in realization of receding horizon control (RHC)

1. The state-space representation of the model is created.
2. Augmented model is created from step 1.
3. Obtain the transformation matrices which help in prediction.
4. Predict the change in input required using the transformation matrices created.

$$\Delta U = (\emptyset^T \emptyset + \bar{R})^{-1} \emptyset^T (R_s - Fx(k_i)) \quad (9)$$

where

\bar{R} is a diagonal matrix with a tuning parameter for desired closed-loop performance.

R_s set point vector $x(k_i)$ gives current state variable information.

5. Use the actual discretized model to obtain output.

4 Simulation Results

4.1 Servo Response

The servo response of QTP with PI, DMC, and RHC for the level h_1 and h_2 for two different step changes is shown below. For the level h_1 , a positive step change of 13.3 % at 500 s and a negative step change of 6.66 % at 800 s are applied and compared in Fig. 2. For the level h_2 , a negative step change of 14.28 % at 500 s and a positive step change of 20 % at 800 s are shown in Fig. 3.

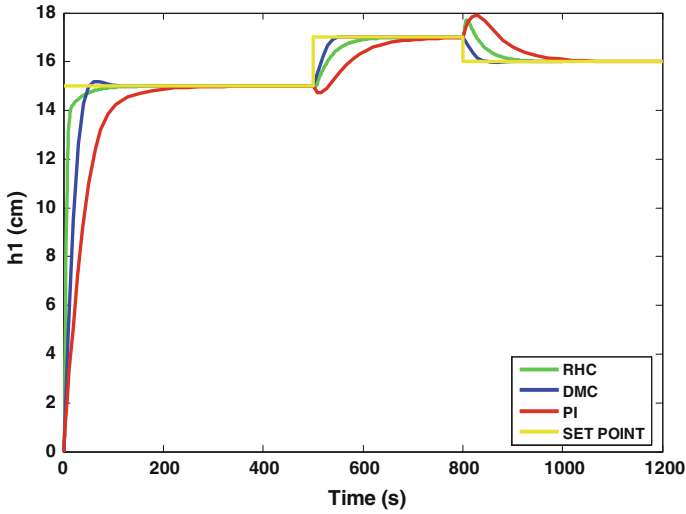


Fig. 2 Servo response for h_1

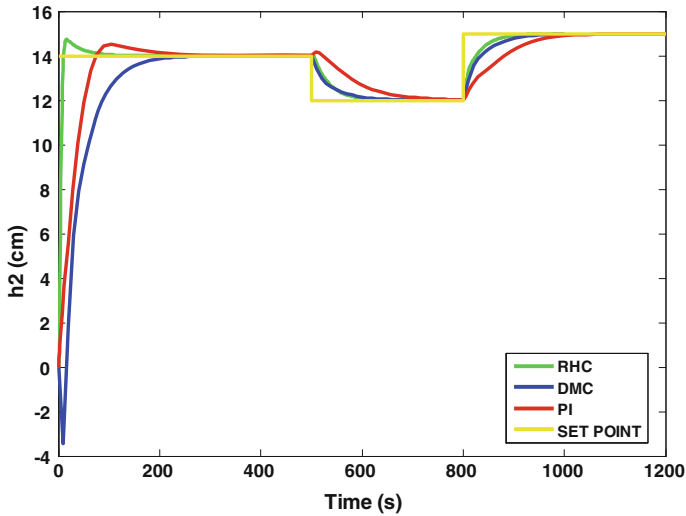


Fig. 3 Servo response for level h_2

4.2 Regulatory Response

The regulatory response of the QTP for the process variable (i.e., level h_1 and h_2) is shown in Figs. 4 and 5, respectively. After reaching a steady state, an input disturbance of 6.66 % is applied at 500 s.

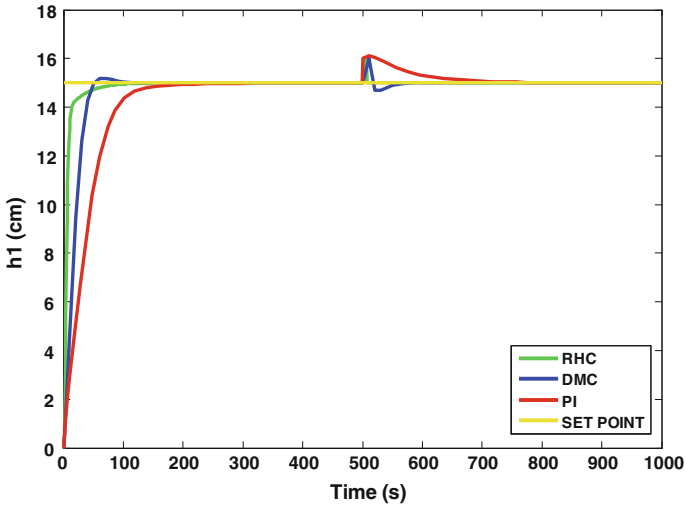


Fig. 4 Regulatory response for level h_1

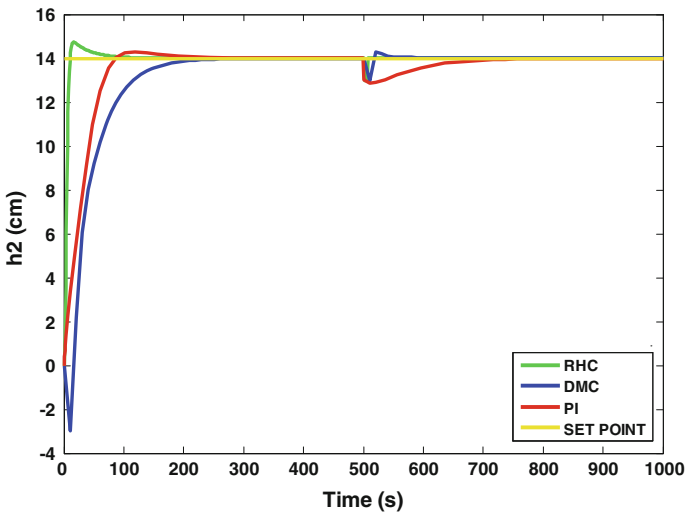


Fig. 5 Regulatory response for level h_2

4.3 Results

A comparison of performance of PI, DMC, and RHC is given in Table 2 in terms of ISE, IAE, and ITAE. It can be observed that settling time and peak overshoot of RHC is lesser than DMC and PI. Similarly, disturbance rejection is more effective in RHC than other two controllers. Also performance indices are lower in RHC compared to response of PI and DMC.

Table 2 Performance indices of controllers

Performance indices	Height of tank	PI	DMC	RHC
ISE	h_1	4,877	998.9	559.6
	h_2	3,511	1,657	398.7
IAE	h_1	581.4	100.3	91.96
	h_2	448	241.1	72.38
ITAE ($\times 10^3$)	h_1	1.9	0.601	0.152
	h_2	1.4	0.515	0.131

5 Conclusion

Model predictive controller is developed to control the level of quadruple tank process. The controller results are compared with that of decentralized PI for servo and regulatory responses. From the servo response, it can be concluded that the system settles faster, and from regulatory response, the disturbance rejection is faster in receding horizon-type predictive controller.

References

1. S. Skogestad, I. Postlethwaite, *Multivariable Feedback Control Analysis and Design*, vol. 2 (Wiley, New York, 2000)
2. B. Wayne Bequette, *Process Control Modeling, Design and Simulation* (Prentice Hall of India, 2004)
3. K.H. Johansson, The quadruple-tank process: a multivariable laboratory process with an adjustable zero. *IEEE Trans. Control Syst. Technol.* **8**, 456–465 (2000)
4. M. Zhuang, D.P. Atherton, PID controller design for a TITO system. *IEE Proc. Control Theory Appl.* **141**, 111–120 (1994)
5. M. Ma et al, Non linear receding horizon control of quadruple tank system and real time implementation. *Int. J. Innovative Comput. Inf. Control* **8**, 7083–7093 (2012)
6. M. Bahrami, B. Ebrahimi, M. Asadi, Robust control of a non linear non minimum phase supersonic flight vehicle based on stable system center. *Aerosp. Sci. Technol.* **25**, 283–291 (2013)
7. L. Wang, *Model Predictive Control System Design* (Springer, New York, 2009)
8. E.F. Camacho, C. Bordons, *Model Predictive Control*, 2nd ed. (Springer, New York, 2004)
9. S. Esaghi, H. Khauati, M.A. Badamchizadeh, I. Hasanzadeh, A predictive controller based on dynamic matrix control for a non-minimum phase robot manipulator. *Int. J. Control Autom. Syst.* **10**, 574–581 (2012)
10. A. O'Dwyer, *Handbook of PI and PID Controller Tuning Rules*, 2nd ed (Imperial College Press, London, 2006)
11. S. Skogestad, Simple analytical rules for model reduction and PID controller tuning. *J. Process. Control* **13**, 291–309 (2003)
12. K.J. Astrom, T. Hagglund, *PID Controllers: Theory, Design and Tuning*, 2nd edn. (Instrument society of America, Research Triangle Park, 1995)
13. R. Vikresh, N. Sivakumar, J.S. Chandra, T.K. Radhakrishnan, A critical study of decentralized controllers for a multivariable system. *Chem. Eng. Technol.* **27**, 880–889 (2004)

14. H. Michalska, D.Q. Mayne, Robust receding horizon control of constrained nonlinear systems. *Trans. Control Syst.* **38**, 1623–1633 (2002)
15. T. Raff, Nonlinear model predictive control of a four tank system—an experimental stability study, in *Proceedings of the 2006 IEEE International Conference on Control Applications* (2006)
16. N. Sivakumaran, T.K. Radhakrishnan, Predictive controller design for non-linear chemical processes. *Indian J. Chem. Technol.* **14**, 341–349 (2007)
17. Y. Zou, Y. Niu, B. Chen, T. Jia, Networked predictive control of constrained linear systems with input quantization. *Int. J. Syst. Sci.* 1970–1982 (2013)