

Lecture Notes in Networks and Systems 244

Padmalaya Nayak
Souvik Pal
Sheng-Lung Peng *Editors*

IoT and Analytics for Sensor Networks

Proceedings of ICWSNUCA 2021

 Springer

Lecture Notes in Networks and Systems

Volume 244

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Advisory Editors

Fernando Gomide, Department of Computer Engineering and Automation—DCA,
School of Electrical and Computer Engineering—FEEC, University of Campinas—
UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering,
Bogazici University, Istanbul, Turkey

Derong Liu, Department of Electrical and Computer Engineering, University
of Illinois at Chicago, Chicago, USA

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering,
University of Alberta, Alberta, Canada

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering,
KIOS Research Center for Intelligent Systems and Networks, University of Cyprus,
Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong,
Kowloon, Hong Kong

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at <http://www.springer.com/series/15179>

Padmalaya Nayak · Souvik Pal · Sheng-Lung Peng
Editors

IoT and Analytics for Sensor Networks

Proceedings of ICWSNUCA 2021

 Springer

Editors

Padmalaya Nayak
Department of Computer Science
and Engineering
Gokaraju Rangaraju Institute
of Engineering and Technology
Hyderabad, Telangana, India

Souvik Pal
Department of Computer Science
and Engineering
Global Institute of Management
and Technology
Nadia, West Bengal, India

Sheng-Lung Peng
Department of Creative Technologies
and Product Design
National Taipei University of Business
Taipei City, T'ai-pei, Taiwan

ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-981-16-2918-1

ISBN 978-981-16-2919-8 (eBook)

<https://doi.org/10.1007/978-981-16-2919-8>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed 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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Organizing Committee and Key Members

Conference Committee Members

Chief Patrons

G. Ganga Raju, President, GRES, India

G. V. K. Ranga Raju, Vice President, GRES, India

Patrons

M. G. Sekharam, CEO, GRES, India

Jandhyala N. Murthy, Director, GRIET, India

J. Praveen, Principal, GRIET, India

General Chair

D. K. Lobiyal, JNU, Delhi, India

Honorary Chair

Sheng-Lung Peng, National Dong Hwa University, Taiwan

Neel Kanth Grover, IKG Punjab Technical University Kapurthala, India

Convenor

Padmalaya Nayak, GRIET, India

Co-Convenor

K. Madhavi, GRIET, Hyderabad, India

Neeraj Mohan, IKG Punjab Technical University Kapurthala, India

Surbhi Gupta, GRIET, Hyderabad, India

Publication Chair

Sheng-Lung Peng, National Taipei University of Business, Taiwan
Padmalaya Nayak, GRIET, Hyderabad, India
Souvik Pal, Global Institute of Management and Technology, India

Program Chair

Siba K. Udgata, University of Hyderabad, Hyderabad, India

Publicity Chair

G. R. Sakthidharan, GRIET, Hyderabad, India
Sachinandan Mohanty, ICFAI, Hyderabad, India
Tanupriya Choudhury, Univ. of Petroleum and Energy Studies, Dehradun, India

Advisory Committee

Ganapati Panda, IIT, Bhubaneswar, India
K. R. Suresh Nair, Chief Technology Officer, NeST Group
V. Kamakshi Prasad, JNTUH, Hyderabad, India
Atul Negi, University of Hyderabad, Hyderabad, India
S. Krishna Kumar, DRDO, Chennai, India
Siba K. Udgata, Hyderabad Central University, Hyderabad, India
Prakash K. Ray, IIIT, Bhubaneswar, India
S. Viswanadha Raju, JNTUH, Hyderabad, India

TPC Chair

Monika Sachdeva, IKG Punjab Technical University Kapurthala, India

Technical Program Committee

Srinivas Chakravarthy, Kettering University, USA
Sayyad Mohiddin, Think Soft, Australia
Manas Ranjan Patra, Berhampur University, Odisha, India
Elsa Estevez, National University of South Argentina
Ismail Saad, University of Malaysia, Sabah, Malaysia
Wan Abul Rahim Wan Mohd Isa, Univ. Technology, MARA, Malaysia
Raghunandan Reddy Alugubelli, MCC, Univ. of South Florida, USA
Mohammad S Hasan, Staffordshire University, United Kingdom
Md. Ahsan Habib, MBST University, Bangladesh
Md. Abdur Razzaque, University of Dhaka, South Korea
Selvakumar Samuel, Asia Pacific University and Technology, Malaysia
E. Balamurugan, Bluecrest College, Accra, Ghana
Brijesh Kumbhani, IIT, Ropar, India
Aryabhhatta Sahu, IIT, Guwahati, India
Prasant Kumar Sahu, IIT, Bhubaneswar, India

Chandrashekhhar Bhende, IIT, Bhubaneswar, India
T. Ramakrishnudu, NIT, Warangal, India
B. N. Bhandari, JNTUH, Hyderabad, India
Debadatta Pati, NIT, Nagaland, India
Pravati Swain, NIT, Goa, India
Subhransu Ranjan Samantray, IIT, Bhubaneswar, India
Aruna Tiwari, IIT, Indore, India
Kameswari Chebrolu, IIT, Bombay, India
Rajib. Kumar Panigrahi, IIT, Roorkee, India
DVLN Somayajulu, NIT, Warangal, India
Asit Mohanty, CET, Bhubaneswar, India
Neeraj Mohan, IKGPTU, Jalandhar, India
Diptendu Sinha Roy, NIT, Meghalaya, India
Gayadhar Panda, NIT, Meghalaya, India
Prabeen Kumar Padhy, IIITDM, Jabalpur, India
Vivekanandan Kaniappan, Bharathiar University, Coimbatore, India
Sraban Kumar Mohanty, IIITDM, Jabalpur, India
K. P. Supreethi, JNTUH, Hyderabad, India
Chapram Sudhakar, NIT, Warangal, India
V. Valli Kumari, Andhra University, India
Karthikeyan S. S., IIITDM, Kanchipuram, India
S. Kanimozhi Suguna, SASTRA DEEMED University, Thanjavur, India
Sudhansu Sekhar Singh, KIIT, Bhubaneswar, India
Niranjan Ray, KITT, Bhubaneswar, India
Ghaida Muttashar Abdulsahib, University of Technology-Iraq, Baghdad, Iraq
M. Nageswar Rao, KL University, India
P. Vidya Sagar, KL University, India
K. Suvarna Vani, VR Siddhartha Engineering College, India
Somya Goyal, SCIT, Manipal University Jaipur, Rajasthan, India

Invited Speaker

Prof. Siba K. Udgata

Professor
School of Computer and Information Sciences
University of Hyderabad (Central University), India

Prof. Sheng-Lung Peng

Professor
Department of Creative Technologies and Product Design
National Taipei University of Business, Taiwan

Dr. Amit K. Mishra

Head
Radar Remote Sensing Group (RRSG)
Electrical Engineering Department
University of Cape Town, South Africa

Dr. S. Krishnakumar

Senior Technical Officer
Defence R&D Organisation (DRDO), Government of India, India

Prof. Manas Ranjan Patra

Professor
Department of Computer Science
Berhampur University, India

Prof. Maninder Singh

Professor and Head
Computer Science and Engineering Department
Thapar Institute of Engineering and Technology, India

Preface and Acknowledgement

The main aim of this proceedings book is to bring together leading academic scientists, researchers, and research scholars to exchange and share their experiences and research results on all aspects of Intelligent ecosystems, data Sciences, and IoT systems. ICWSNUCA 2021 is a multidisciplinary conference organized with the objective of bringing together academic scientists, professors, research scholars, and students working in various fields of engineering and Technology. On ICWSNUCA 2021, you will have the opportunity to meet some of the world's leading researchers, to learn about some innovative research ideas and developments around the world, and to become familiar with emerging trends in Science-Technology. The conference will provide the authors, research scholars, listeners with opportunities for national and international collaboration and networking among universities and institutions for promoting research and developing the technologies globally. This conference aims to promote translation of basic research into institutional and industrial research and convert applied investigation into real-time application.

ICWSNUCA 2021 is hosted by Department of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India. ICWSNUCA 2021 has been held during 26–27th February, 2021, in online mode (ZOOM Platform). The conference brought together researchers from all regions around the world working on a variety of fields and provided a stimulating forum for them to exchange ideas and report on their researches. The proceeding of ICWSNUCA 2021 consists of 45 best-selected papers which were submitted to the conferences, and peer-reviewed by conference committee members and international reviewers. The Presenters have presented through virtual screen. Many distinguished scholars and eminent speakers have joined from different Countries like India, Malaysia, Bangladesh, Sri Lanka, Iraq, Libya, and Taiwan to share their knowledge and experience and to explore better ways of educating our future leaders. This conference became a platform to share the knowledge domain among different countries' research culture. The main and foremost pillar of any academic conference is the authors and the researchers. So, we are thankful to the authors for choosing this conference platform to present their works in this pandemic situation.

We are sincerely thankful to Almighty for supporting and standing at all times with us, whether it's good or tough times and given ways to concede us. Starting

from the Call for papers till the finalization of chapters, all the team members given their contributions amicably, which it's a positive sign of significant team works, the editors and conference organizers are sincerely thankful to all the members of Springer especially Mr. Aninda Bose for the providing constructive inputs and allowing an opportunity to finalize this conference proceeding. We are also thankful to Dr. Thomas Ditzinger, Prof. William Achauer, and Prof. Anil Chandy for their support. We are also grateful to Mr. Daniel Joseph Glarance, Project Coordinator, Springer for his cooperation. We are thankful to all the reviewers who hail from different places in and around the globe, who shared their support and stand firm toward quality chapter submission in this pandemic situation.

Finally, we would like to wish you have good success in your presentations and social networking. Your strong supports are critical to the success of this conference. We hope that the participants not only enjoyed the technical program in conference but also found eminent speakers and delegates in the virtual platform. Wishing you a fruitful and enjoyable ICWSNUCA 2021.

Hyderabad, India
Nadia, India
Taipei City, Taiwan

Padmalaya Nayak
Sheng-Lung Peng
Souvik Pal

About This Book

The conference proceeding book is a depository of knowledge enriched with recent research findings. The main focus of this volume is to bring all the computing and communication-related technologies in a single platform. ICWSNUCA 2021 is aimed at providing a platform for knowledge exchange of the most recent scientific and technological advances in the fields of data analytics and computational Science, to strengthen the links in the scientific community. This event aspires to bring together leading academic scientists, researchers, industry persons, and research scholars to exchange and contribute to their knowledge, experiences and research outcome on all the phases of Computer Science, Information technology, and data analytics. This book focuses on the applications, use-cases, architectures, deployments, and recent advances of Wireless Sensor Networks as well as Ubiquitous Computing. Different Research Topics may be illustrated in this book, like Wireless Sensor Networks for the Internet of Things, IoT Applications for eHealth, smart cities, etc., architectures for WSNs and IoT, WSNs hardware and new devices, Low-power wireless technologies, Wireless ad hoc Sensor networks, Routing and data transfer in WSNs, Multicast communication in WSNs, Security management in WSNs and in IoT systems, and Power consumption optimization in WSNs. In recent years, we have been hearing about Internet of Things, Industrial IoTs (IIoTs), Industry 4.0, Smart Factories, Smart HealthCare, Smart Logistic and Supply Chains, Smart Mobility, and Smart Energy, etc. These technologies deliver huge economy and financial benefits and will continue the same in near future. But surprisingly, sensor nodes are the backbone of these technologies and these nodes can be worn, carried, embedded in the environment that can provide interesting contextual information. We find sensors in our daily life, starting from Mobile devices, Automobiles, Industrial automation, Robotic systems, Vehicular Tracking and Management Systems, etc. A significant increase in real-world event monitoring capability with Wireless Sensor Networks will lead to a further evolution of ubiquitous computing. Wireless Sensor Networks continue to attract a lot of attention from the academia and industry among researchers, industrials, equipment and chip manufacturers and service providers for promoting large-scale deployments in many applications, such as environmental monitoring, military and medical surveillance, health and wellness applications, intelligent traffic management and user's tracking and identification,

including human–computer interaction. It also provides a premier interdisciplinary platform for researchers, practitioners, and educators to present and discuss the most recent innovations, trends, and concerns as well as practical challenges encountered and solutions adopted in the fields of Computer Science. The primary audience for the book incorporates specialists, researchers, graduate understudies, designers, experts, and engineers who are occupied with research and computer science-related issues. The book will be organized in independent chapters to provide readers great readability, adaptability, and flexibility.

Contents

An Interactive Smart Mirror Using Internet of Things and Machine Learning	1
Keval B. Prajapati, Chintan Bhatt, and Hakima Chaouchi	
Cluster Formation Algorithm in WSNs to Optimize the Energy Consumption Using Self-Organizing Map	11
Padmalaya Nayak, GK. Swetha, Priyanka Kaushal, and D. G. Padhan	
CNN-Based Mobile Device Detection Using Still Images	23
Surbhi Gupta, Neela Pravalika, Padmalaya Nayak, and Jaafar Al Ghazo	
E-FFTF: An Extended Framework for Flexible Fault Tolerance in Cloud	35
Moin Hasan, Major Singh Goraya, and Tanya Garg	
Human Abnormal Activity Pattern Analysis in Diverse Background Surveillance Videos Using SVM and ResNet50 Model	47
S. Manjula and K. Lakshmi	
RT-GATE: Concept of Micro Level Polarization in QCA	61
K. Bhagya Lakshmi, D. Ajitha, and K. N. V. S. Vijaya Lakshmi	
Comparative Performance Analysis of Tanh-Apodized Fiber Bragg Grating and Gaussian-Apodized Fiber Bragg Grating as Hybrid Dispersion Compensation Model	71
Baseerat Gul and Faroze Ahmad	
Performance Comparison of Adaptive Mobility Management Scheme with IEEE 802.11s to Handle Internet Traffic	83
Abhishek Majumder and Sudipta Roy	
Automatic Attendance Management System Using Face Detection and Face Recognition	97
M. Varsha and S. Chitra Nair	

ESIT: An Enhanced Lightweight Algorithm for Secure Internet of Things 107
Manoja Kumar Nayak and Prasanta Kumar Swain

A Novel Block Diagonalization Algorithm to Suppress Inter-user Interference in a Multi-user MIMO System 117
Harsha Gurdasani, A. G. Ananth, and N. Thangadurai

Prediction of Chemical Contamination for Water Quality Assurance Using ML-Based Techniques 127
C. Kaleeswari and K. Kuppusamy

Design and Performance Analysis of Two-Port Circularly Polarized MIMO Antenna for UWB Applications 139
Madan Kumar Sharma, Aryan Sachdeva, Ayushi Ojashwi, and Mithilesh Kumar

Intelligent Traffic Control System for Emergency Vehicles 151
Anuj Sachan and Neetesh Kumar

Digital Controller-Based Automated Drainage Water Monitoring and Controlling 161
T. Sairam Vamsi, Ch. Hari Krishna, P. Srinivasaraju, and G. Srinivasarao

Cooperative Agent-Based Location Validation for Vehicular Clouds 171
Shailaja S. Mudengudi and Mahabaleshwar S. Kakkasageri

An Analytical Approach for Traffic Grooming Problems Using Waiting Probability in WDM Networks 183
Priyanka Kaushal, Neeraj Mohan, Surbhi Gupta, and Seifedine Kadry

Flow-Based Detection and Mitigation of Low-Rate DDOS Attack in SDN Environment Using Machine Learning Techniques 193
K. Muthamil Sudar and P. Deepalakshmi

Design and Simulation of MEMS Based Capacitive Accelerometer 207
S. Veena, Newton Rai, Amogh Manjunath Rao Morey, H. L. Suresh, and Habibuddin Shaik

Transport Tracking Using RFID and GSM Based Technique 225
N. Subbulakshmi, R. Chandru, and R. Manimegalai

An Ultra-Wide Band Patch Antenna for Commercial Communication Applications 235
L. Diana Evangeline, G. Shine Let, and C. Benin Pratap

8-Bit Carry Look Ahead Adder Using MGDI Technique 243
P. Ashok Babu, V. Siva Nagaraju, and Rajeev Ratna Vallabhuni

Improved Scientific Workflow Scheduling Algorithm with Distributed Heft Ranking and TBW Scheduling Method 255
Ramandeep Sandhu and Kamlesh Lakhwani

Selection of OLAP Materialized Cube by Using a Fruit Fly Optimization (FFO) Approach: A Multidimensional Data Model 265
Anjana Yadav and Anand Tripathi

Fault Tolerant Multimedia Caching Strategy for Information-Centric Networking 275
Dharamendra Chouhan, Sachinkumar Hegde, N. N. Srinidhi, J. Shreyas, and S. M. Dilip Kumar

Sizing of Wireless Networks with Sensors for Smart Houses with Coverage, Capacity, and Interference Restrictions 285
Jhonatan Fabricio Meza Cartagena, Deepa Jose, and J. S. Prasath

Cloud-Based Parkinson Disease Prediction System Using Expanded Cat Swarm Optimization 299
Ramaprabha Jayaram and T. Senthil Kumar

Electric Vehicle Monitoring System Based on Internet of Things (IoT) Technologies 311
Yogesh Mahadik, Mohan Thakre, and Sachin Kamble

Dynamic Analysis and Projective Synchronization of a New 4D System 323
M. Lellis Thivagar, Ahmed S. Al-Obeidi, B. Tamilarasan, and Abdulsattar Abdullah Hamad

Improving the Protection of Wireless Sensor Network Using a Black Hole Optimization Algorithm (BHOA) on Best Feasible Node Capture Attack 333
Ankur Khare, Rajendra Gupta, and Piyush Kumar Shukla

A Systematic Analysis of the Human Activity Recognition Systems for Video Surveillance 345
Sonika Jindal, Monika Sachdeva, and Alok Kumar Singh Kushwaha

Integrating IoT with Blockchain: A Systematic Review 355
Malvinder Singh Bali, Kamali Gupta, and Swati Malik

Quality Assisted Spectrum Allocation in Cognitive NOMA Networks 371
D. Prasanth Varma and K. Annapurna

Proficient Dual Secure Multi Keyword Search by Top-K Ranking Based on Synonym Index and DNN in Untrusted Cloud 381
Rosy Swami and Prodipto Das

Transfer Learning-Based Detection of Covid-19 Using Chest CT Scan Images 393
Aryaman Chand, Khushi Chandani, and Monika Arora

A Hybridized Machine Learning Model for Optimal Feature Selection and Attack Detection in Cloud SaaS Framework 403
Reddy Saisindhutheja and Gopal K. Shyam

English Master AMMU: Advanced Spoken English Chatbot 415
A. N. Gayathri and V. Viji Rajendran

Investigation of CNN-Based Acoustic Modeling for Continuous Hindi Speech Recognition 425
Tripti Choudhary, Atul Bansal, and Vishal Goyal

Energy Conserving Techniques of Data Mining for Wireless Sensor Networks—A Review 433
Pragati Patil Bedekar, Atul Raut, and Abhimanyu Dutonde

IoT Based Healthcare System for Patient Monitoring 445
S. Saravanan, M. Kalaiyarasi, K. Karunanithi, S. Karthi, S. Pragaspathy, and Kalyan Sagar Kadali

Detection and Classification of Intracranial Brain Hemorrhage 455
K. V. Sharada, Vempaty Prashanthi, and Srinivas Kanakala

Implementation of Efficient Technique to Conduct DDoS Attack Using Client–Server Paradigm 465
Seema Rani and Ritu Nagpal

Design and Development of Retrieval-Based Chatbot Using Sentence Similarity 477
Haritha Akkineni, P. V. S. Lakshmi, and Lasya Sarada

A Pragmatic Study on Movie Recommender Systems Using Hybrid Collaborative Filtering 489
Akhil M. Nair and N. Preethi

An Anatomization of FPGA-Based Neural Networks 495
Anvit Negi, Devansh Saxena, Kunal, and Kriti Suneja

Author Index 507

About the Editors

Padmalaya Nayak is working as Professor in department of Computer Science Engineering in Gokaraju Rangaraju Institute of Engineering Technology, Hyderabad under Jawaharlal Technological University, Hyderabad, since 2009. She completed her doctoral degree from National Institute of Technology, Tiruchirappalli, India in 2010. She has 17 years of teaching and research experience in the area of Ad hoc and Sensor Networks. She has published more than 40 research papers in various International Journals and Conferences. She has also contributed two book chapters to her credit. She has visited many countries to present her research paper in many International Conferences. She has received many National/International Level awards for her academic contributions towards the education system. She is the Reviewer of many IEEE/SPRINGER/ELSEVIER Journals and Conferences. She is the member of IEEE, IETE and IEANG professional bodies. She is also a member of the advisory committee for several conference committees.

Souvik Pal is an Associate Professor and Head of the Computer Science and Engineering Department at the Global Institute of Management and Technology, West Bengal, India. Prior to that, he was associated with Brainware University, Kolkata, India; JIS College of Engineering, Nadia; Elite College of Engineering, Kolkata; and Nalanda Institute of Technology, Bhubaneswar, India. Dr. Pal received his BTech, MTech, and PhD degrees in the field of Computer Science and Engineering. He has more than a decade of academic experience. He is author or co-editor of 12 books from reputed publishers, including Elsevier, Springer, CRC Press, and Wiley, and he holds three patents. He is serving as a series editor for “Advances in Learning Analytics for Intelligent Cloud-IoT Systems”, published by Scrivener Publishing (Scopus-indexed) and “Internet of Things: Data-Centric Intelligent Computing, Informatics, and Communication”, published CRC Press, Taylor & Francis Group, USA. Dr. Pal has published a number of research papers in Scopus / SCI-indexed international journals and conferences. He is the organizing chair of RICE 2019, Vietnam; RICE 2020 Vietnam; ICICIT 2019, Tunisia. He has been invited as a keynote speaker at ICICCT 2019, Turkey, and ICTIDS 2019,2021, Malaysia, MAICT 2020, Iraq; ICWSNUSA 2021. His professional activities include roles as associate editor and

editorial board member for more than 100+ international journals and conferences of high repute and impact. His research area includes cloud computing, big data, internet of things, wireless sensor network, and data analytics. He is a member of many professional organizations, including MIEEE; MCSI; MCSTA/ACM, USA; MIAENG, Hong Kong; MIREA, USA; MACEEE, New Delhi; MIACSIT, Singapore; and MAASCIT, USA.

Sheng-Lung Peng is a Professor and the director (head) of the Department of Creative Technologies and Product Design, National Taipei University of Business, Taiwan. He received the PhD degree in Computer Science from the National Tsing Hua University, Taiwan. He is an honorary Professor of Beijing Information Science and Technology University, China, and a visiting Professor of Ningxia Institute of Science and Technology, China. He is also an adjunct Professor of Mandsaur University, India. Dr. Peng has edited several special issues at journals, such as *Soft Computing*, *Journal of Internet Technology*, *Journal of Real-Time Image Processing*, *International Journal of Knowledge and System Science*, *MDPI Algorithms*, and so on. His research interests are in designing and analyzing algorithms for Bioinformatics, Combinatorics, Data Mining, and Networks areas in which he has published over 100 research papers.

An Interactive Smart Mirror Using Internet of Things and Machine Learning



Keval B. Prajapati, Chintan Bhatt, and Hakima Chaouchi

Abstract With the large-scale improvements in communication technology and easy accessibility of the same has led to the advent of the IoT technology, which in turn has resulted in more, and more devices being IoT enabled. This research paper embarks a next generation smart mirror with the aim of connecting the conventional mirror into the IoT network along with a certain set of features, which were never introduced before. The whole range of features focuses on increasing the utility of daily mirrors. This mirror along with acting as a piece of glass will also be acting as a huge glass notepad upon which the user can leave any handwritten messages which are to be conveyed to others and can never go unnoticed. Keeping in mind the smudging effect and the fingerprint mark that are left behind when the mirror surface is touched, this new feature allows one to write the message on the mirror without touching the mirror surface itself. Additionally, the mirror will also support some of common features as well namely date, weather forecast, news headlines and daily reminders.

Keywords OpenCV · Object detection · Object tracking · Camera · Neural networks · Python

1 Introduction

For more than a decade, the development of applications for the Internet of Things (IoT) [1–5] has mainly concerned the construction of systems for data generation, transmission and analysis for integration into everyday life objects, while the data visualization part has still remained attached to the world of graphical user interfaces,

K. B. Prajapati (✉) · C. Bhatt
Charotar University of Science and Technology, Changa, Gujarat, India
e-mail: 16ce091@charusat.edu.in

C. Bhatt
e-mail: chintanbhatt.ce@charusat.ac.in

H. Chaouchi
Télécom SudParis, Évry, France

executed primarily on personal devices such as laptops, tablets and smartphones. Recently this is no longer true, and there are numerous examples of things that can also be used to display information.

In the contemporary world, people are living an extremely busy life with a very less time in the morning rush hours. Waking up early, getting ready for jobs, catching up the earliest suitable commuting vehicles, which in turn leaves a very less time to get updated with day-to-day changing scenarios of the state, country or the entire world before stepping out of the house. Even though one can be updated by accessing their smartphones, tablets or PCs but for tech-savvy people it just becomes another task. The best time management strategy is to find time where it was never before. Seamless integration of technology into daily lives has made this possible where not only communicating devices like smartphones or tablets play a role for the purpose but also household items partake in the same. Another important point is in a hectic life people often forget about trivial tasks relating to family, friends or colleagues. Even though leaving the message on sticky notes is a good idea but most of the time it goes unnoticed and the task remains incomplete.

This is where the idea of using the mirror surface as a huge notepad comes into the picture. The mirror is such a household item, which is looked upon often either purposefully or accidentally. Taking the advantage of this fact, makes it possible to use the conventional mirror surface as a notepad upon which the user can leave handwritten messages, which are to be conveyed to others, members of the house or for the user himself as a to-do note. The note-taking feature supports all the basic features a regular notepad supports: write, erase and clear all.

The mirror also has an integrated music player. Unlike the conventional music players, this one is controlled via some simple human hand gestures. Introduction of a music player into the mirror enhances the utility of the mirror. For example, during certain occasions, a person might find himself/herself to sit in front of a mirror for a prolonged period in order to dress up. The person can use this music player housed in the mirror instead of playing music on their smartphones or other devices. The feature of hand gestures enhances the multi-tasking of the user as well as allows seamless integration of the same in day-to-day lives. Unlike the other hand gesture detection algorithms, which are based on real-time object detection, the one that is featured in this system is based on a neural network using Keras and TensorFlow, which makes prediction of the hand gestures with the help of a dataset of hand gestures images. This model is fairly more accurate and robust when compared to other ones.

The system also targets a short span of time, which people use to spend in the front of the mirror for dressing up themselves or making/combing their hair. This next generation mirror not only acts as a conventional mirror, which allows one to look in the mirror for dressing up or combing their hair but will also be displaying all the vital, crisp and easily comprehensible information like daily remainders, daily weather forecast, time and date, and top five news headlines. All this information will be displayed on the same piece of glass upon which the user is looking hence increasing the utility of the daily mirror. Apart from that, all the content (weather forecast, news headlines and daily reminder) displayed will be updated on a regular

interval of time so that always most recent information will be available with the system at disposal.

The given set of features will not only increase the utility of conventional mirrors but also will become an effortless and seamless way users can adopt the incremental update in conventional mirrors.

The remainder of this paper is structured as follows. Section 2 presents the research motivation, explaining why this upgrade was necessary. It also shows about the related work that has been done in this field. The design architecture, working and features are described in detail in Sect. 3. Results are discussed in Sect. 4, while Sect. 5 shows conclusion and future enhancement of the paper.

2 Motivation and Related Work

Another version of magic mirror is built by Josep Cumeras I Khan [1], which can be controlled by voice and gestures. This system has its very own OS known as MirrorOS or MOS, which is responsible for all the administration. Along with that it also has a dedicated app which can govern what is to be displayed on the mirror. This mirror uses a set of motion detectors for gesture control. Another project is HomeMirror by Hannah Mitt of Adafruit Industries [2], which uses an old tablet instead of LCD/LED monitor. It can be considered as one of the cheapest smart mirrors. The tablet will be running an android-based application designed to show date, time and other information.

The core concept behind the note-taking feature introduced is based on object detection and tracking the same. All of this is achieved with OpenCV and webcam. Extensive research has been done in this field. Rajeev Thaware, Gouri Patil and Saurabh Joshi of eInfochips [3] have in their research work for object detection and tracking for Effective Traffic Management System using OpenCV has described in simple and lucid manner the entire procedure for object detection.

3 Architecture, Working and Feature

The hardware part involves designing of the wooden frame, which will house all the components of the mirror. The 2-way mirror used is reflective enough so that one can use it as a normal mirror and at a same time is transparent enough to allow light to pass through when the light source is in close proximity to the mirror. The 2-way mirror is sandwiched between the wooden frame and LCD/LED screen. The screen is connected to the Raspberry Pi module, which serves as the brain of the mirror. A webcam is also connected to the Raspberry Pi, which will be helping in the note-taking feature. Another webcam is responsible for hand gesture detection. These hand gestures will in turn control the music player. Working of both the webcams is completely independent. Both the webcams are positioned on the top of the wooden

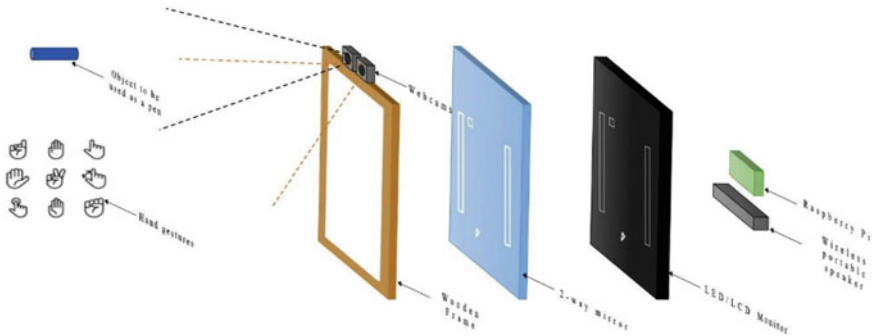


Fig. 1 Hardware model of smart mirror

frame. The system also consists of a wireless speaker, which will be required for music players. Figure 1 shows the hardware design of the next generation mirror.

The software part includes the development of a python application, which will be responsible for displaying the information on to the mirror. The GUI design is mainly done using tkinter library of python. The procedure involved dividing and designing different regions of the screen or mirror for different purposes. The central region of the mirror serves as a notepad. The tkinter library is used to provide software button like write, erase and clear all for this feature. The bottommost region is reserved for news feed, while the top most region is reserved for date, time, weather forecast and remainders. The region between the notepad and news feed is used for displaying the music player. It will consist of three buttons namely six play/pause, next and previous, a marquee text-representing name of current song and a volume slider at bottom of marquee.

3.1 Hand Gesture-Controlled Music Player

This is another notable feature of the mirror. Unlike some common hand gesture detection algorithms which are hardly accurate and fail to produce desired results under various scenarios the one which is running for this system is based on a neural network using Keras and TensorFlow which makes prediction of the hand gestures with the help of given dataset of hand gestures images [6]. Initially this model required a statistical training with certain dataset and later on can do complex prediction accurately in different scenarios as well. One can have their own set of gestures for the music control however they need to train the model for the same first. The entire module can be split into 3 parts:

- (a) Generate and prepare dataset
- (b) Create Model
- (c) Predict with Model and give the commands

The following subsections describe each one of them in detail:

- (a) Generate and prepare dataset.

The system required a unique set of hand gestures, which can vary up to some extent, so it was required to capture varying images of individual gestures multiple times and create a dataset, which will be in future utilized for training and prediction. The images were captured using a regular webcam, which were later resized, converted from RGB to YCbCr colour and marked pixels of hand with white colour and background with black colour. The dataset consists of 1500 plus images of hand gestures (Fig. 2).

Hand gestures include left, right, palm, fist and victory. Figures 3, 4, 5, 6, 7 depict them, respectively.

- (b) Create Model.

It consists of an NN structure with one hidden layer. The size of input layer is nearly 370 k nodes since the size of images is 640 * 576 pixels. The size of the hidden layer is 25 nodes and output will be five nodes (5 types of sign). Figure 8 represents the model. The process includes loading features and labels and randomly initializing NN nodes weights. Then creating cost function, forward propagation and gradient for NN cost function.

- (c) Predict with Model and give the commands.

This part involves testing and predicting whether the model is able to recognize the hand gesture in real-time scenarios and able to predict the hand gestures accurately.



Figs. 2, 3, 4 (2) Right. (3) Palm. (4) Fist

Figs. 5 and 6 (5) Left. (6) Victory



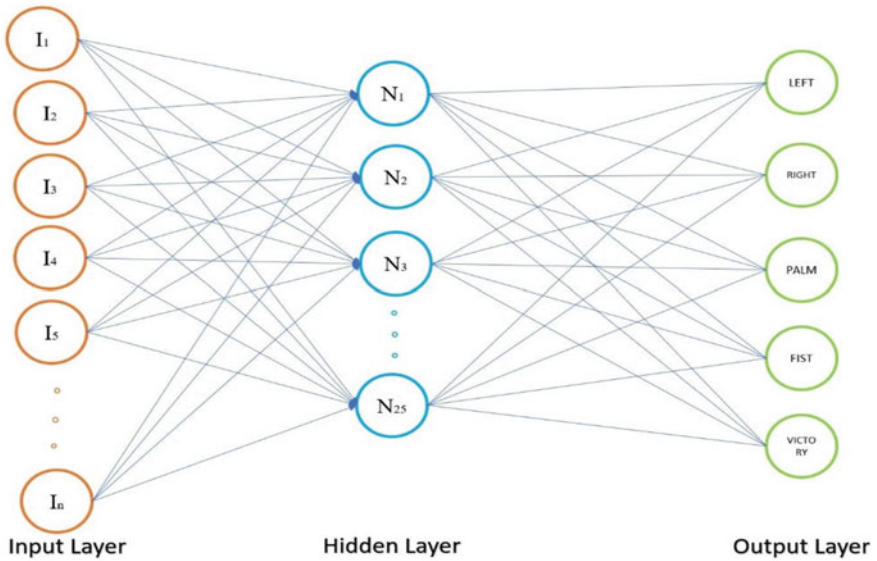


Fig. 7 Neural network model

Once the prediction is done the model should be calling an assigned method in order to control the music player. Here it's

- First to pause the song
- Palm to unpause the song
- Left to play previous song
- Right to play next song
- Victory to increase/decrease the volume.

3.2 Other Content Features

Since the mirror will also be displaying date, time, weather forecast, reminders and news feed this content must be delivered via cloud to the application. Date and time are fetched from the Raspbian OS itself while the weather and news are fetched from the cloud. Apart from fetching this content, a request is sent to the cloud on a regular interval of time for updating the content. The reminders are synced with the calendar of the user's smartphone using google API. Figure 8 shows the overall working architecture of the system.

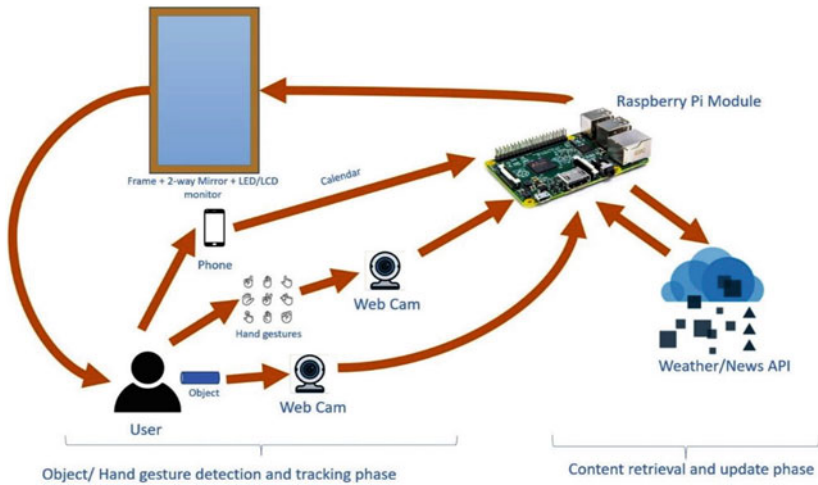


Fig. 8 Working of the next generation mirror

4 Result

With all the components namely Raspberry Pi Module, LCD/LED monitor, two webcams and speakers running all the features of the mirror namely note-taking feature, hand gesture-controlled music player, daily reminders, news feeds, date, time and weather forecast are running perfectly fine. The note-taking features all the components: write, erase and clear all enables its user to use the mirror surface as a huge notepad upon which a handwritten message can be left which is to be conveyed to others. The algorithm for object detection and tracking which are the backbone for note-taking features are working with 100% accuracy. The neural network designed for the hand gesture-controlled music player is working with around 90 to 100% accuracy in predicting real-time hand gestures with help of previously created dataset. Figure 9 shows prediction rate and feature extraction of palm gesture. Once a given hand gesture is predicted an appropriate method is called which in turn governs

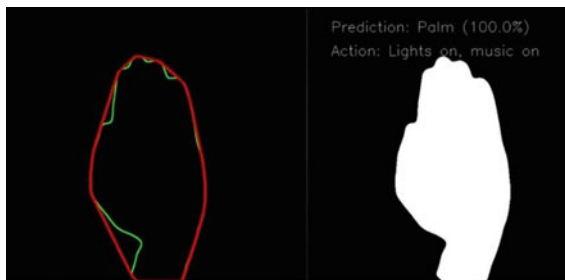


Fig. 9 Palm hand gesture prediction and action

the music player. Apart from this all the basic and common information like daily reminders, date, time, news feed and weather forecast are correct and most recent information. The news feeds are updated every 15 min while the weather forecast is updated every 4 h.

As one can see, the mirror serves as a huge freehand notepad, upon which a message is written. Apart from that, all the three software buttons write, clear all and erase can also be seen. The news feed is present at the bottom while date, time, weather forecast and remainder are present on the top-most region of the mirror.

5 Conclusion and Future Enhancements

This next generation mirror is loaded with features, which are aimed at increasing the utility of the conventional mirror. Apart from that, it also makes it possible to introduce mirrors into the proliferating network of IoT. Since a mirror is invariably, either the best object to catch attention intentionally or accidentally the mirror will be serving as an excellent notepad upon which the user can leave hand-written messages, which are to be conveyed to others or to the person himself as a to-do work. In order to do so all a person needs to do is to hold a pen and start writing in the air, right in front of the mirror without actually touching the mirror surface itself. The camera placed above the mirror will be tracking the movement of the pen and gives the message as an output. This will also keep the mirror free from smudges and fingerprints. The messages once written could also be erased or be entirely wiped off so that the user can write new messages. Keeping a hand gesture-controlled music player not only enhances multi-tasking ability of the user but it also seamlessly integrates the music player into the mirror hence increasing its utility as well. In addition to this, the features of the mirror also focus on effectively utilizing the short span of time people spend in front of the mirror during morning rush hours by displaying vital information in a very precise and crisp format. This information includes date, time, current weather, news headlines and calendar reminders. The user will be updated with all this information in a short span of time.

The future enhancements include introduction of another feature to the mirror, which is voice control. This voice control feature will be governing a certain new set of features such as maps that can be displayed on the mirror, smartphone notifications or playing a video itself. The proposed system either can be achieved with help of third party software like Google Assistant or Amazon Alexa or can develop a less intelligent voice recognition system, which can follow only a certain set of commands. Apart from that, some iterative upgradation can include improvement in prediction rate of hand gesture recognition system.

References

1. Cumeras I Khan, J. (2016). I built a smart mirror with voice and gesture control and a full OS to go with it. <https://imgur.com/gallery/j0BFU>.
2. Adafruit. <https://learn.adafruit.com/android-smart-home-mirror/overview>.
3. Thaware R., Patil, G., & Joshi, S. (eInfochips). Designing an effective traffic management system through vehicle classification and counting techniques. <https://www.design-reuse.com/articles/45224/traffic-management-system-through-vehicle-classification-and-counting-techniques.html>.
4. Dey, N., Hassanien, A. E., Bhatt, C., et al. (2018). *Internet of things and big data analytics toward next-generation intelligence*. Cham: Springer International Publishing.
5. Bhayani, M., Patel, M., & Bhatt, C. (2016). Internet of things (IoT): In a way of smart world. In: S. Satapathy, Y. Bhatt, A. Joshi, D. Mishra (Eds.), *Proceedings of the International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing* (Vol. 438). Singapore: Springer. https://doi.org/10.1007/978-981-10-0767-5_37.
6. Kumari, S., Mathesul, S., Shrivastav, P., & Rambhad, A. (2020). Hand gesture-based recognition for interactive human computer using tensor-flow. *International Journal of Advanced Science and Technology*, 29(7), 14186–14197.

Cluster Formation Algorithm in WSNs to Optimize the Energy Consumption Using Self-Organizing Map



Padmalaya Nayak, GK. Swetha, Priyanka Kaushal, and D. G. Padhan

Abstract Wireless Sensor Networks (WSNs) are considered as one of the most prevailing technologies in today's world due to the diversified applications. These applications are huge in the range such as environmental monitoring, health care, civil and military, disaster management to other surveillance systems. Minimization of energy is one of the most exciting tasks in WSNs as small sensor nodes are battery powered and deploy in remote environments. Clustering is one such imperative technique that can conserve energy more broadly, and evenly which makes the network operational for a longer period. This research paper aims at developing an energy-aware cluster-based routing protocol using Artificial Neural Network (ANN) which finds an optimal number of clusters and rotates the cluster head periodically to balance the energy consumption throughout the network. Typically, the proposed algorithm is developed based on Self-Organizing Map (SOM) to form the clusters, and the K-means algorithm is used to form different sizes of clusters. Finally, an optimal number of clusters is found that impacts the network load and balances the energy consumption. MATLAB is used as a simulation tool for experimental analysis. Simulation results prove that the proposed cluster-based routing protocol "LEACH-SOM" dominates the traditional LEACH routing protocol in terms of minimal energy consumption and makes the network active for long period.

Keywords WSN · SOM · K-means · LEACH

P. Nayak · GK. Swetha (✉) · D. G. Padhan
Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

P. Nayak
e-mail: padmalaya@griet.ac.in

P. Kaushal
Chandigarh Engineering College, Landran, Mohali, India

1 Introduction

Wireless Sensor Networks (WSNs) are considered as an active emerge area because of vital role in numerous applications. These sensor nodes are battery operated and these batteries are required to recharge and replace very often. So, minimal energy consumption of sensor nodes is the most challenging task. Many existing routing protocols discussed in the literature are energy efficient [1–5] rather than simply routing data from source to destination. Majorly of the routing protocols are classified into flat-based, hierarchical-based, query-based, location-based, etc. The hierarchical routing protocols are the most energy-efficient routing protocols [1–5] because they accommodate large-scale sensor nodes by reducing redundant data transmissions and thus, reduce the energy consumption. As our focus on improving the performance of hierarchical routing protocols, we have considered LEACH protocol as the reference protocol and incorporated machine learning techniques in it. Machine Learning (ML) is the subcategory of Artificial Intelligence (AI), which can learn by itself and improve the system performance without being programmed explicitly. ML techniques have been used for many applications in WSNs. Few applications of ML in WSNs are discussed below [8].

- (1) *Scalability*: The routing protocols in WSNs must be scalable because the quantity of sensor nodes varies based on the applications to handle the dynamics of the events. With the help of ML techniques, routing protocols are proved to be scalable.
- (2) *Energy Efficiency*: Energy minimization of sensor nodes is an important task in WSNs because small sensor nodes are battery operated. Even though the nodes are energy harvested nodes, it is very important to know when to harvest. The machine learning techniques help to handle the energy conservation issues in WSNs efficiently.
- (3) *Faulty tolerance*: The operation of WSN should not be disturbed by faulty or failed sensor nodes. ML techniques help to find the faulty nodes in WSNs.
- (4) *Localization*: In WSNs, the sensor nodes are deployed in dynamic environments like underwater, forest, cold places, etc. So, it is very important to identify the location of sensor nodes in any type of dynamic environment. Machine learning techniques help to solve localization issues in WSNs.
- (5) *Quality of service*: In many applications, the quality of service is very important that implies the data should be delivered on time without any noise. The machine learning techniques handle efficiently to achieve the quality of service in WSNs.
- (6) *Routing*: Routing protocols in WSNs not only handles routing but also handles many problems like energy conservations, faulty nodes, heterogeneity of sensor nodes, etc. So, the development of routing protocols is a very important task in WSNs. Machine learning techniques help to make routing protocols energy efficient and to handle these issues easily.

The remaining section of the paper is partitioned as follows. Section 2 discusses the current literature that contains both the traditional clustered-based routing protocols as well as SOM-based routing protocols in WSNs. Section 3 presents the proposed work followed by the experimental results discussed in Sect. 4. Section 5 concludes the paper.

2 Existing Literature

This section gives an overview of traditional hierarchical routing protocols and few ML-based routing protocols. As our proposed work is based on Self-Organizing Map (SOM), we have discussed a few SOM-based routing protocols in WSNs.

(A) Hierarchical Routing Protocol.

The routing protocols in WSNs are categorized in various ways such as based on the network structures, routing operations, and network applications. But majorly of the routing protocols are grouped into flat-based, hierarchical-based, and location-based considering the structure of the network. Hierarchical routing protocols include partition of sensor networks into multiple groups. Each group is controlled by a leader node known as a Cluster Head (CH) as presented in Fig. 1. All the non-cluster nodes that communicate with their CH and CHs are liable to accumulate the sensed data received from the cluster members and communicate with the Base Station (BS). Formation of clusters, selection of CHs, and selection of optimal path are the main challenging tasks in clustered-based routing.

In this regard, LEACH [1] is the oldest hierarchical routing protocol that operates in terms of rounds. Each round carries periodical transmission of messages. It operates in phase manner such as setup phase followed by the transmission phase. Setup phase takes care of the establishment of clusters and cluster head selection. To satisfy this requirement, it uses a probabilistic model. Transmission phase follows

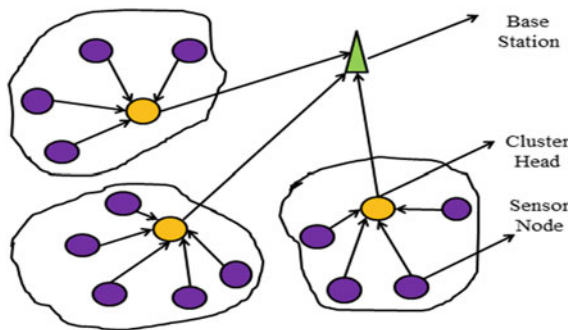


Fig. 1 Basic example of clustered-based WSN

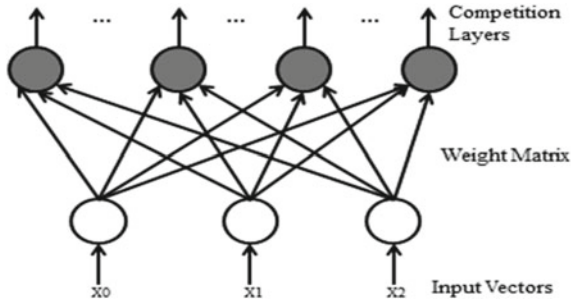


Fig. 2 Architecture of a simple self-organizing map

the setup phase by allowing non-CH nodes to transmit the sensed data to their corresponding CHs. Finally, the CHs take the responsibility to gather the data and deliver it to the Sink node. The authors of [3] have introduced a routing protocol called PEGASIS, where a chain is formed among the sensor nodes through which each sensor node passes the data to the other sensor node till it moves to sink node. TEEN [4] introduces a dynamic routing protocol in which sensor nodes continuously sense the environment and transmits the data till the sensed data exceeds the specified threshold value. APTEEN [5] is the enhanced version of TEEN protocol which is applicable to both reactive and proactive networks. HEED [6] uses a novel clustering approach for routing where the sensor nodes are non-uniformly distributed and has proved its energy efficiency in prolonging the network lifetime compared to other existing protocols.

(B) *SOM-based Routing Protocol.*

Self-Organizing Map (SOM) belongs to unsupervised machine learning technique that allows grouping the unlabeled data according to their similarities, patterns, and differences without any prior training. Unsupervised learning is also known as competitive learning or self-learning developed by Kohonen in 1990. The simple architecture of Self-Organizing Map neural network is shown in Fig. 2. The SOM consists of neurons that are structured in a low-dimensional grid, in which each neuron is connected to other neurons with a neighborhood relation. The SOM consists of the input layer, each input layer neuron is associated with some weight and activities the other neuron in the output layer or competition layer with similarity function called Euclidian distance as shown in Eq. 1. The neuron whose Euclidian distance is minimum becomes a winning neuron.

$$ED_j = \sum_{i=1}^n |Wn_{i,j} - n_i|^2 \quad (1)$$

where n_i is the i th neuron at the input layer, $Wn_{i,j}$ is the weight vector connecting input neuron i to output neuron j , and ED_j is Euclidean distance between the input neuron

and output neuron, respectively. The winning neuron weight is updated according to the neighborhood function as shown in Eq. 2, and this learning process is repeated until the weights are unchanged.

$$Wn_{i,j}^q = Wn_{i,j}^{q-1} + Nh_{i,j}x_i - Wn_{i,j}^{q-1} \quad (2)$$

where $Wn_{i,j}^q$ is new weight vector between neuron i and neuron j . $Wn_{i,j}^{q-1}$ is an old weight vector between i (input layer neuron) and j (output layer neuron). $Nh_{i,j}$ is neighborhood function and x_i is the input sample.

In [7], the authors discuss an Energy-Based Clustering Self-Organization (EBCS) map routing protocol, in which the energy level of nodes and X and Y coordinates of nodes are used as inputs to SOM to make energy balanced clusters. In [8], the author has discussed a routing protocol in which neural networks are used to develop well organized and dynamic topologies to make routing efficient by minimizing the energy utilization in WSNs. In [9], the author has presented an Artificial Neural Network (ANN)-based routing protocol to speed up the lookup table process and make routing easier and efficient in WSNs. The work in [10] discusses a protocol called Adaptive Resonance Theory (ART) neural network that processes high-dimensional complex data into simple low-dimensional data. By doing so, it makes routing efficient in WSNs. Many machine learning-based routing protocols are discussed in [11–16]. Based on this motivation, a SOM-based routing protocol is proposed where we have formed a different number of clusters and identified the optimal number of clusters that balances the energy consumption throughout the network.

3 Proposed Heuristic

The proposed heuristic is developed by incorporating the principle of SOM in WSNs. SOM operates with three inputs such as X coordinates, Y coordinates, and distance to sink node/BS, and the output of SOM is connected to the sensor field. First, the network is partitioned into clusters with the help of winning neurons, and later, the number of clusters is varied with the help of the K-means algorithm. Then, CH is selected based on the centroid of the cluster and the energy level of sensor nodes. The complete procedure is implemented in four phases.

(A) Initialization Phase

- Assume sensor nodes are deployed in a specific area. Consider it as N .

- Each sensor node is equipped with the same energy level and identical by nature.
- Each sensor node is static including the base station.

(B) *Cluster formation Phase*

The X and Y coordinates of sensor nodes, and distance from nodes to the BS are used as inputs to SOM. So that a three-dimensional I matrix is created. The values in the input matrix are normalized using the min-max operation.

$$X_{ij}^j = \frac{x_{ij} - Min_j}{Max_j - Min_j} \quad (3)$$

In Eq. 3, the X_{ij}^j denotes the normalized sensor node value of 'j' column and 'i' row. The Min_j and Max_j represents the minimum and maximum values of 'j' column vector. And the x_{ij} denotes the value of 'j' column and 'i' row of input vector-matrix 'I'. The three-dimensional input matrix 'I' is shown below in Eq. 4.

$$I = \begin{bmatrix} \frac{x_1}{x_{max}} & \frac{y_1}{y_{max}} & \frac{d_1}{d_{max}} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{x_n}{x_{max}} & \frac{y_n}{y_{max}} & \frac{d_n}{d_{max}} \end{bmatrix} \quad (4)$$

In the above input parameters matrix, I , the vector $(x_1 \dots x_n)$ represents X coordinate values of sensor nodes, the vector $(y_1 \dots y_n)$ represents Y coordinates values of sensor nodes, and the vector $(d_1 \dots d_n)$ represents the distance from sensor nodes to BS. From the input matrix I , the weight matrix is derived as shown below in Eq. 5.

$$WI = \begin{bmatrix} \frac{x_1}{x_{max}} & \dots & \frac{x_n}{x_{max}} \\ \frac{y_1}{y_{max}} & \dots & \frac{y_n}{y_{max}} \\ \frac{d_1}{d_{max}} & \dots & \frac{d_n}{d_{max}} \end{bmatrix} \quad (5)$$

In the above Eq. 5, WI represents the weight matrix of the input matrix I. The X and Y coordinates of sensor nodes and distance to BS are used as input parameters of SOM, fed to the sensor field to form the clusters. The sensor nodes are arranged in grid topological structure and weights of input vectors are randomly initialized. The weights of the input vector are changed according to the learning rate and neighborhood function. The output of SOM is fed to the sensor field, and K-means algorithm is employed to formulate the clusters. The K-means algorithm uses Euclidean distance to make clusters as discussed in Eq. 6.

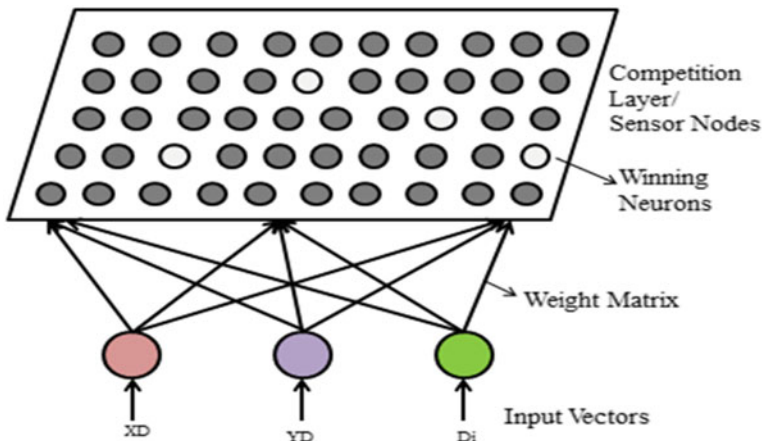


Fig. 3 Proposed WSN clustering model

$$K\text{-means} = \frac{1}{n} \sum_{k=1}^n \sum_{x \in C_k} \|x - Cen_k\|^2. \quad (6)$$

In the above equation, ‘ n ’ denotes the number of clusters, C_k identifies the K th cluster, and Cen_k notifies the centroid of cluster C_k . The proposed model is shown in Fig. 3.

(III) *Cluster Head Selection Phase*

CH is selected after partitioning the network into several clusters. The different number of clusters is formed with the help of the SOM and K-means algorithm. Initially, centroids are found for each cluster, then the CH is selected whose distance to the centroid is low and energy is high. For the entire simulation, the number of clusters remains same, but CHs are rotated based on their energy level.

(IV) *Transmission Phase*

Data transmission phase allows cluster members to transmit the data to their corresponding CHs in each cluster. Then, CHs aggregate the sensor node’s data and transmit it to the sink node/BS. For the entire simulation, the clusters remain same, but cluster heads are rotated for each cluster. The energy model is considered the same as the LEACH protocol.

4 Results and Discussion

MATLAB is used to setup the experiments and our findings are graphically characterized from Figs. 4, 5, 6, 7, 8 and 9. Both the LEACH single hop (LEACH-SH) and LEACH multi-hop (LEACH-MH) are taken as the reference protocols to evaluate

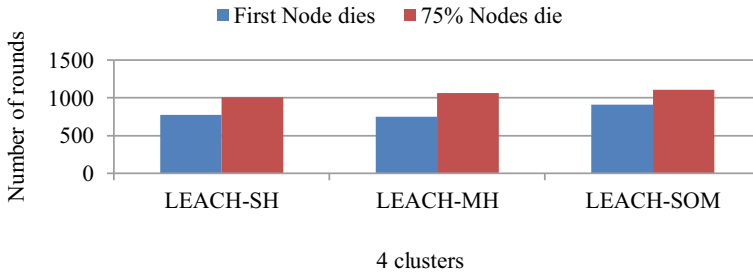


Fig. 4 FND & 75% nodes die w.r.t rounds

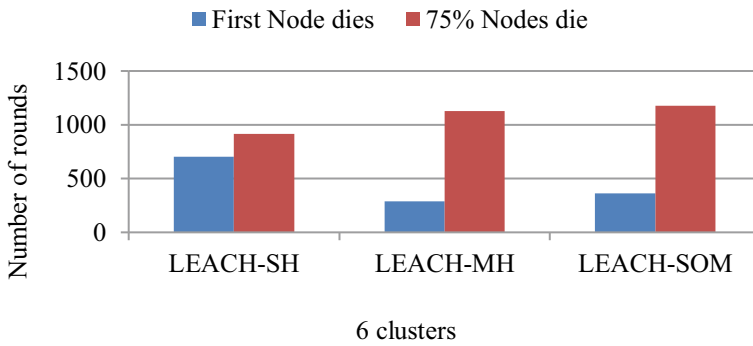


Fig. 5 FND & 75% nodes die w.r.t rounds

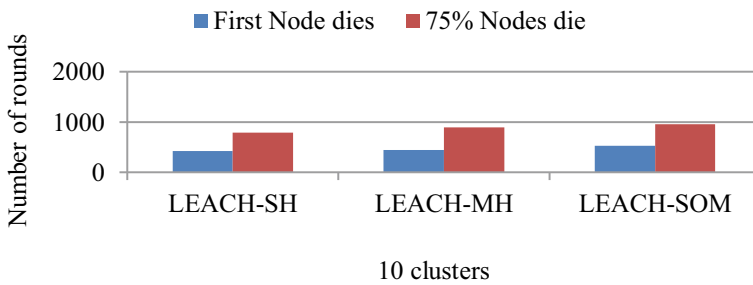


Fig. 6 FND & 75% nodes die w.r.t rounds

the performance of the intended protocol. The simulation parameters are discussed in Table 1. We consider an area of 100*100 m, where 100 sensor nodes are deployed randomly. The base station is placed at 50*200 m. In this approach, the k value is varied to 4, 6, and 10 to form 4, 6, and 10 clusters, respectively. Our experimental analysis concludes that 4 clusters are the optimal number of clusters at which the network performs very well compared to 6 and 10 clusters. The performance metrics like energy consumption, packet delivery ratio, and network lifetime are used to

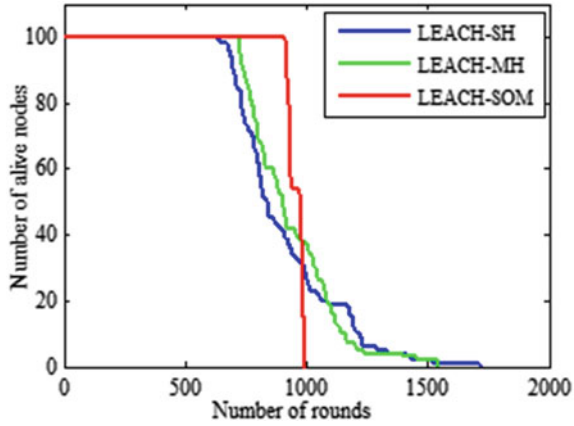


Fig. 7 Network lifetime in terms of alive nodes

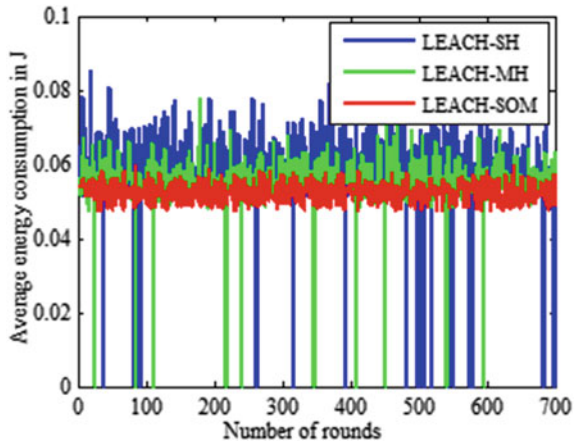


Fig. 8 Average energy consumption versus rounds

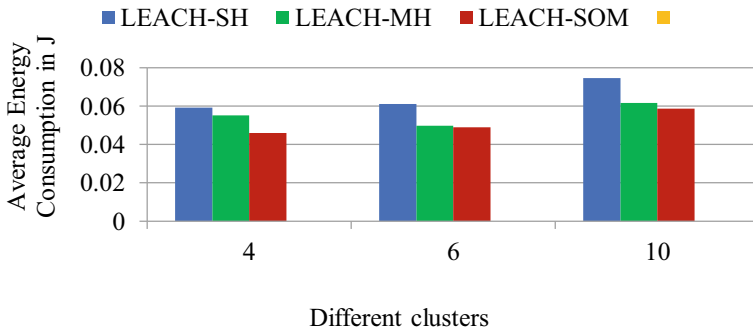


Fig. 9 Average energy consumption versus different size clusters

measure the performance of a network. The figures from 4 to 6 show the time, at what time the First Node Dies (FND) and 75% of nodes die. Figure 4 shows that the first node dies at 896 rounds in LEACH-SOM, whereas First node dies at 664 rounds in LEACH-SH, and it survives up to 638 rounds in LEACH-MH. Usually, the First Node Dies (FND) are the most used performance metric to measure the lifetime of a network. The 75% of nodes are also still alive up to 900 rounds in LEACH-SOM and it is far better compared to LEACH-SH and LEACH-MH. A comparison table about network lifetime in terms of the first node dies, 75% node dies, and the last node dies as shown below in Table 2. Figure 7 is generated through MATLAB which represents the lifetime of the network in terms of the number of alive nodes. We have also generated a MATLAB graph representing average energy consumption per round considering all the three cluster sizes as shown in Fig. 8. The graph shows that the average energy consumption of nodes per round is less in LEACH-SOM related to LEACH-SH and LEACH-MH protocols. Figure 9 depicts the average energy consumption of the network for all sizes of clusters such as four, six, and ten. At cluster size 4, the energy consumption of nodes is less and balanced throughout network compared to other size clusters. So, it is assumed that when the cluster size is large, load is balanced throughout the network.

5 Conclusion

This research work addresses a novel routing protocol based on clustering called “LEACH-SOM” in which effective cluster formation takes place and an accurate number of clusters are formed that balances the energy consumption throughout the network. By doing so, it maintains a healthy network lifetime. The Self-Organizing Map and the K-means algorithm are used to develop this novel cluster-based routing protocol. SOM is applied to divide the network into clusters, and the K-means algorithm is used to vary the size of the clusters. The parameters like X coordinates, Y coordinates, and distance to BS of sensor nodes are fed to the input of the SOM, and the output of SOM is fed to the sensor field. K-means algorithm is applied to the sensor field to vary the number of clusters. This proposed protocol is examined and simulated through MATLAB. The traditional cluster-based routing protocol LEACH is used as a reference to compare the efficiency of the proposed protocol. After extensive simulations, it is concluded that the proposed protocol “LEACH-SOM” efficiently balances the energy consumption and brings a stable lifetime of Wireless Sensor Network as compared to the LEACH protocol.

Table 1 Simulation Parameters

Sensor nodes	100
Simulation terrain	100 × 100 m
Sink or BS	50 × 200 m
$d_{\text{crossover}}$	87 m
Initial energy of sensor nodes	0.5j
E_{elec}	50nj/bit
E_{mp}	0.0013pj/bit/m ⁴
E_{fs}	10pj/bit/m ²
Data packet length	500 bytes
EDA	5nj/bit/signal
Speed of transmission	1 Mbps
Channel representation	Wireless (both directions)
Simulation period	1001 s

Table 2 Time of dead nodes w.r.t 4 clusters

Protocol name	Time in terms of no. of rounds		
	1st node dies	75% nodes die	Last node dies
LEACH-SH	664	836	1801
LEACH-MH	638	888	1533
LEACH-SOM	896	990	1010

References

1. Heinzelman, W. R., Chandrakasan, A., & Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. *IEEE Computer Society Proceedings of the Thirty Third Hawaii International Conference on System Sciences (HICSS '00)*. IEEE, Washington, DC, 2000 (pp. 1–10)
2. Heinzelman, W. R., Chandrakasan, A., Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1(4), 660–670.
3. Lindsey, S., & Raghavendra, C. S. (2002). PEGASIS: Power efficient gathering in sensor information systems. *IEEE Aerospace Conference Proceedings*
4. Manjeshwar, A., & Agrawal, D. P. (2001). TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. *Proceedings of the 15th International Parallel & Distributed Processing Symposium (IPDPS-01)*, San Francisco, CA, April 23–27.
5. Manjeshwar, A., & Agrawal, D. P. (2002). APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. *16th International Parallel and Distributed Processing Symposium (IPDPS 2002)*.
6. Priyadarshi, R. et al. (2018). A novel HEED protocol for wireless sensor networks. *5th International Conference on Signal Processing and Integrated Networks (SPIN)*.
7. Enami, N., Moghadam, R. A., & Ahmadi, K. D. (2010). A new neural network based energy efficient clustering protocol for wireless sensor networks. *IEEE 5th International Conference on Computer Sciences and Convergence Information Technology*.

8. Patra, C., Bhaumik, P., Chattopadhyay, M., & Roy, A. G. (2011). Using self organizing map in wireless sensor network for designing energy-efficient topologies. *IEEE 2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE)*.
9. Turčaník, M. (2012) Network routing by artificial neural network. *IEEE 2012 Military Communications and Information Systems Conference (MCC)*.
10. Mittal, M., & Kumar, K. (2012) Network lifetime enhancement of homogeneous sensor network using ART1 neural network. *IEEE Sixth International Conference on Computational Intelligence and Communication Networks*.
11. Li, L., & Liu, D. (2016). An energy aware distributed clustering routing protocol for energy harvesting wireless sensor networks. *2016 IEEE/CIC International Conference on Communications in China (ICCC)*.
12. Zhao, X., Wei, Z., Cong, Y., Yin, B. (2018). A balances energy consumption clustering routing protocol for a wireless sensor network. *IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC 2018)*.
13. Sanhaji, F., Satori, H., Satori, K. (2019). Cluster head selection based on neural networks in wireless sensor networks. *IEEE 2019 International Conference on Wireless Technologies Embedded and Intelligent Systems (WITs)*.
14. Nayak, P., & Anurag, D. (2016). Distributed fuzzy logic based routing protocol for wireless sensor networks. *IEEE Sensor Journal*, 16(1), 137–144.
15. Nayak, P., & Bhavani, V. (2017). Energy efficient clustering algorithm for multi-hop wireless sensor network using type-2 fuzzy logic. *IEEE Sensor Journal*, 17(14), 4492–4499.
16. Swami, T. J., Ramamurthy, G., & Nayak, P. (2019). Optimal, secure cluster head placement through source coding techniques in wireless sensor networks. *IEEE Communication Letters*, 24(2), 443–446.

CNN-Based Mobile Device Detection Using Still Images



Surbhi Gupta, Neela Pravalika, Padmalaya Nayak, and Jaafar Al Ghazo

Abstract Image forgery has been increased enormously due to the development of software and introduction of new cheaper digital devices like cameras and cellular phones. The introduction of new devices leads to inadvertent capturing of images. A lot of misuse of these devices led to the need for their identification. Source camera identification deals with the issue of identification of the mobile device using their camera through which image has been taken. This empowers the scientific agent to spot mobile device that has been used for capturing the specific image during the study. This is significant as this alphanumeric content is considered as an inaudible observer. In this proposed work, we are reviewing the various approaches for SCI which depend on classical machine learning algorithms feature extraction. Then, a CNN model is proposed for identification of mobile device using vision dataset. High accuracy of 92% is achieved for vision dataset.

Keywords Source camera identification · Convolutional neural network · Sensor fingerprint · Feature extraction · Mobile devices · Classification

1 Introduction

Digital images can be captured through digital cameras or mobile devices. The authentication of questioned smartphone images is very important [20]. Image forgery has been increased rapidly with the growth of digital technologies. Now the pictures can be easily captured without any constraints of network, location, time, etc. Lately, a large amount of multimedia data is produced with various distinct electronic gadgets and exchanged through the internet [7]. This also led to enforcing legal restrictions of capturing the pictures which include places like schools, government places [9]. The two main important domains in multimedia forensics are forgery

S. Gupta · N. Pravalika · P. Nayak

Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India

J. Al Ghazo (✉)

Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

detection and camera documentation [32]. Forgery detection has gained huge attention, and it can be achieved by the identification of source camera. It has many useful applications in different areas like a criminal investigation, the judicial system, military image applications, and news media, etc. Digital images that are captured through cameras, mobile phones, tablets, etc., can be used for illegitimate purposes. A digital image may act as a “proof of occurrence” of an incident [16]. Many times, after a crime, a forensic officer is required to authenticate the images which may be considered as proof. For this purpose, the image source is first identified. The fixing of the image source gives a clue about whether the content of images has tampered or not. Source camera identification and its usage have become very prominent in the present time [33]. It is the process of identification of the mobile device which is used to capture the image. It is considered as evidence in legal terminology. It emphasizes the relationship between mobile cameras and pictures. The foremost aim of this paper is to conduct an assessment of SCI and then propose a better model.

2 Techniques for Source Camera Identification

The main practices to tackle the problem of identification of source cameras for mobile devices are using image headers, sensor fingerprint extraction, and classification, deep learning, etc. Gupta et al. [13] have presented a complete survey of source camera attribution for camera devices. In this paper, we aim at detecting camera device using still images.

2.1 Image Headers

Mullan et al. [24] used JPEG headers and metadata for source camera identification. Digital image metadata is very important to maintain digital image repositories [28]. The image header information includes metadata like parameterization and Exchangeable Image File Format (EXIF) tags. The EXIF metadata recorded by the camera improves the classification accuracy [6]. In this work, source camera identification was done using metadata from an image container. The ordering of the iPhone Model through quantization matrices and EXIF directories was done and accuracy is 65.6% is obtained.

2.2 Images Features

Tsai et al. [30] presented a novel practice for the documentation of source camera for mobile devices. It uses a set of image features for finding the different features of a mobile camera. These features comprise color, quality, and image features and

characteristics of the frequency domain. Numerical image structures also show an prominent part in source camera documentation [15, 22]. The wavelet transform is studied to calculate wavelet domain statistics-based structures. These are categorized using a Support Vector Machine (SVM). They attained accuracy close to 92% for 4 camera models for two different camera brands. McKay et al. [19] stretched the identification of foundation device categories like digital scanners, phone cameras, and mobile cameras, etc.

2.3 Demosaicing Features

Image demosaicing is a process of interpolating a full resolution colored image from Color Filter Array (CFA) samples. In an initial work [19], identification of the source camera model is done blindly by using three types of demosaicing features derived from the given text picture. It is shown that the Eigen demosaicing features function performs remarkably better than other regular features. Nine portable cameras of different prototypes are differentiated reliant upon collected image chunks through Eigen feature regularization and feature reduction [8]. For the given input image, this framework first estimates the demosaicing weights then extracts the demosaicing feature and then regularizes and compresses Eigen features.

2.4 PRNU Feature-Based Approach

Photo Response Non-Uniformity (PRNU) noise is a important technique for SCI. The sensor fingerprint is an exclusive characteristic that is useful for differentiating images taken from the same type of cameras. PRNU has been proved as an effective approach for digital forensics. The drawback of the earlier PRNU-based models [25] is that they are sensitive to random noise. These are not capable of identifying any manipulation or changes in the image or picture. So, a new feature-based PRNU technique for identification of source camera for mobile devices is devised by [14] which chooses the features from the images. These are robust for performing any simple geometric variations or any manipulations on the image such as rotation and scaling of the image. By using the wavelet-based denoising method [1], PRNU noise is obtained from images. This is represented by Higher-Order Wavelet Statistics (HOWS), and this is invariable characteristics for any kind of symmetrical dissimilarities or image manipulations in the image.

2.5 *Mobile Sensor Fingerprint*

Mobile sensor fingerprints demonstrate how several sensors on a smartphone can be utilized to build good hardware pattern of the mobile device. This pattern could be utilized to identify a mobile device as they link to the websites. The accelerometer-based pattern is very fascinating as it is approachable through JavaScript processing in a web browser without the invocation of any authorisation of the user [5]. Several experiments have been conducted by using sensor fingerprint-based properties from ten thousand mobile devices. It is shown that the entropy from sensor fingerprint is solely required to recognize devices from various other devices, with less chance of collision [5].

2.6 *Convolutional Neural Networks and Image Noise Pattern*

Zuo et al. [34] introduced this technique. The dataset from Kaggle competition, named “Camera Model Identification”, is used. CNNs can involuntarily and parallelly extract features and learn to classify during the learning process [31]. It is proved that CNN-based architecture could be utilized for solving the problem of camera detection. In this work, the author proposed a CNN classifier after performing PRNU extraction. PRNU is the most known SCI method in digital forensics [23]. This model is not very good as it is a simple CNN model. Here, only two filters have been used, and the number of epochs used is only 100 for testing purposes. It is difficult to estimate the performance of the model by using a greater number of epochs or filters. In the future, more experiments are to be expected to conclude. The accuracy obtained using this method is 91%. Deep learning depends upon the input data quality for its feature extraction process [11]. Lately, Convolutional Neural Networks (CNNs) have exhibited incredible performance on various projects which include video analysis, image recognition, etc. A CNN comprises a collection of layers in which every layer is an association of different operations or tasks. Mobile camera sensors consist of a noise pattern that can be easily understood by a CNN architecture [18]. This is as well as called camera fingerprint which is used to identify the source mobile device through which image has been captured with good accuracy. The fingerprint is mainly used for two purposes, i.e., smartphone verification and profile linking [4]. PRNU is mostly used for identifying sensor fingerprint [2]. In most of the existing systems using classifier or neural networks, firstly, the features are retrieved from the given image. The dataset is subdivided into testing and training data. Then, a classifier model is obtained through training and evaluation of the model is done through testing. In this paper, we studied the various approaches that have been utilized for the identification of source camera for mobile devices. The several techniques that have been discussed for the identification of source cameras are using image headers, image features, demosiacing features, sensor fingerprint, etc. These are reviewed individually. The comparison of different methods is shown in Table 1.

Table 1 Evaluation of source camera identification techniques for mobile device

Author	Algorithm	Year	Format of image	Number of brands	Number of models	Applied to mobiles	Applied to same brand and different models	Accuracy
Tsai et al. [30]	SVM	2007	JPEG	2	4	Yes	Yes	92%
Cao and Kot [8]	Eigen Feature Regularization	2010	JPEG	4	15	Yes	Yes	99%
Akshatha et al. [1]	SVM	2016	JPEG	7	11	Yes	Yes	96.18%
Zuo et al. [34]	CNN + PRNU	2018	JPEG	3	3	Yes	No	91%
Mullan et al. [1]	JPEG Image Headers	2019	JPEG	1	7	Yes	No	81.70%
Freire-Obregón et al. [11]	CNN	2018	JPEG	2	3	Yes	Yes	98.10%

3 Dataset Used

A dataset is a very important aspect while performing any research-related experiment. There are different types of datasets that are used for source camera identification for mobile devices. Some people create their own dataset while other people prefer using existing datasets. The dataset that we have considered for the proposed model is the VISION dataset. Shullani et al. [27] introduced dataset in which images are taken from several different mobile devices, and it also consists of the images shared from WhatsApp and Facebook. It is a development to be used for mobile forensics. This dataset consists of 34,427 images and 1914 videos captured using 35 mobile devices. Some of them are shown in Fig. 1. Before starting the experiment, the dataset is split as 80% of total data for training and validation and remaining data for testing. Table 2 shows the 11 mobile devices that we have considered from the VISION dataset.

4 Proposed CNN Model and Its Architecture

Convolutional neural networks do not differ much from the other neural network. CNN could generate its own features. It consists of neurons. These are like the neurons in biology. Each neuron has its own parameters, i.e., height and width. Every neuron in the network has a specific operation to perform. These operations differ from one neuron to the other neuron. A CNN is a neural network in which given an input, then it performs a set of operations on the input. It also consists of non-linearity operation, i.e., an activation function for every computational process. At the end of

Fig. 1 Images from VISION dataset

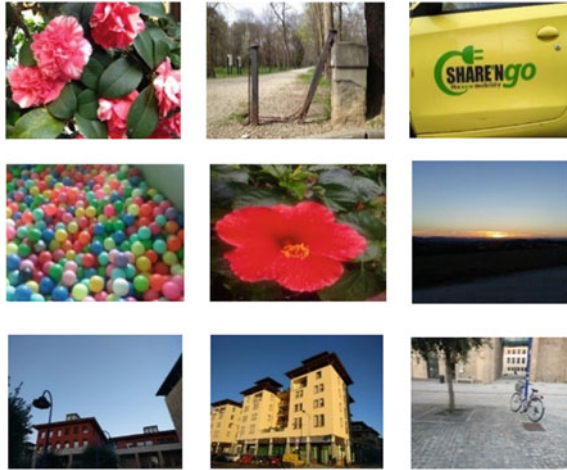


Table 2 Dataset with 11 mobile devices from VISION dataset

S. No	Name of the device	Total images
1	Samsung_GalaxyS3Mini	901
2	Apple_iPhone4S	919
3	Huawei_p9	1066
4	LG_D290	1044
5	Lenovo_P70	809
6	Sony_XperiaZ1Compact	964
7	Microsoft_Lumia640LTE	894
8	WikoRidge4g	2912
9	Asus-Zenfone 2 lazer	957
10	Xiaomi_RedmiNote3	1422
11	OnePlus_A3	1324

the process, it is feed to a loss function. It plays a prominent role in every neuron parameter apprise. CNN considers only image as an input. The idea behind the CNN is to reduce the number of parameters. The different layers that are present in CNN architecture are convolutional, pooling, and fully connected layer. The architecture for the model should be fixed before the start of training model as shown in Fig. 2.

5 Results and Discussions

The accuracy obtained for these 11 mobile device datasets is 92.39%. The accuracy graph obtained for 11 mobile device datasets is shown in Fig. 3. As we can see the

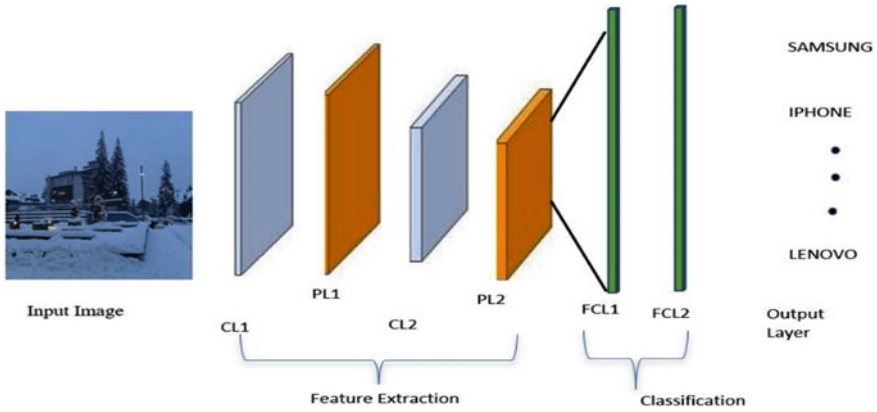


Fig. 2 Architecture for the proposed model

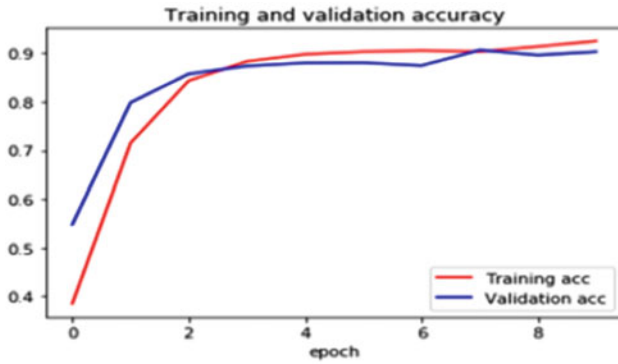


Fig. 3 Graphs obtained for 11 mobile device dataset

validation accuracy is following the trend for training accuracy and is rising after 2nd epoch itself. The experiment has been conducted only for 10 epochs. More epochs will help in improving the accuracy. The confusion matrix for 11 devices has been showed in Table 3. It is evident that almost all the models are correctly classified with 0.99 precision except ‘Asus’ and ‘Wiko’. This is the reason that the overall accuracy dropped by 8%.

Table 4 illustrates the contrast of recommended method with the present methods. Most of the existing techniques used Miche dataset. But this dataset is limited as only three different device images are included. We have used Vision dataset with 11 model. Existing shallow CNN [21] has achieved 95% but we have proposed a lighter CNN with 92.39% accuracy. Moreover, the epochs used are only 10.

Table 3 Confusion matrix for 11 mobile device dataset

	Apple	Asus	Oneplus	Huawei	Lenovo	LG	Microsoft	Redmi	Samsung	Sony	Wiko
Apple	900	0	0	4	2	0	1	4	0	0	8
Asus	0	22	0	0	0	0	0	0	0	0	935
Oneplus	0	0	1320	0	4	0	0	0	0	0	0
Huawei	0	0	1	1065	0	0	0	0	0	0	0
Lenovo	0	0	0	0	809	0	0	0	0	0	1
LG	0	0	0	7	0	1040	0	1	0	0	0
Microsoft	0	0	0	0	4	0	843	2	0	0	0
Redmi	4	0	0	0	0	0	0	1418	0	0	0
Samsung	0	0	10	2	1	1	1	1	811	3	19
Sony	2	0	0	0	0	0	1	0	0	961	0
Wiko	0	22	0	0	0	0	4	4	6	0	2876

Table 4 Contrast with present state of art

Method proposed	Dataset	Image input size	No of the CNN layers employed	No of devices experimented	Achieved accuracy
Barroffio et al. [3]	Miche-I	32×32	7	3	92.3%
Tuama et al. [31]	Miche-I	32×32	7	3	93.4%
Freire-Obregón et al. [11]	Miche-I	32×32	6	3	98.1%
Shallow CNN [21]	Vision	64×64	5	11	95.06%
Proposed shallow CNN	Vision	150×150	4	11	92.39%

6 Conclusion

In this paper, a complete survey of different approaches used for source mobile detection is analyzed. Most of the techniques for the identification of source cameras are using image headers, image features, Demosiacing features, sensor fingerprint, etc. These are reviewed individually. Although these are good at the discrimination of the mobile model, only some methods succeeded in identifying the sensor level camera. More work is required in this direction. CNN model is proposed for identification of mobile device using vision dataset. High accuracy is achieved. In future, ensemble CNN architecture-based SCI technique can be proposed for better accuracy. The same model can be explored for other computer vision applications [12, 17].

References

1. Akshatha, K. R., Karunakar, A. K., Anitha, H., Raghavendra, U., & Shetty, D. (2016). Digital camera identification using PRNU: A feature based approach. *Digital Investigation*, 19, 69–77.
2. Amerini, I., Caldelli, R., Cappellini, V., Picchioni, F., & Piva, A. (2009). Analysis of denoising filters for photo response non uniformity noise extraction in source camera identification. In *2009 16th International Conference on Digital Signal Processing* (pp. 1–7). IEEE.
3. Barroffio, L., Bondi, L., Bestagini, P., Tubaro, S. (2016). Camera identification with deep convolutional networks, *CoRR* abs/1603.01068.
4. Bertini, F., Sharma, R., Ianni, A., & Montesi, D. (2015). Smartphone verification and user profiles linking across social networks by camera fingerprinting. In *International Conference on Digital Forensics and Cyber Crime* (pp. 176–186). Springer, Cham.
5. Bojinov, H., Michalevsky, Y., Nakibly, G., & Boneh, D. (2014). Mobile device identification via sensor fingerprinting. arXiv preprint [arXiv:1408.1416](https://arxiv.org/abs/1408.1416).
6. Boutell, M., & Luo, J. (2004). Photo classification by integrating image content and camera metadata. In *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004.*, Vol. 4, pp. 901–904. IEEE.
7. Caldelli, R., Becarelli, R., & Amerini, I. (2017). Image origin classification based on social network provenance. *IEEE Transactions on Information Forensics and Security*, 12(6), 1299–1308.

8. Cao, H., & Kot, A. C. (2010, May). Mobile camera identification using demosaicing features. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (pp. 1683–1686). IEEE.
9. Corripio, J. R., González, D. A., Orozco, A. S., Villalba, L. G., Hernandez-Castro, J., & Gibson, S. J. (2013). Source smartphone identification using sensor pattern noise and wavelet transform
10. De Marsico, M., Nappi, M., Riccio, D., & Wechsler, H. (2015). Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recognition Letters*, 57, 17–23.
11. Freire-Obregón, D., Narducci, F., Barra, S., & Castrillón-Santana, M. (2019). Deep learning for source camera identification on mobile devices. *Pattern Recognition Letters*, 126, 86–91.
12. Gupta, S., & Kumar, M. (2020). Forensic document examination system using boosting and bagging methodologies. *Soft Computing*, 24(7), 5409–5426.
13. Gupta, S., Mohan, N. & Kumar, M. A. (2020). Study on source device attribution using still images. *Arch Computat Methods Eng* (2020). <https://doi.org/10.1007/s11831-020-09452-y>.
14. Darvish Morshedi Hosseini, M., & Goljan, M. (2019). Camera identification from HDR images. In *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security* (pp. 69–76).
15. Hu, Y., Li, C. T., & Zhou, C. (2010) Selecting forensic features for robust source camera identification. In *2010 International Computer Symposium (ICS2010)* (pp. 506–511). IEEE.
16. Kharrazi, M., Sencar, H. T., & Memon, N. (2004). Blind source camera identification. In *2004 International Conference on Image Processing, ICIP'04.*, Vol. 1, pp. 709–712). IEEE.
17. Kumar, M., Gupta, S., Gao, X. Z., & Singh, A. (2019). Plant species recognition using morphological features and adaptive boosting methodology. *IEEE Access*, 7, 163912–163918.
18. Lawgaly, A., Khelifi, F., & Bouridane, A. (2014). Weighted averaging-based sensor pattern noise estimation for source camera identification. In *2014 IEEE International Conference on Image Processing (ICIP)* (pp. 5357–5361). IEEE.
19. Li, X., Gunturk, B., & Zhang, L. (2008) Image demosaicing: A systematic survey. In *Visual Communications and Image Processing 2008* (Vol. 6822, p. 68221J). International Society for Optics and Photonics.
20. Liu, Q., Li, X., Chen, L., Cho, H., Cooper, P. A., Chen, Z., Sung, A. H. (2012). Identification of smartphone-image source and manipulation. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 262–271). Springer, Berlin, Heidelberg.
21. Marra, F., Gragnaniello, D., & Verdoliva, L. (2018). On the vulnerability of deep learning to adversarial attacks for camera model identification. *Signal Processing: Image Communication*, 65, 240–248.
22. McKay, C., Swaminathan, A., Gou, H., & Wu, M. (2008) Image acquisition forensics: Forensic analysis to identify imaging source. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1657–1660). IEEE.
23. Mohanty, M., Zhang, M., Asghar, M. R., & Russello, G. (2018). PANDORA: Preserving privacy in PRNU-based source camera attribution. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 1202–1207). IEEE.
24. Mullan, P., Riess, C., & Freiling, F. (2019). Forensic source identification using JPEG image headers: The case of smartphones. *Digital Investigation*, 28, S68–S76.
25. Peng, F., Shi, J., & Long, M. (2013). Comparison and analysis of the performance of PRNU extraction methods in source camera identification. *The Journal of Computer Information Systems*, 9(14), 5585–5592.
26. Shaya, O. A., Yang, P., Ni, R., Zhao, Y., & Piva, A. (2018). A new dataset for source identification of high dynamic range images. *Sensors*, 18(11), 3801.
27. Shullani, D., Fontani, M., Iuliani, M., Al Shaya, O., & Piva, A. (2017). Vision: A video and image dataset for source identification. *EURASIP Journal on Information Security*, 2017(1), 15.
28. Tesic, J. (2005). Metadata practices for consumer photos. *IEEE Multimedia*, 12(3), 86–92.

29. Tian, H., Xiao, Y., Cao, G., Zhang, Y., Xu, Z., & Zhao, Y. (2019). Daxing smartphone identification dataset. *IEEE Access*, 7, 101046–101053.
30. Tsai, M. J., Lai, C. L., & Liu, J. (2007) Camera/mobile phone source identification for digital forensics. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, Vol. 2, pp. II-221. IEEE.
31. Tuama, A., Comby, F., & Chaumont, M. (2016). Camera model identification with the use of deep convolutional neural networks. In *2016 IEEE International workshop on information forensics and security (WIFS)* (pp. 1–6). IEEE.
32. Van Lanh, T., Chong, K. S., Emmanuel, S., & Kankanhalli, M. S. (2007). A survey on digital camera image forensic methods. In *2007 IEEE international conference on multimedia and expo* (pp. 16–19). IEEE.
33. Zeng, H., & Kang, X. (2016). Fast source camera identification using content adaptive guided image filter. *Journal of forensic sciences*, 61(2), 520–526.
34. Zuo, Z. (2018). Camera model identification with convolutional neural networks and image noise pattern.

E-FFTF: An Extended Framework for Flexible Fault Tolerance in Cloud



Moin Hasan, Major Singh Goraya, and Tanya Garg

Abstract Fault tolerance provisioning is extremely important in cloud. Literature suggests that the existing frameworks are service provider centric. Being a pay-per-use model, fault tolerance in the cloud should be flexible with respect to the users' requirement. Our previously proposed framework FFTF (Flexible Fault Tolerance Framework) is further extended to E-FFTF in the present paper to provide user transparency in the selection of service category based on the task completion deadline and slack time. E-FFTF successfully establishes task execution price savings through flexible fault tolerance. E-FFTF is proved to be beneficial for both the cloud service providers and service users as per the obtained experimental results.

Keywords Cloud computing · Fault tolerance · Flexible fault tolerance · Reliability · Service price saving

1 Introduction

Fault tolerance is among the most imperative issues in cloud due to its dynamic architecture [1, 2]. Cloud follows the service-oriented architecture where computing services are provisioned on demand basis in a pay-per-use manner [3]. Therefore, provisioning fault tolerance in a flexible manner where users can avail it as per their budgetary constraints will consequently strengthen the service orientation of cloud.

In our previous research work [4], FFTF (flexible fault tolerance framework) is proposed in cloud using cooperative resource group based on the popular replication technique [5, 6]. Replicated resource groups are organized with respect to the

M. Hasan (✉)
Lovely Professional University, Phagwara, India
e-mail: moin.25676@lpu.co.in

M. S. Goraya
Sant Longowal Institute of Engineering and Technology, Sangrur, India

T. Garg
Thapar Institute of Engineering & Technology, Patiala, India
e-mail: tanya.garg@thapar.edu

quantized reliability of the cloud hosts quantized into five fuzzy sets. Flexible fault tolerance is then realized through three service categories (premium, standard, and economy) by organizing the replicated groups from these cloud hosts. Premium category services are highly fault-tolerant, suitable for the users with strict task deadlines. Standard and economy category services offer relatively lower fault tolerance. They are suitable for the users with softer task deadlines. Although, FFTF achieves flexibility in fault tolerance, still following are the important issues which needs further exploration: (a) Service selection is largely user centric where users select the service as per their budgetary constraints. In such case, it is possible that inexperienced users may choose inappropriate service category. (b) FFTF does not give an insight into price saving through flexible fault tolerance. In E-FFTF, we endeavour to address the above issues with the following contributions:

- *Developing a service selection scheme:* We propose three service classes (A, B, and C) in the decreasing order of reliability. E-FFTF selects suitable class for the users corresponding to the criticality of their tasks based on their respective deadlines. Slack time of a task is used to define the level of fault tolerance.
- *Price saving evaluation through flexible fault tolerance:* Reliability is an important parameter of fault tolerance [7]. The prominent cloud providers claim their reliability above 99 percent.¹ Taking this as benchmark, we evaluate the final price saving through flexible fault tolerance in E-FFTF.

E-FFTF is beneficial for both the provider as well as the user. Service provider can extend the low reliability resources to the tasks with low level of fault tolerance. Thus, the low reliability resources may not be instantly required to be pushed to the obsolete state. Cloud users will have the chance of price saving for the extended task deadline.

In rest of the paper, Sect. 2 describes the proposed E-FFTF explaining the system model and task execution model. Experimental evaluation is given in Sect. 3 and the paper is finally concluded in Sect. 4.

2 Proposed E-FFTF

This Section describes E-FFTF by giving an insight into the system model, fault model, and task execution model.

2.1 System Model

Cloud data center is considered as a set $DC = \{PM_1, PM_2, PM_3, \dots, PM_m\}$ of m physical machines (PMs). PM_i in the DC is modelled by its processing power

¹ <https://www.entrepriseai.news/2015/01/06/aws-rates-highest-cloud-reliability/>.

(PM_i^{pp}) evaluated in Million Instructions per Second (MIPS), working state (PM_i^{ws}) which can be failed or repaired, operational mode (PM_i^{om}) which can be active or sleep, and quantized reliability (PM_i^{qr}) . The raw reliability of PM_i for time x is calculated as a single module repairable system through Gamma distribution as

$$PM_i^{rr}(x) = \int_x^\infty \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-x\beta} dx \quad (1)$$

Reliability of PMs is quantified as Very Low ($VL = 0.1$), Slightly Low ($SL = 0.3$), Moderate ($MD = 0.5$), Slightly High ($SH = 0.7$), Very High ($VH = 0.9$) using fuzzy logic [4]. $PM_i = \{VM_{i1}, VM_{i2}, VM_{i3}, \dots, VM_{in}\}$ represents n number of virtual machines (VMs) deployed on demand over PM_i . VM_{ij} is modelled by its processing power (VM_{ij}^{pp}) evaluated in MIPS and quantized reliability VM_{ij}^{qr} considered equivalent to PM_i^{qr} .

PMs observe failures due to various faults in cloud. Multiple PMs may fail simultaneously, but independently. Failure of VMs are considered totally dependent on their respective parent PMs . All the VMs deployed over PM_i will be automatically undeployed at the event PM_i fails in the data center [8].

2.2 Task Execution Model

A finite set of p independent tasks $T = \{t_1, t_2, t_3, \dots, t_p\}$ is considered for execution [4]. Tasks are broadly categorized as delay-sensitive and delay-insensitive. Delay-sensitive tasks are high priority tasks with some specific completion deadline and require a strong fault tolerance mechanism. A delay-sensitive task t_k is modelled by its arrival time (t_k^{at}), task size (t_k^{ts}) in Million Instructions (MI), deadline (t_k^{dl}), and service class (t_k^{sc}). Delay-insensitive are the low priority tasks with no specific deadline. These tasks improve the resource utilization of the system. A delay-insensitive task t_k is modelled by its arrival time (t_k^{at}) and task size (t_k^{ts}). If t_k is allocated for execution to VM_{ij} , its execution time t_k^{et} and expected finish time t_k^{ft} are calculated as

$$t_k^{et} = \frac{t_k^{ts}}{VM_{ij}^{pp}} \quad (2)$$

$$t_k^{ft} = t_k^{st} + t_k^{et} \quad (3)$$

where, t_k^{st} is the starting time of the task execution. If the deadline of a delay-sensitive task falls beyond its expected finish time based on the system performance, then its slack time (t_k^{sl}) is evaluated as

Table 1 Service Classes and their reliability

Service Class	Reliability
Class A	0.962
Class B	0.814
Class C	0.624

$$t_k^{sl} = t_k^{dl} - t_k^{ft} \quad (4)$$

Slack time is an important parameter while selecting the service. Large value of slack time corresponds to softer deadline where the task can be executed in a low reliability service class resulting in the price saving.

To execute a delay-sensitive task t_k , replication group $G_k = \{VM_{i_1j_1}, VM_{i_2j_2}, VM_{i_3j_3}\}$, consisting of three VMs deployed over different PMs , is organized in the data center based on the quantified reliability of VMs . Among quantified reliability values, three VMs can be selected in thirty-five ways [9]. Each replication group is considered as a multi-module repairable system where the reliability is calculated as

$$Rel(G_k) = 1 - \prod_{y=1}^3 \{1 - VM_{i_yj_y}^{qr}\} \quad (5)$$

Thirty-five combinations are clustered further into classes (A, B, and C) using k-means clustering [10]. Class A represents high, Class B represents moderate, and Class C represents relatively lower fault tolerance. Table 1 enlists the three service classes. Reliability lesser than 0.5 is neglected while realizing the service classes.

Corresponding to the slack time, group G_k is organized for t_k in a suitable service category as follows:

- Let t_k^{sl} be the slack time of task t_k , then the execution time in terms of its slack time and deadline will be given as $t_k^{et} = t_k^{dl} - t_k^{sl}$.
- For each service category (in the increasing order of reliability), if $(t_k^{dl} \times service\ reliability > t_k^{et})$, organize the G_k in the current service category.

For a task t_k , the price of accessing a service is calculated as follows: Let the price of provisioning 250MIPS for 100seconds with reliability approaching to 1 be the base price equal to 1unit. Then, to execute t_k for t_k^{dl} seconds on VM_{ij} in G_k , the execution price t_k^{ep} is evaluated as

$$t_k^{ep} = (VM_{ij}^{pp} \times t_k^{dl} \times Rel(G_k)) / 25000 \quad (6)$$

After the group organization, delay-sensitive task execution is started on the maximum reliability VM (nominated as primary) in the group. The remaining VMs in the group are nominated as backups.

On the other hand, to execute a delay-insensitive task t_k , an idle backup VM in the existing replication groups is searched to initiate the task. Delay-insensitive tasks' execution on the backup VMs considerably increases the resource utilization of the replication groups.

Tasks' execution states are periodically checkpointed, so that minimum rollback would be required in the event of a failure. Let the time interval between two successive checkpoint operations be denoted by Int_{CP} , then the updating in the task size t_k^{ts} of the task t_k allocated to VM_{ij} after a checkpoint interval is given as follows:

$$t_k^{ts_upd} = t_k^{ts_prev} - (VM_{ij}^{pp} \times Int_{CP}) \quad (7)$$

where $t_k^{ts_upd}$ and $t_k^{ts_prev}$ represent updated task size and previous task size, respectively. Due to the failure of a PM in the data center, its deployed VMs will also fail. If the failed VM is primary, the delay-sensitive task is migrated to one of the two backup VMs in the decreasing order of their reliability. In case, the backup VM selected for the task migration is already executing a delay-insensitive task, its task is suspended to make the backup VM available for the delay-sensitive task. A new idle backup VM is searched for the suspended task. Task migration subjected to some delay depends upon utilized memory, dirty memory page production rate, and transmission rate, etc. [11]. In E-FFTF, task migration delay $t_k^{del_mig}$ is computed using pre-copy migration as

$$t_k^{del_mig} = \frac{mem_{util}}{rate_{trn}} \left(\frac{1 - R^{num_{itr}+1}}{1 - R} \right) \quad (8)$$

where mem_{util} is the amount of utilized memory, $rate_{trn}$ is the rate of transmission, R is the ratio of dirty memory page production rate to transmission rate, and num_{itr} is the number of iterations required for migration. Considering the $t_k^{del_mig}$, t_k^{ft} is modelled as

$$t_k^{ft} = t_k^{ft} + t_k^{del_mig} \quad (9)$$

3 Experimental Evaluation

This Section evaluates E-FFTF through simulation experiments in terms of its fault tolerance capability, resource utilization, and price savings on both artificial and real workload. The experimental setup and simulation results are discussed as follows.

3.1 Simulation Settings

CloudSim [12] is used for the system performance evaluation with following simulation settings:

- Cloud environment consists of 10 data centers with 10000PMs each.
- PM failure is modelled on Gamma distribution [13] keeping $\alpha = 0.29$ and $\beta = 119.60$ like in [8].
- Eight types of VMs are considered with processing power as 250 MIPS, 500 MIPS,, 2000 MIPS.
- Time to activate a PM and deploy a VM is equal to 90 and 15 seconds, respectively [14].
- 50000, 100000, and 150000 independent delay-sensitive tasks are considered for execution.
- Google Workload Tracelog [15] is used to calculate the task size for real workload as

$$t_k^{ts} = (time_{finish} - time_{schedule}) \times u_{avg} \times p_{CPU} \quad (10)$$

Where $time_{schedule}$, $time_{finish}$, and u_{avg} represent the schedule time, finish time, and average usage of CPU for the task, respectively. p_{CPU} (1500 MIPS [15]) is the CPU processing power.

- Task size for artificial workload is generated through uniform distribution across real Workload.
- For delay-sensitive tasks, deadlines are estimated corresponding to the VMs' average processing power. For delay-insensitive tasks, no deadline of completion is considered.
- Task arrival in artificial workload is modelled through Poisson distribution keeping average inter-arrival time as 2, 4, or 6 seconds. Task arrival in real workload is considered as per the Google Tracelog.
- Checkpoint interval equal to 60 minutes, 90 minutes, or 120 minutes are considered to periodically save the execution states of running tasks.

To establish the fault tolerance capability, we evaluate its task deadline guarantee, and the resource utilization is realized through the group utilization index.

- *Task Deadline Guarantee (TDG)*: TDG is calculated as the percentage of delay-sensitive tasks successfully finished within their respective specified deadline.

$$TDG = \left(\frac{\text{Number of delay - sensitive tasks meet their deadline}}{\text{Total number of delay - sensitive tasks}} \times 100 \right) \% \quad (11)$$

- *Group Utilization Index (GUI)*: GUI is calculated as the percentage of average time for which a replication group executes the assigned tasks (delay-sensitive + delay-insensitive) to the average time for which a replication group exists in the

data center.

$$GUI = \left(\frac{\text{Average execution time of CRG}}{\text{Average life time of CRG}} \times 100 \right) \% \quad (12)$$

- *Average Execution Price (AEP)*: AEP measures the average price paid by the users for the task execution under a specific service class. For example, the AEP for Class A tasks is calculated as follows:

$$AEP_{ClassA} = \frac{\sum_{l=1}^{No. \text{ of Class A tasks}} \{t_l^{ep}\}}{No. \text{ of Class A tasks}} \quad (13)$$

3.2 Results and Discussion

Experiments are conducted to obtain the results in terms of TDG and GUI by altering one of the three parameters, viz. task count, task inter-arrival time, and checkpoint interval. Performance of E-FFTF is compared with FFTF in terms of fault tolerance (TDG) and resource utilization (GUI). Figures 1 and 2 compare the performance of E-FFTF and FFTF for the parameter TDG on artificial and real workload, respectively.

TDG of E-FFTF is comparatively higher than that of FFTF because service selection in E-FFTF is performed by the system based on the task slack time, whereas in FFTF, it is done by the user itself. Therefore, in FFTF a user may select a low fault tolerance service for price saving but end up in the task rejection due to low performance. Corresponding to the alteration in task count, TDG being above 95% proves its stability in varying situations. Along increasing checkpoint interval, TDG slightly decreases because at higher intervals, long rollbacks are required in the event of failures and the probability of missed deadlines increases. Along increasing task

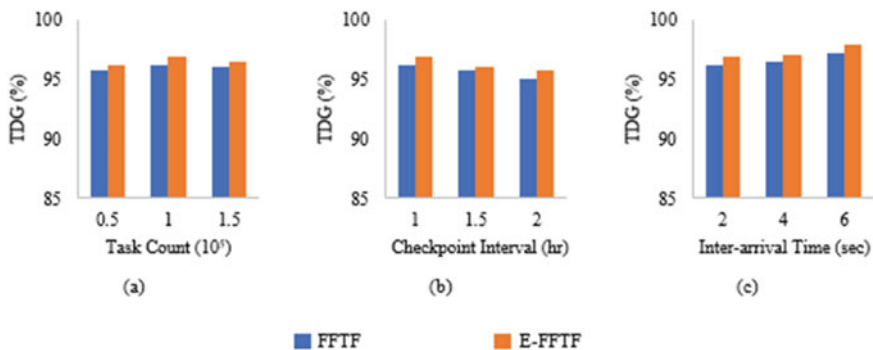


Fig. 1 TDG comparison on artificial workload w.r.t **a** task count, **b** checkpoint interval, and **c** inter-arrival

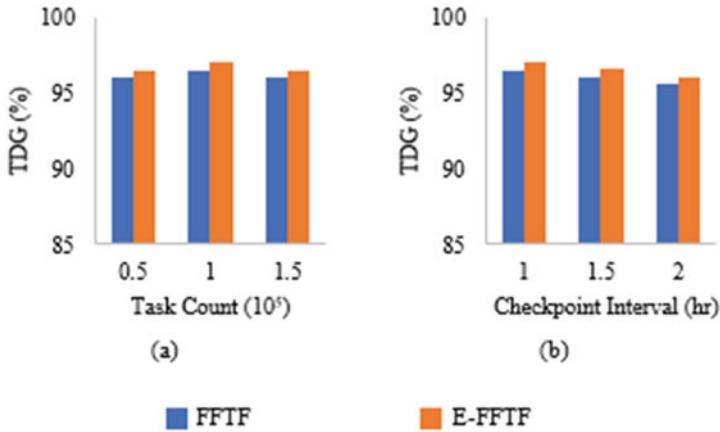


Fig. 2 TDG comparison on real workload w.r.t **a** task count and **b** checkpoint interval

inter-arrival time, TDG also increases because the system gets enough time to make the resources available.

Figures 3 and 4 compare the performance for GUI. In terms of resource utilization, both the frameworks achieve similar performance. As shown in figures, GUI increases along increasing checkpoint interval. It is explainable because at higher interval, long rollbacks are required which consequently increases the task execution time. On real workload, GUI shows little fluctuations which occur due to the large variation in the task size.

Figure 5 gives the AEP of E-FFTF for different service classes with respect to task count. AEP decreases along the decreasing reliability of service and hence justifies the price flexibility of the framework. Since, AEP is calculated considering the deadline of the task and reliability of the executing resources, therefore, in this

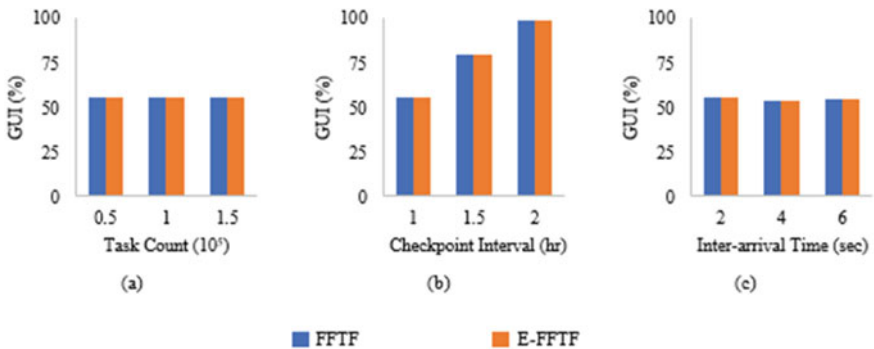


Fig. 3 GUI comparison on artificial workload w.r.t **a** task count, **b** checkpoint interval, and **c** inter-arrival time

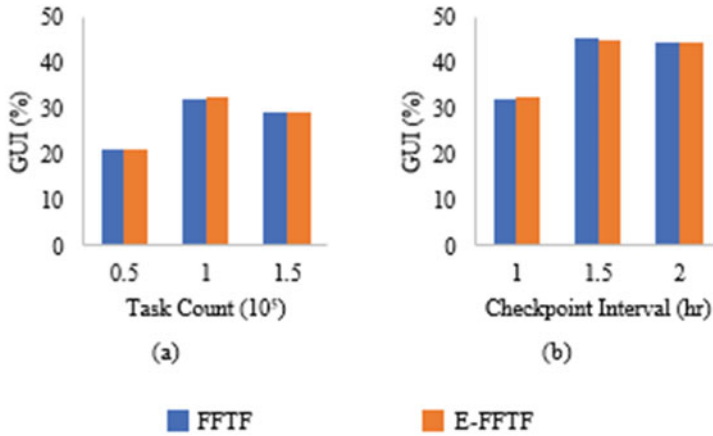


Fig. 4 GUI comparison on real workload w.r.t a task count and b checkpoint interval

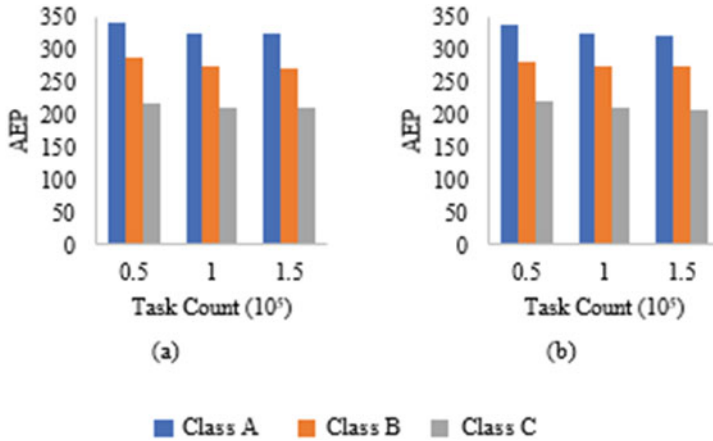


Fig. 5 AEP for different service classes on a artificial and b real workload

experiment, inter-arrival time and checkpoint interval are not varied as AEP is not dependent on them.

Figure 6 ascertains the AEP of E-FFTF and presents a comparative view with the non-flexible fault tolerance scenario. For comparison, the service reliability of 99.999% (reported for Amazon, Google and Microsoft) is extended to the users irrespective of their task deadline/slack time. Accordingly, the task execution price is computed for the high reliability service. In E-FFTF, comparatively low task execution price is obtained. Its prime reason is the saving in task execution price by selecting resources with reliability value directed by its slack time. Execution price of the tasks with enough slack time is reduced by selecting comparatively low reliability resources because their delayed execution may be accounted for its slack.

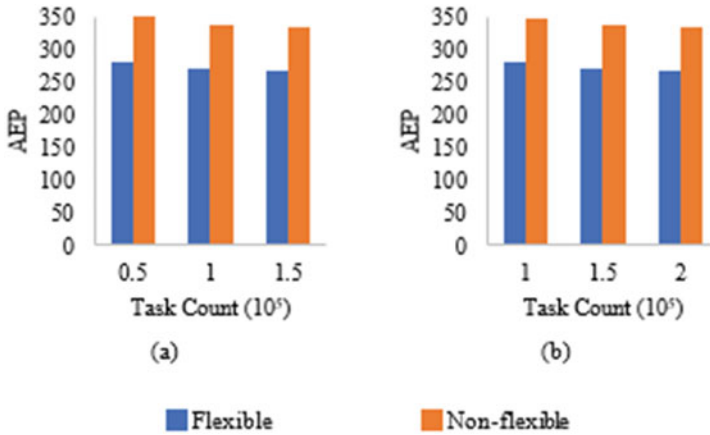


Fig. 6 AEP of flexible and non-flexible fault tolerance on **a** artificial and **b** real workload

Results shows that E-FFTF saves nearly 20% of the execution price as compared to the non-flexible fault tolerance without compromising the system performance.

4 Conclusions and Future Research Scope

In this research, FFTF is further extended to E-FFTF to provide the user transparent selection of service class and to establish its price saving capability while maintaining the performance. To prove the usefulness of E-FFTF, experiments are conducted over synthetic and real workload and obtained better results than FFTF. Task execution price saving is proved by comparing the total price invested in the flexible service selection through E-FFTF with the other well-established cloud service providers which intends to ensure high reliability service irrespective of the specified task deadline.

Prime outcomes from E-FFTF are: (a) Service selection is made transparent as class (A/B/C) is devised from the attributes such deadline and slack time. (b) Cloud user is benefitted by executing the task with softer deadline with low level of fault tolerance, which, in turn, levy low charges. (c) Service provider is benefitted by extending the services from low reliability resources to the user task with softer deadline. For the future research, we plan to apply the backup overloading approach within the replication groups along with executing delay-insensitive tasks to further enhance the resource utilization.

References

1. Fagg, G., Dongarra, J. (2000). FT-MPI: Fault tolerant MPI, supporting dynamic applications in a dynamic world. In *Proceedings of the Recent Advances in Parallel Virtual Machine and Message Passing Interface*. pp. 1–8
2. Abid, A., Khemakhem, M. T., Marzouk, S., Jemaa, M. B., Monteil, T., & Drira, K. (2014). Toward antifragile cloud computing infrastructures. *Procedia Comput. Sci.*, 32, 850–855.
3. Yadav, N., & Goraya, M. S. (2018). Two-way ranking based service mapping in cloud environment. *Future Generation Computer Systems*, 81, 53–66.
4. Hasan, M., & Goraya, M. S. (2019). Flexible fault tolerance in cloud through replicated cooperative resource group. *Computer Communications*, 145, 176–192.
5. Goraya, M. S., & Kaur, L. (2013). Fault tolerance task execution through cooperative computing in grid. *Parallel Process. Lett.*, 23, 1–20.
6. Malik, S., Huet, F., 2011. Adaptive fault tolerance in real time cloud computing. In *Proceedings of the IEEE World Congress on Services*. pp. 280–287.
7. Varasteh, A., Tashtarian, F., Goudarzi, M. (2017). On reliability-aware server consolidation in cloud datacenters. In *Proceedings of the 16th International Symposium on Parallel and Distributed Computing*. pp. 95–101.
8. Birke, R., Giurgiu, I., Chen, L.Y., Wiesmann, D., Engbersen, T. (2014) Failure analysis of virtual and physical machines: Patterns, causes and characteristics. In *Proceedings of the Int. Conf. Dependable Syst. Networks*. pp. 1–12.
9. More Permutations and Combinations. <https://www.cs.sfu.ca/~ggbaker/zju/math/perm-comb-more.html>
10. Hartigan, A., Wong, M.A. (1979). A K-Means clustering algorithm. *J. R. Stat. Soc.*
11. Sun, G., Liao, D., Zhao, D., Xu, Z., & Yu, H. (2018). Live migration for multiple correlated virtual machines in cloud-based data centers. *IEEE Transactions on Services Computing*, 11, 279–291.
12. Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A. F., & Buyya, R. (2011). CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Softw. - Pract. Exp.*, 41, 23–50.
13. Schroeder, B., & Gibson, G. A. (2010). A large-scale study of failures in high-performance computing systems. *IEEE Trans. Dependable Secur. Comput.*, 7, 337–350.
14. Hermenier, F., Lorca, X., Menaud, J., Muller, G., Lawall, J. (2009). Entropy: a consolidation manager for clusters. In *Proceedings of the 5th International Conference on Virtual Execution Environments*. pp. 41–50.
15. Wang, J., Bao, W., Zhu, X., Yang, T., & Xiang, Y. (2015). FESTAL: Fault-tolerant elastic scheduling algorithm for real-time tasks in virtualized cloud. *IEEE Transactions on Computers*, 64, 2545–2558.

Human Abnormal Activity Pattern Analysis in Diverse Background Surveillance Videos Using SVM and ResNet50 Model



S. Manjula and K. Lakshmi

Abstract Today, almost all the places are observed by surveillance cameras. The aim is to monitor the activities in and around, especially for abnormal activities. But it requires manual assistance to watch/monitor. However, manual inspection is a monotonous job, that reflects on information lost. It shows the importance of an automatic abnormal activity detection system. This is considered a difficult task because of object overlapping, light variation, clutter background, camera angle, presence of various activity, and posture variations. It manipulates the human actions in videos. Hence, it is challenging to recognize abnormal human actions. This work examines the concert of the classical *SVM* and *ResNet50* model among four datasets: The 'SAIAZ' (*Students Activities In Academic Zone*)-Corridor, 'SAIAZ'-Open space, Classroom Violence from YouTube (cc), and Mixed Background Dataset (MBD). The 'SAIAZ' was created by student volunteers of our Institution, and MBD is a collection of selected frames from various videos. Abnormal actions like slapping, punching, kicking, running, and fighting is commonly extant in these datasets. Here, MBD is assumed to adopt various real-world situations. The *SVM* achieved the classification accuracy of 85%, 92%, 60%, and 44% on SAIAZ-Corridor, SAIAZ-Open space, Classroom Violence, MBD respectively. The *ResNet50* achieved significant improvements in all the given datasets.

Keywords Abnormal Activity detection · Surveillance videos · *SVM* Classifier · *ResNet50* · Diverse background

S. Manjula (✉) · K. Lakshmi

Periyar Maniammai Institute of Science and Technology, Thanjavur, Tamil Nadu, India

e-mail: manjula_se@pmu.edu

K. Lakshmi

e-mail: lakshmi@pmu.edu

1 Introduction

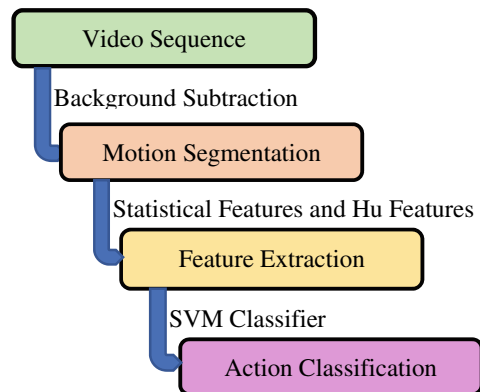
With the incremental growth of the population, most of the public places are occupied by the people. It is forced to monitor the places and manage the crowd activity. It recommends increasing public safety and provide protection to the physical assets. Nowadays the cost and maintenance of the surveillance system are drastically reduced, and the rate of unwanted activities has increased that influence many places that are under surveillance. The surveillance system to be monitored continuously or read by security personnel, if necessary. However, analyzing the video manually is a tedious task because of the larger size of the videos and involving different kinds of activities. To avoid manual intervention, the system requires the automatic detection of abnormal activities [1]. So, video-based abnormal activity detection becomes an active research area [2]. The real-world actions are dissimilar in nature, thus could not define all the abnormal activities.

The aim of this work is to identify the abnormal activities like slapping, kicking, punching, running, and throwing a stone in the following datasets such as ‘SAIAZ’ (*Student Activities In Academic Zone*)-Corridor, ‘SAIAZ’-Open Space, *Classroom Violence dataset*, and *Mixed Background Dataset (MBD)*. The ‘SAIAZ’ dataset is taken from our university premises. The actions in ‘SAIAZ’ was performed by a group of 6 to 7 student volunteers. Classroom Violence is a video clip taken from YouTube (cc). The MBD is a collection of selected frames from different videos. The reason to have MBD is to assess the system performance in a heterogeneous background.

This work is intended to study the performance of the classical *SVM* model and *ResNet50* approach in the datasets ‘SAIAZ’, *Classroom Violence dataset*, and in *MBD*. The *SVM* classification model (Fig. 1) gets the input from the video sequences, then segmentation is performed by background subtraction after that statistical & Hu features are extracted [3], finally, classification is performed by *SVM* classifier.

Deep learning architectures are the most effective approaches for object detection and classification [4]. The CNN architectures have a deep structure, which learns

Fig. 1 SVM classification model



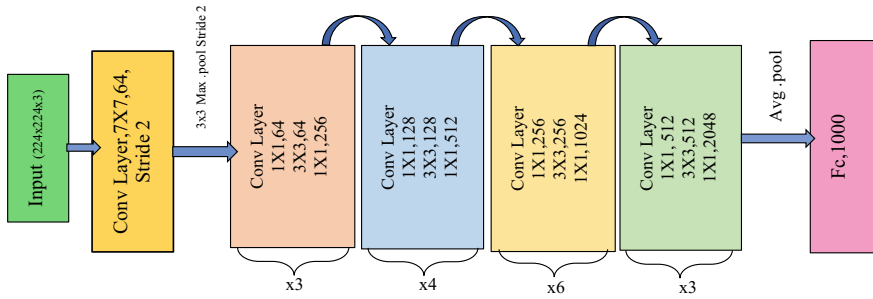


Fig. 2 Architecture of pre-trained *ResNet50* model

high-level structures of input data. Deep learning algorithms achieved a great success in classifying actions [5]. The *ResNet50* deep learning model is used here. This model uses skip connections (Residual Connections) to avoid information loss during the training period. This network is a combination of 49 convolutional layers and 1 fully connected layer. Figure 2 shows the *ResNet50* architecture. This is a regular feed-forward neural network with a residual connection. This *SVM* and *ResNet50* aims to investigate the classification accuracy on various datasets. The paper is organized as follows. In Sect. 2, a review made on state-of-the-art abnormal action detection systems. Section 3 describes the working methodology. Section 4, Illustrate the experimental results of video sequences. Finally, Sect. 5 concludes and presents the future work.

2 Related Work

The amount of surveillance data is enormously growing in recent years. A well-organized and consistent video monitoring system is required [6] to handle this video data. In [7], a histogram of optical flow method was used as an image/frame descriptor. It is more reliable to notify the abnormal activities and the PETS benchmark dataset was utilized to validate the effectiveness of the system. In classification phase, a single and multi-view SVM classifier was used to classify the actions and it was found that running action as abnormal activity. The parameters such as ROC and AUC were used to evaluate the system performance. The system [8] considers the poor-quality, occluded images and maintains the smoothness of the event.

The method [9] was equipped with temporal dependencies between the actions; it ensures the event detection in videos, and it was designed to handle *null events*. The Dataset Hollywood and TRECVID Surveillance Event Detection (SED) was considered to evaluate the system performance. STIP and Fisher Vector features are adopted and a multiclass SVM classifier was encouraged to classify the actions. The author in [10] was designed a Fuzzy Semantic Petri Nets (FSPN) system. It aims to detect the concurrently occurring events, and gives a summary of the situation,

of *what's going on*. The author in [11] was proposed a system using Generative Adversarial Nets (GAN), and it was skilled with normal events only not in abnormal events. During test time, the abnormalities are identified by checking the deviations from the normal actions. These can be assessed by frame-level and pixel-level fashion. In this system, a walking pedestrian is considered as a normal one and a person using a skateboard is treated as an abnormal activity. The standard UMN and UCSD dataset was used and Quantitative results like ROC (Receiver Operating Characteristics), AUC (Area Under Curve), and EER (Equal Error Rate) are employed here.

The author in [12] modeled a new FCN architecture to handle the abnormal regions. The kernels of this model are educated with the training video. This system is intended for speedier and precise system for anomaly detection in video data. The person walking in opposite direction is considered as an abnormal activity, walking pedestrian is a normal one.

In the system [13], a Temporal CNN Pattern was found the anomalous activity in a video data set. The walking pedestrian is considered as normal, movement of cars, bicycles, and skateboard is assumed as an abnormal activity. The UCSD and UMN standard dataset was used and the evaluation is performed on pixel level and frame level. This system's performance was evaluated by ROC, AUC, and EER.

The system [14] developed a new abnormal activity detection framework. It proposed Tracklets to formulae the salient points in each frame and the linear SVM was used to classify the actions in standard datasets like UMN, UCSD, and BEHAVE. The performance of the system was measured by ROC, AUC, and EER. It describes walking is a normal activity and people running away is an abnormal activity.

In this proposed Action Recognition with Image-Based CNN Features based system [15], where feature representation for videos is achieved by a hierarchical structure of CNN features such as CNN global features and *fully connected AlexNet features*. And it captures exact *AlexNet* features by using Caffe [16]. This proposed system achieves better performance on action recognition in the pioneered datasets such as KTH, UCF-11, and UCF-Sport action dataset. The system [17] 3D ResNet with Ranking Loss Function used to detect the abnormal activities in video data. This proposed 3D deep multiple instances learning with *ResNet* (MILR) model attains the best performance on the UCF-Crime dataset. ROC and AUC metrics are used to evaluate the system.

The author Hu [18] constructed a model to select the suspicious object region and the first-level features are extracted and fed into the deep network for deep feature extraction. The least-square SVM classification model is used for classification. The embedded chip Hi3531 was used and shows the high detection rate than the DLHS, DLPD, CTLD, Yolo-SD, and HOD-CCL methods. It detects the most abnormal behaviors and passes the alarming message in real-time surveillance.

3 Methodology

This proposed work is to study the performance of the *SVM* classification model and *ResNet50* model. And this is tested with *SAIAZ*, *Classroom Violence*, and *Mixed background Dataset (MBD)*. Each video dataset is converted to frames. The frames are labeled as a normal and abnormal activities. This presented system is trained with normal and abnormal frames and tested with test frames.

3.1 SVM Classifier Model

This work proposes to compare the activity identification using the classic *SVM* classifier and *ResNet50 model*. Well and profound *SVM* classifier is adapted and skilled to execute the linear and nonlinear classification [19, 20]. This *SVM* action classification system has the following stages 1. Obtaining video sequences, 2. Segmenting foreground objects by using background subtraction method, 3. Extracting features like statistical and Hu features by Hu invariant moments [3], 4. Classifying actions by *SVM* classifier. The part of this *SVM* work was published and available at: <https://srn.com/abstract=3453542> [21].

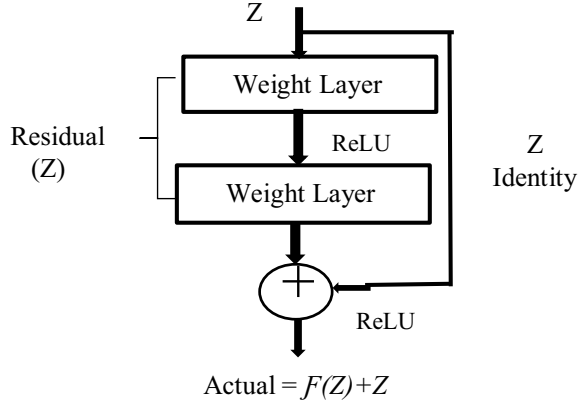
3.2 ResNet50 Model

The convolutional neural networks (CNNs) are an optimal choice for analyzing and detecting an abnormal action in video sequence. The CNN architecture learns low-mid-high-level features from an image. The CNNs automatically extract features from the data. The CNN is a set of convolution, activation, pooling, and fully connected layers. The recent CNNs are deeper as compared to the shallow network. And the recent CNNs has complex connections among alternative layers like *ResNet* [17]. Deep learning algorithms attained prodigious success in detection and classification actions [22].

The Pre-trained *ResNet50* model is used to perform the classification task in this proposed work. The popular *Resnet50* model is a combination of 49 convolution layers and one fully connected layer [23]. The Fig. 2 shows the pre-trained *ResNet50* architecture. The *ResNet* was designed with ultra-deep network to handle the vanishing gradient problem that earlier model suffers [23]. The *Resnet* has a residual block is shown in Fig. 3.

The Data augmentation is accomplished to resize the image as 224×224 and the standard color augmentation is performed in the *ResNet50* model. As per the literature, the *LeNet* [24] is used for binary images of handwritten digits, whereas other pre-trained models are prepared to adopt to color images. After each convolution operation and before the activation function a batch normalization is performed,

Fig. 3 Residual block



and the ReLU activation function is adopted here. The Mini-Batch and Stochastic Gradient Descent-based approaches are applied to train the data with a batch size 256 and along with the backpropagation mechanism. And this *ResNet50* model follows feed-forward neural network with *shortcut* connections. The *shortcut* connections of *ResNet50* model performs *identity* mapping and its output is added to the stacked layers.

Residual is an error [22] and this is the amount or number to change the prediction to meet out the actual value. The residual layer output is based on the outputs of $(l - 1)$ th layer. It is achieved by the previous layer of Z_{l-1} and $F(Z_{l-1})$ is the output obtained by performing convolution operation with different sizes of filters. The predicted value ' Z ' to be equal to the actual value. The residual function will compute and produce the residual of the model to match the predicted value with the actual value. By the mechanism of transfer learning, the same (pre-trained) deep learning approaches can be applied for different applications or with different types of data sets [23]. Therefore, *ResNet50* pre-trained model is used to find the abnormal activities in the above-mentioned datasets.

4 Experimental Analysis

In this work, MATLAB 2018a is used to perform the experiments. The performance of *SVM* and *ResNet50* model is compared with the datasets mentioned earlier. A set of normal and abnormal activity frames are obtained from the video data set. Four different types of video datasets were used here. *SAIAZ (Student Activities In Academic Zone)*—Corridor, *SAIAZ-Open Space*, *Classroom Violence*, *MBD* was utilized to test the performance of *SVM* Classifier and *ResNet50* deep learning model. The *SAIAZ* video dataset was created at our university premises by a group of six to seven student volunteers. It has two sets of video data: one is obtained from corridor and another one is from open-space. The *SAIAZ* dataset has 926 frames

with the following activities like slapping, kicking, punching, hitting, and fighting. And it was obtained from a top angled single static camera with a homogeneous background and the background illumination is uniform.

The *Classroom Violence* video dataset was taken from the YouTube (CC) and converted into frames. This video was taken from left, right, and front angled camera with different illumination condition. And this dataset consists 2433 frames with the activities like punching, kicking, slapping, pushing, and fighting. The MBD has 40 frames, it was formed by a set of selected frames from the different background videos. This MBD dataset has different kinds of background, illumination condition, variety of actions, involving a numerous number of persons with different posture.

All the video sequences are resized to 256×256 pixels. These four different kinds of data sets were tested with *SVM* classifier and *ResNet50* model. Figure 6 shows the *ResNet50* architecture obtained from MatLab. For simplicity, a specific region of *ResNet50* first layer is visualized. Figure 7, shows the low-level feature weights of first layer which was shown in the Fig. 6.

The dataset is classified as training set and test set. The training set has 100 frames, comprising of fifty normal activity frames and fifty abnormal activity frames, this was chosen by the presence of activity in the frame. The test set has 25 frames with 15 normal and 10 abnormal action frames. The systems performance is evaluated by the following performance metrics: True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), False Negative Rate (FNR) (Table 2), Precision, Recall, F1 Score, and accuracy of action detection rate are used.

For *ResNet50* model, the input image is resized as $224 \times 224 \times 3$ in data augmentation process and it is normalized with ‘zerocenter’ function. Here, the mean image is subtracted from an input image, to have a zero for mean and 1 for standard deviation. This ensures the smooth and faster optimization process during training period. Fc1000 is the output classification layer and it classifies the normal and abnormal actions. Figure 4a to 4h represents the sample normal and abnormal activity frames of four different kinds of datasets.

5 Result and Discussion

A portion of this work using *SVM* classification on *SAIAZ-Corridor* was published in [21], may refer for further details. This proposed work discusses about, how *SVM* and *ResNet50* works on different types of dataset. In the *ResNet50*, the residual block is a basic structure of network. In the training period, a certain residual block is enabled, and the input streams are passed through the identity shortcut and the weight layers, otherwise the input flows only through the identity shortcut. Figure 7, shows the low-level feature weights (edges, curves) of the first layer. During training time, the layer has a “existence possibility” and it is randomly dropped. At the time of testing, all blocks are kept active and re-calibrated according to its existence possibility. The camera angle, illumination, background, and posture-based features of images



Fig. 4 a, b—SAIAZ-corridor: normal and abnormal action frames; c, d—SAIAZ-open space: normal and abnormal action frames; e, f—classroom -violence—normal and abnormal action frames; g, h—mixed background—normal and abnormal action frames

are used for the classification task. The classification accuracy is presented in the Table 1.

In the dataset SAIAZ-Corridor and SAIAZ-Open space, the proposed model SVM and ResNet50 works well when compared to the remaining two datasets. The observation presented from the above video dataset is taken from the top angled static camera and there is no angle transformation in it. It is also observed that the illumination is uniformly distributed and the background is identical all over the video. The accuracy of the SVM on SAIAZ-Corridor is relatively lower than the SAIAZ-Openspace, Why, because the training set of SAIAZ-open space was clearly defined. The accuracy was brought down in SAIAZ Corridor due to ambiguity in a selected training set. Where the deep learning model gets better results on it. Looking into the Classroom-violence dataset, the detection rate of SVM is reduced because the dataset is taken from a different angle camera. Here, the illumination is diverse and more objects are involved; these are few notable reasons why it brings down the detection rate. Even in the uncertainty of datasets, the ResNet50 works well. In the above three cases, each training dataset and testing dataset is taken from the same video. But in the case of MBD, the system is trained with any one of the above three dataset, but the test case is from MBD, hence the detection rate of the system is reduced. In

Table 1 Classification accuracy obtained on various dataset

Dataset	SAIAZ-Corridor (SC)		SAIAZ-OpenSpace (SO)		Classroom Violence (CV)		Mixed Background (MBD)	
	SVM	ResNet50	SVM	ResNet50	SVM	ResNet50	SVM	ResNet50
% of Accuracy	85	96	92	96	60	92	44	68

Table 2 Performance metrics

Dataset	SAIAZ- Corridor (SC)		SAIAZ—OpenSpace (SO)		Classroom-Violence (CV)		Mixed Background Dataset (MBD)	
	SVM	ResNet50	SVM	ResNet50	SVM	ResNet50	SVM	ResNet50
TPR	1	1	0.9	1	0.6	0.9	0.75	0.8
FNR	0	0	0.1	0	0.4	0.1	0.25	0.2
TNR	0.73	0.93	0.93	0.93	0.6	0.93	0.16	0.6
FPR	0.33	0.07	0.06	0.07	0.4	0.06	0.84	0.4

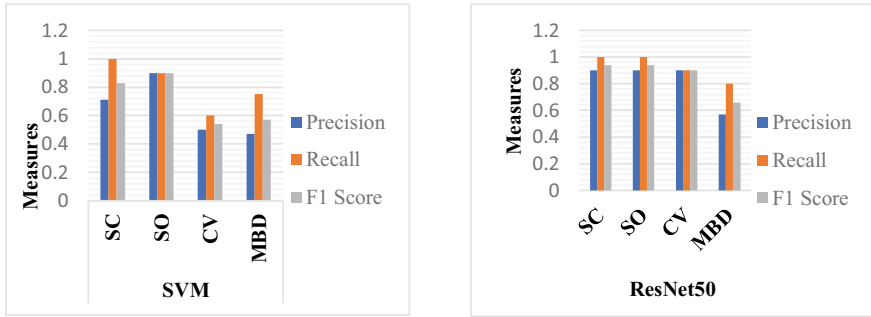


Fig. 5 a Performance Measures on SVM. b Performance Measure on ResNet50

the fourth case as well, deeper layers have learned the features and have produced positive results than the SVM.

In Table 2, the performance of the proposed model is displayed. The True Positive Rate (TPR), False Negative Rate (FNR), True Negative Rate (TNR), and False Positive Rate (FPR) is obtained. In Fig. 5a and 5b, the performance metrics precision, recall, and F1 Score are exhibited. Figure 8 shows some of the mixed background resultant video frames. As per the results, comparatively *ResNet50* shows the better performance on all type of datasets (Fig. 5a and 5b).

$$TPR = TP / (TP + FN) \quad (1)$$

$$FNR = FN / (TP + FN) \quad (2)$$

$$TNR = TN / (TN + FP) \quad (3)$$

$$FPR = FP / (FP + TN) \quad (4)$$

$$\text{Precision} = TP / (TP + FP) \quad (5)$$

$$\text{Recall} = TP / (TP + FN) \quad (6)$$

$$F1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{Accuracy} = TP + TN / N \quad (8)$$

TP—True Positive; FN—Fales Negative; TN—True Negative; FP—False Positive; N—Total number of frames.

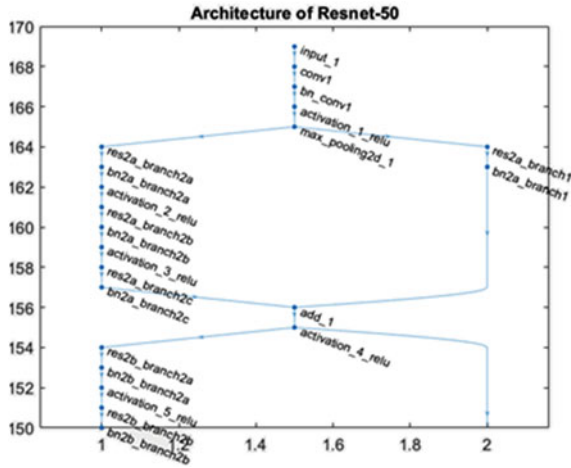


Fig. 6 First Layer of ResNet50 on Visualization of region 50–170 obtained through MATLAB

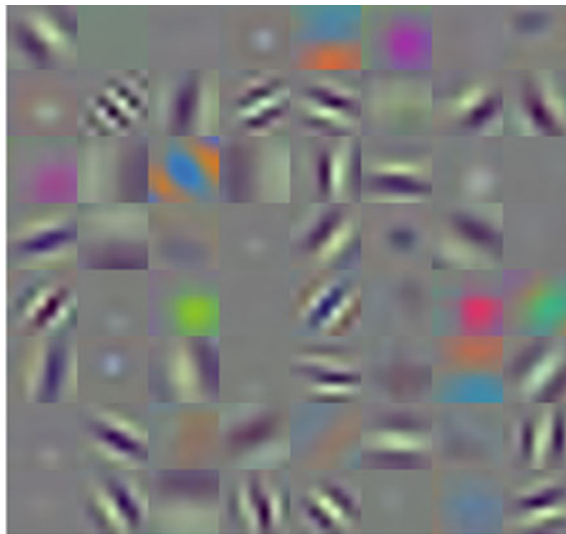


Fig. 7 Visualization of low-level feature weights (edges, curves) of first layer of region 150–170 obtained through MATLAB

6 Conclusion

This system evaluates the classification accuracy in classical SVM classifier and ResNet50 model. The SVM gives better classification accuracy on SAIAZ, where



Fig. 8 Resultant abnormal action frames from mixed background Dataset

SAIAZ video taken from the top angled single static camera in uniform illumination condition with simple background. Both Classroom violence and MBD dataset have a complex background with different illumination conditions when placed at different camera angles. Thus, the classification accuracy of *SVM* classifier on classroom violence data set and MBD set is not up to the level. Therefore, the *ResNet50* model was adopted to overcome those problems. The deep learning *ResNet50* model achieves better performance than *SVM* on given four types of datasets. But, in both cases, irrespectively, the result is pull down in MBD. The reason behind for the above case is that the training and testing frames are totally different. To test and improve the complexity of the system, earlier condition was entertained. To accomplish the above situation, the system is to be modified. The *ResNet50* shows good results on *SAIAZ* and *Classroom Violence* dataset. But, significantly decreases on mixed heterogeneous background. To handle these situations, the system will be modified in our future work.

References

1. Ke, S.-R., Thuc, H. L. U., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H., & Choi, K.-H. (2013). A review on video-based human activity recognition. *Computers*, 88–131. <https://doi.org/10.3390/computers202088ComputersISSN2073-431Xwww.mdpi.com/journal/computer>.
2. Popoola, O. P., & Wang, K. (2012). Video-based abnormal human behaviour recognition - a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 865–878.
3. Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory, IEEE Explore*, 179–187.
4. Lecun, Y., Behgio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
5. Zheng, G. (2017). Fruit and vegetable classification system using image saliency and convolutional neural network. In *IEEE 3rd Information Technology and Mechatronics Engineering Conference* (pp. 613–617).
6. Zhang, J., Wu, C., & Wang, Y., et.al. (2018). Detection of abnormal behaviour in narrow scene with perspective distortion. *Journal of Machine Vision and Applications*, 1–12.

7. Wang, T., Chen, J., Honeine, P., & Snoussi, H. (2015). Abnormal event detection via multikernel learning for distributed camera networks. *International Journal of Distributed Sensor Networks*, 2015, 1–9. Article Id: 989450. <https://doi.org/10.1155/2015/989450>.
8. Tran, D., Yuan, J., & Forsyth, D. (2014). Video event detection: From subvolume localization to spatiotemporal path search. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 36(2), 404–416.
9. Cheng, Y., & Fan, Q., Pankanti, S., & Choudhary, A. (2014). Temporal sequence modeling for video event detection. In *CVPR2014 Open Access version of Computer Vision Foundation*.
10. Szwed, P., & Komorkiewicz, M. (2013). Object tracking and video event recognition with fuzzy semantic petri nets. In *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems* (pp. 167–174). IEEE. 978-1-4673-4471-5/\$25.00 c.
11. Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L., Regazzoni, C., & Sebe, N. (2018). Abnormal event detection in videos using generative adversarial nets. In *Proceedings - International Conference on Image Processing, ICIP* (pp. 1577–1581).
12. Sabokrou, M., Fayyaz, M., Fathy, M., Moayed, Z., & Klette, R. (2018). Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. *Journal of computer vision and Image Processing*, 172, 88–97.
13. Ravanbakhsh, M., Nabi, M., Mousavi, H., Enver, S., & Sebe, N. (2018, January). Plug-and-Play CNN for crowd motion analysis: An application in abnormal event detection. In *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018* (pp. 1689–1698).
14. Mousavi, H., Nabi, M., Galoogahi, H. K., Perina, A., & Murino, V. (2015). Abnormality detection with improved histogram of oriented tracklets. In *International Conference on Image Analysis and Processing* (pp. 722–732). Springer.
15. Ravanbakhsh, M., Mousavi, H., Rastegari, M., Murino, V., Davis, L. S. (2015). Action recognition with image based CNN features. <http://arxiv.org/abs/1512.03980>.
16. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*.
17. Dubey, S., Boragule, A., & Jeon M. (2019). 3D ResNet with ranking loss function for abnormal activity detection in videos. In *2019 International Conference on Control, Automation and Information Sciences (ICCAIS)* (pp. 1–6). <https://doi.org/10.1109/ICCAIS46528.2019.9074586>.
18. Hu, Y. (2020). Design and implementation of abnormal behavior detection based on deep intelligent analysis algorithms in massive video surveillance. *Journal of Grid Computing*, 18, 227–237. <https://doi.org/10.1007/s10723-020-09506-2>.
19. Fu, Z., Robles-Kelly, A., & Zhou, J. (2010, December). Mixing linear SVMs for nonlinear classification. *IEEE Transactions on Neural Networks*, 21(12), 1963–1975.
20. Yu, H., & Kim, S. (2012, January). SVM tutorial: Classification, regression, and ranking. https://doi.org/10.1007/978-3-540-92910-9_15.
21. Ravichandran, M., & Krishnamoorthy, L. (2019, September 14). Detection and recognition of abnormal behaviour patterns in surveillance videos using SVM classifier. In *Proceedings of International Conference on Recent Trends in Computing, Communication & Networking Technologies (ICRTCCNT) 2019*. SSRN. <https://ssrn.com/abstract=3453542>, <https://doi.org/10.2139/ssrn.3453542>.
22. He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. [arXiv:1512.03385v1](https://arxiv.org/abs/1512.03385v1) [cs.CV].
23. Zahangir Alom, Md., et al. (2018). The history began from AlexNet: A comprehensive survey on deep learning approaches. *Computer Vision and Pattern Recognition* (cs.CV). [arXiv:1803.01164](https://arxiv.org/abs/1803.01164) [cs.CV].
24. Le Cun, Y., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1989). Handwritten digit recognition with a back-propagation network. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)* (pp. 396–404).

RT-GATE: Concept of Micro Level Polarization in QCA



K. Bhagya Lakshmi, D. Ajitha, and K. N. V. S. Vijaya Lakshmi

Abstract Quantum-dot Cellular Automata is an evaluation paradigm in which transistors are not used and viable candidate for replacing the CMOS based technology. QCA is one of the boosting nanotechnology devices with the aim to replace the CMOS technology. QCA is implemented by utilizing the tunneling of the electrons with the given potential within the quantum cell. We made an attempt to suggest a multiplexer architecture in QCA using micro-level polarization. The proposed multiplexer design saves **16.67%** of effective area compared to the best designs reported till date. In this paper, new designs of universal gates are proposed. The proposed NOR and NAND gates requires less number of quantum cells which results in less effective area compared to the conventional majority gate based designs. By using these micro-level polarized gates, the multiplexer which is proposed in this paper is implemented. The proposed multiplexer is designed and simulated by QCA Designer.

Keywords Quantum-dot Cellular Automata (QCA) · RT-Gate (Rotate T-Gate) · Universal gates · Majority Voter Gate (MVG) · 2×1 Multiplexer · Operation cost (O-Cost)

1 Introduction

QCA is a proposed improvement on CMOS technology. *Moore's law* stated that the count of transistors present in an integrated circuit doubles about every eighteen months. But in the recent era, the size of the MOSFET is very minute. To satisfy the Moore's law, the size of the MOSFET must be very much shortened by reducing

K. Bhagya Lakshmi (✉)

Sasi Institute of Technology and Engineering, Tadepalligudem, India

D. Ajitha

Sreenidhi Institute of Science and Technology, Hyderabad, India

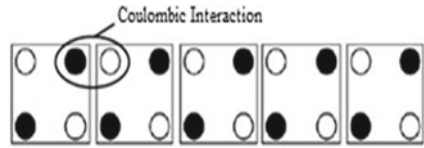
e-mail: ajithad@sreenidhi.edu.in

K. N. V. S. Vijaya Lakshmi

Sri Vasavi Engineering College, Tadepalligudem, India

e-mail: vijayalakshmi.kakaraparthi@srivasaviengg.ac.in

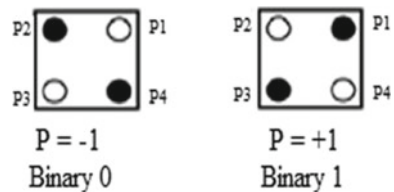
Fig. 1 Communication between the adjacent cells in a QCA



the length of its channel. By reducing the length of MOSFET’s channel beyond certain limits, short channel effects come to existence. There comes the QCA technology which gives transistor less designs. The speed, density of devices and power consumption of QCA devices are at the forefront of the technologies of semiconductor like CMOS. In Fig. 1, the columbic repulsion rule indicates the interaction between the two neighbor cells in QCA. Based on the adjacent cell’s electron repulsions, the position of electrons will be set in the quantum cells. This way, the message will be transferred from input side to output side. The different combinations of the quantum cells give rise to different devices.

In QCA technology, the wire is in the appearance of an array, a mixture of quantum cells. There are four corners of each quantum cell that contain four quantum dots. The electrons still occupy diagonal positions due to the Coulomb repulsion force. The polarization $P = +1$ is defined when one of the electrons is positioned at left corner lower side and the other electron is positioned at the right upper corner. In opposite case, it is defined as $P = -1$. These representations are shown in Fig. 2. Polarization of a quantum cell is given by “ $P_i = ((P1 + P3) - (P2 + P4))/(P1 + P2 + P3 + P4)$ ” [1]. If two electrons reside in the opposite quantum dots of a quantum cell, then the polarization exits which defines logic 1 and logic 0. According to this, in Fig. 2, the first quantum cell polarization is $P = -1$ and $P = +1$ for the second quantum cell. In this paper, new designs of universal gates are proposed by changing the fixed polarization levels. The basic gates and the existing T-gate structure are discussed in the subsequent section. The proposed universal gates structure is discussed in Sect. 3 and the multiplexer implementation is discussed in Sect. 4 followed by the conclusion in Sect. 5.

Fig. 2 QCA cell polarization states



2 QCA Basic Elements

The wire and Majority Voter Gate (MVG) are the basic elements [2] for logic implementation in QCA. Figure 3a is a QCA wire, an array of quantum cells. Figure 3b is QCA MVG whose logic is identical to Boolean expression $WX + XY + YW$ where W, X, Y are inputs. MVG implementation is shown in Fig. 3c. Cells A, B, and C are input cells, D is the output cell. The polarization of D is same as the majority of the input cells polarized value.

The basic gates AND & OR gates are implemented using majority voter gate shown in Fig. 4a. The majority voter gate has the drawback of having no provision for an inverter and has to base on displaced wire chain for inversion and it is shown in Fig. 4b.

2.1 Universal Gates Based on Partial Polarization Technique

Partial polarization is the action associated between two mutually orthogonal diagonal components [2]. The T-Gate can be implemented using 4 quantum cells [3]. However, to improve output polarization energy, one more quantum cell is to be added at the output. The maximum and minimum amplitude of output-simulated waveform is $9.52e-001e$ and $-9.52e-001$, respectively. Table 1 gives the T-Gate implementation of NAND gate and NOR gate with 5 quantum cells. The existing T-gate structure was constructed with normal cells and 0.20 polarization value. Architecture for NAND/NOR is proposed using partial polarization of ± 0.75 [4]. The NAND/NOR requires 5 quantum cells with 2 cells partially polarized [4]. However, it requires less number of cells a new gate structure is proposed to achieve better AUF factor

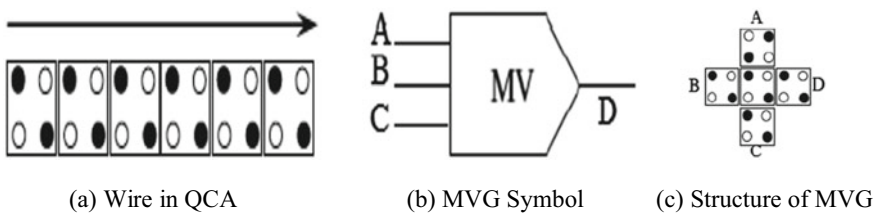

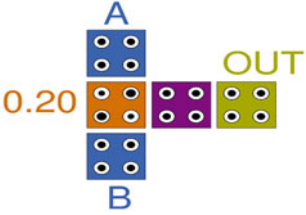

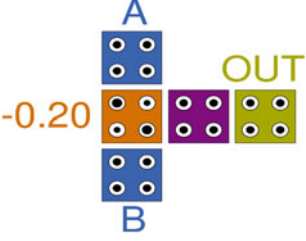


Fig. 3 Basic QCA logic circuits



Fig. 4 a AND & OR gates. b NOT gate

Table 1 Universal gates using partially polarized T-gate

Logic gate	Symbol	QCA layout
$RT^+(A, B) = \overline{A \cdot B}$		
$RT^-(A, B) = \overline{A + B}$		

and discussed in Sect. 3. The clocking signal is a clipped signal sine wave that is a function of time is used as a default clock signal [5, 6].

3 Proposed RT-Gate

Quantum dots are part of a semiconductor device. The intrinsic electrostatics in the semiconductor device allows a “puddle” of electrons to form, near at semiconductor insulator boundary. The silicon metal-oxide-semiconductor (MOS) system is often used, with an additional level of gates in the insulator to form the quantum dot. In the presence of random cell rotation, it is assessed [7] that QCA devices are highly reliable. In the suggested RT-Gate, we used all quantum cells in rotated mode. The low-level fixed polarization ranges from ± 0.001 to ± 0.023 . With the negative partial polarization NOR gate and by positive partial polarization NAND gate functionality can be achieved. The minimum waveform amplitude and maximum waveform output amplitude are $-9.76e-001$ and $9.76e-001$, respectively. The proposed RT-Gate is shown as NAND Gate and NOR Gate, respectively, in Figs. 5 and 6, in which both are implemented with 5 quantum cells.

Fundamental logic gates are developed using the proposed structure of universal logic gates. To get AND gate functionality from the proposed NAND gate, one more rotated cell is added to the output. The implementation of proposed AND gate is shown in Fig. 7. Likewise, to get OR gate functionality from the partially polarized NOR gate, one more rotated cell is added at the output. The implementation of proposed OR gate is shown in Fig. 8.

Fig. 5 Implementation of RT gate as NAND gate

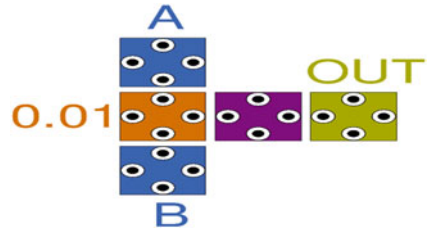


Fig. 6 Implementation of RT gate as NOR gate

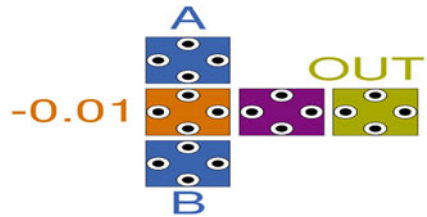


Fig. 7 Implementation of AND gate

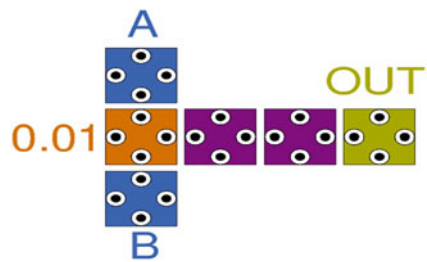
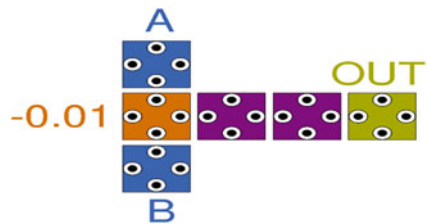


Fig. 8 Implementation of OR gate



It is possible to enforce the NOT Gate structure by adding the same logic value to both of the inputs in the projected universal gate, and the same logic can also be extended to both NOR / NAND Logic. The effective area of the structure which is proposed is calculated to be $3,600 \text{ nm}^2$ against the T-gate [3] is 21769 nm^2 for effective implementation of the T-gate, the effective area required is $3,600 \text{ nm}^2$. The area utilization factor is 2.22. The area is calculated using the AUF formula method [8]. Using two clock zones, the output for the implemented gate is examined for minimum complexity. The power drop for proposed RT-Gate is 54.55 nWatts. The

power drop of the proposed gate is calculated by using the algorithm that is used for computing the power drop in a coplanar wire [8]. As compared with the existing gate structures presented to design the serial adders and universal gate, the proposed gates will occupy less area with micro polarization [14, 15].

4 Multiplexer

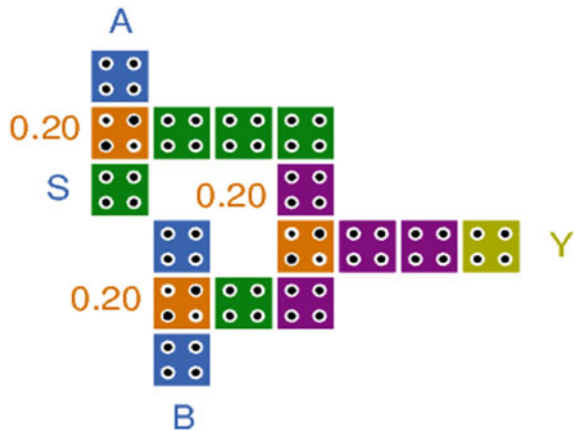
Multiplexer is one of the key elements in digital circuit designs and widely used in the implementation of an FPGA circuits. The multiplexer design shown in Fig. 9 is proposed by Soudip Sinha Roy [3] using T-Gate requires 16 Quantum-dot cells and the maximum and minimum amplitude of the output waveform is simulated with $9.54e-001$ and $-9.54e-001$, respectively.

4.1 Proposed Multiplexer

This paper explores the implementation of an area efficient design of a 2×1 multiplexer in QCA. The proposed multiplexer design and its simulated output are shown in Figs. 10 and 11 correspondingly. With low value polarizations and rotated cells, the proposed structure of 2:1 multiplexer is implemented.

The performance of the anticipated 2:1 multiplexer is compared with the preceding designs with the effective area and O-Cost. The analysis is shown in Table 2. The proposed multiplexer has been shown to be successful by using a smaller number of QCA cells. The existing multiplexer design [3] can be implemented by using 16 QCA. 2 to 1 multiplexer digital design [9] is implemented by using 34 QCA cells. QCA Designer tool is used to examine the design of QCA which is implemented [10].

Fig. 9 Existing Multiplexer design



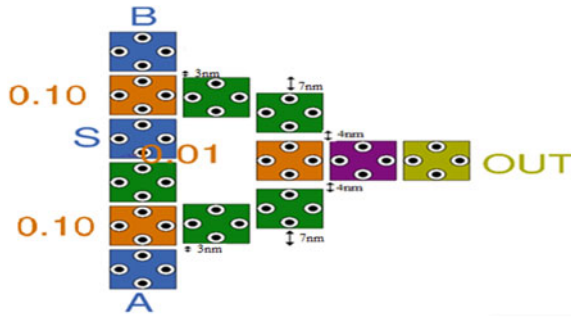


Fig. 10 Proposed multiplexer design

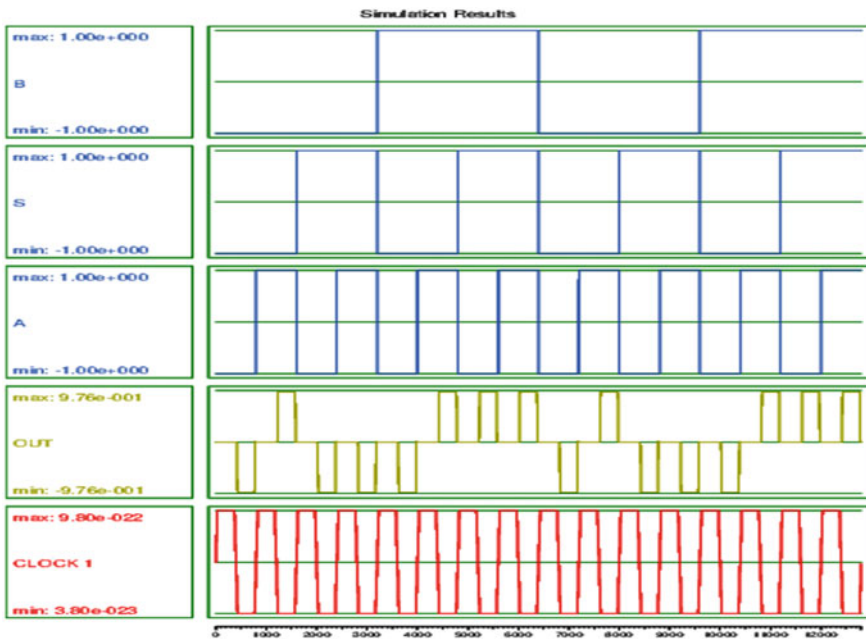


Fig. 11 Output waveform of proposed multiplexer

The constructed 2 to 1 multiplexer layout [11] is created by the QCA-LG method, which consists of 75 cells. The versatile architecture of the 2 to 1 multiplexer [12] comprises 41 quantum cells.

A novel design for 2:1 MUX which is implemented in [13] consists of 23 quantum cells, where as the proposed design requires only 13 cells. Length and breadth are the two dimensions of this layout. The normal spacing of 2 nm between two quantum cells is the QCA meaning. 1 nm is the separation of the cells from the cell bed's terminal point. Because of the equal length and width of 18 nm resembles a perfect

Table 2 Analysis of 2:1 Multiplexer compared with the existing designs

S. No	No. of quantum cells for design implementation	Operation cost	Effective area in nm ²	Area improvement achieved in the projected RT-Gate
1	2 × 1 MUX [11]	75	121,764	90.14%
2	2 × 1 MUX [12]	41	45,924	73.86%
3	2 × 1 MUX [9]	34	40,764	70.56%
4	2 × 1 MUX [13]	23	23,360	55.03%
5	2 × 1 MUX [3]	16	21,769	44.87%
6	2 × 1 MUX [8]	14	14,400	16.67%
7	Proposed RT-mux	13	12,000	–

square, the cells are similar in terms of area. The region is, therefore, $18\text{--}18\text{ n}^2\text{m}^2$. $L = (1 + 18 + 2 + 18 + 2 + 18 + 2 + 18 + 2 + 18 + 18 + 2 + 18 + 18 + 1) = 100\text{ nm}$ is the estimated length of the proposed MUX. $W = (1 + 18 + 2 + 18 + 2 + 18 + 2 + 18 + 2 + 18 + 2 + 18 + 2 + 18 + 18 + 1) = 120\text{ nm}$ is the width of the layout. Consequently the area of the circuit is 12000 (nm)^2 . The effective area occupied by the proposed 2:1 multiplexer is **12,000 (nm)²**. In the architecture, only single layer is used and there are no wire crossings. The multiplexer output can be observed in clock 1, i.e., only 2 clock zones are needed as shown in Fig. 11. The amplitude (Max and Min) of the output waveform here are $9.76\text{e-}001$ and $-9.76\text{e-}001$, respectively.

5 Conclusion

Multiplexers are essential components for many logical and digital circuits. A 2×1 multiplexer in QCA with unique design is presented in this paper. In contrast to the present design, the proposed multiplexer design saves effective area by **16.67%**. By using this proposed design, the area and speed is optimized. Any combinational circuit can be implemented in an efficient manner with this proposed design. It can be considered that the anticipated design is more stable than the previous models, based on the implementation performance.

References

1. Lent, C. S., & Tougaw, P. D. (1997). A device architecture for computing with quantum dots. *Proceedings of the IEEE*, 85(4), 541–557.
2. Lahiri, M., Hochrainer, A., Lapkiewicz, R., Lemos, G. B., & Zeilinger, A. (2017). Partial polarization by quantum distinguishability. *Physical Review A*, 95(3), 033816.

3. Mukherjee, C., Roy, S. S., Panda, S., & Maji, B. (2016, May 24). T-gate concept of partial polarization in quantum dot cellular automata. *Proceedings of 20th International Symposium on VLSI Design and Test (VDATE)* (pp. 1–6). IEEE.
4. Ghosal, S., & Biswas, D. (2013). Study and defect characterization of a universal QCA gate. *International Journal of Computer Applications*, 975, 8887.
5. Lu, Y., Liu, M., & Lent, C. (2007). Molecular quantum-dot cellular automata: From molecular structure to circuit dynamics. *Journal of applied physics*, 102(3), 034311.
6. Blair, E. P., & Lent, C. S. (2003, September). Quantum-dot cellular automata: an architecture for molecular computing. In *Proceedings of International Conference on Simulation of Semiconductor Processes and Devices, 2003. SISPAD 2003* (pp. 14–18). IEEE.
7. Yang, X., Cai, L., Wang, S., Wang, Z., & Feng, C. (2012). Reliability and performance evaluation of QCA devices with rotation cell defect. *IEEE Transactions on Nanotechnology*, 11(5), 1009–1018.
8. Roy, S.S., 2017. pGate: An Introduction to A Novel Universal Gate and Power Drop Calculation of QCA circuits. *International Journal Of Engineering And Computer Science (IJECS)*, pp.20967–20972.
9. Askari, M., Taghizadeh, M., & Fardad, K. (2008, May 13). Digital design using quantum-dot cellular automata (a nanotechnology method). In *Proceedings of International Conference on Computer and Communication Engineering* (pp. 952–955). IEEE.
10. Walus, K., & Jullien, G. A. (2006). Design tools for an emerging SoC technology: Quantum-dot cellular automata. *Proceedings of the IEEE*, 94(6), 1225–1244.
11. Srivastava, S., & Bhanja, S. (2007). Hierarchical probabilistic macro modeling for QCA circuits. *IEEE Transactions on Computers*, 56(2), 174–190.
12. Teodósio, T., & Sousa, L. (2007, November 19). QCA-LG: A tool for the automatic layout generation of QCA combinational circuits. In *Proceedings of Norchip conference* (pp. 1–5). IEEE.
13. Kim, K., Wu, K., & Karri, R. (2006). The robust QCA adder designs using composable QCA building blocks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 26(1), 176–183.
14. Mostafaei, A., & Rezaei, A. (2017). Novel optimized designs for QCA serial adders. *International Journal of Information Technology and Computer Science*, 2(2), 38–46.
15. Bilal, B., Ahmed, S., & Kakkar, V. (2017). Optimal realization of universality of Peres gate using explicit interaction of cells in quantum dot cellular automata nanotechnology. *International Journal of Intelligent Systems and Applications*, 11(6), 75.

Comparative Performance Analysis of Tanh-Apodized Fiber Bragg Grating and Gaussian-Apodized Fiber Bragg Grating as Hybrid Dispersion Compensation Model



Baseerat Gul and Faroze Ahmad

Abstract Fiber optic systems are used for the prolonged reach transmission systems, but by increasing the bit rate which is the main requirement of the current time, dispersion gets arisen which results in intersymbol interference. Compensation of dispersion to improve the transmission capability of the fiber optic system provides a vast field for research. From the literature survey done, use of Dispersion compensation fiber has been found as the most reliable method for compensating the dispersion, but it becomes expensive as the length of Dispersion compensation fiber is increased for long distance transmission. The Fiber Bragg Grating is also used as a dispersion compensation module as reported in previous works but has been found inefficient method. However, the Performance of the Fiber Bragg Grating can be enhanced by adapting optimum Chirping technique and Apodization profile. From the previous reported works, Tanh-Apodized Fiber Bragg grating and Gaussian-Apodized Fiber Bragg grating are found to have optimum performance characteristics in terms of side lobe suppression and maximum reflectivity, which motivates us to analyze the respective Fiber Bragg Gratings for compensating the dispersion at various chirping techniques and variable grating lengths. In this work, Tanh-Apodized Fiber Bragg grating and Gaussian-Apodized Fiber Bragg grating are analyzed and simulated in various chirping techniques individually, as well as along with the Dispersion compensation fiber, in the hybrid model of dispersion compensation for a 100 km long optical fiber link at the data rate of 10Gbps. The simulation software used is optisystem. Also, the grating length has been varied and the different performance characteristics like Q-factor, BER, and Eye diagram are analyzed and compared. It has been observed that the Gaussian-Apodized quadratic-chirped Fiber Bragg Grating at the grating length of 26.6 mm along with the 11 km long Dispersion compensation fiber makes the cheaper dispersion compensation module with the finest performance.

Keywords Optical fiber communication · Dispersion · Dispersion compensation fiber · Fiber bragg grating

B. Gul (✉) · F. Ahmad

Department of Electronics and Communication Engineering, Islamic University of Science and Technology, Awantipora, J&K, India

e-mail: baseerat.gul@islamicuniversity.edu.in

1 Introduction

In the current era of Optical Fiber communication (OFC), the main issues with the increasing bit rate for the Conventional Single mode fiber (CSMF) are the Group velocity dispersion (GVD), Attenuation, and Non-linear effects [1]. However, the Attenuation problem has been solved by the introduction of Erbium doped fiber amplifier (EDFA) [2, 3] and the effect of non-linearity can be reduced by keeping the associated power-level below certain limits [2]. GVD, however, becomes negligible for bit rates below 2 Gbps, but for higher bit rates, it reduces the quality parameters of the signal and needs to be compensated. Abundant research work has been reported in the past regarding the compensation of dispersion to improve the performance characteristics of high speed optical communication systems. Dispersion compensation fiber (DCF) and Fiber Bragg grating (FBG) are most widely used for compensating the dispersion. DCF has shown superior performance in compensating the dispersion to a great extent but has been proven as an expensive method for compensating the dispersion and becomes costlier with its increasing length. FBG has shown inefficient performance as compared to DCF, but its performance can be improved by adapting proper Chirping technique and Apodization profile. From the work reported in paper [4], Gaussian-Apodized FBG and Tanh-Apodized FBG have shown optimum performance in terms of side lobe suppression and the maximum reflectivity. Paper [5] reports that the bandwidth of the compensated spectrum can be leveled up to optimal value by choosing the proper grating length and the chirp rate of the CFBG. In this paper, comparative study and analysis of five different dispersion compensation models have been made for 100 km long optical fiber communication systems at the data rate of 10 Gbps. In these dispersion compensation models, various chirping techniques of Tanh-Apodized FBG and Gaussian-Apodized FBG have been investigated individually and as a part of hybrid model along with the DCF to discover the dispersion compensation model which shows the finest performance both in terms of cost reduction and quality improvement. The lower order of the GVD, i.e., Chromatic dispersion is taken into consideration while ignoring the higher orders of GVD which have negligible effect at the data rate of 10Gbps and becomes prominent at the data rate of 40Gbps or above [2, 3, 6].

2 Theory

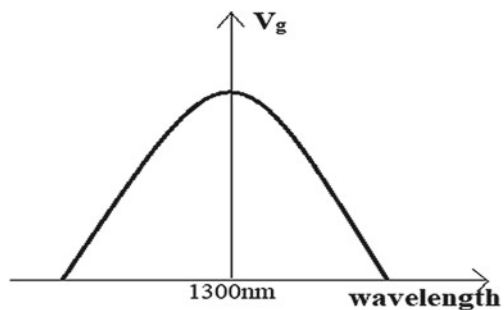
The Chromatic dispersion is caused due to the finite spectral width of the optical source and the non-linear variation in the refractive index of the core of CSMF with respect to the variation in wavelength parameter. Dispersion arisen in CSMF results in pulse broadening, which increases with the increase in transmission distance and the bit rate, hence making the high data rate long-haul optical communication difficult [1], which at present has the highest demand in the telecommunication field.

Although the chromatic dispersion is seen zero at Zero Dispersion wavelength (ZDW = 1300 nm), but the preferable wavelength parameter for the optical transmission is 1550 nm due to the least attenuation and other favorable parameters [2, 3]. The effect of dispersion at the 1550 nm wavelength window is the major constraint [3]. DCF and FBG are the extensively adapted methods for compensating the dispersion by connecting either one or both in cascade with the CSMF [3]. From the literature survey done, the use of DCF has been found the most reliable but expensive method for compensating the dispersion, and the use of FBG has been found an inefficient method [1–3, 7]. Performance of FBG as a dispersion compensation model, however, can be enhanced by adapting optimum Chirping techniques and Apodization profiles [1, 4, 8].

2.1 Dispersion Compensation Fiber

At ZDW (1300 nm), Group velocity (V_g) associated with the signal propagating through CSMF is seen maximum and decreases monotonically on the either side of wavelength change as shown in Fig. 1. If the operating wavelength chosen is 1550 nm, longer wavelength components in CSMF propagate with the slower speed and the shorter wavelength components propagate with the faster speed which results in dispersion. DCF has been designed in such a way by the variation in refractive index such that it allows longer wavelength components to propagate at the faster speed and vice versa, that is why it is said that the DCF is associated with negative dispersion [6]. DCF with negative dispersion when connected in cascade with the CSMF nullifies the dispersion arisen in CSMF. Negative dispersion in DCF can be increased by increasing the doping level of core with GeO_2 , however, higher doping in turn increases the losses [6], hence the negative dispersion associated with the DCF cannot be increased beyond its optimal value which is approximately in the range of -70 to -80 ps/nm/km. For compensating higher dispersion produced by increasing length of CSMF, length of the DCF needs to be increased accordingly. However, increasing the length of DCF increases the cost associated with it.

Fig. 1 Group velocity versus wavelength



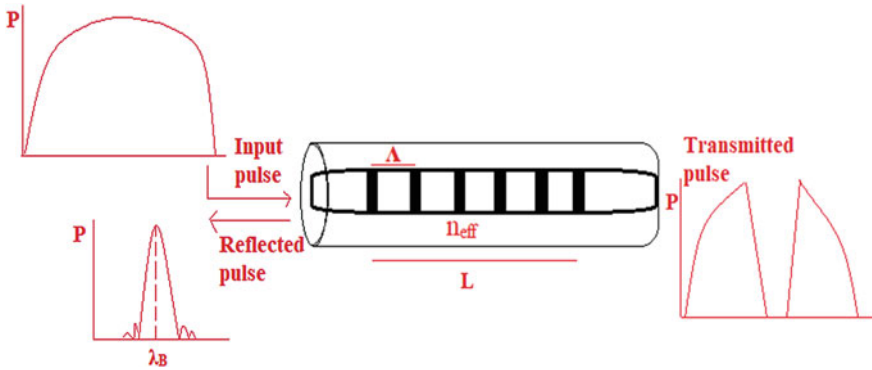


Fig. 2 Fiber Bragg Grating

2.2 Fiber Bragg Grating

FBG designed on a short segment of optical fiber, as shown in Fig. 2, is analogous to a stop band filter centered at wavelength called as Bragg's Wavelength (λ_B). It is a type of Bragg's reflector which works on the principle of Fresnel reflection [3]. FBG is fabricated by modulating the refractive index periodically and the respective period is named as Grating period (Δ). For a specific temperature and strain, Bragg's wavelength could be controlled by controlling the effective refractive index (n_{eff}) and the grating period [1, 9], as represented by the Eq. (1)

$$\lambda_B = 2 \Delta n_{\text{eff}}. \quad (1)$$

If along with the Grating length (L) refractive index modulation is varied, bandwidth of the reflected spectrum broadens as each portion of the grating reflects different spectrum [1, 5, 7]. FBG of such kind is named as Chirped FBG (CFBG) and the technique adapted is known as Chirping. The various chirping techniques that are mostly used are linear chirping, quadratic chirping, square root chirping, and cubic root chirping. By controlling the chirp rate, peak reflectance can be controlled directly [5]. A CFBG provides longer delay to quick moving wavelength components and the shorter delay to fast moving wavelength components [1, 9, 10], hence finds a major application in dispersion compensation in long-haul optical communication. The reflection spectrum of the FBG exhibits side lobes which arises due to the rapid change in index modulation and can be removed by ramping up and down the strength of modulation smoothly [1], such an FBG is termed as Apodized FBG and the technique adapted is named as Apodization. To achieve a certain peak reflectance, overall grating length needs to be increased [4]. There is a trade-off between the side lobe suppression and the maximum peak reflectance for some restricted grating length [4], but both can be leveled to an optimum value by adapting optimum chirping and apodization techniques as shown in the current work.

3 Experimental Results

The reflected spectrum of the FBG is associated with the side lobes which can be suppressed by apodization. Also, the maximum reflectivity can be increased by increasing the grating length. For some constant grating length, side lobe suppression and peak reflectance shows trade-off and needs to be set out at the optimum value. From the work reported in paper [4], Gaussian-Apodized FBG and Tanh-Apodized FBG have shown best performance when we want to have both side lobe suppression, as well as the maximum reflectivity, at the optimum level. To use the FBG for dispersion compensation, bandwidth of the reflected spectrum needs to be increased which can be done by chirping. Paper [5] reports that the bandwidth of the reflected spectrum and the maximum reflectivity can be leveled up to optimum value by choosing the proper grating length and the chirp rate of the CFBG. Which chirping technique gives the best performance for the respective apodization techniques haven't been analyzed mostly? So our main aim in this work is to determine the chirping technique and the grating length of the FBG which gives the finest performance for compensating the dispersion in Gaussian-Apodized FBG and Tanh-Apodized FBG. In the current work, five dispersion compensation models have been implemented and simulated for a 100 km long optical fiber link at the data rate of 10Gbps as shown in Fig. 3 at the parameters described in Table 1. The simulation software used is OptiSystem, in which the performance characteristics of the designed links are observed and compared in terms of the Eye diagram. The Eye diagrams for the five simulated links are shown from Figs. 8 to 12, in which the red curve represents the Q-factor and alongside the various analytical parameters like Maximum Q-Factor, Minimum BER, Eye height, etc., are defined, so that the performance of the links can be analyzed and compared (Table 2).

In the link-1, DCF with the attenuation of 0.2 dB/km and dispersion of -70 ps/nm/km is used as a dispersion compensation module and is connected in cascade with the 100 km long CSMF. The link is analyzed at variable DCF lengths. It has been observed that the highest Q-factor of 36.02 comes into origin when the length of DCF chosen is 24.28 km. To compensate the losses introduced by the DCF itself, an extra EDFA of gain 4 dB is added to the link. The Eye diagram of the simulated link-1 is shown in Fig. 8.

In the link-2 and link-3, Tanh-Apodized FBG and Gaussian-Apodized FBG are used individually as dispersion compensation modules for compensating the dispersion of 100 km long optical fiber link. Both the links are observed for various chirping techniques to discover the optimum chirping technique for the respective FBG(s). Also, the grating lengths of the respective FBG(s) are varied. The analyzed response of the Q-factor at various chirping techniques verses the grating length for both the links are shown in Figs. 4 and 5. It can be seen that the link-2 gives the highest Q-factor of 23.71 for linear chirping at the grating length of 16 mm and the link-3 shows the highest Q-factor of 31.74 for quadratic chirping at the grating length of 41.8 mm. The Eye diagrams for the simulated respective links at the optimum performance characteristics are shown in Figs. 9 and 10.

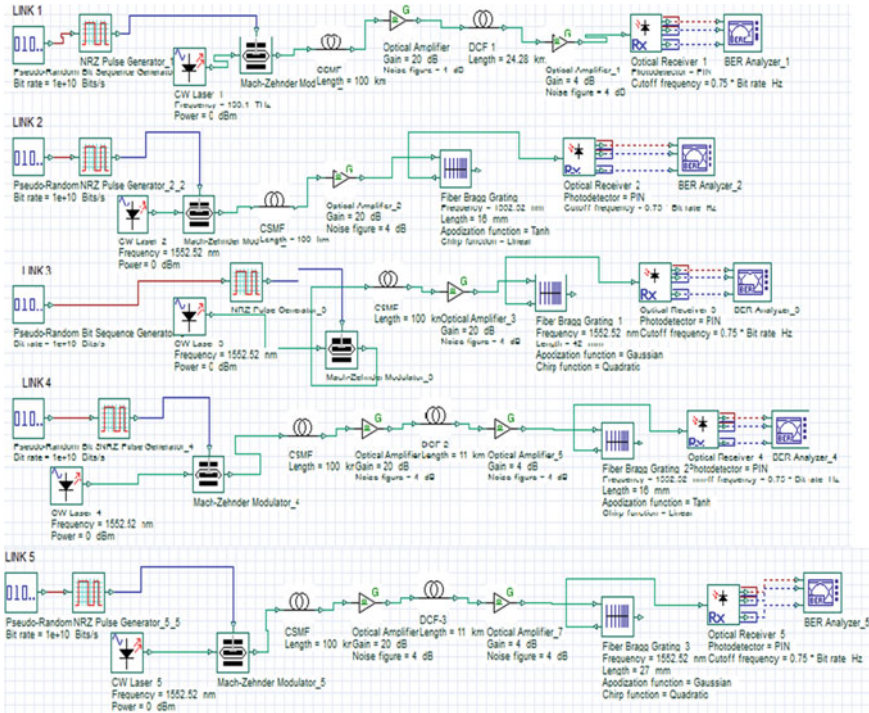


Fig. 3 Simulation set up of five dispersion compensation modules

In the link-4 and link-5, Tanh-Apodized-linear-chirped-FBG and Gaussian-Apodized-Quadratic-chirped-FBG along with the DCF are connected in cascade with the 100 km long CSMF as a hybrid dispersion compensation model. 10, 11, 12, and 13 km long DCF(s) are observed successively at various grating lengths of the respective FBGs as shown in Figs. 6 and 7. It has been observed that the Hybrid dispersion compensation model of 12 km long DCF and Tanh-Apodized linear-chirped FBG gives the highest Q-factor of 19.44 at the grating length of 16.2 mm for the link-4 and the Hybrid dispersion compensation model of 11 km long DCF and Gaussian-Apodized Quadratic-chirped FBG gives the highest Q-factor of 35.22 at the grating length of 26.6 mm for the link-5. The Eye diagrams for the simulated link-4 and link-5 at the optimum performance characteristics are shown in Figs. 11 and 12.

4 Conclusion

The dispersion compensation models used in the five simulated links and their performance are summarized in Table 2. In terms of performance characteristics, it can be

Table 1 Components, parameters and values

S.no	Components	Parameters and values
1	Data source	Bit rate 10Gbps
2	Pulse generator	NRZ
3	CW laser	Frequency 193.1THz Power 0 dBm Line width 10 MHz Noise bandwidth 0THz Noise threshold -100 dB Noise dynamic 3 dB
4	Mach-zender modulator (MZM)	Extinction ratio 30 dB
5	CSMF	Reference wavelength 1552 nm Length 100 km Attenuation 0.2 dB/km Dispersion 16.75 ps/nm/km
6	EDFA	Gain 20 dB Noise Fig. 4 dB
7	Optical receiver	Photodetector PIN Gain 3 Ionization ratio 0.9 Responsivity 1A/W Dark current 10nA Cutoff frequency 0.75* Bit rate

Table 2 Performance analysis of five simulated links

	Dispersion compensation module		Q-factor
	DCF length	FBG used	
Link-1	24.28 km	Not used	36.02
Link-2	0 km	Tanh-Apodized linear-chirped FBG with grating length of 16 mm	23.71
Link-3	0 km	Gaussian-Apodized Quadratic-chirped FBG with grating length of 41.8 mm	31.74
Link-4	12 km	Tanh-Apodized linear-chirped FBG with grating length of 16.2 mm	19.44
Link-5	11 km	Gaussian-Apodized Quadratic-chirped FBG with grating length of 26.6 mm	35.22

concluded that the Link-1, in which the 24.28 km long DCF alone is used as a dispersion compensation module and the Link-5, in which the 11 km long DCF along with Gaussian-Apodized Quadratic-chirped FBG at grating length of 26.6 mm is used as a dispersion compensation module gives the finest eye diagram and the highest Q-factor. The high cost of DCF is always an issue and it can be seen that the link-1 uses almost 2.2 times longer DCF as compared to the link-5. So in terms of cost effectiveness, it can be concluded that Link-5, in which the 11 km long DCF along

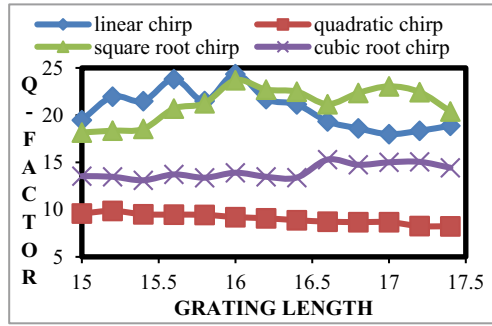


Fig. 4 Q-factor verses grating length for various chirping techniques in Tanh-Apodized FBG

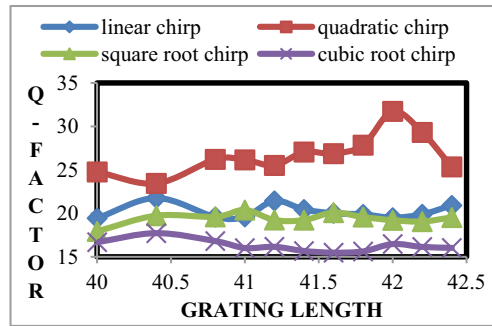


Fig. 5 Q-factor verses grating length for various chirping techniques in Gaussian-Apodized FBG

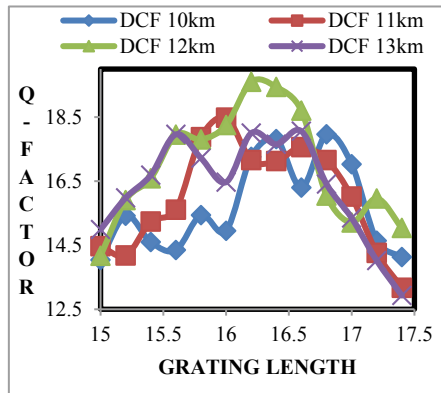


Fig. 6 Q-factor verses grating length for the hybrid model of DCF with Tanh-Apodized Linear-Chirped FBG

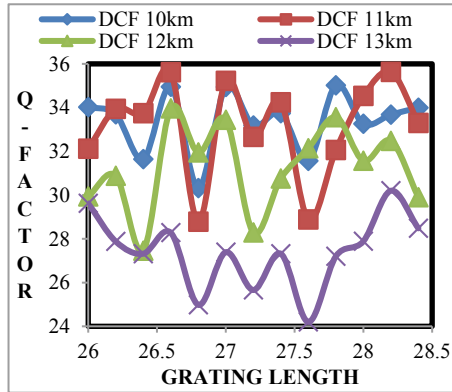


Fig. 7 Q-factor versus grating length for the hybrid model of DCF with Gaussian-Apodized Quadratic-chirped FBG

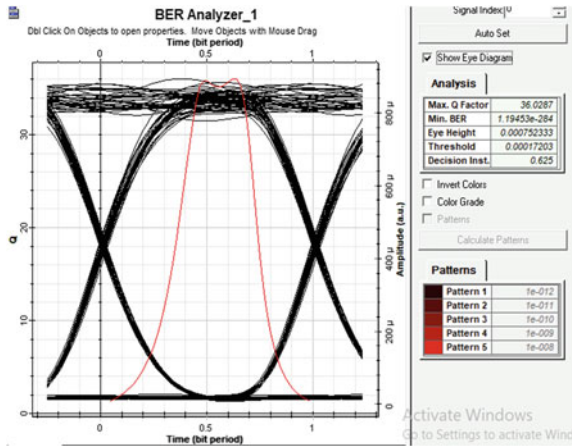


Fig. 8 Eye diagram of link-1

with Gaussian-Apodized Quadratic-chirped FBG at grating length of 26.6 mm is used as a dispersion compensation module makes the link cheaper as compared to Link 1. Hence, it can be concluded that the Gaussian-Apodized Quadratic-chirped FBG with grating length of 26.6 mm along with the 11 km long DCF makes the most efficient dispersion compensation model with the cost reduction of almost 54% as compared to using DCF alone for the 100 km long optical fiber link at the data rate of 10 Gbps Figs. 8, 9, 10, 11 and 12.

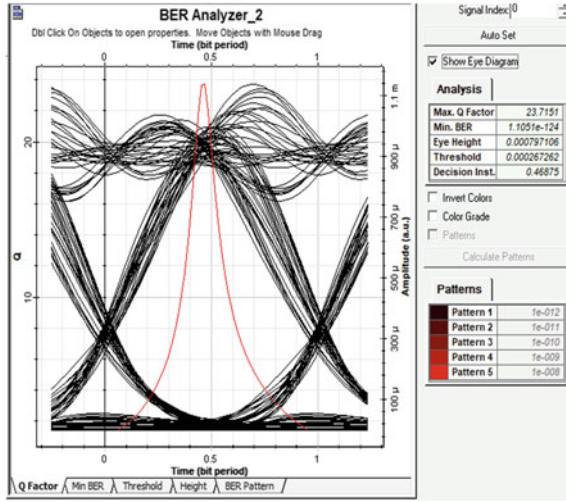


Fig. 9 Eye diagram of link-2

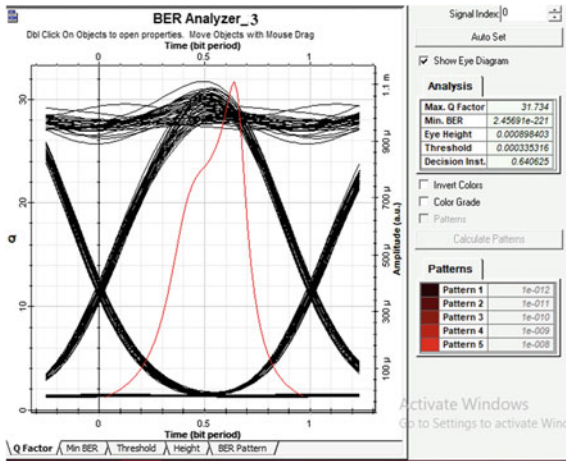


Fig. 10 Eye diagram of link-3

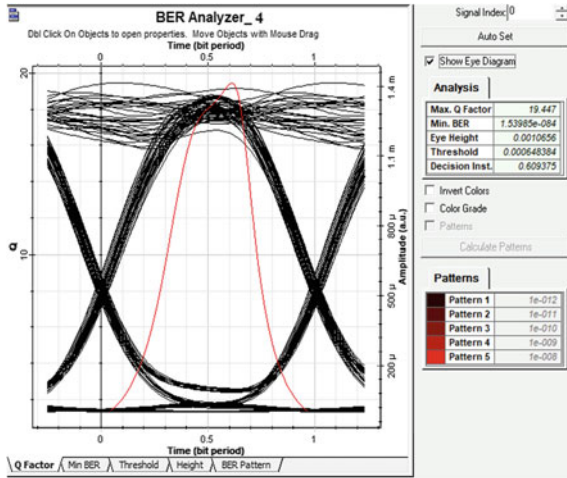


Fig. 11 Eye diagram of link-4

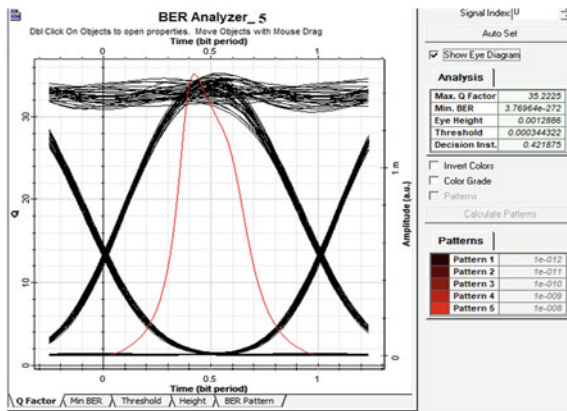


Fig. 12 Eye diagram of link-5

References

1. Mohammed, N. A., Solaiman, M., & Aly, M. H. (2014). Design and performance evaluation of a dispersion compensation unit using several chirping functions in a tanh apodized FBG and comparison with dispersion compensation fiber. *Applied Optics*, 53(9), 10.
2. Ahmed, R. K., & Mahmood H. A. (2017). Performance valuation of high data rate optical communication system utilizing FBG compensated Dispersion schemes under different Modulation techniques. *Diyala Journal of Engineering Sciences Iraq*, 10(02), 94–106.
3. Sharma, A., Singh, I., Bhattacharya, S., & Sharma, S. (2019). Performance comparison of DCF and FBG as dispersion compensation techniques at 100 Gbps over 120 km using SMF. In *Springer Nature Singapore Pte Ltd. 2019 V. Nath and J. K. Mandal (eds.), Nanoelectronics, circuits and communication systems, Lecture notes in electrical engineering 511*.

4. Ashry, I., Elrashidi, A., Mahros, A., Alhaddad, M., & Elleithy, K. (2014). Investigating the performance of apodized Fiber bragg gratings for sensing applications. In *IEEE Proceedings of the Zone 1 Conference of the American Society for Engineering Education*.
5. Tosi, D. (2018). Review of chirped fiber bragg grating (CFBG) fiber-optic sensors and their applications. *Sensors*, 18, 2147.
6. Pal, B. P., & Pande, K. (2002). Optimization of a dual-core dispersion slope compensating fiber for DWDM transmission in the 1480–1610 nm band through G.652 single-mode fibers. *Elsevier, Optics Communications*, 201, 335–344.
7. Hussein, T. F., Rizk, M. R. M., Aly, M. H. (2019). A hybrid DCF/FBG scheme for dispersion compensation over a 300 km SMF. In *Optical and quantum electronics*, Springer Science + Business Media, LLC, part of Springer Nature.
8. Khan, S. S. A., & Islam, M. S. (2012). Determination of the best apodization function and grating length of linearly chirped fiber bragg grating for dispersion Compensation. *Journal of communications*, 7(11).
9. Meena, M. L., & Gupta, R. K. (2019). Design and comparative performance evaluation of chirped FBG dispersion compensation with DCF technique for DWDM optical transmission systems. *Optik—International Journal for Light and Electron Optics*, 188, 212–224.
10. Spolitis, S., & Ivanovs, G. (2011). Extending the reach of DWDM-PON access network using chromatic dispersion compensation. *IEEE*, pp. 29–33.

Performance Comparison of Adaptive Mobility Management Scheme with IEEE 802.11s to Handle Internet Traffic



Abhishek Majumder and Sudipta Roy

Abstract Wireless Mesh Network (WMN) is one type of mobile ad hoc network. So, if there is a link breakage, it can heal and reorganize the network by itself. Portals, Mesh Gates and Access Points, Stations (STAs) and external station (STA) are main components of WMN. Portals are used to connect the network to the internet. STAs are the nodes through which routing of packets take place. External STAs are the users of WMN. Mesh Gates and Access Points are the stations which route the packets and also act as Access Points for the external STAs. In case of WMN, IEEE 802.11 s is the standard. When the external STA moves in the WMN, it goes out of coverage of a Mesh Gate and Access Point and enters into the coverage of another one. Hybrid Wireless Mesh Protocol (HWMP), routing protocol of IEEE 802.11s, does not support external Station's (STA's) mobility. To integrate mobility, Adaptive mobility management technique had been proposed. The scheme considers session and mobility activities of the external STA by taking into account its session-to-mobility ratio (SMR). Based on SMR value, it makes the decision about transmission of route management packets. This paper presents numerical analysis of adaptive mobility management scheme and HWMP. Simulation of both techniques are carried out using NS-2. Performance comparison shows that adaptive mobility management scheme outperforms HWMP. Moreover, a comparison between the numerical analysis and simulation results have also been performed. It has been observed that results in case of HWMP are very close.

Keywords Portal · Mesh station · Mesh gateway · IEEE 802.11s · External station

A. Majumder (✉)

Department of Computer Science & Engineering, Tripura University, Suryamaninagar, Tripura, India

S. Roy

Department of Computer Science & Engineering, Assam University, Silchar, Assam, India

1 Introduction

Portal, Mesh station (STA), external station (STA), Mesh Gate and Access Point form Wireless Mesh Network (WMN) [1–5]. Portals are the mesh nodes that connect the WMN to the internet. Mesh STAs receive and send packets. They also relay packets of other nodes. Mesh gates and access points (APMG) are the nodes which serve network connection to clients. External STAs are the client nodes to whom network services of WMN are provided. The internet packets are sent and received to and from the remote nodes.

In WMN, the users using external STA are free to move. Therefore, handoff process is performed. During this process, the online activities of external STA gets interrupted. So, it is necessary to have seamless mobility management. It is desirable to have minimum route management packet propagation for better performance. Reduced routing overhead will result in enhanced performance.

For IEEE 802.11s [6–10] the routing protocol is Hybrid Wireless Mesh Protocol (HWMP) [9, 11–13]. Uplink packets are transmitted by external STA towards APMG which are subsequently forwarded towards portal. APMG receives downstream internet packets from portal and forwards them towards external STA. Mobility is not supported in IEEE 802.11s [14]. Mobility [7, 14] management becomes an important challenge with increase in external STA's number and mobility. For mitigating the challenge, Adaptive mobility management technique [15] had been designed. In this work, adaptive mobility management technique and IEEE 802.11s are numerically analyzed considering only internet traffic. Simulation results of both the schemes are also presented under varying data rate and node speed.

In this paper, Sect. 2 presents working of HWMP. In Sect. 3, working of adaptive mobility management scheme is presented. Section 4 discusses numerical analysis of both schemes. Section 5, discusses the simulation. In Sect. 6, comparison of results of simulation and numerical analysis is discussed. Conclusion has been presented in Sect. 7.

2 Hybrid Wireless Mesh Protocol

The scheme [9, 12, 13] utilizes the following message: path error (PERR), path request (PREQ), route announcement (RANN), path reply (PREP), proxy update confirmation (PXUC), and proxy update (PXU). Portal advertises its presence through broadcast of RANN. PREQ is broadcasted by source mesh node for finding route to destination. The destination in response sends back PREP message. For updating proxy information the mesh gate transmits PXU message. In response, the PXUC is sent back. PERR is broadcasted to propagate the information that some route is broken.

In proactive RANN mode following steps are carried out during handoff process:

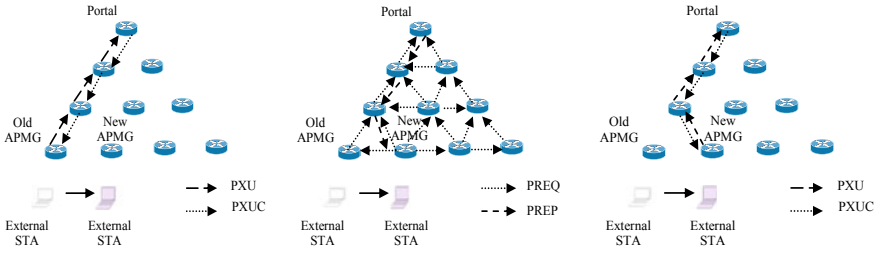


Fig. 1 a Previous APMG sending PXU to portal. b Process for setting up route when no route present towards destination. c Process for setting up route when route is present to the destination

- i. Old APMG transmits PXU towards portal. Portal removes external STA’s old proxy table entry and transmits back PXUC. Figure 1a shows the operation.
- ii. When handoff occurs, external STA searches the route towards destination. If no route is found, PREQ is broadcasted. If portal receives PREQ, PREP is sent back. Figure 1b shows the process. If it has a route, PXU is transmitted to the portal. In response, PXUC is sent back towards source APMG. Figure 1c shows the process.

3 Adaptive Mobility Management Technique

Adaptive mobility management technique [15, 16] is presented in this section. It has been designed for handling internet traffic only. Routing and mobility management are its major components.

If handoff occurs, the external STA computes session-to-mobility ratio (SMR) [17] value. Then it sends SMR value with re-association message. When SMR value (SMR_{ESTA}) is received by APMG, it compares the value with threshold δth .

- If $SMR_{APMG} \leq \delta th$, new external STA’s entry will be created in routing table of the new APMG. PREQ will be sent towards old APMG. When it receives PREQ, it updates external STA’s routing table and next hop will be set to new APMG’s address.
- If $SMR_{APMG} > \delta th$, in current APMG’s proxy table external STA’s information will be appended. PXU will be transmitted by new APMG towards portal. Portal updates its proxy table if it receives PXU message. So, forward chain gets reset.

Since RANN is periodically transmitted by the portal each APMG has route towards portal. Upstream packet is sent towards serving APMG by source external STA. It forwards packet to the portal. Downlink packet is received by portal. After that, it forwards received packet towards serving APMG of external STA. When packet is received, in proxy table of serving APMG, search is carried out for external STA’s entry. If it is present, packet is transmitted towards external STA. Else, packet will be transmitted towards APMG which is next hop to reach external STA.

Fig. 2 Example scenario when $SMR_{APMG} \geq \delta th$

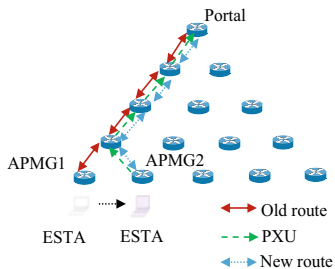


Fig. 3 Example scenario when $SMR_{APMG} < \delta th$

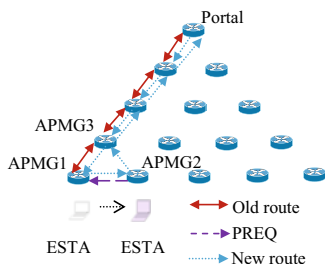


Figure 2 shows example of routing and mobility management. Let external STA (ESTA) moves from the coverage of APMG1 to APMG2. ESTA sends its current SMR value to APMG2 with re-association message. Let, external STA's SMR is greater than δth . So, APMG2 sends PXU message to portal. Portal updates proxy information of source external STA. Therefore, a path is established between source ESTA and the portal.

Figure 3 shows another scenario. This scenario is similar to the previous one. Only difference is that, here, the value of SMR is less than δth . When handoff takes place, APMG2 sends PREQ to APMG1. ESTA's entry in routing table of APMG1 will be updated and APMG2 is set as next hop. Upstream packets of internet are transmitted by APMG2 via APMG3. Portal sends downstream internet packets towards APMG1. It transmits the packet towards APMG2.

4 Numerical Analysis

Let, exponential distribution with rate λ_s [18] is followed by external STA's subnet residence time t_s . It is also considered that the inter session departure and arrival time follows exponential distribution having rates λ_d and λ_a , respectively [19]. Let t_a and t_b be the inter session arrival and departure time. Let the number of external STA is M . Between portal and arbitrary APMG, average hop count distance is considered as α . Let, γ is communication latency per hop. Value of threshold SMR is considered

as δth . Let percentage of downstream internet packet is r_{inter} . In a session, average packet number per time unit is considered as λ_p [17].

4.1 Handoff Cost

In adaptive mobility management scheme, SMR is calculated by external STA during handoff. SMR is calculated as [17]

$$SMR_{APMG} = \frac{t_s}{t_a + t_d} \quad (1)$$

Therefore, probability of SMR_{APMG} be less than δth can be calculated as [17],

$$P_{th} = P(SMR_{APMG} < \delta th) = \begin{cases} \frac{\lambda_a \lambda_d \lambda_s \delta th}{\lambda_d - \lambda_a} \times \left[\frac{1}{\lambda_a (\delta th \lambda_s + \lambda_a)} - \frac{1}{\lambda_d (\delta th \lambda_s + \lambda_d)} \right] & \text{if } \lambda_a \neq \lambda_d \\ 1 - \frac{\lambda_s^2}{(\lambda_s \delta th + \lambda)^2} & \text{if } \lambda_a = \lambda_d = \lambda \end{cases} \quad (2)$$

In adaptive mobility management technique PXU is sent towards portal if external STA's SMR value becomes greater than or equal to δth . Otherwise, PREQ is sent towards old APMG by new APMG. Therefore, cost for this process is

$$C_{hSMR_{newAPMG}} = \gamma \text{ if } SMR_{APMG} < \delta th \\ = \alpha \times \gamma \text{ if } SMR_{APMG} \geq \delta th \quad (3)$$

For this process number of messages transferred is

$$CM_{hSMR_{newAPMG}} = 1 \text{ if } SMR_{APMG} < \delta th \\ = \alpha \text{ if } SMR_{APMG} \geq \delta th \quad (4)$$

Therefore, per handoff cost is computed as

$$C_{hSMR} = \gamma \times P(SMR_{APMG} < \delta th) + \alpha \times \gamma \times P(SMR_{APMG} \geq \delta th) \\ = \gamma \times P_{th} + \alpha \times \gamma \times (1 - P_{th}) \quad (5)$$

Number of messages transmitted per handoff is

$$CM_{hSMR} = 1 \times P(SMR_{APMG} < \delta th) + \alpha \times P(SMR_{APMG} \geq \delta th) \\ = P_{th} + \alpha \times (1 - P_{th}) \quad (6)$$

In this work, internet traffic is considered only. So, in HWMP during handoff, PXU is sent to portal by old APMG. The portal responds by sending PXUC to old APMG. The cost for PXU and PXUC message is [20]

$$C_{\text{hHWMPoldAPMG}} = 2 \times \alpha \times \gamma \quad (7)$$

The number of message transferred during the procedure is

$$CM_{\text{hHWMPoldAPMG}} = 2\alpha \quad (8)$$

The new APMG transmits PXU toward portal when there is proxy information of destination or route to destination. PXUC is transmitted by portal in response. Otherwise, PREQ is broadcasted by APMG for finding route towards destination. Intermediate APMG rebroadcasts PREQ message. Let the probability of having proxy entry or route to the destination node in the new APMG is P_d . Therefore, cost incurred for transmission and reception of messages to and from the new APMG is

$$C_{\text{hHWMPnewAPMG}} = \{(1 - P_d) \times (M + \alpha) + P_d \times 2 \times \alpha\} \times \gamma \quad (9)$$

Number of messages transferred during this process is

$$C_{\text{hHWMPnewAPMG}} = (1 - P_d) \times (M + \alpha) + P_d \times 2 \times \alpha \quad (10)$$

So, per handoff cost can be calculated as

$$C_{\text{hHWMP}} = (C_{\text{hHWMPoldAPMG}} + C_{\text{hHWMPnewAPMG}}) \quad (11)$$

In case of HWMP, number of messages transferred during each handoff is

$$CM_{\text{hHWMP}} = (CM_{\text{hHWMPoldAPMG}} + CM_{\text{hHWMPnewAPMG}}) \quad (12)$$

4.2 Packet Delivery Cost

With respect of external STA's serving APMG, let its average displacement per APMG association be c_{move} [16]. For downlink internet, traffic packet delivery cost is [16]

$$C_{\text{pdSMR}} = \alpha \times \gamma + \gamma \times c_{\text{move}} \times P_{\text{th}} \times \left[\frac{1 - P_{\text{th}}^{\left(\frac{M-1}{c_{\text{move}}}\right)}}{1 - P_{\text{th}}} \right] \quad (13)$$

Uplink internet traffic is sent to portal without tunneling. So, the cost of packet delivery for uplink traffic is computed as [16]

$$C_{puSMR} = \alpha \times \gamma \quad (14)$$

So, per second packet delivery cost is

$$C_{pSMR} = \left\{ \frac{r_{inter}}{100} \times C_{pdSMR} + \left(1 - \frac{r_{inter}}{100} \right) \times C_{puSMR} \right\} \times (\lambda_a + \lambda_d) \times \lambda_p \quad (15)$$

In case of HWMP, downlink internet packet is directly transferred to external STA's current APMG. Similarly, uplink packets are transferred directly towards the portal by external STA's current APMG. So, in case of uplink and downlink traffic, the cost for packet delivery is computed as

$$C_{puHWMP} = C_{pdHWMP} = \alpha \times \gamma \quad (16)$$

So, cost per second for packet delivery is computed as

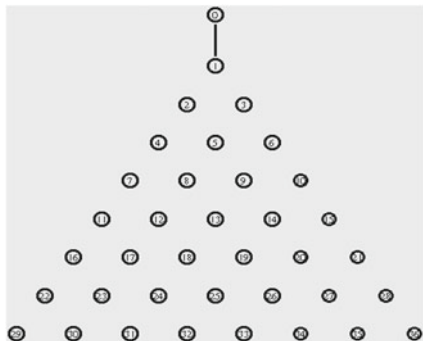
$$C_{pHWMP} = \alpha \times \gamma \times (\lambda_a + \lambda_d) \times \lambda_p \quad (17)$$

5 Comparison and Simulation

Simulation results have been presented in this section. NS-2 [21] is used for carrying out simulation. Random waypoint mobility model is used as mobility model of external STA. MAC layer uses IEEE 802.11. 10 times simulation has been repeated. Figure 4 shows the topology.

δ th is set to 0.5. Packet size is considered as 258. Node 1 is the portal of the WMN shown in Fig. 4. CBR traffic is considered. External STA's pause time is 0 s. External

Fig. 4 Mesh STAs' topology used



STA's number in the topology is 36. 40 m has been set as two neighboring mesh STA's distance. Radio range of each of the mesh STA is 41.04 m. The simulation has been carried out for 100 s. During simulation, the maximum number of connections at a time is 10.

5.1 Impact of Network Load

External STA's maximum speed is considered as 10 m/s. Data rate is considered as 32, 64, 128, 256, 384, 512, 640, 768, 896, and 1024 kbps. External STA's number has been set as 10.

Figure 5 shows data rate's impact on throughput of IEEE 802.11s as well as adaptive mobility management technique. According to the figure, both scheme's throughput initially increases with rise in data rate, but reduces in later part. Through network, more data packets are sent, as data rate increases. Therefore, at initial part, as data rate increases, there is an increase in throughput also. If data rate is increased further, network's congestion also increases and throughput of network decreases. Since large quantity of PXUC, PXU, and PREQ messages are transmitted in IEEE 802.11s, congestion will be more. Therefore, as data rate increases throughput of IEEE 802.11s will increase at lower rate compared to adaptive mobility management scheme.

Data rate's impact upon normalized overhead for routing is shown in Fig. 6. As data rate increases, network gets congested. So, route management packets' retransmission will increase. In IEEE 802.11s, PPREQ, PXUC, and PPREQ packets are transmitted in large number. Therefore, retransmission of great quantity of route management packets also takes place. So, in IEEE 802.11s, since the data rate increases, normalized routing overhead will also be raised. Number of PXU and PREQ packets transmitted is limited in adaptive mobility management technique.

It can be seen from Fig. 7 that for IEEE 802.11s, as rate of data increases, the average cost of handoff also increases. In adaptive mobility management technique, route management packets' retransmission is limited. So, average handoff cost

Fig. 5 Throughput versus data rate

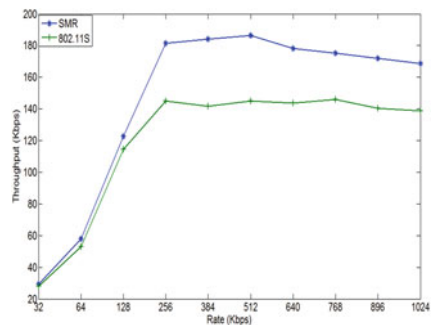


Fig. 6 Normalized routing overhead versus data rate

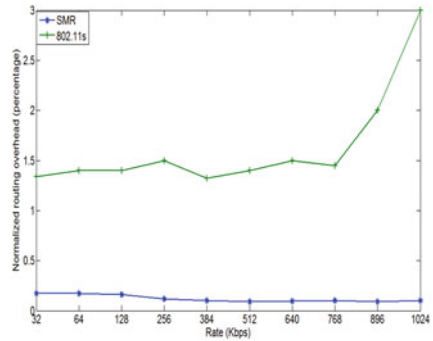
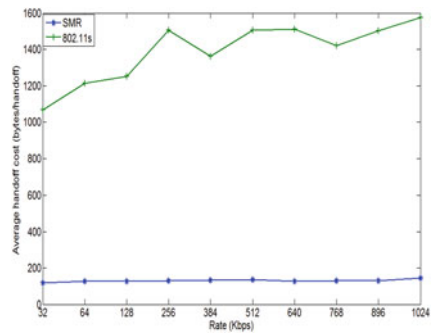
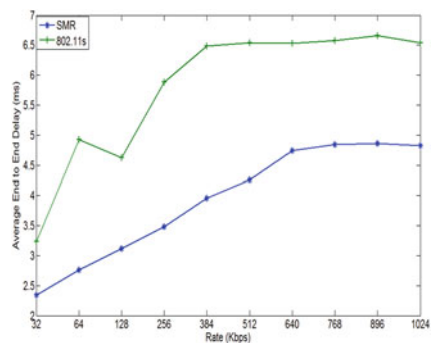


Fig. 7 Avg. handoff cost versus rate of data



continues to be steady as data rate increases average. It can be seen from Fig. 8 that in IEEE 802.11s as rate of data rises, the average end-to-end delay increases. Reason behind this is the congestion caused by increasing rate of data and retransmission of great quantity of route management packets. In adaptive mobility management technique, route management packets are less propagated although the congestion increases as rate of data increases. So, average end-to-end delay also becomes higher as data rate increases. The increase is less in comparison with IEEE 802.11s.

Fig. 8 Avg. end-to-end delay versus rate of data



5.2 Impact of External STA Speed

The section discusses the impact of speed of external STA upon performance of adaptive mobility management technique and HWMP. 512 kbps is set as data rate. 10 has been set as external STA's number. External STA's maximum speed is set between 10 and 100 m/sec.

Figure 9 shows impact of node speed on normalized routing overhead. Handoff occurs more frequently if external STA's speed becomes high. Therefore, more PXUC, PXU, and PREQ messages are transmitted in IEEE 802.11s as external STA's speed increases. It leads to rise in normalized routing overhead. In the adaptive mobility management technique, there is very limited propagation of PXU and PREQ messages.

Figure 10 presents impact of node speed upon average cost for handoff. In IEEE 802.11s transmission of great quantity of management packets occur when handover takes place frequently due to increase in external STA's speed. This causes congestion and route management packets' retransmission in the network. Therefore, as external STAs' maximum speed increase, average cost for handoff also increases. In adaptive mobility management scheme, route management packet's propagation is very less.

Fig. 9 Normalized routing overhead versus external STA's maximum speed

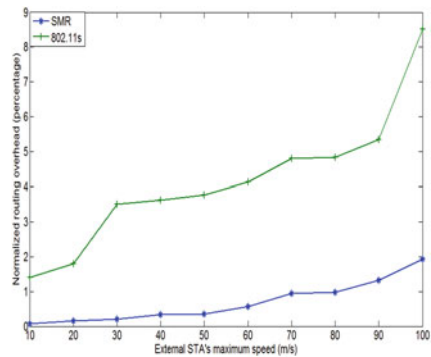


Fig. 10 Average handoff cost versus external STA's maximum speed

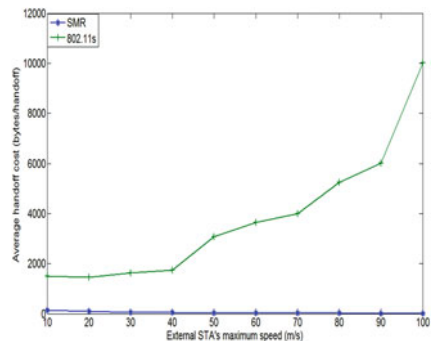
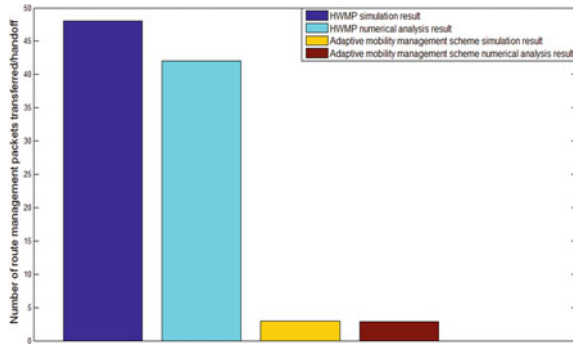


Fig. 11 Number of route management packets transmitted per handoff obtained from simulation and numerical analysis



This results in less congestion. So, with rise in external STA’s maximum speed handoff cost remains stable.

6 Comparison of Numerical Results and Simulation

Figure 11 depicts comparison of numerical analysis and simulation results. Considering topology given in Fig. 4 the value of M and α are considered as 36 and 4, respectively. The value of $\lambda_s, \lambda_a, \lambda_d, \delta th,$ and P_d have been considered as, 0.025882, 0.02, 0.02, 0.5, and 0.2, respectively. For minimizing the effect of congestion, data rate and maximum node speed have been set as 512 Kbps and 10 m/s, respectively. External STA’s number is considered as 10. In SMR based scheme, results are very close. In HWMP, results are not that close. The main reason is packet dropping and congestion.

7 Conclusion and Future Work

Comparative study has been carried out in this paper between adaptive mobility management technique and IEEE 802.11s. Both the schemes are simulated using NS-2 and numerically analyzed. The schemes have been simulated in NS-2. Simulation results have been compared considering average handoff cost, normalized routing overhead, avg. end-to-end delay and throughput with respect to rate of data and external STA’s speed. From the comparison, it was observed that the adaptive mobility management technique outperforms IEEE 802.11s. A comparison of results of numerical analysis and simulation was carried out. It has been found that in case of adaptive mobility management scheme, results obtained from simulation and numerical are very close. In HWMP, the results are not that close. In future comparison of adaptive mobility management technique will be carried out with other baselines considering more parameters.

References

1. Zachos, C. K., & Loo, J. (2016). Wireless mesh network: design, modeling, simulation, and analysis. In J. Loo, J. L. Mauri, J. H. Urtiz (eds.), *Mobile Ad Hoc networks: Current status and future trends*, CRC Press, pp. 449–466.
2. Akyildiz, I. F., Wang, X., & Wang, W. (2005). Wireless mesh networks: A survey. *Computer Networks*, 47(4), 445–487.
3. Akyildiz, F. (2005). A survey on wireless mesh networks. *IEEE Communication Magazine*, 43(9), S23–S30.
4. Zhang, Y., Luo, J., & Hu, H. (2006). *Wireless mesh networking architectures*. Protocols and standards: Auerbach Publications.
5. Benyamina, D., Hafid, A. S., & Gendreau, M. (2012). Wireless mesh networks design—A survey. *IEEE Communications Surveys and Tutorials*, 14(2), 299–310.
6. Hiertz, G. R., Denteneer, D., Max, S., Taori, R., Cardona, J., & Berlemann, L., et al. (2010). IEEE 802.11s: The WLAN mesh standard. *IEEE Wireless Communication*, 17(1), 104–111.
7. IEEE Std 802.11-2012. (2010). Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements—Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications.
8. IEEE P802.11s/D1.07-2007. (2007). Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements—Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications.
9. Camp, J. D., & Knightly, E. W. (2008). The IEEE 802.11s extended service set mesh networking standard. *IEEE Communications Magazine*, 46(8), 120–126.
10. Wang, X., & Lim, A. O. (2008). IEEE 802.11s wireless mesh networks: Framework and challenges. *Ad Hoc Networks*, 6(6), 970–984.
11. Sooriyaarachchi, S. J., Fernando, W. A. C., & Gamage, C. D. (2015). Evaluation of scalability of Hybrid Wireless Mesh Protocol in IEEE 802.11. In *Proceedings of the Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)* (pp. 152–159), Colombo, Sri Lanka.
12. Bari, S. M. S., Anwar, F., & Masud, M. H. (2012). Performance study of hybrid wireless mesh protocol (HWMP) for IEEE 802.11 s WLAN mesh networks. In *Proceedings of the International Conference on Computer and Communication Engineering* (pp. 712–716), Kuala Lumpur, Malaysia.
13. Bahr, M. (2006). Proposed routing for IEEE 802.11s WLAN mesh networks. In *Proceedings of the 2nd Annual International Workshop on Wireless Internet* (p. 5), Boston, USA.
14. Sampaio, S., Soutp, P., & Vasques, F. (2015). A review of scalability and topological stability issues in IEEE 802.11s wireless mesh networks deployments. *International Journal of Communication Systems*, 29(4), 671–693.
15. Majumder, A., & Roy, S. (2020). Implementation of adaptive mobility management technique for Wireless Mesh Network to Handle Internet Packets. In *Proceedings of the 2nd International Conference on Communication, Devices and Computing* (pp 97–107), Haldia, India.
16. Majumder, A., Roy, S., & Dhar, K. K., (2013). Design and analysis of an adaptive mobility management scheme for handling internet traffic in wireless mesh network. In *Proceedings of the Annual International Conference on Emerging Research Areas* (pp. 1–6), Kanjirapally, India.
17. Majumder, A., & Roy, S. (2013). Design and analysis of a dynamic mobility management scheme for wireless mesh network. *Scientific World Journal, Article ID, 656259*, 1–16.
18. Pack, S., Jung, H., Kwon, T., & Choi, Y. (2005). SNC: a selective neighbor caching scheme for fast handoff in ieee 802.11 wireless networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(4), 39–49.

19. Kwon, S. J., Nam, S. Y., Hwang, H. Y., & Sung, D. K. (2004). Analysis of a mobility management scheme considering battery power conservation in IP-based mobile networks. *IEEE Transactions on Vehicular Technology*, 53(6), 1882–1890.
20. Majumder, A., & Roy, S. (2015). Implementation of forward pointer-based routing scheme for wireless mesh network. *Arabian Journal for Science and Engineering*, 41(3), 1109–1127.
21. Fall, K., & Varadhan, K. (2007). The network simulator (ns-2). <http://www.isi.edu/nsnam/ns>.

Automatic Attendance Management System Using Face Detection and Face Recognition



M. Varsha and S. Chitra Nair

Abstract Attendance plays a major role in every education system. Taking attendance of students manually can be a great burden for teachers. It may cause many problems like loss of time, repetition, incorrect markings, and difficulties in marking them. To avoid this, there is a need to design an automatic system that overcomes the issues with the traditional attendance system. There are many automatic methods available for this purpose like fingerprint systems, RFID systems, face recognition systems and iris recognition systems, etc. But among these Face Recognition proved to be more efficient. The main objective of this paper is to propose a model that captures images from videos, detect and recognize the faces, predict the recognized face, and then marks attendance. In this work, a basic step has been performed which uses fifteen classes from LFW Dataset and faces are detected, recognized, and then prediction is done on a randomly selected image from the used dataset. This system uses a combination of Multi Task Cascaded Neural Network (MTCNN) algorithm along with FaceNet that can be used to detect faces and extract facial features from images. SVM is used to predict the face of the person from the image. The proposed system obtains a classification accuracy of 99.177% for the training set and 100% on the test set. The accuracy, precision, recall, and F1-score are computed.

Keywords Face detection · Face recognition · MTCNN · FaceNet · SVM

1 Introduction

Students attendance plays a crucial role in every education system. When attendance is taken manually, it should be marked in a sheet which causes time loss, repetition, and incorrect marking. This may cause a great burden for teachers. So, there is a requirement for an automatic attendance management system. There are various automatic techniques available now like Face Recognition, Finger-print verification, RFID systems, Iris Recognition systems, Voice and Speech recognition systems [1, 2]. Among this, face recognition proved to be more efficient [2]. This is because in

M. Varsha (✉) · S. Chitra Nair
NSS College of Engineering, Palakkad, India

the case of the fingerprint verification system, the students have to form a queue to scan their thumb on the scanning device to mark their attendance [3], which causes time loss. In the case of RFID systems, the students will have to carry RFID cards, which is a weakness. There is a chance for the students to forget to take the RFID cards [4]. In the case of iris recognition systems, even though this method has very high accuracy, it is not practically possible. This is because there will be many pupils in the class and iris detection of each student is not feasible [3]. The major drawback of the voice/speech recognition system is that the speech features are sensitive to certain factors like background noise, and also the voice change that happens with age. These systems are not reliable and also may not work accurately if the students are suffering from throat infection [1].

Face detection is a technique of locating and localizing the faces in an image by creating a bounding box through the extent. A facial recognition is a technology that identifies and verifies a person from images or video frames. It is also specified as a Biometric Artificial Intelligence-based application. This is because, it will uniquely recognize a person by analyzing their facial features. There are many methods in which facial recognition works, but mainly they work by comparing the facial features with the known faces in a database to find a match.

In this work, an automatic system is developed using image processing techniques that can mark attendance automatically. An effective face recognition algorithm is proposed which can identify students efficiently. For image processing, an effective platform is used. The objective of this work is to propose a model that captures images from videos, detect and recognize the faces, predict the recognized face, and then marks attendance. This paper aims to build an automatic attendance management system that detects the faces of students from images using MTCNN, extracts the facial features using FaceNet model and classifies using SVM. After recognizing the faces, the attendance is marked. Initially, we worked on LFW dataset and the future work is to work on the image captured on real-time videos. The Sect. 2 of this paper presents related work, Sect. 3 presents motivation, Sect. 4 demonstrates the proposed system, Sect. 5 shows the Experimental results, and Sect. 6 shows the conclusion and future work.

2 Related Works

Over the last few decades, the research area is mostly focused on face detection and recognition. This section provides a detailed overview of the different methods used for implementing automatic attendance management systems.

2.1 Attendance System Using Image Processing Techniques

The idea behind this work [5], was to develop an automatic system which will handle the attendance of students using image processing techniques. Their aim was to develop a system that detects and recognizes the students faces using frames from videos and then recording their attendance by identifying them from their variant features. For this, they used Viola Jones algorithm and Fisher Face algorithm for detecting and recognizing faces, respectively. The proposed system achieved an accuracy of 45–50%. They were able to overcome the drawback of manual attendance systems. The efficiency of this system can be increased by improving the training process.

2.2 Attendance System Using Face Detection and Face Recognition

In this work [6], they used Haar filtered Adaboost for detecting faces and Principal Component Analysis (PCA) and Local Binary Pattern Histogram (LBPH) algorithms for identifying the detected faces. This system proposed an approach that provides better results than traditional attendance systems and other automatic attendance systems like biometric fingerprint and RFID systems. They obtained an accuracy in the range 75–90% during the experiments.

2.3 Attendance System Using Deep Learning Framework

In this work [7], they proposed a system using deep learning frameworks. They have used a state-of-the-art face detection model and a novel recognition architecture for detecting and recognizing faces respectively. CNN was used to develop the automated attendance system. They proposed a system that achieved an accuracy of 98.67% on LFW datasets and 100% on classroom datasets. They used a spatial transformer network to learn the alignment of faces which led to greater facial verification accuracy. They have mainly used the frontal face images of students.

2.4 Deep Learning Paradigm for Attendance Systems

In this work [8], they proposed a system using Convolutional Neural Networks. The captured frames were passed to the SRNet (Single Image Super-Resolution Network) for the image super resolution which results in SR images. The faces were detected from the SR image and cropped around the bounding boxes using MTCNN. Then

FaceNet was used for identifying the detected faces. They used RAISE and DIV2K for SRNet, VGGface2 for FaceNet and LFW and their own dataset for testing and validation. They produced an accuracy of 96.80% on LFW datasets.

2.5 Attendance System Using FaceNet and SVM

In this work [9], their objective was to get an improved accuracy for multi-face recognition. The proposed system uses FaceNet for feature extraction and SVM as classifier. They achieved an accuracy of 99.6% for multi-face recognition. The overall accuracy which they obtained by using CNN model was less than the accuracy which they obtained using FaceNet and SVM.

3 Motivation

In recent years, Image processing played a unique role in technological advancement as it deals with extracting useful information from a digital image. As the applications of these techniques are increasing day by day, a lot of research are being done in this field. Image capturing plays an important role in the educational field, robotics, the medical field, and in smart phones. A classroom consists of a huge number of students and marking their attendance manually wastes a lot of time. So, there is a need for an automatic system that can be built using current technology. Face recognition, which is an application derived from image processing, is one of the best methods used for human detection. It helps in detecting the face of each student. Face is multi-dimensional structure and it requires good computational analysis for recognition.

4 Proposed System

An automatic attendance management system using face detection and recognition are proposed in this paper. The proposed system aims to take the images of the students from classroom videos, to detect and identify their faces from images and upon successful recognition to mark their attendance automatically. Figure 1 illustrates the model representation of the proposed system. The system takes real-time classroom videos as input. These videos are captured and frames are created in the pre-processing phase. The faces of students are detected from the frames. After detection of a face, the system crops the face and performs certain image processing techniques. Then the features are extracted from the detected faces. Lastly, the features of the detected faces and the test faces are compared in the classifier and the faces are recognized. Then the prediction is generated.

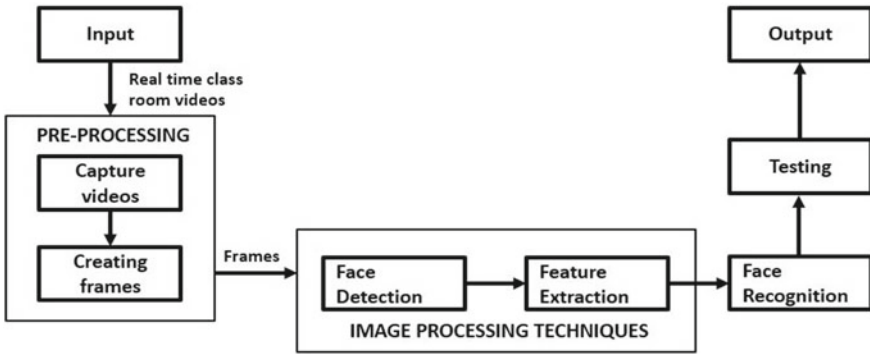


Fig. 1 Model representation

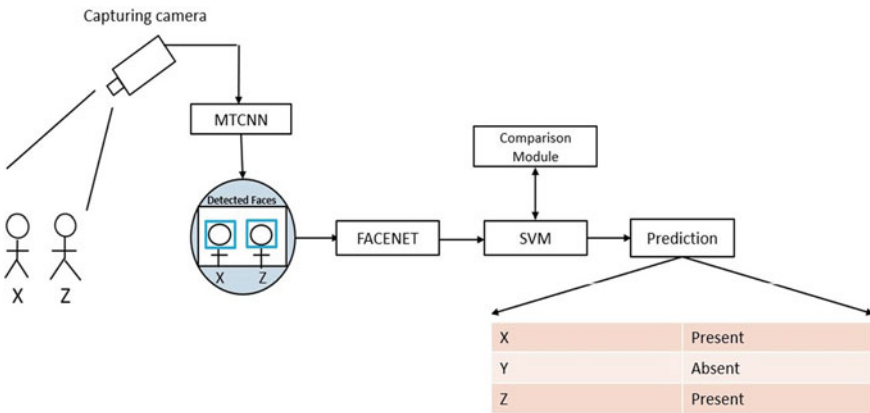


Fig. 2 System design

The proposed system requires a camera mount to a surface in the classroom at a point where it could capture the images of all the students in the classroom. The Fig. 2 illustrates the detailed system design of the proposed model. Here, MTCNN is used for face detection [8, 10]. FaceNet is used for creating face embeddings for a given image and SVM is used as a classifier for classification [9, 10]. The working of the proposed model is explained in brief below.

4.1 Input

The input to the model is real-time classroom videos of the students. Camera should be placed in the classroom in such a way that it captures all the students and their

faces effectively. This camera needs to be interfaced to the computer system for further processing either through a wired or a wireless network.

4.2 *Image Pre-Processing*

In this phase, frames are created from videos. For this work, a sample video was taken from the internet and the frames were created. The size of the sample video was 0.34 s. The frames were created as one frame per 0.07 s.

4.3 *Image Processing*

In this phase, the faces are detected from the images and face embeddings are created from the detected faces. These features are used to recognize the faces of the people to mark their attendance. In the proposed system, MTCNN is used for creating Face Detector and FaceNet model is used for creating face embeddings for each detected face. And SVM classifier is used to predict the name of the given face [4]. The sub-steps are explained in brief below:

Face Detection. Face detection is the process of identifying and extracting the faces from images. MTCNN is used for face detection. This is a state-of-the-art deep learning model for face detection, described in [10, 11]. It helped in localizing the faces from images and created bounding boxes around their extent. The first step was to detect faces from images and to reduce the dataset to a set of faces. For this, the images were loaded as Numpy array and converted to RGB if in case the image is black and white. Then the MTCNN Face Detector was created. This detected the faces from the image and new dataset was obtained [4]. Figure 3 depicts the reduced dataset. Here, the faces are detected from each image of one class and the dataset is reduced to a series of faces only. The images are resized to $(160 \times 160 \times 3)$ which is the input shape of FaceNet. All the images in the train and test sets are loaded and the faces are extracted.

Feature Extraction. In this section, face embeddings are created. Face embeddings are vectors that represent the features extracted from the face [4]. These embeddings are then compared with the vectors created for other faces. The FaceNet model will create the face embeddings for a face [12]. This model will return the face embeddings of the train and test datasets. Each face embedding is comprised of 128 vectors.

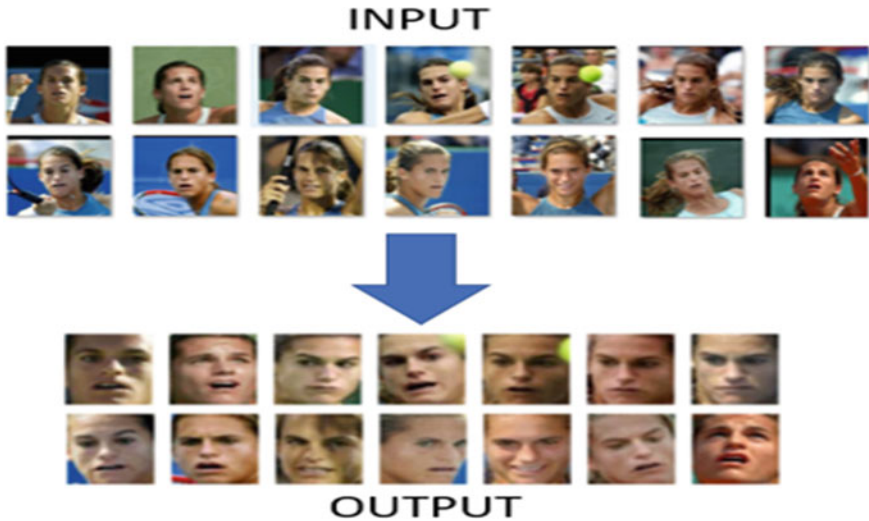


Fig. 3 Reduced dataset

4.4 Face Classification

The next step is to develop a model for classifying the face embeddings. First, the face embedding vectors are normalized. Here, SVM is used to work with face embedding. This is because SVM is efficient in separating the face embedding vectors [4]. Next, the model is evaluated and then the classification accuracy is calculated.

4.5 Testing

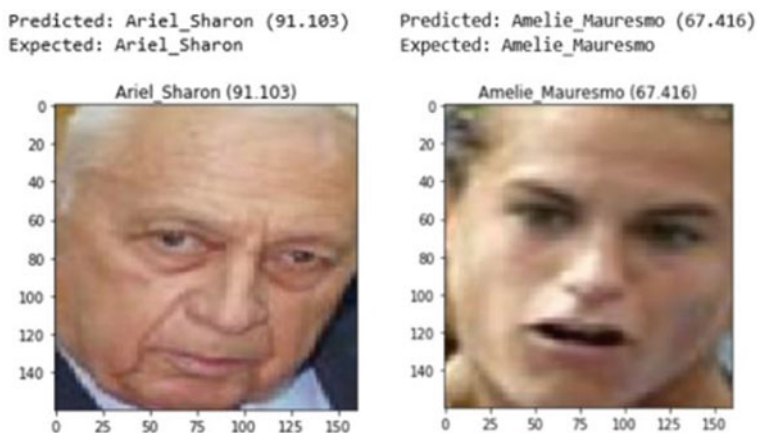
For testing, we selected a random image from the test dataset. The face embeddings are used as input to make predictions. The expected class name, predicted class name, and probability of the prediction is obtained.

5 Experimental Results

The dataset used here is LFW datasets [13]. For this work, 15 classes of image data were taken from the LFW datasets for the implementation of this system. The dataset was divided into two sets: Train and test sets. For training 80% of images were used and for testing 20% of the images were used. Table 1, describes the classification accuracy obtained with a different number of classes. This system produced a classification accuracy of 99.177% for training and 100% for testing

Table 1 Classification accuracy

LFW dataset	Train accuracy (%)	Test accuracy (%)
5 Classes	99.401	100
10 Classes	99.099	100
15 Classes	99.177	100

**Fig. 4** Prediction done by SVM classifier [13]

with 15 classes. And in case of 10 classes, it provided an accuracy of 99.99% for training and 100% for testing.

FaceNet extracts high quality features from faces and creates facial embeddings. For prediction, a random image was selected from the test set. The system correctly recognized the faces from the images. The proposed system produced a good accuracy in recognizing the faces which were randomly taken from the test set [9]. The Fig. 4 depicts the face recognized from the random image taken from the test set. A plot of the selected face along with its expected name, predicted name, and probability is given. The accuracy, precision, recall, and F1-score of the system are also computed. The classification performance for the classes showing precision, recall and F1-score is shown in Table 2. This model provided an accuracy_score of 1.

6 Conclusion and Future Work

Automatic Attendance System has been proposed to reduce the drawbacks of traditional (manual) systems. The proposed system produced a good result with provided datasets. It saves time, especially during a lecture with a large strength of pupils. This attendance system illustrates the use of image processing techniques in a classroom. The system can be improved by implementing it in real time by capturing the images

Table 2 Classification performance

Classes	Precision	Recall	F1-score	Classes	Precision	Recall	F1-score
0	1.00	1.00	1.00	7	1.00	1.00	1.00
1	1.00	1.00	1.00	8	1.00	1.00	1.00
2	1.00	1.00	1.00	9	1.00	1.00	1.00
3	1.00	1.00	1.00	10	1.00	1.00	1.00
4	1.00	1.00	1.00	11	1.00	1.00	1.00
5	1.00	1.00	1.00	12	1.00	1.00	1.00
6	1.00	1.00	1.00	13	1.00	1.00	1.00

from real-time videos, detect faces from the created frames, recognize the faces, and then marks the attendance which is the future work.

Sentimental analysis and emotional analysis can also be implemented in the future. Sentimental analysis provides the feedback of a class which shows on which topics and at what time attracts more concentration of the pupils. Emotional analysis will help the faculties to acquire the feedback of their class and this helps them to change or improve their pedagogy. These developments can improve the applications of the work.

References

- Jayant, N. K., & Borra, S. (2016). Attendance management system using hybrid face recognition techniques. In *2016 Conference on advances in signal processing (CASP)*, June 9, pp. 412–417.
- Islam, M. S., Mahmud, A., Papeya, A. A., Onny, I. S., & Uddin, J. (2017). A combined feature extraction method for automated face recognition in classroom environment. In *International symposium on signal processing and intelligent recognition systems*, September, pp. 417–426.
- Varadharajan, E., Dharani, R., Jeevitha, S., Kavinmathi, B., & Hemalatha, S. (2016). Automatic attendance management system using face detection. In *2016 Online international conference on green engineering and technologies (IC-GET)*, November 19, pp. 1–3.
- Arsenovic, M., Sladojevic, S., Anderla, A., & Stefanovic, D. (2017). FaceTime—Deep learning based face recognition attendance system. In *2017 IEEE 15th International symposium on intelligent systems and informatics (SISY)*, September 14, pp. 000053–000058.
- Hapani, S., Prabhu, N., Parakhiya, N., & Paghdal, M. (2018). Automated attendance system using image processing. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, August 16, pp. 1–5.
- Sanli, O., & Ilgen, B. (2018). Face detection and recognition for automatic attendance system. In *Proceedings of SAI intelligent systems conference*, September 6, pp. 237–245.
- Sarkar, P. R., Mishra, D., & Subhramanyam, G. R. S. (2019). Automatic attendance system using deep learning framework. In *Machine intelligence and signal analysis*, pp. 335–346.
- Gupta, R. K., Lakhani, S., Khedawala, Z., Chudasama, V., Upla, K. P. (2018). A deep learning paradigm for automated face attendance. In *Workshop on computer vision applications*, December 18, pp. 39–50.
- Nyein, T., Oo, A. N. (2019). University classroom attendance system using facenet and support vector machine. In *2019 International conference on advanced information technologies (ICAIT)*, November 6, pp. 171–176.

10. Brownlee, J. (2020). Machine learning algorithms in python, machine learning mastery. <https://machinelearningmastery.com/machine-learning-with-python/>. Accessed 8 June 2020.
11. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10), 1499–1503.
12. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*; pp. 815–823.
13. Huang, G. B., Mattar, M., Berg, T., Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*, October.

ESIT: An Enhanced Lightweight Algorithm for Secure Internet of Things



Manoja Kumar Nayak and Prasanta Kumar Swain

Abstract With the increasing use of sensors and intelligent devices, Internet of Things (IoT) becomes an important area of research to establish connectivity among connected devices. Traditional algorithms used for encryption are found to be highly complex and with higher number of rounds for encryption which is computationally expensive. In an IoT network, communicating nodes adapt fewer complex algorithms for maintaining security. Hence, lightweight security algorithms are used for this purpose. In this paper, we have proposed a method for encryption and decryption process called Enhanced Secure IoT (ESIT) of an image using shift (\ll or \gg) and bitwise binary modulo 2 ($+_2$) operation for data transmission. It is a block cipher that accepts a 64-bit key which performs Left shift (\ll) and ($+_2$) function for encryption, and right shift (\gg) and ($+_2$) decryption. It is the normal bitwise left and right shift by removing the sequence of q -bits from the left and right side, respectively. The performance of the proposed method is established by correlation, entropy, and image histogram. Experimental evaluation on four available images clearly shows the advantages of the given approach.

Keywords IoT · Security · Encryption · Decryption

1 Introduction

The idea of IoT was coined by Kevin Ashton at Auto-ID center of MIT in 1999 [1]. In the twenty-first century, IoT becomes an important part in the lifestyle of the modern human-being to make their life smart and easy. It impacts on different fields of human life to make life smart, such as education, manufacturing automation, retail, financial technology, healthcare, governance, power utilities, agriculture, home, city, companies, and individuals. So, IoT has been a buzzword of the present days in communication technology due to intelligent computation. To provide services from Device to Devices (D to Ds), Devices to Device (Ds to D), Devices to Devices (Ds

M. K. Nayak · P. K. Swain (✉)

Department of Computer Application, Mharaja Sriram Chandra Bhanja Deo University, Baripada, India

to Ds) in an organization or to individual users, IoT employs millions of things like sensors, actuators, smart sensors, modern vehicles, bar-codes, smart cards, and RFID tags that are connected and providing benefits in various situations to exchange data among them in internet. These things will be responsible to gather and transfer related information to the gateway or cluster head or cloud potentially which are thousands of miles away [2]. The IoT allows “People Things” to be connected every time, every place with any things, and anyone using every services [3]. At present, smart military camp at eastern Ladhak is ready for Indian army people due to the technological innovation of IoT. As more and more people use the social media application and internet, there are number of challenges in IoT platform such as communication, management security, and monitoring. Among all these challenges, security is the top-most one. For sensitivity to IoT applications, it is desirable that devices are authenticated to ensure the data source. The sensible devices are known as resource constraint devices as they have low computation power, small size, less memory, low bandwidth, and limited battery life [4].

The use of tiny devices in IoT that are sensing heterogeneous access, data processing, services, and applications are prime parts in the platform. So IoT platforms are affected from information destruction level attacks such as interruption attack, eavesdropping attack, modification attack, fabrication, man-in-middle attack, and message Replay attack, and so on [5]. Due to above basis of attacks to information, the current state of research on key technology including encryption, secure communication, sensor data protection, and cryptographic procedure are the discussed challenges. To protect data from malicious event, it is a significant challenge for IoT technology. Nowadays, IoT is a smart technology, as peer-to-peer communication occurs efficiently due to the use of lightweight cryptographic algorithms that play a key role in achieving the security and privacy with constraint resources [6]. To reduce weakness of information transmission in IoT environment, the Enhanced lightweight cryptographic (ESIT) procedure is highly essential.

2 Related Works

A block cipher comprises of three sections. (1) Encryption process (2) Decryption process and (3) Key generation process. The large sub-keys are generated from the user secret key using the key generation algorithm. Through sub-keys, this algorithm allows all the secret key bits to impact encryption process in each round. Most secret key generation requires, once to serve both encryption and decryption process. Parameter of lightweight block ciphers which are found to be important are (1) block size (2) key length (3) operations (4) key scheduling (5) minimum number of rounds [7]. The lightweight ciphers are selected for resource constraint devices, and therefore, software and hardware implementation should be optimized. The major purpose of lightweight block cipher is to secure information confidentiality, integrity against security threat in IoT environment. By considering key parameters, such as size of block, key size, the architecture, and the number of rounds, we started an effort

to include some of the light weight block ciphers as explained in the literature. A modern block cipher with a moderate code size requirement is proposed by Rijndael [8]. It is resistant against linear crypto analysis and differential attacks. Here, the key attack is possible when number of rounds is less than 9. With optimum parameters, it acts as an energy-efficient cipher. Authors propose an algorithm called Skipjack in [9], that is, the block cipher of 64-bit data operates on 80-bits key. Due to its shortest extended key and need of less memory and code, the space requirement is minimum. This uses energy efficiently than Rijndael, but it is found to be a more efficient cipher when speed is optimized. The fact is that it is not declared as secure by NSA. In 1999, Skipjack best attacked with lower than 32 rounds are known as Biham. RC5 is a cipher that supports both hardware and software implementation. It is represented as RC5-w/r/b where w denotes the word size, r denotes number of rounds, and b is the key length [10]. RC5 has poor energy efficiency due to Archille's heel of MSP430F149 used for multiplication and rotations. RC5 requires very small code memory. In 1998, it was attacked which breaks RC5-32/12/16 with just 244 chosen plain texts. Here, it is prescribed that the number of rounds must be greater than or equal to 18 for the security. Camellia is a Feistel structure cipher on 6 rounds that needs 13×28 plain text, with complexity of 2112 for cipher executions [11]. RC5 and RC6 are less energy-efficient than Camellia. This required large code size and more memory space. The attack to the Camellia is named as sequential attack. The resource constraint application for the lightweight algorithms like HEIGHT, TEA, KLEIN, and KATAN are executed on AVR Atmel. It is a tiny micro controller to assess memory efficiency, consumption of energy, and to evaluate level of confusion and diffusion for security investigation in [12]. Authors in [13] proposed an ultra-lightweight cipher named PRESENT which is a SP network having 31 rounds. The encryption algorithm DES is used based on 16 round Feistel cipher [14].

Due to key size of 56-bits, it is attacked by exhaustive attack that can break DES. So, DES is not secure today. 3-DES or Triple-DES is found to be more secure as it is of 48 rounds. In [15], the new block cipher SIT is a lightweight block cipher that uses 64-bits information with 64-bits key. To combine Shannon's confusion and di-

fusion principle, substitution-permutation (SP) network is used in AES block cipher. The Feistel structure is used by block ciphers Blowfish and DES taking the advantages of same encryption and decryption operations. SIT combines both Feistel Structure and SP network using principles of both to provide security. This procedure has two parts: Key generation and encryption. The use of key of 64-bits is segmented into 4-blocks and fed into SP network (Here F-function) organized in 4×4 matrices and generates new five unique keys.

3 Proposed Framework

We have proposed an enhanced version of lightweight algorithm for security in IoT data communication entitled Enhanced Secure IoT (ESIT) algorithm. Here, we have taken a plain text and 64-bit keys to generate a 64-bit block cipher and used in ESIT.

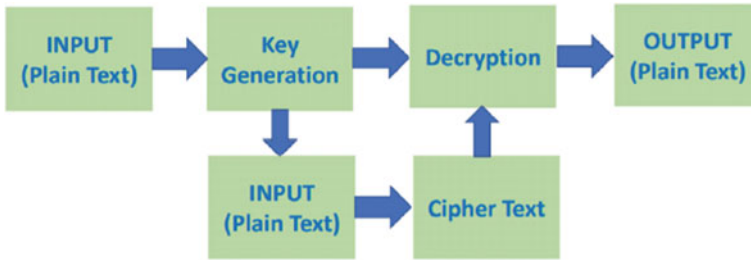


Fig. 1 Overview of the proposed ESIT framework

The proposed work follows the sequence such as: (1) Key Generation, (2) Encryption, (3) Decryption. Figure 1 depicts our proposed model. Each component of the process is described below.

3.1 Key Generation

Figure 2 shows the structure of the key generation. The detailed procedure of key generation is described as below:

In this case, a 64-bit key (K_c) is input to the key generation block and five unique keys are generated. Key generation block (as shown in Fig. 2b) uses Khazad block

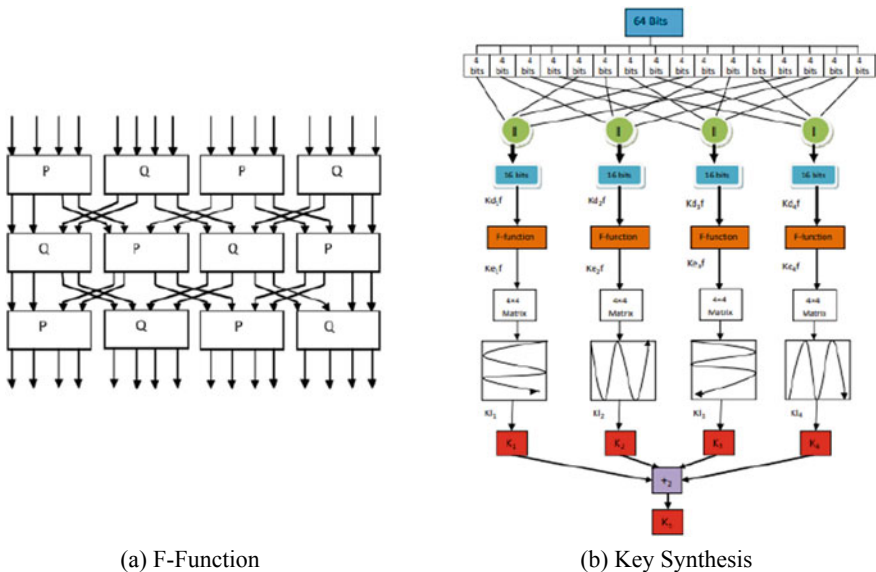


Fig. 2 Flowchart of key synthesis

cipher [16]. To obtain multiple keys, 64-bits key (K_c) is divided into equal 4-bit units. The F-function (as shown in Fig. 2a) is implemented on 16-bits data. Here, all the 4 F-function blocks are used and can be represented as:

$$k_{(d_i)}f = \prod_{j=1}^4 k_{c(j-1)+i} \quad (1)$$

where i deviates from 1 to 4, in case of initial four round as shown in Fig. 2a. In the next step, the 16-bits of $K_{d_i}f$ is used with the f -function to get K_{e_i} and it is defined as:

$$k_{e_i}f = f(k_{d_i}f) \quad (2)$$

The F-function, the confusion, and diffusion are created and explained in [15] to transform linear and non-linear functions. Here, P and Q are represented from Khazads mini-box and presented in tables (a) and (b) (as explained in [16]). To prove F-function, a 4×4 matrix is used, called K_m as explained in [15].

The value of initial four round keys K_1 , K_2 , K_3 , and K_4 is obtained by converting their matrices into 4 arrays of 16-bit. This is called as special key of round keys (Kr). Then $+_2$ operations are conducted among the K_1 , K_2 , K_3 , and K_4 , to get key K_5 and it is defined as:

$$K_5 = +_{2_{i=1}}^4 K_i \quad (3)$$

3.2 Encryption

In the proposed method, the round keys that are generated from key generation process are applied for encryption. A plain text of size 64-bit (P_i) is segmented into four 16-bits text in the first round. Each of these can be noted as P_{X0-15} , P_{X16-31} , P_{X32-47} , P_{X48-63} . In each round, with the passing of bits to decrease data originality, the order of bits is swapped, and then the level of confusion increases in the originated cipher text. The respective round key K_i perform the bitwise XNOR to obtain key from the key generation process much before time. P_{X0-15} and key is executed with K_i and P_{X48-63} to obtain the respective results Ro_{11} and Ro_{14} . The result of XNOR is input to the F-function and output generated as Ef_{11} and Ef_{r1} and shown in Fig. 2. Left shift function (\ll) and bitwise binary modulo 2 ($+_2$) function is used between Ef_{11} and P_{X32-47} to get Ro_{12} , and Ef_{r1} and P_{X16-31} to get Ro_{13} and is given by:

$$\mathbf{R}_{o_{i,j}} = \begin{cases} P_{x_{i,j}} \odot K_i & \text{for } j = 1 \text{ and } 4 \\ (P_{x_{i,j+1}} \ll q) + 2E_{f_{li}} & \text{for } j = 2 \\ (P_{x_{i,j-1}} \ll q) + 2E_{f_{ri}} & \text{for } j = 3 \end{cases} \quad (4)$$

Similar process is imitated for the rest of the rounds by Eq. (4). The output of the first round added to generate Cipher Text (C_i) is given by (Fig. 3):

$$C_i = R_{51} + R_{52} + R_{53} + R_{54}$$

3.3 Decryption

In our proposed method, the symmetric key is used for encryption and decryption. So, the Decryption is the overturn process, i.e., on the move unintelligible cipher image back to plain image. It makes use of accurate keys (above five distinctive keys from the key generation part), where the diffusion function is primarily applied, afterward substitution operations, XNOR, right shift operation (\gg) with Bitwise binary modulo2 ($+_2$), swapping, and left shifting operations in the same order.

4 Observations

Here, an experiment has been conducted with 4 images as input as in SIT [15]. The format of the pictures used are Portable Network Graphics (PNG) and each image with dimension of 256 X 256 pixels. The output is measured on the basis correlation, entropy, and histogram. All simulations are conducted on a PC with Intel core-i5-3.10 GHz and 8 GB of memory.

4.1 Evaluation Parameters

- **Correlation:** In our proposed algorithm, correlation coefficient is determined for both the real and encrypted images. The output shows that for a plain image, the nearby pixels have better correlation. Hence, the encryption process of proposed (ESIT) algorithm is exciting. Table 1 shows the correlation and correlation coefficient with high delay.
- **Image Entropy:** Image entropy is used to express number of data coded by an encryption algorithm. After image encryption, it is measured with higher entropy for higher security of encryption method as shown in Table 1.

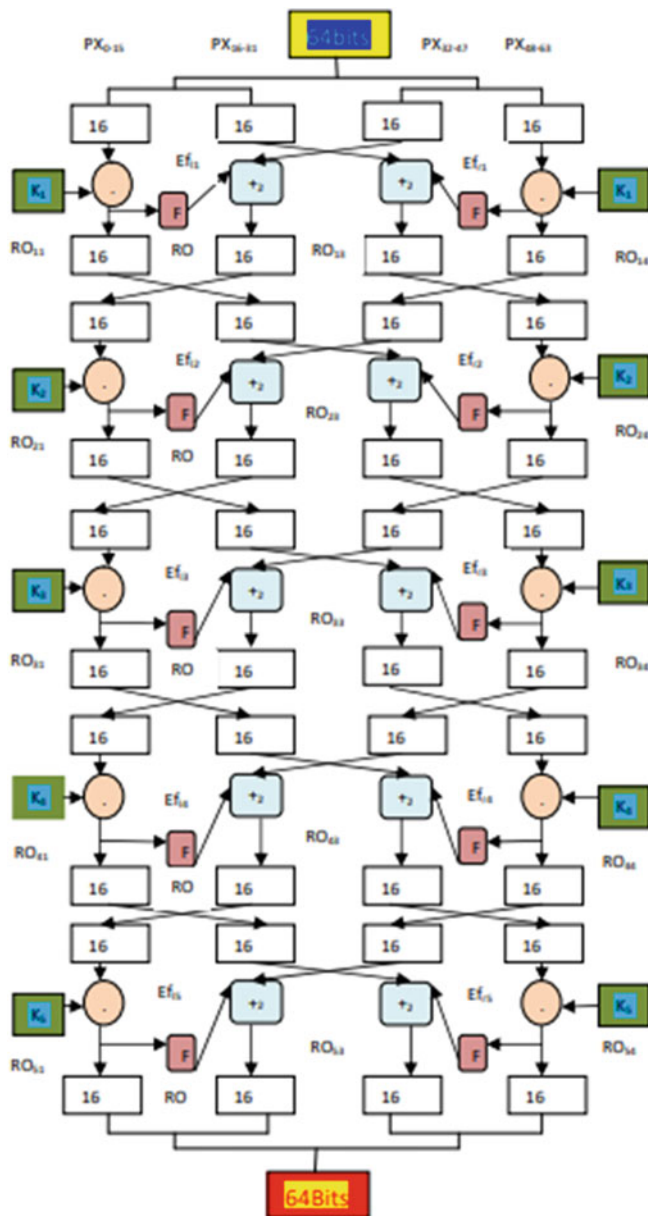


Fig. 3 Flowchart of encryption process

Table 1 Comparative study of Encrypted images with respect to correlation and entropy

Image	Measured Parameters	Usman et al. [15]	ESIT (ours)
Lena	Correlation	0.0012	0.0012
	entropy	7.9973	7.9979
Baboon	Correlation	0.0023	0.0031
	entropy	7.9972	7.9976
Cameraman	Correlation	0.0012	0.0012
	entropy	7.9973	7.9977
Panda	Correlation	0.0022	0.0031
	entropy	7.9971	7.9978

- Image Histogram:** The security strength of a lightweight algorithm is detected through a powerful technique named image histogram. As shown in Fig. 4, the steep lines points to the number of pixels, and the straight line points to the intensity value for histogram. After encryption (attacks are not possible), pixels are equally speeded in the encrypted images (original images of Lena and Cameraman) in

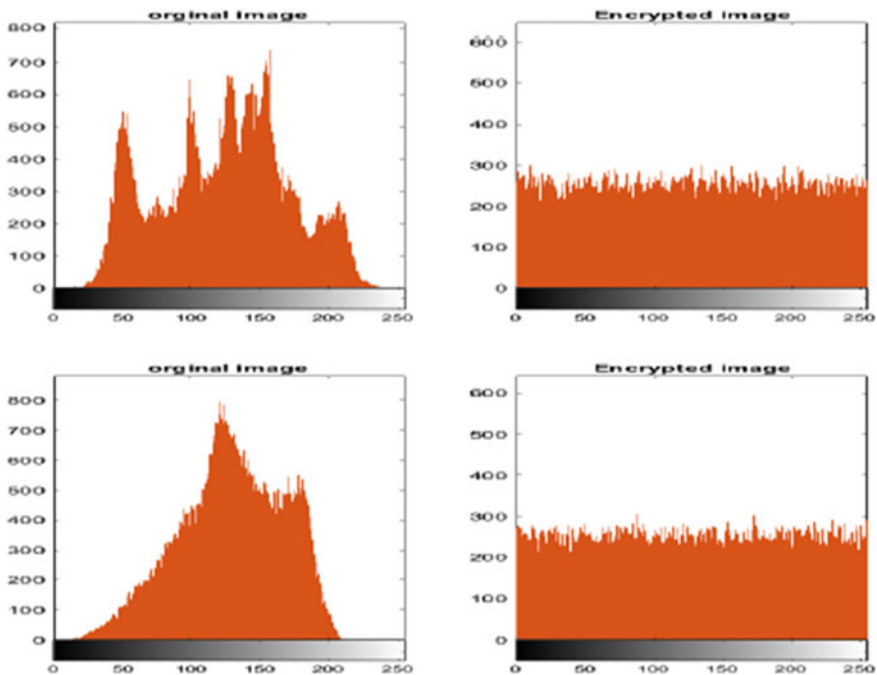


Fig. 4 Image histogram: X-axis represents intensity value and Y-axis represents number of pixels for image of Lena (first row) and Baboon (second row), First column (original Image) and second column (encrypted image)

left-hand side and corresponding encryption image in right end. Which indicates lightweight algorithms secured as the encryption is uniform.

5 Conclusion

This paper presents a lightweight cryptography technique, which is one of the important part to secure cipher for resource constraint applications in IoT system. With new developments, research proceeds with increasing implementation of IoT enabled systems; there is a significant demand of strong security for smart things, services, and applications. Therefore, our proposed lightweight block cipher technique, ESIT, will provide security and privacy in data communication in resource constraint applications of IoT platform.

References

1. Sundmaeker, H., Guillemin, P., Friess, P., & Woelffle, S. (2010). Vision and challenges for realising the internet of things. *Cluster of European Research Project on the Internet of Things, European Commission*, 3(3), 34–36.
2. Singh, S., Sharma, P. K., Moon, S. Y., & Park, J. H. (2017). Advanced lightweight encryption algorithms for IoT devices: Survey, challenges and solutions. *Journal of Ambient Intelligence and Humanized Computing*, pp.1–18.
3. Sadeghi, A. R., Wachsmann, C., & Waidner, M. (2015). Security and privacy challenges in industrial internet of things. In *Proceedings of 52nd ACM/EDAC/IEEE design automation conference (DAC)*, pp. 1–6.
4. Rana, S., Hossain, S., Shoun, H. I., & Kashem, M. A. (2018). An effective lightweight cryptographic algorithm to secure resource-constrained devices. *Spectrum*, 9(11), 1–9.
5. Hossain, M., Hasan, R., & Skjellum, A. (2017). Securing the internet of things: A meta-study of challenges, approaches, and open problems. In *Proceedings of IEEE 37th international conference on distributed computing systems workshops (ICDCSW)*, pp. 220–225.
6. Dhanda, S. S., Singh, B., & Jindal, P. (2020). Lightweight cryptography: A solution to secure IoT. *Wireless Personal Communications*, 112(3), 1947–1980.
7. Cazorla, M., Marquet, K., & Minier, M. (2013) Survey and benchmark of lightweight block ciphers for wireless sensor networks. In *Proceedings international conference on security and cryptography (SECRYPT)*, pp. 1–6.
8. Daemen, J., & Rijmen, V. (1999). *AES proposal: Rijndael*.
9. Biham, E., Biryukov, A., Shamir, A. (1999). Cryptanalysis of skipjack reduced to 31 rounds using impossible differentials. In *Proceedings international conference on the theory and applications of cryptographic techniques* (pp. 12–23). Berlin, Heidelberg: Springer.
10. Biryukov, A., & Kushilevitz, E. (1998). Improved cryptanalysis of RC5. In *Proceedings of international conference on the theory and applications of cryptographic techniques* (pp. 85–99). Berlin, Heidelberg: Springer.
11. He, Y., & Qing, S. (2001). Square attack on reduced camellia cipher. In *Proceedings of international conference on information and communications security* (pp. 238–245). Berlin, Heidelberg: Springer.
12. Jha, V. K. (2011). Cryptanalysis of lightweight block ciphers. Aalto University School of Science Degree Programme of Computer Science and Engineering, Master's Thesis.

13. Bogdanov, A., Knudsen, L. R., Leander, G., Paar, C., Poschmann, A., Robshaw, M. J., Seurin, Y., & Vikkelsoe, C. (2007) PRESENT: An ultra-lightweight block cipher. In *Proceedings of international workshop on cryptographic hardware and embedded systems* (pp. 450–466). Berlin, Heidelberg: Springer.
14. Blaze, M., Diffie, W., Rivest, R. L., Schneier, B., & Shimomura, T. (1996). Minimal key lengths for symmetric ciphers to provide adequate commercial security. *A Report by an Ad Hoc Group of Cryptographers and Computer Scientists*. Information Assurance Technology Analysis Centre Falls Church VA.
15. Usman, M., Ahmed, I., Aslam, M. I., Khan, S., & Shah, U. A. (2017). SIT: A lightweight encryption algorithm for secure internet of things. [arXiv:1704.08688](https://arxiv.org/abs/1704.08688).
16. Barreto, P. S. L. M., & Rijmen, V. (2000). The Khazad legacy-level block cipher. Primitive Submitted to NESSIE. 97, p.106.

A Novel Block Diagonalization Algorithm to Suppress Inter-user Interference in a Multi-user MIMO System



Harsha Gurdasani, A. G. Ananth, and N. Thangadurai

Abstract Diversity in MIMO applications tend to ameliorate the architecture of system for compensation with upgraded hardware and software requirements. Single-user MIMO systems allow one user to be serviced per transmission interval. This maximizes the throughput of a single user, but its disadvantage is that it does not take advantage of multi-user diversity. Multi-user MIMO systems (MU-MIMO) have become the main technique for meeting the requirements. In MU-MIMO, one of the biggest issues to deal with is eliminating co-channel interference. The block diagonalization (BD) is a linear precoding method used in broadcast channels of MU-MIMO that has been effective in removing multi-user interference (MUI), but is not computationally efficient. To counter this, we have developed a novel optimization algorithm (Bacterial Foraging Optimization), and implemented it with variation in the number of users and modulation orders. This paper exhibits the improvement in the diversity, and proves the efficiency of the proposed algorithm.

Keywords BFO · Block diagonalization · MIMO · MU · CSI

1 Introduction

Cellular systems have evolved a lot since the emergence of MIMO technology, but still limited by the amount of frequency bands allocated in the cell. If someone wants to communicate with several users within a cell, but not in the conventional way, i.e. using several channels. This time, communication is to be done with K users in the same band. This, in turn gives rise to co-channel interference within the same cell. With the innovation of MIMO technology, the cellular system has found a way to increase its data rates by applying MIMO technology, in addition to achieving high spectral efficiency within cellular networks, achieving great evolution of this system [1, 2]. MIMO technology consists of having more than one antenna

H. Gurdasani (✉) · N. Thangadurai
JAIN (Deemed-to-be University), Bangalore, India

A. G. Ananth
NMAMIT, Nitte, Mangalore, India

in the transmitter and the receiver [3]. MIMO increases the spectral efficiency of a wireless communication system by using the spatial domain [4]. In MU-MIMO, one of the biggest issues to deal with is eliminating co-channel interference [5]. This is because we will concentrate different users within the same frequency, increasing the interference between users [6]. If we manage to eliminate inter-user interference within the cell, then we are increasing the capacity of the system, the main objective of our algorithm to study. It is very important to have the Channel Status Information (CSI), since the development of our algorithm depends on this [7]. The solutions implemented so far, counteract the co-channel interference generated by different users, eliminating it almost entirely, but at the cost of additional signal processing algorithms on the transmitting and receiving sections [8–10].

Section 2 of the paper presents the proposed system model to suppress inter-user interference, Sect. 3, the results of the simulations, and finally, in Sect. 4, the conclusions.

2 Proposed Methodology

2.1 System Model

For our work, we will implement this model on multiple number of users. Also, with respect to users, it will be assumed that they are very far apart, so the user's channel will be independent with respect to the rest. We know that Channel Status Information (CSI) is very important in the transmitter, so it will be considered an ideal feedback channel between base station and user. Finally, the number of transmitting antennas would be equal to the product of the number of receiving antennas with the number of users, respectively.

Let us begin with an assumption of a multi-user communication system in which one base station supports several mobile stations.

Assuming that N_B and N_M are the base station and mobile station antennas, respectively.

As k independent user, $k.N_M$ antennas communicate with base station N_B antennas, where end-to-end communication for downlink is considered as $(k.N_M) \times N_B$ mimo system (Fig. 1).

'In a multiuser communication system, multiple antennas facilitate the base station to transmit multiple user data streams to be decoded by each user in downlink channel' [1].

By considering k independent users, where $X \in C^{N_B \times 1}$ is the transmit signal from the BS and $R_u \in C^{N_M \times 1}$ with received signal at the u^{th} user, where, $u = 1, 2, 3, \dots, k$ [3].

Let $Ch_u \in C^{N_M \times N_B}$ represent the channel gain between BS and the u th user. The signal received at the u th user position is expressed as:

$$R_u = Ch_u X + Z_u \quad (1)$$

$$u = 1, 2, 3, \dots, k$$

where $Z_u \in C^{N_M \times 1}$ is the additive zero mean circular complex Gaussian random vector [.] for all user.

Where X is the set of transmitted signal (Ch_1, \dots, Ch_k).

However, the main hindrance in data transmission in Broadcast Channel (BC) is that the signal detection on the receiver side is not linear, and therefore, interference cancellation becomes necessary. This paper utilizes Block Diagonalization precoding technique as the solution for the above stated problem.

2.2 Broadcast Channel Transmission via Block Diagonalization

Among the various solutions available to mitigate the problem of co-channel interference produced inside a cell by communicating several users in the same frequency band, we will deal with Block Diagonalization (BD) Technique. It uses a linear precoding matrix to transmit multiple frames of data to each user while removing inter-user interference. In the next section, we realize that multiplying the precoding matrix, with the channels of the interfering users, inter-user interference will be cancelled instantaneously, transmitting only the signal to the desired user. And, even if the interference from the other user is cancelled, inter-symbolic interference (ISI) will still be present in the receiver.

The channel inversion method is effective on its parts in clipping the interferences (any signal else the target signal). But it also introduces considerable noise enhancement in signals [11–14]. Block Diagonalization on other hand cancels only interferences of other user's signals at the stage of precoding. The inter-interference of signals from antenna, if occurred, could be tackled by various detection algorithm on-rolled in a MIMO network.

Let $N_{M,u}$ denote the number of antennas for the u^{th} consumer. Where $u = 1, 2, 3, \dots, k$.

For the u th signal $\tilde{x}_u \in C^{N_{M,u} \times 1}$, the received signal, $R_u \in C^{N_{M,u} \times 1}$ can be expressed as:

$$\begin{aligned} R_u &= Ch_u \sum_{k=1}^K p_k \tilde{x}_k + Z_u \\ &= Ch_u p_u \tilde{x}_u + \sum_{k=1, k \neq u}^K Ch_k p_k \tilde{x}_k + Z_u \end{aligned} \quad (2)$$

where $Ch_u \in C^{N_{M,u} \times N_B}$ is channel matrix between BS and u th user.

$w_u \in C^{N_a \times N_{M,u}}$ is the precoded matrix for the u th user and Z_u denotes the noise vector.

From Eq. (2), $\{Ch_u p_k\}_{u \neq k}$ increases interference to u th user unless,

$$Ch_u p_k = O_{N_M \times N_{M,u}}, \quad \forall u \neq k \quad (3)$$

where $O_{N_M \times N_{M,u}}$ is a zero matrix.

To meet the total p

$$\tilde{Ch}_u p_u = O_{(N_{M,total} - N_{M,u}) \times N_{M,u}}$$

ower constraints, the precoder $p \in C^{N_B \times N_{M,u}}$ must be unitary, $u = 1, 2, 3, \dots, k$.

From Eq. (3), the interference free received signal is,

$$R_u = Ch_u p_u \tilde{x}_u + Z_u \quad (4)$$

$$u = 1, 2, 3, \dots, k$$

For obtaining the value of \tilde{x}_u , various signal detection algorithms now can be employed for estimation.

To obtain $[P_k]_{k=1}^K$, let us take channel matrix of all users except u th user.

$$\tilde{Ch}_u = [(Ch_1)^{Ch} \dots (Ch_{u-1})^{Ch} (Ch_{u+1})^{Ch} \dots Ch_k]^{Ch} \quad (5)$$

where $N_{M,total} = \sum_{u=1}^k N_{M,u} = N_B$

$$\tilde{Ch}_u p_u = O_{(N_{M,total} - N_{M,u}) \times N_{M,u}} \quad (6)$$

$$u = 1, 2, 3, \dots, k$$

Hence, precoding matrix $P_u \in C^{N_B \times N_{M,u}}$ should exist in null space of \tilde{Ch}_u and precoders should satisfy the Eq. (6). For this, the singular value decomposition (SVD) \tilde{V}_u of \tilde{Ch}_u is expressed in terms of non-zero and zero singular values and zero singular values [1].

$$\tilde{Ch}_u = \tilde{U}_u \tilde{\Lambda}_u [\tilde{V}_u^{non\ zero} \tilde{V}_u^{zero}]^{Ch} \quad (7)$$

Where $\tilde{V}_u^{non\ zero} \in C^{(N_{M,total}-N_{m,u}) \times N_B}$ **and** $\tilde{V}_u^{zero} \in C^{N_{m,u} \times N_B}$ are comprising of right singular vectors that are denoted by non-zero singular values and zero singular values, respectively.

From Eq. (7), multiplying $\tilde{C}h_u$ with \tilde{V}_u^{zero} , we get following term,

$$\tilde{C}h_u \tilde{V}_u^{zero} = 0 \tag{8}$$

Multiplication of both the terms, i.e. channel gain and SVD results in zero. The zero received signals at destination end indicate the minimization of interference in signals. Thus, $P_u = \tilde{V}_u$ can be employed to precode the signal of u th user.

From Eq. (9), It can be seen that precoding matrix $P_u = \tilde{V}_u$ for the u th user. Where \tilde{V}_u is composed of zeros and non-zero singular values. Size of \tilde{V}_u depends on size of $\tilde{C}h_u$. If \tilde{V}_u is a large matrix than it consist greater number of non-zeros singular values which causes Eq. (3) to be:

$$\tilde{C}h_u P_K > 0 \quad \forall u \neq k \tag{9}$$

If \tilde{V}_u is a smaller matrix than it consist less number of non-zeros singular values which causes Eq. (3) to be:

$$\tilde{C}h_u P_K < 0 \quad \forall u \neq k \tag{10}$$

Both the cases cause a significant co-channel interference for user u , since channel matrix is not completely block diagonalization. Thus, size of \tilde{V}_u should be optimum for a better performance.

Since the size of \tilde{V}_u depends on size of $\tilde{C}h_u$, we can influence the magnitude of $\tilde{C}h_u$ by configuring the number of receiving antennas N_{RX} for each user.

To find an optimal value of N_{RX} , an objective function can be drawn as follows:

$$\min f(\tilde{C}h_u) = |\tilde{C}h_u P_k| \tag{11}$$

where,

$$\tilde{C}h_u = f(N_{RX}) \tag{12}$$

The value of N_{RX} can be found as,

$$N_{RX} = G(N_{RX}) \tag{13}$$

Where G is an operator which is optimized by Bacterial Foraging Optimization.

In order to explain the food search strategy of bacteria (specifically Escherichia Coli, present in the intestines of humans), four processes are used:

- (i) Chemotaxis
- (ii) Swarming
- (iii) Reproduction
- (iv) Dispersion

Bacteria have a set of rigid flagella that allow movement. This movement used in the search for food can be classified in two ways or modes of movement: Swimming and the fall together called Chemotaxis. A movement in a certain direction is called Swimming or Running, while Chemotaxis is the movement of the set of Bacteria in different directions. During the life of the bacteria, both modes of movement alternate. Clockwise rotation of the flagella results in Chemotaxis, and counterclockwise rotational movement results in Swimming.

Successively, after many stages of Chemotaxis, the original set of bacteria is fit for reproduction. The reproduction process consists of dividing a Bacterium into two identical ones. This is reflected during the optimization process, replacing half of the bacteria with poorer fitness function, with the ones with better fitness function. During the evolution process, however, the population of bacteria remains stable.

The dispersal of a group of bacteria to a new location results in an alteration to the natural process of evolution [9]. This concept is used to replace bacteria close to the location of the food to avoid stagnation at local minimums.

Initialization of the total number of Bacteria (numBact) with $(N + 1)/2$ solutions (variables to optimize) that satisfy the restrictions of maximum and minimum values of the variables (var_{max} and var_{min}), their aptitudes are calculated, and the bacteria with the best aptitude as $var_{global}^{(0)}$.

Scatter-kill consists of going through and selecting the bacteria from the population to apply scatter-killing with a P_{el} probability. To do this, these bacteria are eliminated and new ones are dispersed in a random location, maintaining the population size.

The breeding process kills half the population of bacteria with the worst fitness, and doubles the half best adapted to maintain population size.

Swarming is applied to the abilities of the bacteria by the following equation [4].

$$J_i^{(j+1,k,l)} = J_i^{(j,k,l)} - \sum_{numBact} d_{attract} x \left(e^{-w_{attract} x \sigma_i^{(j+1,k,l)}} \right) + \sum_{numBact} d_{repellant} x \left(e^{-w_{repellant} x \sigma_i^{(j+1,k,l)}} \right) \quad (14)$$

where,

$$\sigma_i^{(j+1,k,l)} = \sigma_i^{(j,k,l)} + \sum_{numBact} \left(var_{global}^{(j,k,l)} - var_i^{(j,k,l)} \right)^2 \quad (15)$$

The chemotaxis movement that is applied to the variables (solutions) is as follows:

The variables of the next iteration are calculated imposing the established limits of maximum and minimum value (var_{max} and var_{min}) as:

$$var_i^{(j+1,k,l)} = var_i^{(j,k,l)} + (c^{var_i} - d_2) \times \frac{\Delta^{(var_i)}}{\Delta\Delta} \quad (16)$$

where,

$$\Delta^{(var_i)} = (2 \cdot round(rand(.)) - 1) \bullet rand(\bullet);$$

$$\Delta\Delta = \sqrt{\Delta^{(var_i)} \times \Delta^{(var_i)}} \quad (17)$$

$$c^{var_i} = c_{max} - (c_{max} - c_{min}) \times \frac{j \times k}{max_{cycle}} \quad (18)$$

Swarming is applied obtaining $\Delta^{(var)}$ and $\Delta\Delta$ from Eq. (17), and the new variables are calculated with their restrictions:

$$var_i^{(j+1,k,l)} = var_i^{(j,k,l)} + (c^{var_i} + d_1) \times \frac{\Delta^{(var_i)}}{\Delta\Delta} \quad (19)$$

3 Simulation and Results

To verify the proposed Bacterial Foraging Algorithm in combination with Block Diagonalization Method, extensive and comprehensive simulations were carried out in MATLAB 2017a environment, and the following results were achieved.

Figure 2 exhibits the comparative performance of proposed approach using MMSE and ZF equalizers under Rayleigh Channel conditions.

The simulation result in Fig. 3 shows the performance comparison of SNR vs BER plots of different number of users (K) using the proposed approach, BFO optimized block diagonalization technique.

From the above graph, it is understood that due to the use of Block Diagonalization technique, the probability of occurrence of errors drops sharply for less number of users, but becomes almost constant as the no. of users increase. This can be a potential disadvantage.

The simulation result in Fig. 4 shows the performance comparison of SNR vs BER plots of different modulation orders (M) using the proposed approach.

The detailed analysis of the plots demonstrate that:

- To achieve a BER of 10^{-2} , the SNR required is 4 dB for a modulation order $M = 2$, which increases to 9 dB to achieve a BER 10^{-3} .

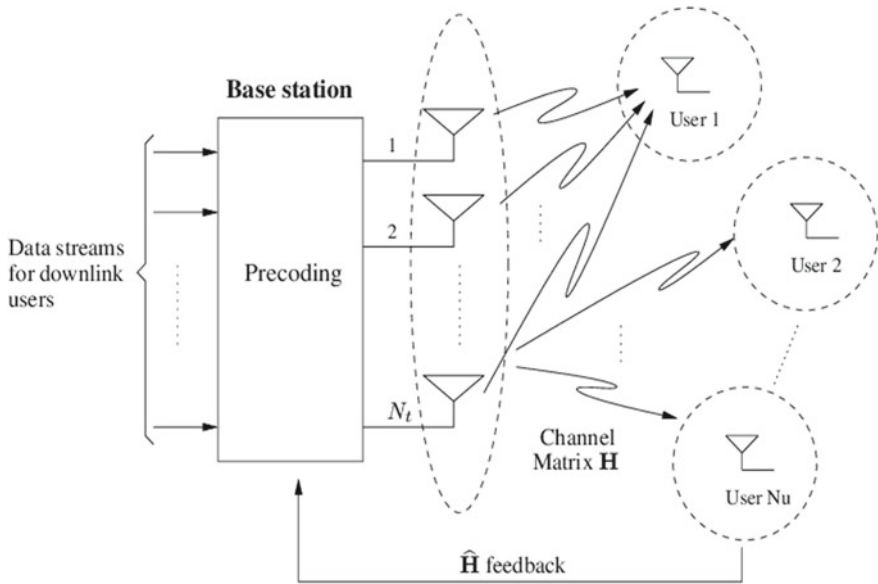


Fig. 1 Multi-user MIMO framework downlink channel model [1]

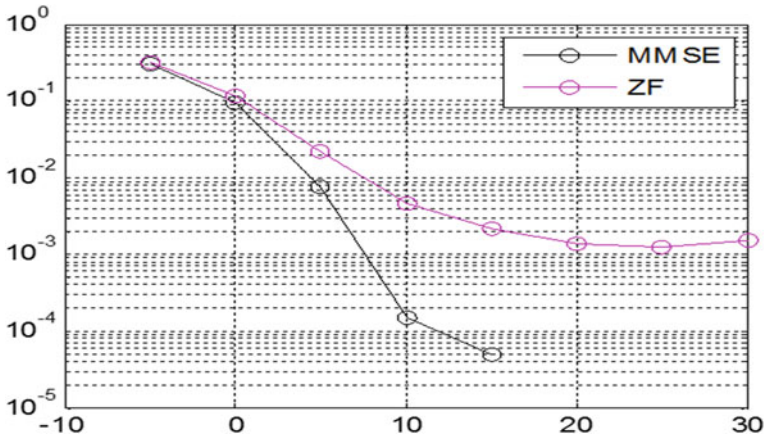


Fig. 2 MMSE and ZF equalizations under Rayleigh channel conditions

- Similarly, to achieve a BER of 10^{-2} , the SNR required is 8 dB for a modulation order $M = 4$, which increases to 11 dB to achieve a BER 10^{-3} .
- Also, to achieve a BER of 10^{-2} , the SNR required is 14 dB for a modulation order $M = 8$, which increases to 18 dB to achieve a BER 10^{-3} .
- Finally, to achieve a BER of 10^{-2} , the SNR required is 19 dB for a modulation order $M = 16$, which increases to 24 dB to achieve a BER 10^{-3} .

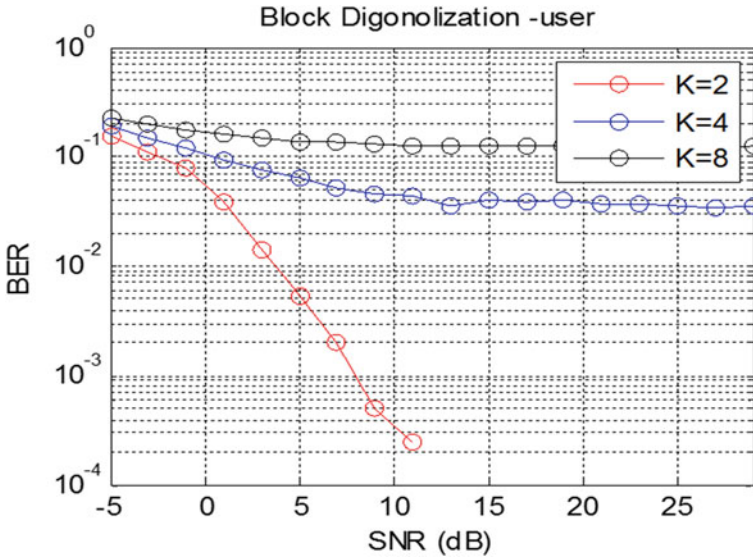


Fig. 3 Block diagonalization for various number of users

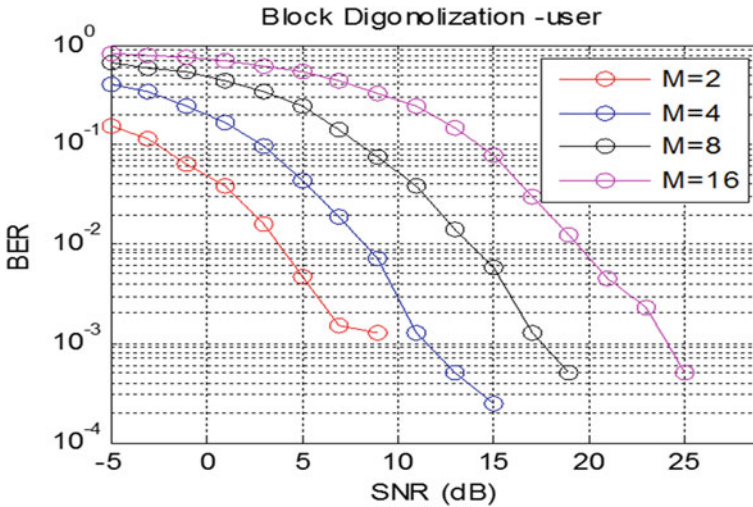


Fig. 4 BER curve with BFO optimized BD for different modulation order

4 Conclusion

From the above simulation results, the following can be inferred that the average value of BER is best achieved when the number of users is less, that is, the average

value of BER is best achieved when the modulation order $M = 2$. Also, the SNR vs BER average value using MMSE equalizer is better than that of ZF equalizer.

References

1. Saxena, S., & Singh, R. P. (2018). Optimal block diagonalization precoding scheme for MU-MIMO downlink channel using particle swarm optimization. *International Journal of Engineering & Technology*
2. Ni, W., & Dong, X. (2016). Hybrid block diagonalization for massive multiuser MIMO systems. *IEEE Transactions on Communications*, 64(1), 201–211.
3. Murti, F. W., Astuti, R. P., & Nugroho, B. S. (2015). Enhancing performance of block diagonalization precoding in multi user MIMO (MU-MIMO) downlink. In *3rd International Conference on Information and Communication Technology (ICoICT), 2015* (pp. 102–106). IEEE.
4. Liang, L., Xu, W., & Dong, X. (2014). Low-complexity hybrid precoding in massive multiuser MIMO systems. *IEEE Wireless Communications Letters*, 3(6), 653–656.
5. Khan, M. H. A., Cho, K. M., Lee, M. H., & Chung, J. G. (2014). A simple block diagonal precoding for multi-user MIMO broadcast channels. *EURASIP Journal on Wireless Communications and Networking*, 2014(1), 1–8.
6. Larsson, E. G., Edfors, O., Tufvesson, F., & Marzetta, T. L. (2014). Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2), 186–195.
7. Saha, S. K., Kar, R., Mandal, D., Ghoshal, S. P. (2013). Bacteria foraging optimisation algorithm for optimal FIR filter design. *International Journal of BioInspired Computation*.
8. Caire, G., Jindal, N., Kobayashi, M., & Ravindran, N. (2010). Multiuser MIMO achievable rates with downlink training and channel state feedback. *IEEE Transactions on Information Theory*, 56(6), 2845–2866.
9. Li, W., & Latva-aho, M. (2011). An efficient channel block diagonalization method for generalized zero forcing assisted MIMO broadcasting systems. *IEEE Transactions on Wireless Communications*, 10(3), 739–744.
10. Park, J., Lee, B., & Shim, B. (2012). A MMSE vector precoding with block diagonalization for multiuser MIMO downlink. *IEEE transactions on communications*, 60(2), 569–577.
11. Gupta, A., Saini, G., Sachan, S. B. L. (2013). An efficient linear precoding scheme based on block diagonalization for multiuser MIMO downlink system. *International Journal of Engineering Research & Technology (IJERT)*, 2(10). ISSN: 2278-0181
12. Kurniawan, E., Madhukumar, A. S., & Chin, F. (2009). Heuristic antenna selection algorithm for multiuser multi-antenna downlink. In *IEEE 69th Vehicular Technology Conference, 2009. VTC Spring 2009* (pp. 1–5). IEEE.
13. Kim, Y. H., & Kim, J. Y. (2011). Performance analysis of multi user MIMO system with block-diagonalization precoding scheme. Supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0029329) and MKE (The Ministry of Knowledge Economy), Korea, 5, p. 6.
14. Khan, S., Chockalingam, A., Sundar Rajan, B. (2008). A low-complexity precoder for large multiuser MISO systems. In *Proceedings of the 67th IEEE vehicular technology conference, VTC Spring 2008, 11–14 May 2008, Singapore*.

Prediction of Chemical Contamination for Water Quality Assurance Using ML-Based Techniques



C. Kaleeswari and K. Kuppusamy

Abstract Big Data is used in a spacious, distinct set of information that is continuously evolving estimates. Excessively vast array of tools can be computed to reveal models, patterns, and relationships, exclusively linked to human attitudes and communication. Its applications are applied in a range of real-time applications such as agriculture, weather forecasting, healthcare, and water resource management. Each Big Data research-oriented process can provide the outcome of the prediction or forecasting process. Analyzing the quality of water is still so important to a human being, because human beings cannot live without water. Smart Cities have been growing in the recent period, but many rural development areas are not yet emerging and do not have a lot of facilities like a smart city, and these areas are affected by the chemical contaminations like Arsenic, Cadmium, and some heavy metals. In this research work, Prediction of Chemical contamination for Water Quality Assurance using ML-based Techniques is proposed. This article offers a brief overview of Machine Learning Algorithms such as Decision Tree, Support Vector Machine, Random Forest, Naïve bayes, K-Nearest Neighbor. These algorithms are used for classification and prediction using datasets to analyze which Algorithm provides the best Predicting Accuracy as seen in this paper.

Keywords Machine learning · Prediction · Contamination · Classification · Water · Quality assurance

1 Introduction

Predictive Analytics is a field of futuristic data used to prove things about anonymous future events. Predictive Analytics has used more methods from learning-based techniques like Machine Learning and Deep Learning, Statistics, Artificial Intelligence, Data Mining, and remote sensing to inspect prevailing analysis to predict the future [1–3, 5, 8, 9]. Generally, real data is used to frame an empirical model with a significant tendency. It also has received a broad range of concentration in recent

C. Kaleeswari (✉) · K. Kuppusamy
Department of Computational Logistics, Alagappa University, Tamilnadu, India

times due to current trends in technology support, especially in the field of Machine Learning and Big Data [5, 8, 9]. The Water Quality Assurance Framework is one of the applications of urban development. Nobody can define the term “smart city,” but it can be used to make a normal city into a smart city with the right environmental setup using emerging technologies. Analyzing the water situation is still so important to a human being, because human beings are not living without water [10]. They use rivers, surface water, seawater, drinking water, ground water, reservoirs, ponds, and large aquatic areas. These services are impaired by some industrial exposures, such as chemical pollution, human trafficking, and lower chlorine levels. We know that seawater as well as its underlying water bodies are also salty. The Water Quality Assurance Framework is used to evaluate the use of water in agriculture and replicate it for drinking using the Systematic Approach. Rural Communities are not aware of the use of water and of the diseases induced by the use of these forms of water supplies polluted with chemicals such as Fluoride, Nitrate, Arsenic, Iron, Cadmium, Chromium, and Heavy Metals [2, 10–12]. In the study, the authors have established a smart city system for urban areas and several smart city initiatives are underway. Ability to predict the influence of water use in rural areas based on the quality of the water is more important. It is critical to determine whether the water is suitable for drinking or not, whether it is suitable for agricultural activities or not, and whether it can be recycled for routine use [6]. The study of the impacts faced by rural people can be used to recognize and raise awareness in the use of water.

In this research work, it is suggested that the Chemical Pollution Method be developed for the predictive analysis of water quality using Machine Learning. In the state-of-the-art, samples of water were collected from lakes and rivers only in smart cities. Instead, in this research work, the Water Sample Datasets obtained from the Water Board will be used as state-of-the-art water quality analysis in India. Urban areas have a lot of government-run services, but rural areas do not have an excessive development infrastructure resulting in certain health problems being posed and have little knowledge of them [9].

2 Related Works

Abba et al. [1], proposed the hybrid models GP-ELM and XGB-ELM. These two models were combined with various algorithms such as Extreme Gradient Boosting (XGB), Genetic Programming (GP), Extreme Learning Machine (ELM), Linear Regression (LR), and Stepwise Linear Regression (SLR). These hybrid models provide the highest accuracy for the minimum number of input variables. Libralato et al. [2], reported rare earth elements, nanomaterials, micro-and nano-plastics, PFAS-Perfluoroalkyl and polyfluoroalkyl substances, PPCPs-Pharmaceuticals and Personal Care Products, ionic liquids, artificial sweeteners, flame retardants, and illicit drugs contaminated with water. Most of these elements are discharged by sewage water, and the only solution to this problem is a wastewater treatment plant.

Miao et al. [3], have introduced the CWQI (Canadian Water Quality Index) which helps to assess the quality of water in a spacious and objective manner and to deter waste in water. This work also applied a remote sensing technique combined with CWQI using satellite data. Fatemi [4], noted the strategies and policies available for managing the quality of water in the context of environmental change. The author discusses the major impacts of water quality regions, such as soil erosion, floods, natural hazards, sediment loading, land use, and suggests some methods for problems such as phytoremediation, organic farming, wastewater treatment plants, etc. Chen et al. [5], said they were trying for the first time to achieve something that could be examined in contrast to the prediction of water quality. There are 7 traditional and 3 set models used in big data research (33,612 attentions) from the well-known rivers in China (2012–2018). Performance Evaluation of the F1-Score metrics, recall, precision and potential key exploration. On comparing the seven algorithms, Random Forest, Deep Cascade Forest, and Decision Tree provide slightly better forecasting performance. Zhang et al. [6], sought to improve the water quality index and the HHR (Human Health Risk) survey in the semi-arid regions of China with 51 samples collected. From the results, it has been shown that the collected water is significantly alkaline and provides a range of total dissolved solids. There are 94.12% of the samples that are suitable for drinking, and the major polluted parameters are NO₂, NO₃, F. The HHR reported that children are mainly affected by contaminated water compared to the male and female categories. Nenang et al. [7], used the DNA-based approach in their research and PCR–RFLP (Polymerase Chain Reaction–Restriction Fragment Length Polymorphism) as a biological indicator for regional weather. In the case of environmental pollution, the chemical parameters related to RFLP are determined by the PCA (Principal Component Analysis). Lerios and Villarica [8], reported the use of Machine Learning Algorithms such as Naïve bayes, Decision Tree, Random Forest, and Deep Learning Algorithm. Decision Tree Algorithm provides the best Accuracy and Precision compared with other Algorithms. Authors collected datasets using seven stations to evaluate seven quality attributes in a single reservoir named Laguna.

3 Study Site

The approach of the research work involves different stages. Dataset collection via Data.gov.in, Data.world.com, Kaggle, USDA Agricultural Research Service, and Water Board is one of the features of the system. Journal articles are collected and studied from various websites such as Springer, Elsevier, Research Gate, Wiley Publications, Bentham Science Publishers, Elsevier, Google Scholar, Taylor & Francis, and Emerald Publications from all around the world [4] (Fig. 1).



Fig. 1 Satellite image of the dataset description, India

4 Materials and Methods

Classification is used to forecast the result of a given data sampling when the variable of the outcome is in the pattern of categories. This model may look at the given data and to forecast labels like “Contaminated” or “Fresh.” In this work, datasets are collected from various sources and applied for classification to predict the chemical contaminated accuracy using Machine Learning. For data integrity, unstructured format data is transformed to a structured format using pre-processing or filtering principles. Training data for pre-processing was used for the classification of Machine Learning Algorithms such as Decision Trees, Support Vector Machine, Naïve bayes, Random Forest, and K-Nearest Neighbor [5]. These ML-based algorithms are ideal for predictive analytics in the recent era and also have improved results and are comparable to other classification techniques. This scientific investigation has not been carried out recently, and a few alternative research papers have also been focused on this topic. The datasets are collected at first decision and preprocessed for the classification to be applied. The proportion of missing pre-processing values carried out by Anaconda 3 is shown in Fig. 2.

4.1 Machine Learning Algorithms

4.1.1 Naïve Bayes Classification

Naive Bayes is a classification algorithm developed on the basis of theorem of Bayes with the presumption of autonomy between forecasters [5, 8]. This classifier conjectures that the presence of a particular characteristic in a class is unrelated to the existence of all other characteristics. For example, a bird can be considered a peacock in

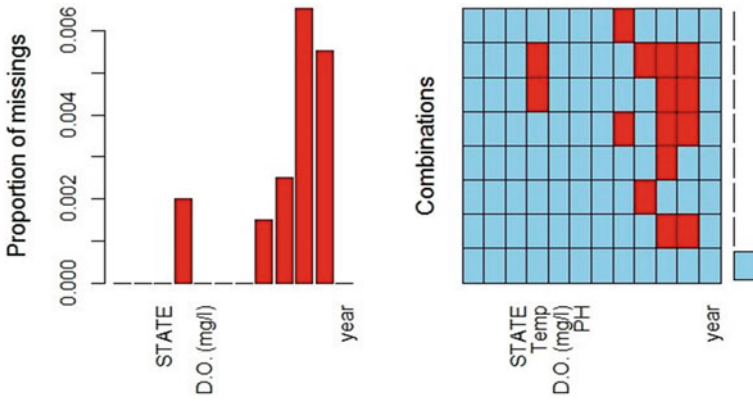


Fig. 2 Proportion of missing in preprocessing

the combined color of blue, green with long feathers and often spread its feathers. These features and properties clearly indicate the possibility that this is the bird and that is why it is known as Naïve. This model is very simple to develop and is especially useful for large volumes and a variety of datasets such as CSV files.

Figure 3 shows the result based on the Bayes theorem, which gives a path of scheming to manipulate Posterior Probability $P_1(C_1|X_1)$ from $P_1(C_1)$, $P_1(X_1)$, and $P_1(X_1|C_1)$. The following formula (1) is given below:

$$P_1(C_1|X_1) = \frac{P_1(X_1|C_1)P_1(C_1)}{P_1(X_1)} \tag{1}$$

- $P_1(C_1|X_1)$ Rear Probability or Posterior Probability of class (C_1 , target point) given forecaster (X_1 , parameters).
- $P_1(C_1)$ Previous or Prior Probability of a Class.
- $P_1(X_1|C_1)$ Likelihood which is the Probability of Prediction of a given class.

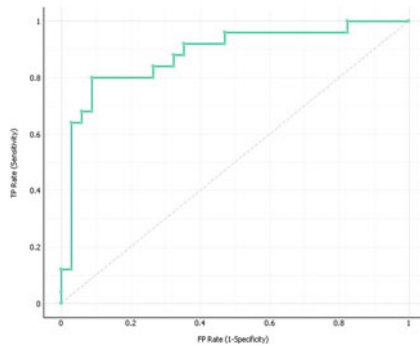


Fig. 3 ROC curve of Naïve Bayes

$P_1(X_1)$ Prior or Previous Probability of forecaster.

4.1.2 K-Nearest Neighbor Classification

KNN could be used for regression and classification time series prediction. But, it is also commonly applicable to classification problems of modern applications such as industry, social media applications for sentimental analysis, and also environmental monitoring. If KNN is used for classification, we can see the result as a lot of times when similar objects are connected to each and every data point [5, 8]. This technique captures a similarity plan called distance, range, and proximity or range. Figure 5 shows the ROC analysis based on the false positive rate and true positive rate following the below procedure.

1. Upload the dataset.
2. Boot K to selected number of instances called neighbors.
3. In each instance in the data
 - 3.1 Manipulate the range of data between the query sample and the present sample.
 - 3.2 Include the range and the ratio of the example to the samples collected.
4. Sort the values of the interval and range in ascending order.
5. Click on the first K entry in the sorted data order.
6. Take the labels from the selected k entries.
7. If Classification is required, provide the mode values of K.
8. If Regression is necessary, provide the mean value of K.

4.1.3 Random Forest Algorithm

Random Forest classification technique contains a set of smooth and simple prediction algorithm, and each one is effective in responding to a set of forecasting outcome values [5, 8]. The reaction of those belonging to a category as a class for classification issues, which collaborates or categorizes a set of individual forecasting outcomes with one of the divisions, resides in the variable. This model achieves the result like a large plant enclosed in bark and leaves. There are an inequitable number of simple trees, whatever the ultimate outcome can determine. The assembly technique of a simple tree defines the classification issues for the most prominent level. The use of tree collections must have a major evolution in the accuracy of forecasts. Each tree reacts by one relying upon this range of constraints selected individually (with a substitute) and the tree dispersion in the forest, independent of an array that is part of the massive reference scenario throughout the generic data set as shown in Fig. 4.

The optimal range for predicting variables is indicated as $\log_2 M_1 + 1$, where “ M_1 ” will be the range of the input variables. Random Forest predictions are reported as mediocre tree predictions. Random Forest Prediction is calculated using the formula (2).

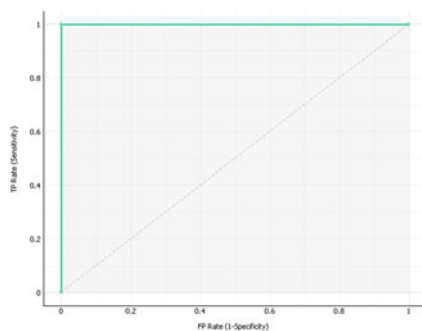


Fig. 4 ROC curve of random forest

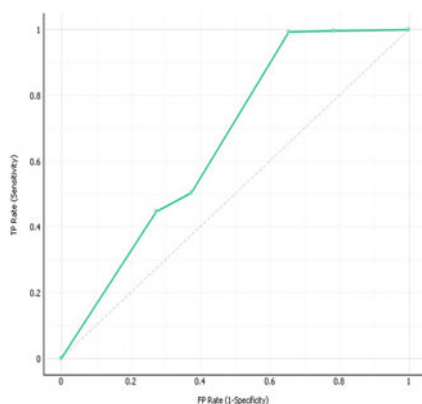


Fig. 5 ROC curve of K-nearest neighbor

RF Predict:

$$P_1 = \frac{1}{K_1} \sum_{K_1=1}^{K_1} K_1^{\text{th}} \text{TreeReaction} \tag{2}$$

where “K₁” has distinctive trees in the broad area like forest.

4.1.4 Decision Tree Algorithm

Decision Tree, as the name comes with it, appears like a tree structure operating on the key fundamentals of traditional legislation and laws. It is effective and has strong predictive analysis algorithms [5, 8, 10]. It has especially parameterized that of internal nodes, branches, and terminal nodes. Instead, every internal node has a

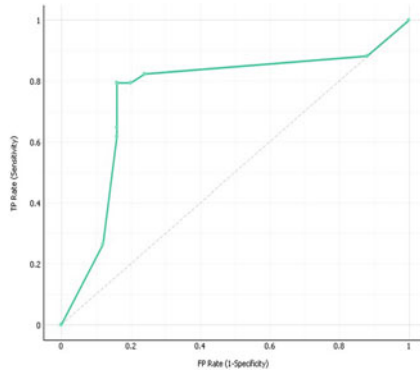


Fig. 6 ROC curve of decision tree

grip on a “test” on a function, branches have a test result, and so each leaf is a class tag as shown in Fig. 6. The main objective of using a DT would be to generate a training model that is being used to determine that hash value obtained by training basic decision rules implicit in the training data shown in Fig. 6.

4.1.5 Support Vector Machine

Support Vector Machine seems to have been a popular ML-based algorithm that conducts both regression and classification tasks [10]. This algorithm would be to generate an excellent outer boundary line or can detect n-dimensional space within classes to the eluded to as feature vector, which is why the method is described as SVM (Support Vector Machine). It is categorized into two forms, Linear SVM and Non-Linear SVM, and has applications such as Face Discovery Method, Image Categorization, Text Mining, and Classification. The ROC analysis of SVM is shown in Fig. 7.

5 Results and Discussion

5.1 Performance Evaluation

In this section, the performance metrics are analyzed for the estimation of the results shown by the individual algorithm in this research work, and comparative analysis of various machine learning approaches for predicting best accuracy and precision in the Python Framework is done. ROC (Receiver Operator Characteristics) curve plotting is given below based on the ML Techniques. False Positive Rate and True Positive Rate with threshold values are used for the calculation of the ROC. Proposed

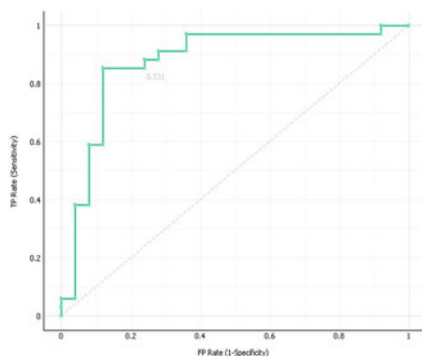


Fig. 7 ROC curve of support vector machine

work provides the perfect Prediction Results, depending on the Precision. Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest, K-Nearest Neighbor are used in this work to estimate the accuracy of the classification [8, 9]. Random Forest has the highest Accuracy (98%) and Precision (98.04%) compared to other Algorithms shown in Table 2.

6 Conclusion

In this work, Machine Learning-based Water Quality Assurance framework has been formulated for measuring the chemical contamination and quality attributes in a specific station in state-wise in India using the datasets collected from various web sources. The collected datasets are preprocessed and classified by Machine Learning Algorithms [5, 8, 9]. This proposed methodology has been verified for its ability to measure Water Quality with high accuracy and precision. Chemical parameters of contamination levels estimated in this research, such as Arsenic, Fluoride, Cadmium, Chromium, Heavy Metals, etc., are shown in Table 1 [2, 10]. This method gives the good Prediction Results depending on the Precision. Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest, K-Nearest Neighbor are applied in this work to predict the Classification Accuracy [8, 9]. Random Forest provides the best accuracy (98%) and precision (98.04%) compared with other algorithms as shown in Table 2.

Table 1 Dataset description

S. no	Attributes	Description
1	Locations	The location where samples of water are obtained
2	Date	Date for monitoring of water samples
3	State	The location of the region as in which state
4	pH	Potential Hydrogen Values ranging from 6.5 to 9.0
5	Temperature	14 ⁰ C to 36 ⁰ C
6	DO	Dissolved Oxygen mg/L (5)
7	EC	Property or power to conduct heat, electricity, or sound (EC above 3000 micro mhos/ cm)
8	(NO ₃ -H)	Nitrate (above 45 mg/l)
9	Salinity	Salty (above 900 mg/L)
10	F	Fluoride (above 1.5 mg/l)
11	As	Arsenic (above 0.01 mg/l)
12	Iron	Iron (above 1 mg/l)
13	Heavy Metals	Lead (above 0.01 mg/l)
14	Cr	Cadmium (above 0.003 mg/l)
15	Ch	Chromium (above 0.05 mg/l)

Table 2 Algorithm-wise classification accuracy

Algorithm	Accuracy (%)	Precision (%)
Naïve Bayes	72.58	73.26
Random forest	98	98.04
Decision tree	88.98	89
KNN	85.29	85.3
SVM	84.23	84.23

Acknowledgements This work was financially supported by Rashtriya Uchchar Shiksha Abhiyan (RUSA, Phase 2.0) grant sanctioned vide letter No.F.24-51 / 2014-U, policy (TNMulti-Gen), Dept. of Edn. Govt. of India, Dt 09.10.2018.

References

1. Abba, S. I., Hadi, S. J., Sammen, S. S., Salih, S. Q., Abdulkadir, R. A., Pham, Q. B., Yaseen, Z. M. (2020). Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. *Journal of Hydrology*, 587 (2020) 124974. <https://doi.org/10.1016/j.jhydrol.2020.124974>.
2. Libralato, G., Freitas, R., Buttino, I., Arukwe, A., & Torre, C. D. (2020). Special issue on challenges in emerging environmental contaminants CEEC19. *Environmental Science and Pollution Research*, 2020(27), 30903–30906. <https://doi.org/10.1007/s11356-020-09539-w>.
3. Miao, S., Liu, C., Qian, B., & Miao, Q. (2020). Remote Sensing-based water quality assessment for urban rivers: A study in linyi development area. *Environmental Science and Pollution Research*, 2020(27), 34586–34595. <https://doi.org/10.1007/s11356-018-4038-z>.
4. Fatemi, A. (2020). Strategies and policies for water quality management of Gharasou River, Kermanshah, Iran: A review. *Environmental Earth Sciences* (2020), 79, 254. <https://doi.org/10.1007/s12665-020-08997-2>.
5. Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, Da., Chen, X., Deng, Y., & Ren, H. (2020). Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Research*, 171(2020), 115–454. <https://doi.org/10.1016/j.watres.2019.115454>.
6. Zhang, Q., Panpan, X., & Qian, H. (2020). Groundwater quality assessment using improved water quality index (WQI) and human health risk (hhr) evaluation in a semi-arid region of Northwest China. *Exposure and Health*, 2020(12), 487–500. <https://doi.org/10.1007/s12403-020-00345-w>.
7. Neneng, L., Nugroho, R. A., Komai, Y., Takayama, N., & Kawamura, K. (2019) Water quality measurements with a simple molecular analysis (PCR-RFLP) of the microbiome in a metropolitan river system in Japan. *Walailak Journal of Science and Technology* 2020, 17(3), 257–268.
8. Leros, J. L., & Villarica, M. V. (2019). Pattern extraction of water quality prediction using machine learning algorithms of water reservoir. *International Journal of Mechanical Engineering and Robotics Research*, 18(6).
9. Koditala, N. K., & Pandey, P. S. (2018). Water quality monitoring system using IoT and machine learning. In *IEEE*.
10. Karthick, T., Gayatri, D., Kohli, T. S., & Pandey, S. (2018). Prediction of Water quality and smart water quality monitoring system in IoT environment. *International Journal of Pure and Applied Mathematics*, 118(20). ISSN: 1314–3395 (On-line Version).
11. Kumar, V., Sharma, A., Chawla, A., Bhardwaj, R., Kumar, A., & Thukral, A. K. (2016). Water Quality Assessment of River Beas, India, using Multivariate and Remote Sensing Techniques. *Environment Monit Assess*.
12. Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors*, 16, 1298. <https://doi.org/10.3390/s16081298>.

Design and Performance Analysis of Two-Port Circularly Polarized MIMO Antenna for UWB Applications



Madan Kumar Sharma, Aryan Sachdeva, Ayushi Ojashwi,
and Mithilesh Kumar

Abstract It proposes and simulates a new Ultra-wide (UWB) Multiple Input Multiple Output (MIMO) circular polarization (CP) antenna. The developed antenna's configuration is uncomplicated and incorporates dual T-shaped feed ports at left and right edges to achieve both LHCP and RHCP at same frequency band. A wide circular slot with a hexagonal stub is created in the ground plane in order to achieve better bandwidth for the realization of UWB. The evolution measures of the antenna put forward are presented ahead in this research paper, and its parameters are designed to obtain the optimal return loss, isolation, polarization, and ARBW. Also, a parametric study is done on various types of materials chosen as substrate. The parametric study results, portrays that substrate with higher dielectric tends to give poor ARBW. Operating bandwidth ($S_{11} < -10\text{dB}$) of the planned antenna is obtained to be 6.7 GHz (3.4 to 10.1 GHz). An isolation of more than 15 dB is also achieved. 3 dB ARBW is obtained in the extent of 3.5–5.5 GHz. Furthermore, the circular polarization features of the antenna proposed are favorable. The surface current and radiation patterns obtained illustrates the circular polarization. Both RHCP and LHCP are obtained at the same frequency in this proposed antenna design. This proposed antenna is a solution to orientation problems of transmitter and receiver, complicated designing, multipath fading, and weather penetration. The simulated antenna is a low profile, compact sized antenna simulated for UWB circular polarization and next generation applications.

Keywords MIMO antennas · Ultra-wide band · Circular polarization · Axial ratio

1 Introduction

Nowadays, with the increase in the demand of the data, the high data transfer and transmission rate is of importance. In order to achieve this, ultra-wide bandwidth is

M. K. Sharma (✉) · A. Sachdeva · A. Ojashwi
Galgotias College of Engineering and Technology, Greater Noida, India

M. Kumar
Rajasthan Technical University, Kota, India

required. In modern usage, MIMO is widely used for its ability to convey and obtain more than one data signal concurrently over the identical channel by using multipath propagation.

At present, wireless networks are linearly polarized due to which they are less resistant to signal degradation and face multipath fading. In the communication system, the proper positioning of the antenna components, the transmitter, and receiver is of much importance and also a difficult task. Thus, by using a circularly polarized (CP) antenna, this problem can be solved, as till the time the transmitter and receiver is placed within the obtained range, it will observe equal amount of radiation in all directions. In comparison to linearly polarized antennas, the CP antenna has numerous advantages [1], such as higher performance, mobility, and good penetration strength. Apart from that, the multipath fading problem is also reduced using CP antenna. CP antennas are now more popular amongst other antennas for its better performance and benefits of the impact of depolarization in the modern communication system.

Different antennas had been studied which obtained broadband circular polarization by various techniques. Over the past years, antennas with different kinds of slots, such as Crescent moon shaped [2], C-shaped slot antenna [3], had been produced. A MIMO microstrip-fed antenna with both rectangular shapes cut on the ground on either side of the feed in [4] had been demonstrated for wideband circular polarization applications. A coplanar waveguide (CPW) feed antenna with three parallel coplanar ground planes was also reported in [5]. In [6], microstrip-fed monopole antenna has been claimed to CP mode excitation, an inverted L-shape parasitic strip has been used to achieve the CP. A square slot antenna with dual circular polarization was proposed in [7] using two asymmetric T-shaped feed lines. In [8], CPW-based monopole wideband antenna with an asymmetrical ground plane and a square-ring with an opening at the bottom edge was reported. A regular-hexagonal wide-slot antenna structure for broadband CP with CPW feed had been filed in [9], which also had slot protruded into an L-shaped monopole patch. By etching an L-slot, the dual-band function is obtained in the radiating patch of [10]. The characteristic mode theory is used in [11] to direct antenna design and evaluate performance. MIMO antenna comprising uneven antenna elements and two-sided shorted ground had been contemplated in [12]. Two MIMO antenna devices, one for 4G band and other for a possible 5G band, were analyzed in [13]. Authors in [14] have described the ways to reduce electrical coupling of the MIMO antenna. In [15], to prevent the current flow between the two ports of the same antenna, rectangular slots were engraved on either side of ground. The defected ground structure of a two-element orthogonally arranged MIMO had been presented in [16]. The MIMO efficiency analysis obtained by the use of orthogonal CP radiators in contrast to orthogonal linearly polarized radiators had been presented in [17]. The two salient performance parameters of the MIMO antenna, namely, isolation and bandwidth had been observed in [18]. On the ground plane, two ground stubs had been inscribed and that led to enhancement of isolation and increased impedance bandwidth in [19]. This concept helped us understand ways to attain better isolation for our built antenna.

Table 1 Comparative analysis of the proposed antenna

Ref. no	Antenna size (mm*mm)	Operating BW(GHz)	Axial Ratio	Isolation (dB)	Polarization
Midya et al. [1]	48 * 48	3–4.2	3–4.2	≤ -15	RHCP/LHCP
Saini and Dwari [7]	60 * 60	2–4.76	2–3.7	Below 15	RHCP/LHCP
Xu et al. [8]	90 * 100	0.88–3.08 3.64–5.58	0.98–2.68 3.69–5.5	–	RHCP/LHCP
Zhou et al. [9]	62 * 62	1.80–4.6	2.5–4.1	Below 15	RHCP/LHCP
Proposed antenna	48 * 48	3.4–10.1	3.5–5.5	Below 15	RHCP/LHCP

Although properties such as wide bandwidth and circular polarization had been observed and studied over in the above literature survey, they tend to face certain problems such as complex designs, complicated geometry, relatively narrow axial ratio bandwidth (ARBW), and larger size.

Table 1 shows the results of a comparison of the proposed antenna with quoted antenna in the literature. It can be clearly noticed from the table that the antenna suggested has comparatively better CP features, axial ratio bandwidth (ARBW), and an ultra-wide bandwidth with better size optimization.

Based on the paper [2], a novel UWB-CP antenna is proposed in the carried out work. A T-shaped feed is used in place of the rectangular feed to attain a better ARBW. Also, in proposed antenna, a hexagonal stub has been created which further helps in widening the antenna bandwidth. The CST microwave studio tool was used to build and configure the proposed antenna. This antenna is apt for next gen wireless applications and ultra-wide bandwidth circular polarization applications.

2 Antenna Design and Configuration

The proposed antenna conformation is outlined in Fig. 1. While designing the antenna, 1.6 mm depth of the substrate, Rogers RT5880LZ with loss tangent of 0.02, and relative permittivity of 2.0 is used. Measurements of the proposed antenna are as stated in Table 2. It is developed over a square substrate having the size of $48.0 \times 48.0 \text{ mm}^2$.

At the lower edge, the antenna comprises a ground plane and two microstrip feed lines at the upper portion of the dielectric substrate. On the ground surface, a large circular shaped slot with two strip lines connecting the two edges is established to obtain ultra-wide bandwidth. A decagon has been added in the slot which acts as a hexagonal stub and initiates better result.

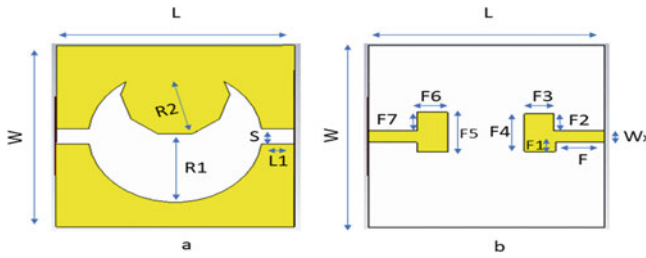


Fig. 1 Proposed antenna configuration

Table 2 Measurements of proposed antenna (in mm)

Parameter	Value	Parameter	Value	Parameter	Value
W	48	L1	6.61	F3	6
L	48	W _x	3	F4	10
R1	17.5	F	10.0	F5	10.5
R2	11	F1	2.5	F6	6.25
S	4	F2	4.5	F7	5

The hexagonal-shaped stub protruding from the ground surface helps in achieving CP. A 50 Ω feed line is used to excite the antenna. It is located from both the sides in the precinct of the hexagonal stub to broaden the ARBW of the single element. Also, the rectangular slots created in the ground thus decreases the feed-ground mutual coupling and helps in getting better parametric results.

2.1 Evolution Process of the Proposed Antenna

In Fig. 2, the four progression steps of the proposed antenna have been shown. Firstly, a two-feedline antenna with a circular slot is depicted as Ant1. It provides operating bands approximately from 4.38 to 9.5 GHz. Poor isolation above 15dB is obtained throughout in this case. Secondly, the decagon is added in the complete circular slot, thereby creating a stub as shown in Ant2. It has the return loss in the range of 3.64–7.9 GHz. However, ARBW below 3dB is not achieved in this antenna. Thirdly, in the ground plane of Ant2, two rectangular slots are integrated to decrease the mutual



Fig. 2 Progression steps of the developed antenna

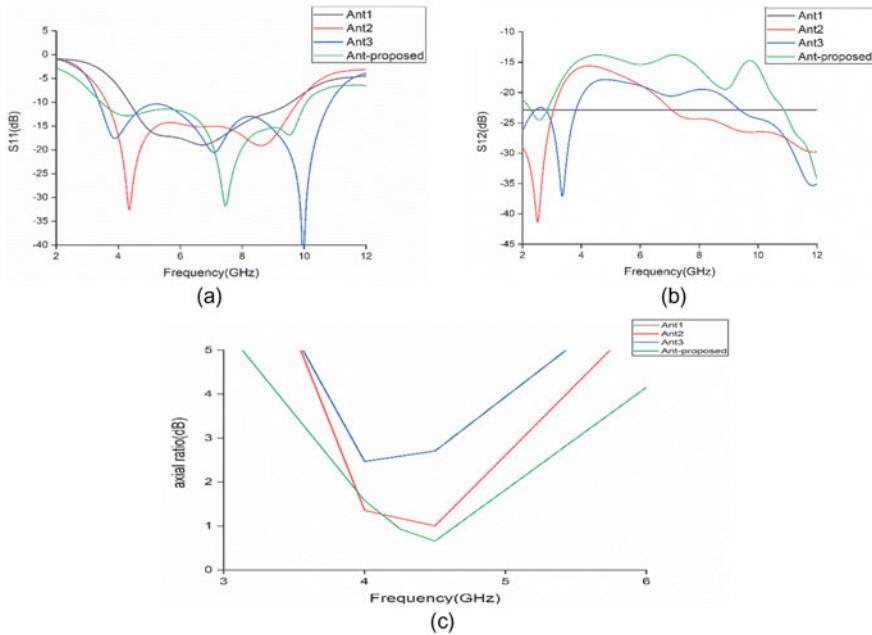


Fig. 3 Antenna performance parameters **a** Magnitude of S_{11} , **b** Magnitude of S_{12} , **c** Axial ratio

coupling between feed and ground. This is structured as shown in Ant3. Finally, to raise Axial ratio, the rectangular feed is designed as T-shaped feedline, parallelly facing each other.

The plots for s-parameter and ARBW with respect to every evolution change of antennas are shown in Fig. 3. Antenna 1 is named as Ant 1, and similarly Antenna 2 and 3 are named as Ant 2 and Ant 3, respectively.

2.2 Parametric Study

It is undoubtedly essential for us to research the performance of the developed antenna parameters. We shall therefore understand the variability of S_{11} , S_{21} , and AR in account of other materials with different value of permittivity.

2.2.1 Effect of the Substrate Permittivity

The S_{11} , S_{21} , and axial ratio variability with various other material type of the proposed antenna's substrate are depicted in Fig. 4. It had been seen that when

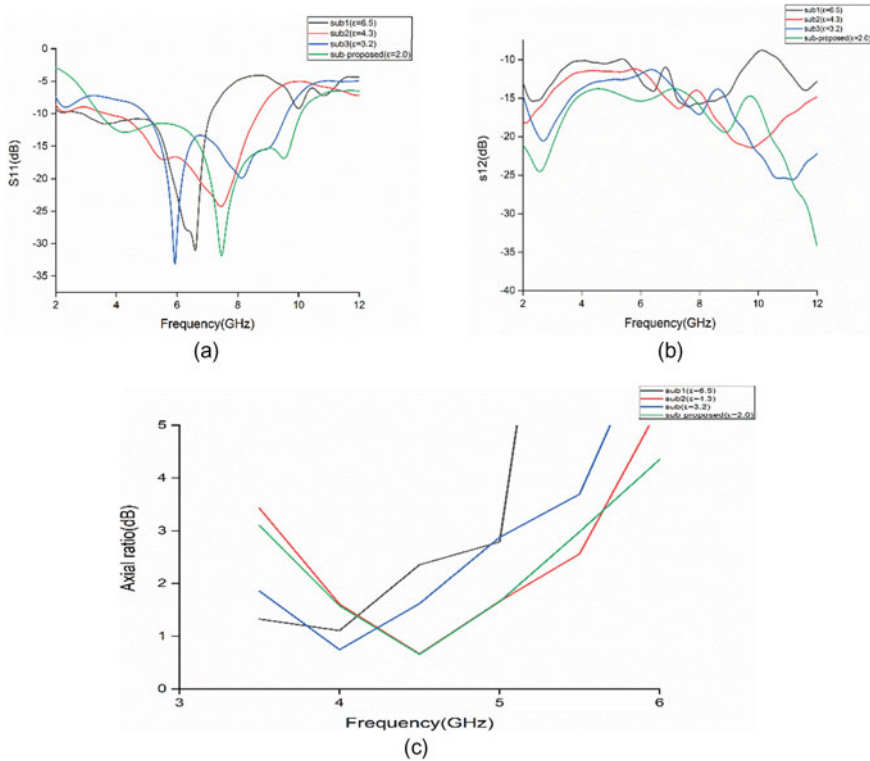


Fig. 4 Effect of substrate on **a** magnitude of S11 **b** magnitude of S12 **c** axial ratio

Rogers RO3006 with a permittivity of 6.5 had been used as substrate, the bandwidth gets limited to approximately 4.5 GHz. Also, it suffered with lesser ARBW. When the substrate had been changed to FR-4 (permittivity = 4.3), comparatively an increase in bandwidth had been observed. A better ARBW had been obtained but at the cost of poor isolation of the antenna. Rogers RO4830 substrate with a permittivity of 3.2 had been then selected as the substrate material. A better return loss, isolation as well as ARBW had been obtained in this case although the bandwidth achieved had been comparatively less than the proposed antenna. The best return loss had been observed in the proposed antenna where the substrate Rogers RT5880LZ had been used having a permittivity of 2.0. The ARBW, isolation, and bandwidth had been maintained in this case. Considering the results obtained, it had been made clear that the substrate with lower dielectric gives better return loss and isolation. In addition to this, substrate material has a big role to play in the ARBW. Material with higher dielectric value leads to poor ARBW in many cases.

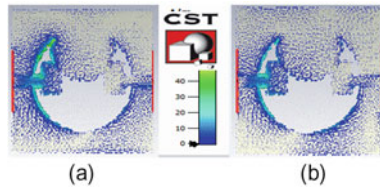


Fig. 5 Surface current at a 4.5 GHz b 5.5 GHz when port1 is excited

2.3 Surface Current

Figure 5 shows the simulated induced current distribution at the resonant frequency of 4.5 and 5.5 GHz. It should be remembered that at the same temporal moment, all the flows had been presented. When one port had been stimulated, the other port had been terminated through 50ohm matched load. Figure 6 clearly illustrates that when port1 is excited, the current density at port2 is lesser and vice versa. This antenna is thus said to have less mutual coupling. It was found by the effects of the surface current that antenna port which is not excited has low induced current. However, excited port has opposite rotation of the current for every step. This corresponded in achieving various modes of CP.

3 Results and Discussion

Figure 6 describes the obtained s-parameter of the antenna developed. The built antenna can be said to have an impedance bandwidth in the range of 3.4–10.1 GHz. In addition to this, the antenna provides adequate isolation (mutual coupling) less than -15dB within a range of 3.1–10.1 GHz. The axial ratio of the simulated antenna is put forward in Fig. 7. From axial ratio graph, it is clear that

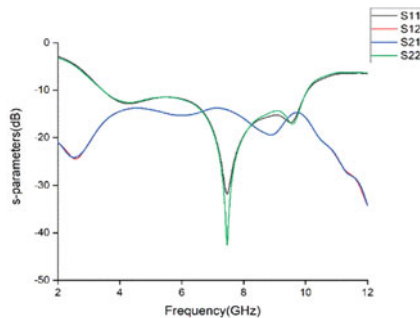


Fig. 6 Simulated S-parameters of the built antenna

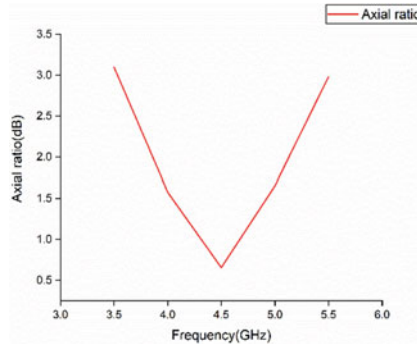


Fig. 7 Simulated AR of the built antenna

AR is below 3dB for the frequency spectrum 3.5–5.5 GHz, therefore it is offered CP characteristics for the wide band more than 2 GHz.

Radiation patterns at interested frequency of 3.5 GHz, 4.5 GHz, and 5.5 GHz are shown in Fig. 8. The results verify the correctness of the designed CP antenna with ultra-wide band. This antenna can exhibit LHCP and RHCP at the same frequency.

The gain and efficiency plot of the developed antenna is illustrated in Fig. 9 with frequency on the X-axis. The antenna is advantageous as it has peak gain of more than 4 dBi within the interested frequency range. The radiation efficiency of the antenna is also more than 70 %.

4 Conclusion

A new UWB-MIMO antenna with circular polarization is presented and implemented using CST successfully in this study. It has a very broad impedance bandwidth, a strong 3dB ARBW with good circular polarization, and port isolation performance. By using a circular slot with hexagonal stub, a wideband CP has been established. After the addition of the T-shaped feedline, the 3dB ARBW has been further expanded. The antenna is competent working in both modes of LHCP and RHCP at the very same frequency. This antenna is an ultra-wide bandwidth CP in nature, and therefore, will play significant role for ultra-wide bandwidth applications. Its expanded ARBW and good polarization makes it propitious for next gen applications.

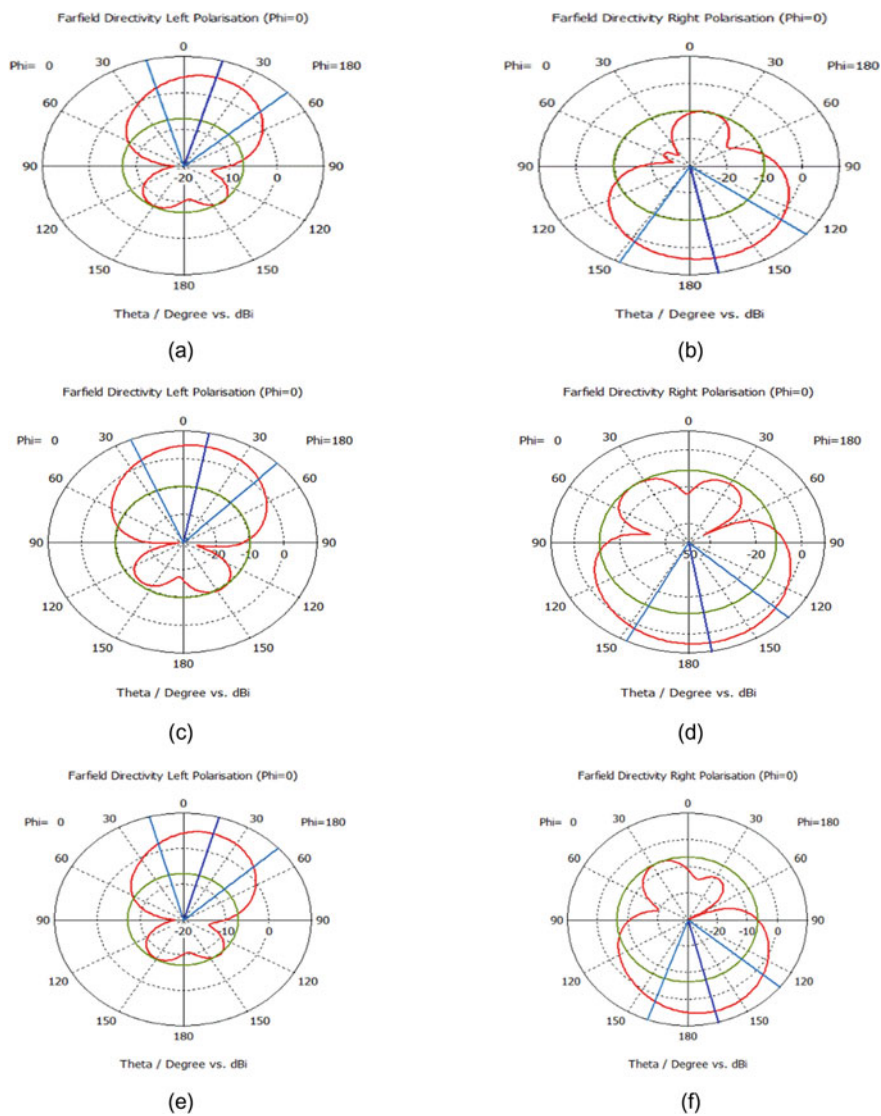


Fig. 8 Radiation patterns of the built antenna for LHCP and RHCP at the frequency **a** 3.5 GHz **b** 4.5 GHz **c** 5.5 GHz

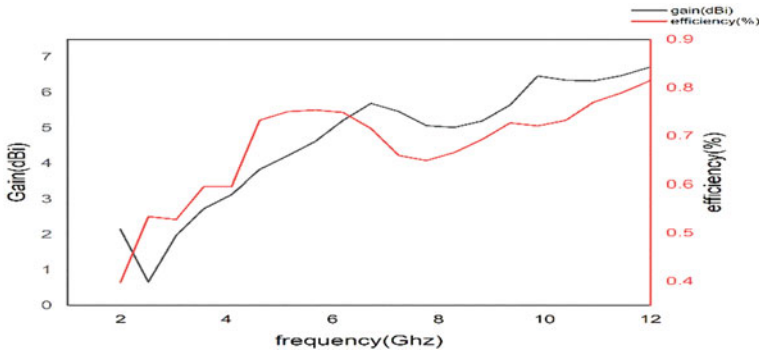


Fig. 9 Plot of gain and efficiency versus frequency

References

1. Midya, M., Bhattacharjee, S., & Mitra, M. (2019). Broadband circularly polarized planar monopole antenna with G-shaped parasitic strip. *IEEE Antennas and Wireless Propagation Letters*, 18(4), 581–585.
2. Ojaroudi Parchin, N., Basherlou, H. J., & Abd-Alhameed, R. A. (2020). Dual circularly polarized crescent-shaped slot antenna for 5G front-end systems. *Progress In Electromagnetics Research*, 91, 41–48.
3. Xu, R., Li, J. Y., & Liu, J. (2017). A design of broadband circularly polarized C-shaped slot antenna with sword-shaped radiator and its array for L/S-band applications. *IEEE Access*, 6, 5891–5896.
4. Ullah, U., Mabrouk, I. B., Koziel, S., & Al-Hasan, M. (2020). Implementation of spatial/polarization diversity for improved-performance circularly polarized multiple-input-multiple-output ultra-wideband antenna. *IEEE Access*, 8, 64112–64119.
5. Ullah, U., Al-Hasan, M., Koziel, S., & Mabrouk, I. B. (2020). Circular polarization diversity implementation for correlation reduction in wideband low-cost multiple-input-multiple-output antenna. *IEEE Access*, 8, 95585–95593.
6. Ullah, U., & Koziel, S. (2018). Design and optimization of a novel miniaturized low-profile circularly polarized wide-slot antenna. *Journal of Electromagnetic Waves and Applications*, 32(16), 2099–2109.
7. Saini, R. K., & Dwari, S. (2015). A broadband dual circularly polarized square slot antenna. *IEEE Transactions on Antennas and Propagation*, 64(1), 290–294.
8. Xu, R., Li, J. Y., Liu, J., Zhou, S. G., & Wei, K. (2018). Design of spiral slot-based dual-wideband dual-sense CP antenna. *IET Microwaves, Antennas and Propagation*, 13(1), 76–81.
9. Zhou, S. W., Li, P. H., Wang, Y., Feng, W. H., & Liu, Z. Q. (2011). A CPW-fed broadband circularly polarized regular-hexagonal slot antenna with L-shape monopole. *IEEE Antennas and Wireless Propagation Letters*, 10, 1182–1185.
10. Boukarkar, A., Lin, X. Q., Jiang, Y., Nie, L. Y., Mei, P., & Yu, Y. Q. (2017). A miniaturized extremely close-spaced four-element dual-band MIMO antenna system with polarization and pattern diversity. *IEEE Antennas and Wireless Propagation Letters*, 17(1), 134–137.
11. Li, H., Sun, S., Wang, B., & Wu, F. (2018). Design of compact single-layer textile MIMO antenna for wearable applications. *IEEE Transactions on Antennas and Propagation*, 66(6), 3136–3141.
12. Luo, X. S., Weng, Z. B., Zhang, W. J., & Yang, L. (2018). Compact planar multiband MIMO antenna based on composite right/left-handed transmission line for mobile phone applications. *Microwave and Optical Technology Letters*, 60(6), 1505–1511.

13. Sharawi, M. S., Ikram, M., & Shamim, A. (2017). A two concentric slot loop based connected array MIMO antenna system for 4G/5G terminals. *IEEE Transactions on Antennas and Propagation*, 65(12), 6679–6686.
14. Nadeem, I., & Choi, D. Y. (2018). Study on mutual coupling reduction technique for MIMO antennas. *IEEE Access*, 7, 563–586.
15. Chattha, H. T. (2019). 4-port 2-element MIMO antenna for 5G portable applications. *IEEE Access*, 7, 96516–96520.
16. Kumar J. P., & Karunakar G. (2018). Circular slotted ring shaped mimo antenna for ultra wideband applications. In *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, vol. 13, pp. 1–5.
17. Dicandia, F. A., Genovesi, S., & Monorchio, A. (2017). Analysis of the performance enhancement of MIMO systems employing circular polarization. *IEEE Transactions on Antennas and Propagation*, 65(9), 4824–4835.
18. Sharma, M. K., Kumar, M., Saini, J. P., & Singh, S. P. (2020). Computationally optimized MIMO antenna with improved isolation and extended bandwidth for UWB applications. *Arabian Journal for Science and Engineering*, 45(3), 1333–1343.
19. Sharma, M. K., Kumar, M., & Saini, J. P. (2015). UWB-MIMO antenna with enhanced isolation for breast cancer detection. In *2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, (pp. 787–790).

Intelligent Traffic Control System for Emergency Vehicles



Anuj Sachan and Neetesh Kumar

Abstract Traffic Control System has faced many issues, such as high waiting time, resulting in the emission of CO₂ (carbon) gas, which results in environment degradation and the loss of crucial time for the emergency vehicles, which increases impatience in drivers, which could be one of the primary causes of increase in accidents. There is also a considerable delay in vehicles' arrival time, particularly emergency vehicles that get stuck in the long queue. It is a matter of concern, as these vehicles are being categorized as priority vehicles, so their waiting time should be minimum. The current Traffic Control System does not use any specific algorithm to control the traffic for priority vehicles such as fire brigade, ambulance. These drawbacks can be resolved by replacing the existing Traffic Control System with a more intelligent Traffic Control System. In this work, we proposed an Intelligent Traffic Control System for Emergency Vehicles model (ITCSEV), which can identify the vehicles and separate them based on their types like "emergency," "passenger." It can change the duration and traffic signal accordingly. This model has been tested by us through a practical simulation using an open-source simulator, i.e., SUMO (Simulation of Urban MObility), on the map of India's Gwalior city, which demonstrates the efficiency and reliability of the model.

Keywords Traffic Control System · Emergency vehicles · Arrival time · Waiting time · SUMO

1 Introduction

The Traffic Control System is essential for organizing vehicles' movement and managing the chaotic situations of traffic congestion in cities' junctions. The continuous rise in vehicles' density has raised many concerns like air pollution, which causes environmental degradation, and noise pollution, which causes mental and

A. Sachan (✉) · N. Kumar

Indian Institute of Technology-Roorkee, Roorkee, Uttarakhand 247667, India

N. Kumar

e-mail: neetesh@cs.iitr.ac.in

physical stress on drivers. The current Traffic System divides the phase duration in an equal number of the time frame and then changes the Traffic Lights to Red/Green/Yellow accordingly. Often in case of malfunction of traffic signals, the light traffic intersections must be personally controlled by traffic police using hand signals. However, that is not the case every time. A particular lane does not always have congestion, so there is no need to fix any lane's time duration. It should be dynamic, real-time, and relative to the number of vehicles present at any amount of time in that lane.

Traffic Control System should be enhanced to establish a smart city in the future in the generation of advanced technology [1–7]. For this, the existing standard way of traffic control would not work [8, 9]. Thus, we have to shift to an Intelligent Traffic Control System which would sense the conditions and adjust relatively to the situation. ITCSEV (Intelligent Traffic Control System for Emergency Vehicles) model will be used to inform the traffic signals to change the time duration and color of the traffic signal accordingly when an emergency vehicle is approaching the junction. It will also result in less waiting time, reduced wastage of fuel, decreased noise and air pollution, and the cutback in road accidents. The rapid development in networking and data collection has permitted the execution of an Intelligent Traffic Control System that can have accurate records of the real-time traffic data daily of different types of vehicles, i.e., car, bus, police, fire brigade, etc. [4]. The proposed Intelligent Traffic Control System for Emergency Vehicle (ITCSEV) model detects the emergency vehicle from the vehicular traffic network through the induction loop detectors. It changes the traffic light accordingly to Red/Yellow/Green state. A functional simulation is being generated using an open-source simulator tool to check the feasibility of the proposed model, i.e., Simulation of Urban Mobility (SUMO) [10–12] on the map of Gwalior City, India. Also, the proposed ITCSEV model results are compared with the standard simulation, emergency simulation, and real-time traffic situation.

2 Related Work

Plenty of researches are being done to monitor the traffic lights in recent times efficiently and effectively. Previously, the researchers focused mainly on linear programming, fuzzy logic, deep reinforcement, etc. [10, 11]. However, all these techniques tried to handle the traffic scenario with a piece of finite and outdated knowledge, and none of them had any method to detect an emergency vehicle on the vehicular network [1, 2].

In this ITCSEV model, an algorithm detects an emergency vehicle through an induction loop detector that identifies the vehicle type. If the detected vehicle turns out to be an emergency vehicle [7], that lane's traffic light turns "Green" accordingly. This model also stores the vehicles' records like waiting time, CO₂ (carbon) emissions, average queue length, etc. Earlier, there was no way to have these records because

of the unavailability of advanced cameras and sensors. Many issues in the vehicular network could not be solved due to the absence of this critical information.

There are mainly two drawbacks in the previous works. Firstly, most researches test the simulations in a simple traffic scenario that lacks reality, and secondly, there was no uniqueness given to any emergency vehicle. In contrast, in the ITCSEV model, the simulation is being tested on the map of Gwalior city of India and assigned siren to the emergency vehicle for proper identification [5, 6].

3 The Proposed Protocol

3.1 Problem Statement

This model focuses on prioritizing the vehicle based on vehicle types like the car, bus, emergency vehicle (ambulance, police, and fire brigade), and then dynamically decides to control the traffic suitably. The proposed ITCSEV model operates in two stages as shown in Fig. 1. In the first stage, the Traffic Control System collects various information of different vehicles from the vehicular network with the help of induction loop detectors, which work as a sensor. The second stage uses the captured data to decide whether to pass the particular vehicle by changing the traffic light.

The traffic light has three signals, i.e., Red/Yellow/Green, so a single traffic light cannot control the traffic, especially when the vehicles are coming from different directions at one junction. In such a case, an emergency vehicle cannot find its

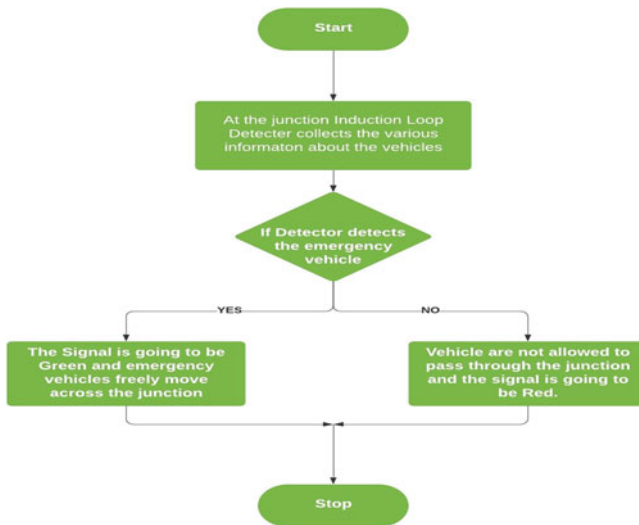


Fig. 1 Flowchart of recommended method

way out of that chaos. So, there should be a proper algorithm that can identify the emergency vehicle and let it pass from that congestion.

The time duration of stay at one specific traffic signal, i.e., Red, Green, Yellow, is known as a single phase. The number of potential signals mark the total number of phases at that junction. At that intersection, all phases begin to shift in a repetitive cyclic fashion. This technique aims to set the phase of the signal process dynamically. So, if the signal is green (G), the vehicle will travel freely through the junction, and if the traffic signal is Red (R) at any junction, then the vehicles are not permitted to drive through.

In Fig. 2, the traffic signal is Red as there are no emergency vehicles currently at the junction. The induction loop detector detects the vehicle and stores its information. Now in Fig. 3, an emergency vehicle “Ambulance” is detected by the induction loop. So, the traffic light at that junction will be changed to Green, and free passage will be given to that vehicle.

The Intelligent Traffic Control System for Emergency Vehicles (ITCSEV) model learns to identify an emergency vehicle and respond with a better decision, which allows the Traffic Control System to manage the traffic situations on the vehicular network and to save a significant amount of time.



Fig. 2 Before ambulance is detected



Fig. 3 After ambulance is detected

3.2 Intelligent Traffic Control Model

ITCSEV model is different from previous models, as in this model, there are induction loop detectors that keep the record of the traffic from the vehicular network. Numerous dictionaries of junction nodes are created in this model, which stores the information of the traffic light phases on different lanes like, for example, N1 = “R6”: “GgggrrrGrrr,” “R1”: “GrrrrRgGGGg,” “R4”: “GrrgGGgGrrr,” so basically, here N1 is a dictionary which is used as a junction node, R1, R4, R6 are the keys which represent the lanes that intersect to form a junction, and “G,” “g,” “R,” “r” are different phases of the traffic light.

Two parameters are used to measure the state of traffic at the intersection, i.e., the location and velocity of a vehicle in the vehicle network. The entire traffic intersection is broken into square grids of the same size. Two values (position, velocity) of the vehicle within the grid describe the state-space for each grid. The position is a binary value where zero (0) means that inside the grid there is no vehicle, and one (1) means the presence of the vehicle. The velocity dimension is a float value that shows the vehicle’s current velocity in a meter/second. In Tables 1 and 2, the traffic conversion into state-space is shown.

The model is smart enough to pick a suitable action based on the traffic scenario for the smooth flow of the vehicles at the junction. The model scans the state of the traffic in this system and chooses one of the two acts:

- 0-does not change the phase of traffic signal, and
- 1-turn on the green lights sequentially for the next traffic light.

Firstly, the model calculates the load (weight) of the vehicles on the vehicular network using the function cal weight. In Algorithm 1, the conditional loop for takes in a vehicle id from a list of vehicles on the vehicular network list1. Then, the type of vehicle is initialized and compared with different vehicle types, i.e., police, ambulance, bus, etc., and weight is set accordingly to the respective vehicle types.

Table 1 Presence of the vehicle in the traffic grid

1	0	0	0	0	1	1
0	0	1	0	0	1	1
0	0	0	1	0	0	1
0	1	0	0	0	1	1

Table 2 Velocity of the vehicles in the traffic grid

0.9	0	0	0	0	0	0
0	0	0.6	0	0	0	0
0	0	0	0.5	0	0	0
0	0.5	0	0	0	0	0

Algorithm 1 Find weight**Input:** list1 (list of vehicles)**Output:** Load (weight) of different vehicles

```

1. Function cal_weight(list1)
2.   weight = 0;
3.   for vehicles in list1 do
4.     t = current vehicle type
5.     if t == police then
6.       | weight = weight+3;
7.     end if
8.     if t == bus then
9.       | weight = weight+5.5;
10.    end if
11.    if t == truck then
12.      | weight = weight+6;
13.    end if
14.    if t == motorcycle then
15.      | weight = weight+0.5;
16.    end if
17.    if t == ambu then
18.      | weight = weight+5;
19.    end if
20.    if t == fire then
21.      | weight = weight+5;
22.    else
23.      | weight = weight+3;
24.    end if
25.  end for
26.  return weight
27. end

```

Algorithm 2 Find vehicle between detector and traffic signal**Input:** list of detector id**Output:** dict1 (contains detector ids)

```

1. Function detector()
2.   for detector id(i) in list of detectors do
3.     value = list of vehicle that were on the induction loop
4.     lane1 = id of lane the induction loop is on.
5.     if value == True then
6.       | lane2 = id of lane the named vehicle was at the last step
7.       | if lane1 = lane2 then
8.         | add vehicle id of named vehicle to detector dictionary
9.         | apply set() on detector dictionary
10.        | convert detector dictionary to list (dict1)
11.      end if
12.    end if
13.  end for
14.
15.  for detector id(i) in dict1 do
16.    lane1 = id of lane the induction loop is on
17.    for vehicle id(k) in dict1[i] do
18.      | lane2 = id of lane the vehicle id(k) was at the last step
19.      | if lane1 != lane2 then
20.        | remove vehicle id(k) from dict1[i]
21.      end if
22.    end for
23.  end for
24.  return dict1
25. end

```

Now, the model finds the vehicle on each lane between the detector and traffic light with the help of induction loop detectors using function detector in Algorithm 2. The first conditional loop traverses the different detectors in the detector list. If the lane id of that lane on which the induction loop is on equivalent to the lane id of that lane on which the last vehicle was on then, the vehicle id of that vehicle is added to the dictionary of detectors and is converted to a list. The second conditional loop traverses the appended detectors list, and the lane is initialized with the id of the lane the induction loop is on. If the lane on which the induction loop is not equivalent to the lane of the last step vehicle, then the vehicle id is removed from the list of detectors.

Next, the model traverses the list of vehicles to find the type of detected vehicle through the detectors in Algorithm 3, and a counter is initialized to zero (0). If the type of vehicle is emergency like ambulance, then the counter is increased to one (1), and if the condition fails, the counter remains zero (0). Further in Algorithm 4, the waiting time and the total number of emergency vehicles are initialized to zero (0). The list of detectors is traversed, and the list of vehicles list1 is initialized with the value of the current detector id. Now that list of vehicles list1 is traversed and a variable is initialized for the type of vehicle that is detected, if that vehicle type matches an emergency vehicle then the total number of emergency vehicle is incremented by one (1), and the waiting time of that emergency vehicle is added to the total waiting time. Now, if a total number of emergency vehicles is greater than zero (0), the integer value of the average waiting time of the emergency vehicle is returned.

Algorithm 3 Find emergency vehicle**Input:** list1 (list of vehicles)**Output:** type of vehicle (emergency or other vehicle)

```

1. Function emergency_vehicle
2.   counter = 0
3.   for vehicle id in list1 do
4.     type = vehicle type of detected vehicle
5.     if type = vehicle (emergency) then
6.       counter = counter + 1
7.     else
8.       counter = 0
9.     end if
10.  end for
11.  return counter
12. end

```

Algorithm 4 Find waiting time of emergency vehicle**Input:** dict1 (contains detector ids), list1 (list of vehicles)**Output:** waiting time of emergency vehicle

```

1. Function avg_wt_vehicle
2.   waiting_time = 0
3.   total_e_vehicle = 0
4.   for detector id(i) in dict1 do
5.     list1 = detector id(i)
6.     for vehicle id in list1 do
7.       type = vehicle type of detected vehicle
8.       if type = emergency vehicle (ambulance) then
9.         total_e_vehicle = total_e_vehicle + 1
10.        waiting_time = waiting_time + (waiting time of detected vehicle)
11.       end if
12.     end for
13.   end for
14.   if total_e_vehicle > 0 then
15.     return integer value of (waiting time / total_e_vehicle)
16.   else
17.     return 0
18.   end if
19. end

```

4 Experimental Analysis

4.1 Simulation Study

The experimental study of the proposed ITSCEV through a practical simulation is discussed in this section. The simulation is prepared using the open-source simulator Simulation of Urban Mobility on the real map of Gwalior City in India (SUMO). The performance of the proposed ITSCEV model is evaluated in Normal Simulation, Emergency Simulation, and Real-time Simulation on the Waiting Time of the emergency vehicle. The simulated results demonstrate the feasibility of the model and how the waiting time of emergency vehicles on the vehicle network can be minimized.

4.2 Evaluation Methodology and Parameters

Using Python, the proposed ITSCEV model algorithms are implemented, and the simulation was developed in an open-source Urban MObility Simulation simulator tool (SUMO). In this model, the Fuzzy inference module is explicitly used to select an execution mode by observing the current status of vehicular traffic scenarios, and the model uses a free and open-source data flow software library, i.e., TensorFlow.

The simulation is carried out on the map of the Gwalior city of India by using a practical map that is converted into roads and traffic lights with the aid of SUMO conversion software, i.e., netconvert. The length of the vehicle depends on the type of vehicle present in the vehicles’ network. With over 1000 vehicles in the entire vehicle network, the vehicles’ arrival in the network follows randomly. SUMO also provides an interface to display the traffic flow in real time, i.e., TCI (Traffic Control Interface), and is also compatible with the Python API (Application Programming Interface), which is primarily used to retrieve vehicle’s information from the traffic light control junction and to control the duration of the traffic light accordingly.

In addition, the combination of Python and SUMO is very powerful with the aid of the Python API (TCI), enabling the user to monitor the SUMO traffic flow in Python through written algorithms. The length of yellow signal in the Traffic Control System is set at 5 s, which is transition time from green to red for the signal to change. The goal of our model is to minimize the waiting time of the emergency vehicle by changing the traffic signal to green whenever emergency vehicle is detected via the induction loop detectors.

4.3 Experimental Results

This subsection presents the details of the experimental results of the ITSCEV model proposed along with the comparison of different parameters with the other model.

The figures show the comparative results of the ITSCEV model based on average waiting time, impatience level, speed, and time loss of emergency vehicles through a graphical representation. The X-axis represents the number of simulation steps, also known as episodes. The Y-axis indicates various parameters as (a) average waiting time in seconds (s), (b) impatience level, (c) speed (m/s), and (d) time loss (s) of the emergency vehicle, respectively. The model is being run on three different types of simulations, i.e., Standard simulation in which the phase duration of traffic lights at junctions is fixed, Emergency Simulation in which the emergency vehicle is given free passage on traffic lights, and Real-Time Simulation, which runs on a real-world traffic light scenario.

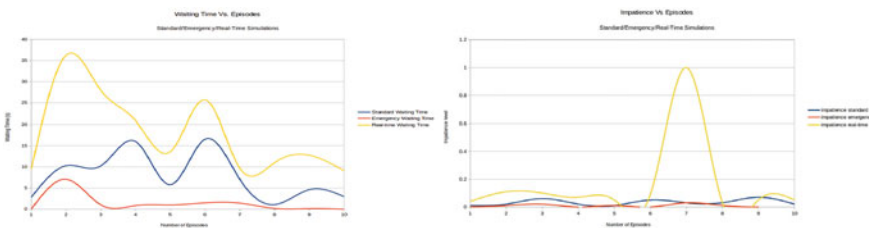


Fig. 4 Waiting time and impatience level

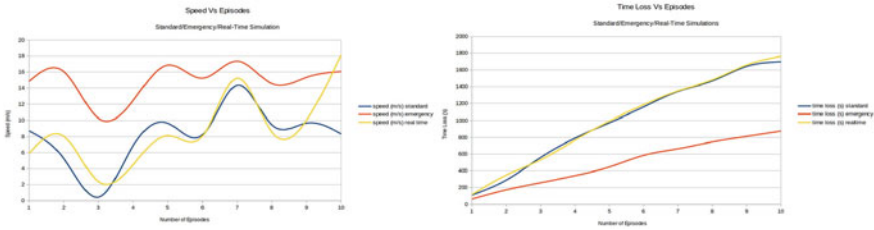


Fig. 5 Speed and time loss

In Fig. 4, the graph represents the average waiting time of the emergency vehicle at a junction. It is clearly visible that the emergency vehicle’s waiting time in the ITSCEV model is meager compared to another scenario. In Fig. 4, the graph depicts drivers’ impatience level on-road, which is very low in the emergency simulation compared to other scenarios. This will benefit the drivers as high impatience levels can cause stress and anxiety and can even lead to road accidents endangering people’s lives.

In Fig. 5, the graph represents the speed of the emergency vehicle on the vehicular network. So, if an emergency vehicle has to reach its destination on time, it should advance at a certain speed. In this model, since the priority vehicles are given a free passage at the traffic junction, the vehicle’s speed would be much higher than in other scenarios. In Fig. 5, the graph illustrates the time loss of emergency vehicles on the vehicular network in different scenarios. It can be seen that the time loss in the ITSCEV model is low compared to other simulations. This can perhaps result in saving many lives as the emergency vehicle can reach the destination on time.

It is evaluated through the series of experiments that the proposed model successfully reduces the average waiting time, the degree of impatience of drivers, and the time loss of emergency vehicles. It can also be used for high priority vehicles such as ambulances, police cars, and fire brigades in the future. Since the experiment consists of multiple vehicles, including emergency vehicles and high priority vehicles, the ITSCEV dramatically decreases the total waiting time.

5 Conclusion and Acknowledgement

The introduced model, i.e., Intelligent Traffic Control System for Emergency Vehicle (ITCSEV), explains the working of the induction loop detectors. This model will take data from different vehicles that will accordingly minimize the waiting time of emergency vehicles to a bare minimum. A practical simulation is generated using an open-source simulator tool, i.e., Simulation of Urban Mobility (SUMO) on the map of Gwalior city of India, to check the efficiency of the ITCSEV model. Normal simulation, emergency simulation, and real-time traffic simulation are compared with the results of the ITCSEV model.

Acknowledgements The authors would like to acknowledge Indian Institute of Technology-Roorkee, India for sponsoring the conference and the Council of Scientific and Industrial Research (CSIR), Govt. of India as this work is carried out under the project titled “Emergency Aware Intelligent Dynamic Traffic signal System for smart cities” with grant id 22(0817)/19/EMR-II).

References

1. Du, X., & Liang X. (2019). A deep reinforcement learning network for traffic lights’ cycle control in vehicular network. *IEEE Transactions on Vehicular Technology*.
2. Azimirad, E., Pariz, N., & Sistani, M.B. (2010). A novel fuzzy model and control of single intersection at urban traffic network. *IEEE Systems Journal*, 4 107–111.
3. Hammi, B., Khatoun, R., Zeadally, S., Fayad, A., & Khoukhi, L. (2017). IoT technologies for smart cities. *IET Networks*, 7(1), 1–13.
4. Shah, S.H., & Nahrstedt, K. (2002) Predictive location-based QoS routing in mobile ad hoc networks. In *2002 IEEE International Conference on Communications. Conference Proceedings. ICC 2002 (Cat. No.02CH37333)*, vol. 2 (pp. 1022–1027). New York, NY, USA.
5. Kompella, V. P., Pasquale, J. C., & Polyzos, G. C. (1993). Multicast routing for multimedia communication. *IEEE/ACM Transactions on Networking*, 1(3), 286–292.
6. Roy, A., & Das, S. K. (2004). QM2RP: A QoS-based mobile multicast routing protocol using multi-objective genetic algorithm. *Wireless Networks*, 10, 271–286.
7. Kumar, V., & Kumar, S. (2016). Energy balanced position-based routing for lifetime maximization of wireless sensor networks. *Ad Hoc Networks*, 52, 117–129, ISSN 1570-8705.
8. Haghghat, A.T., Faez, K., Dehghan, M., Mowlaei, A., & Ghahremani, Y. (2003). GA-based heuristic algorithms for QoS based multicast routing. *Knowledge-Based Systems*, 16(5–6), 305–312, ISSN 0950-7051.
9. Meraihi, Y., Acheli, D., & Ramdane-Cherif, A. (2019). QoS multicast routing for wireless mesh network based on a modified binary bat algorithm. *Neural Computing and Applications*, 31, 3057–3073.
10. Fahad, M., Aadil, F., Rehman, Z.U., Khan, S., Shah, P.A., Muhammad, K., Lloret, J., Wang, H., Lee, J.W., & Mehmood, I. (2018). Grey wolf optimization based clustering algorithm for vehicular adhoc networks. *Computers and Electrical Engineering*, 70, 853–870, ISSN 0045-7906.
11. Genders, W., & Razavi, S. (2016). Using a deep reinforcement learning agent for traffic signal control. arXiv preprint [arXiv:1611.01142](https://arxiv.org/abs/1611.01142).
12. Kumar, N., & Rahman, S.S. (2019). Deep reinforcement learning with vehicle heterogeneity based traffic light control for intelligent transportation system. In *2019 IEEE International Conference on Industrial Internet (ICII)*, IEEE, pp. 28–33.

Digital Controller-Based Automated Drainage Water Monitoring and Controlling



T. Sairam Vamsi, Ch. Hari Krishna, P. Srinivasaraju, and G. Srinivasarao

Abstract The main purpose of this paper is to replace the manual work in sewage cleaning and monitoring with automation. In this era, there is a lack of utilization of technology for sewage disposal tasks. At present, manual effort has been used for drainage cleaning, which leads to loss of human life while cleaning the blockage in pipes. In order to overcome these types of problems, “Digital Controller based automated drainage water monitoring and controlling” will be helpful. This project includes PLC (Programmable Logic Controller), which plays a vital role in solving real-time problems. Using the resource in this paper, it is possible to monitor and control the sewage system continuously without using manual work.

Keywords PLC · Sewage disposal · Automation

1 Introduction

Drainage is the system by which excess water or other liquids are drained from a place. These are also in place to remove wastewater effectively, and this is referred to as a sewer system. One of the main functions of the drainage system is to collect and transport waste materials along with water [1]. Due to the solid waste, the flow of water will get blocked in drainages, these blockages lead to improper flow of drainage system. The drainage system must be cleaned time to time. In this present era, though there is high development in technology, cleaning of these drainages is done manually. The person who enters to clean the drainage is affected due to hazardous gases and faces serious health issues. Sometimes, there may be loss of

T. S. Vamsi (✉) · Ch. Hari Krishna · P. Srinivasaraju · G. Srinivasarao
SVECW, Bhimavaram, India

Ch. Hari Krishna
e-mail: harikrishnach@svecw.edu.in

P. Srinivasaraju
e-mail: viceprincipal@svecw.edu.in

G. Srinivasarao
e-mail: principal@svecw.edu.in

human life. So, there is a need to use automation in drainage systems to remove manual work.

There have been many methods used till now for cleaning of drainage systems but there is no completely automated method [2]. Here, in this paper, we are going to eliminate manual work entirely for total working of cleaning the drainage system. The method which is implemented in this paper helps in preventing the blockages in drainages automatically. This method is used to extract the solid waste materials from the wastewater. A series of mesh layers are employed at the flow of drainage water in order to collect the solid waste materials. The level of wastewater in the water tank is monitored using the level sensors. The collected solid waste will be dumped into a garbage bin with the help of mechanical motion based on the output from the sensors. There the waste materials will end up into tiny particles. The pumping of water into the mesh layers is controlled by the Level and Ultrasonic sensors [3].

2 Block Diagram

The block diagram for the proposed solution is shown in the Fig. 1.

1. The project utilizes two Level sensors, one Ultrasonic sensor, and one pH indicator.
2. One Level sensor is placed at the low level and one is at high level in the water tank to measure the level of the water [4].

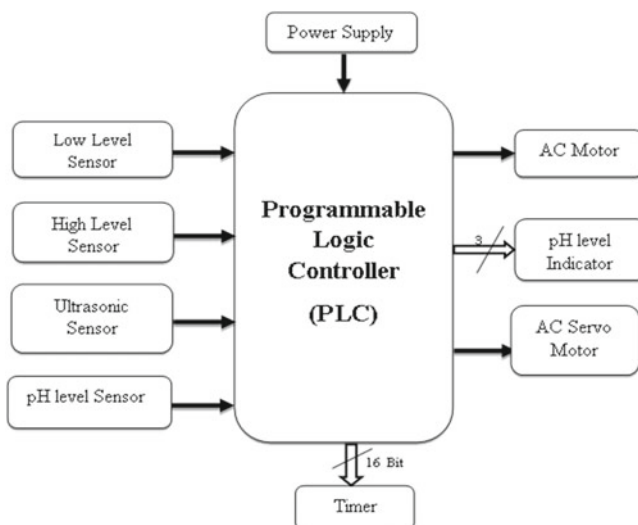


Fig. 1 Block diagram of the proposed solution

3. In the net collecting solid waste, an Ultrasonic sensor is placed in order to check the solid waste level.
4. The pH indicator is used to get the acidic and basic nature of the water, thereby LEDs are used to indicate the current range of pH [5].
5. By taking the input from the sensors through Programmable logic Controller (PLC), AC motor and AC servo motor will be swiveled [6].
6. Timer will be used to provide delay between the tasks.

3 Implementation of the Project Using PLC

The project “PLC Implementation of Control Logic for Monitoring and Controlling of Drainage Water System” is implemented in two stages [7].

1. Software Implementation.
2. Hardware Implementation.

The Software Implementation of the project is shown in Fig. 2. Initially, the values from the corresponding sensors were taken. If waste level sensor is in OFF state, then the slider will be in rest state, then depending on low-level sensor and high-level sensor conditions, motor will be either in OFF or ON state [8]. Otherwise, the waste level sensor will be in ON state, then simultaneously the slider is in ON state, and timer also starts working. After the timer stops, the waste level indicator will go to OFF state. The pH sensor will be continuously in working state irrespective

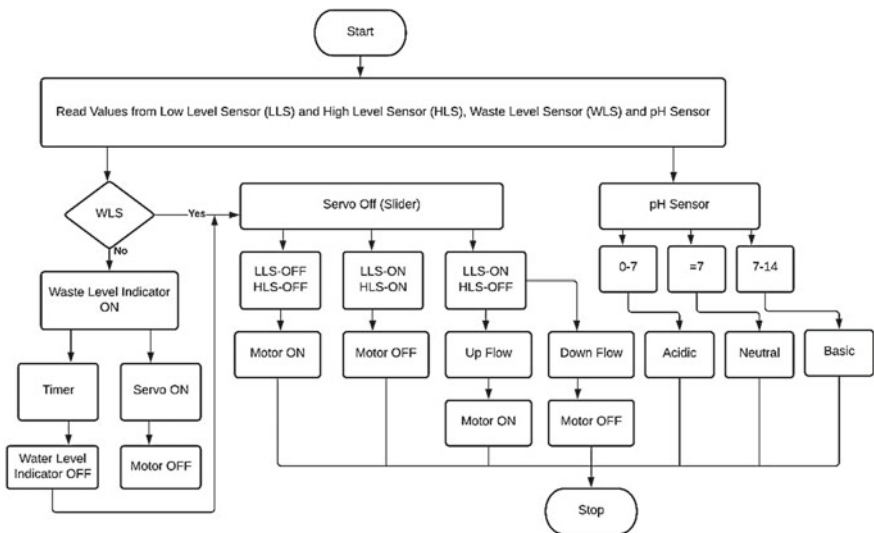


Fig. 2 Flow chart of software implementation

Table 1 Hardware components and its specifications

S. No	Name of the component	Specification
1	PLC	DVP-14SS11R2
2	8 channel relay cards	Single changeover
3	Terminal block	–
4	SMPS	230VAC–(19–26)V DC
5	Miniature circuit breaker (MCB)	5SL61027RC
6	Level sensor	–
7	Normally open and normally closed switches	–
8	DC motor	12 V
9	Inductive sensor	–
10	Buzzer	12 V DC
11	Indicator	Pilot Indicators

of the remaining sensors and always note corresponding pH values which will feed respective indicators (Table 1) [9].

The hardware implementation of the project is shown in the following Fig. 3.

In the Fig. 3, NC Switch, NO Switch, Inductive Sensor, Level Sensor are used as inputs, and Buzzer, DC Motor, Pivot Indicator are used as outputs. Here, NC Switch is used as low-level sensor to detect the wastewater level in the water tank, Level Sensor is used as High-level sensor to identify the water level at the top of the tank [10]. Inductive sensor acts as Mesh Level Sensor for identifying the level of waste in the mesh layer. In this prototype, DC Motor replaces the operation of AC Motor to pump the water. Based on the pH values of water, when the pH level is in between

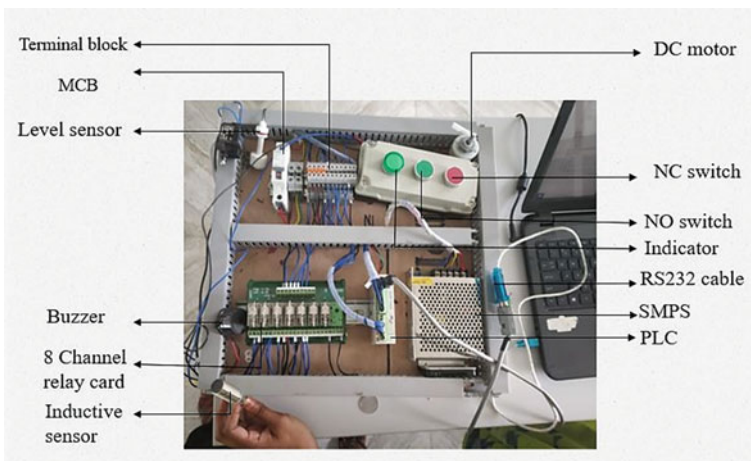


Fig. 3 Hardware implementation of the project

0 and 7, we will be able to hear a buzzing sound to indicate the nature of water as acidic. If the pH is 7, Pivot indicator will be ON to represent that the water is in neutral nature [11]. If the pH level is above 7 and below 14, both buzzer and pivot light will be in OFF condition to indicate the nature of water as Basic.

4 Results and Discussions

Initially, all the sensors are in their respective positions. Based on the inputs given by the sensors (Level sensors, pH sensor), the operation of ac motor and ac servo motor will be functioning. Table 2 shows sensor functionality with respect to servo motor with two different levels, Low and High.

4.1 Detailed Description of Stages

The following cases show the working stages of sensors and motor with respect to software implementation.

Case 1: If the Low-Level Sensor and High-Level Sensor are in OFF stage, then the motor will be in ON stage. Similarly, when LLS or HLS is OFF, then Motor will be in ON Stage. Figures 4 and 5 show the PLC outputs and inputs when the LLS and HLS are in different conditions. X0, X1 indicates LLS and MMS, Y0 indicates output, i.e., Motor.

Note: In figures, Color indicates OFF and Normal indicates ON (Input Side), Color Indicates ON and Normal indicates OFF (Output Side).

Table 2 Stages of sensors and motor functioning based on their conditions

AC Servo motor	Low-level sensor (LLS)	High-level sensor (HLS)	AC motor
ON	X	X	OFF
<i>Upward flow</i>			
OFF	OFF	OFF	ON
OFF	ON	OFF	ON
OFF	OFF	OFF	ON
OFF	ON	ON	OFF
<i>Downward flow</i>			
OFF	ON	OFF	OFF
OFF	OFF	OFF	ON

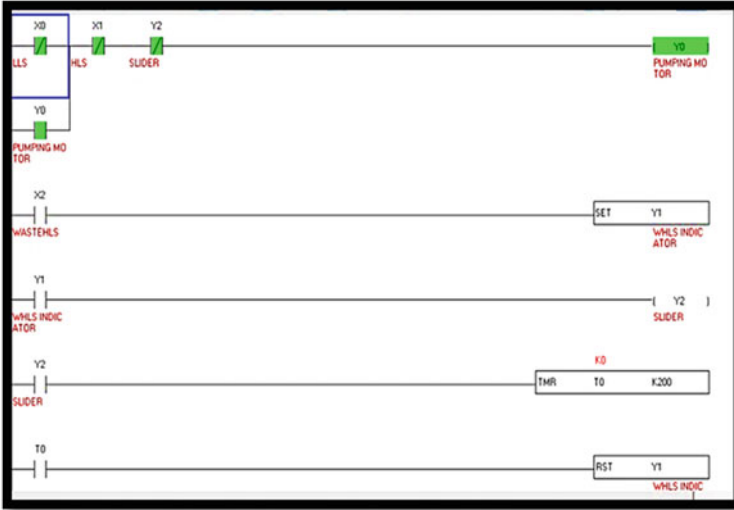


Fig. 4 PLC output when LLS-OFF, HLS-OFF

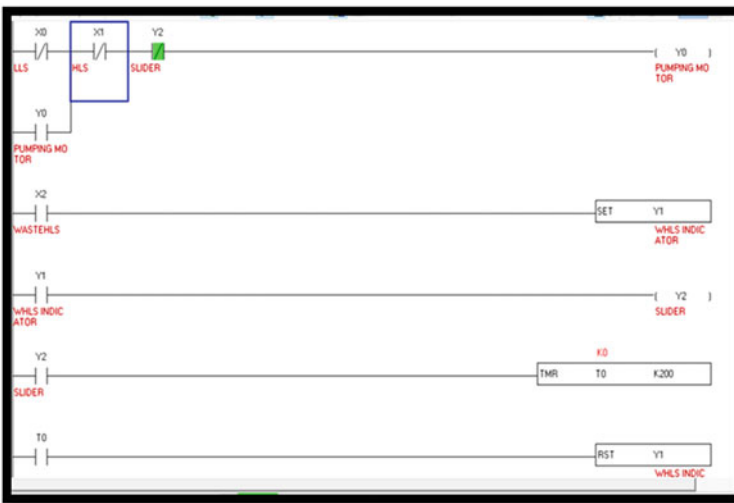


Fig. 5 PLC output when LLS-ON, HLS-ON

Case 2: Figure 6 shows the PLC output when servo motor is ON and Low-level Sensor, Servo Motor (Slider) are in ON Stage and High-level Sensor is in OFF Stage, then the Motor will be in OFF stage.

Case 3: If the Waste Level Sensor is in ON stage, then the Servo Motor (Slider) will be ON stage which drives the Motor into OFF stage automatically.

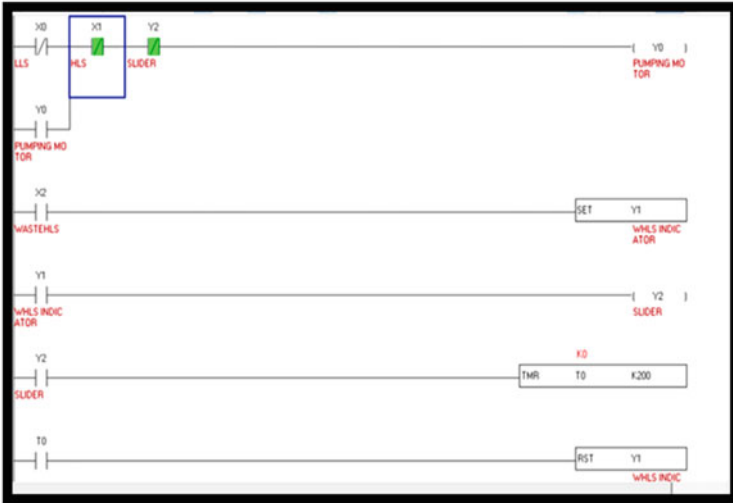


Fig. 6 PLC output when LLS-ON, HLS-OFF, servo motor-ON

Case 4: If the waste level sensor is in OFF stage and the Servo motor is in ON stage, then the Motor will be in OFF stage.

The above Figs. 7 and 8 show the PLC output when controlling with Waste level sensor which is indicated as X2.



Fig. 7 PLC output when WLS-ON, servo motor-ON

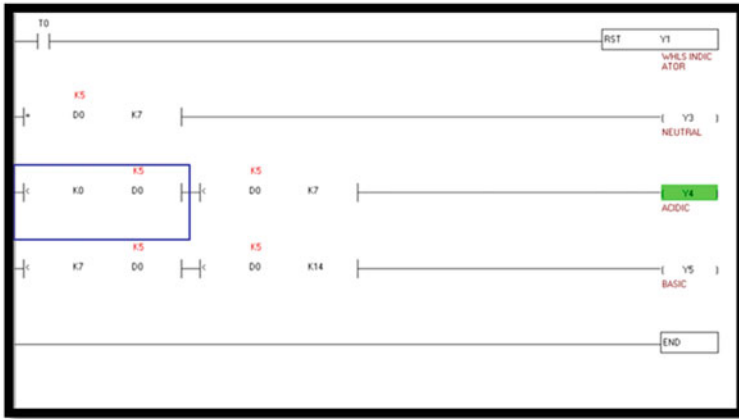


Fig. 10 pH acidic indication

Case 1: If the pH level of the water is equal to 7, then the Neutral Indicator will be ON.

Case 2: If the pH level of the water is below 7, then the Acidic Indicator will be ON, and If pH level of water is above 7, then the Basic Indicator will be ON.

5 Conclusions

The drainage system usually gets blocked because of the solid waste present in water. These get jammed in small holes in the path of water flow, which leads to back flow of water on to the roads. There are many cleaning techniques for the sewage disposal tasks. The project implemented in this paper is more reliable and fully automatic. This is mainly used in large workplaces like industries, hospitals, etc. It can separate the solid waste completely. A PLC logic for drainage system monitoring and controlling is designed in DELTA WPL Soft 2.49 in order to collect and separate the solid waste from drain water. The collected waste is made into tiny pieces at the garbage bin.

Acknowledgements This work is carried by Applied Robotic Control (ARC) Lab in Shri Vishnu Engineering College for Women (A).

References

1. Sathiyakala, R., et al. (2016). Smart sewage cleaning system. *International Journal of Innovative Research in Computer and Communication Engineering*, 4.
2. Xu, J., & Wang, W.B. Design of sewage treatment system based on arm controller. In *2016 28th Chinese Control and Decision Conference (CCDC)*.
3. Haswani, N.G. et al. (2018). Web-based realtime underground drainage or sewage monitoring system using wireless sensor networks. In *Fourth International Conference on Computing Communication Control and Automation*.
4. Deshpande, V. S., & Vibhute, A. S. (2014). Home automation using PLC and Amit S. Vibhute SCADA. *Multidisciplinary Journal of Research in Engineering and Technology*, 1(1), 111–118.
5. Nandgave, A., Deshbhratar, H., Khandare, S., & Heda, L. (2014). Industrial drives & automation using PLC. *International Journal Of Engineering Research & Technology (IJERT)*, 3(2).
6. Sairam Vamsi, T., & Srinivasa Raju, G.R.L.V.N. (2018). ARM-based industrial circuitous like robotic implementation using CAN bus for surveillance. In S. Satapathy, V. Bhateja, P. Chowdary, V. Chakravarthy, & J. Anguera (Eds.), *Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications*. Lecture Notes in Electrical Engineering, vol 434. Springer, Singapore. https://doi.org/10.1007/978-981-10-4280-5_28.
7. Sairam Vamsi, T., Viswanadam, R., & Praveen kumar K. An industrial surveillance sensoric robot controlling through Arduino and GSM. *International Journal of Control Theory and Applications (IJCTE)* Scopus Indexed with ISSN:0974-5572, 9(23), 87–96.
8. Du, Z.B. (2014). Study on treatment of wastewater system based on ARM. *Advanced Materials Research*. <https://doi.org/10.4028/www.scientific.net/AMR.1006-1007.732>.
9. Zhang, Q., Wang, X., & Fang, Z. Water treating control system based on predictive control theory of multi-model switching algorithm. *Proceedings of the 5 World Congress on intelligent Control and Automation*, 4(5), 633–636.
10. Peng, X., Xiao, L., Mo, Z., & Liu, G. (2009). The variable frequency and speed regulation constant pressure water supply system based on PLC and fuzzy control. *IEEE International Conference on Measuring Technology and Mechatronics Automation*, 910–913.
11. Xinping, W. (2007). Application of PLC program control technology in the field of industrial water treatment. *Automatic Instrument*, 28, 168–171.

Cooperative Agent-Based Location Validation for Vehicular Clouds



Shailaja S. Mudengudi and Mahabaleshwar S. Kakkasageri

Abstract The evolution of Intelligent Technology Systems (ITS) has made possible the designing of vehicles with more sophisticated computing, communicating, and sensing capabilities. Vehicular Cloud (VC) magnifies these capabilities by cost-effectively sharing their resources. Location information about a node plays a prime role in many VC services which are based on the location of the user node. In this paper, we present a cooperative agent-based location verification framework for VC. Central Authority (CA), a trusted entity that coordinates the activities in VC, takes care of the location verification process with the help of special nodes known as verifier nodes. Verifier nodes are distributed in the network randomly and secretly. The presence of such nodes is known only to the CA, which makes it stronger in terms of security.

Keywords Vehicular Cloud · Cloud computing · Public key encryption · Software agent

1 Introduction

Vehicular Ad-hoc Network (VANET) is an area of mobile ad-hoc networks intending to provide safety for its users. The increase in the number of smart vehicles on road and evolution in VANET technology enables provision for applications, such as driver safety, enhanced traffic management, and infotainment. The most important of them is the safety application where resources with the vehicles are used effectively to increase the safety of the user in VANET. The vehicle nodes are converted to network devices with communicating and computing capabilities to help the user with safety messages like collision avoidance, obstacle detection, turn, and speed limit warnings. Along with safety, there is a huge demand for applications that ease the driving experience and provide entertainment [1, 2].

S. S. Mudengudi (✉)
Tontadarya College of Engineering, Gadag 582101, India

M. S. Kakkasageri
Basaveshwar Engineering College, Bagalkot 587102, India

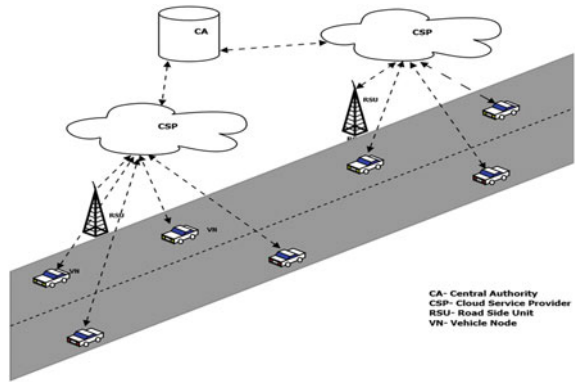
The VANET infrastructure consists of Road Side Unit (RSU) installed at places like sides of the road, and On Board Unit (OBU) which are deployed in the vehicle nodes. OBU and RSU communication facilitate safety applications and other services. As safety related information contributes to the lives of humans participating in VANET, security is of utmost crucial importance [2]. A comprehensive and novel design where resources in VANET are effectively used in collaboration with the cloud is presented in [3], which is termed Autonomous Vehicular Clouds (AVC). Resources in AVC are dynamically pooled to form the cloud and provide services to the authorized user. The aim is the effective utilization of resources, with greater elasticity and from a commercial perspective [4]. The resources from the vehicles, RSU, and conventional cloud together form the Vehicular Cloud (VC). Fundamentals of VC, designing of VC architecture, operations, and services of VC are discussed in [5]. The author also further states, the VC is still under the primitive stage, and issues such as resource discovery, fault tolerance, or management of service-related content storage, and many others still need to be addressed. VC architecture presented in [6] is classified into on board unit, communication, and cloud service layers. The on board unit is installed on the vehicle node which consists of sensors, storage units, etc., with computing capabilities. The communication layer is responsible for the vehicle node to infrastructure and vehicle node to a vehicle node. The cloud computing layer is subdivided into resources, virtualization layer, services, application programming interface, and application layer. VC facilitates several services that may require or may be based on the location of a node. Several methods are present to get location information. It may be done by the node itself using technologies like GPS (Global Positioning System) or by neighboring nodes. VC services are economy-based on a greedy or malicious node that may manipulate its location information to enjoy more facilities than allotted or subscribed. Attacks such as position-spoofing intend to manipulate the position-related information of the node [7].

To the best of our knowledge, until now, there is no mechanism to prove or detect a vehicle nodes position in the VC network. We present an agent-based location verification service, which verifies the location of the vehicle node using verifier nodes deployed randomly in the network. A vehicle node may be denying about its position during a dispute after utilizing the cloud resources available at that location. In case of dispute, the following actions are being actuated:

- The Central authority which is the most trusted and secured entity sends a query regarding the location of the particular vehicle node at a required time.
- The location validation agency verifies the location of the node with the help of verifier nodes that are deployed randomly in the network.
- The knowledge about the presence of the verifier nodes is known only to CA.

This elevates the attacks which are node targeted and increases the security of VC. The VC architecture is shown in Fig. 1. VC consists of Vehicle Nodes (VN), Road Side Unit (RSU), Cloud Service Provider (CSP), conventional cloud, and Central Authority (CA). The vehicle nodes are capable of communicating with CSP, neighboring VN's, and RSUs. Some of the vehicle nodes or the RSU are chosen as verifier nodes by the CA. CA is the trusted entity that coordinated all the activities of VC.

Fig. 1 Vehicular cloud network



It generates the public and private key pair for the verifier nodes and delivers it to respective nodes securely. The following assumptions are made:

- The verifier nodes are chosen by CA and known only to CA.
- The verifier nodes and CA have high storage and computational capabilities.

The paper is organized as follows. Section 2 presents related work, Sect. 3 presents the proposed cooperative agent-based location validation method, the conclusion is presented in Sect. 4.

2 Related Work

VANET is being attractive to elevate issues such as traffic congestion and accidents along with providing security with increased efficiency. Besides, it also faces attacks targeted towards the security and privacy of the user [8, 9]. Aiming to preserve the privacy of the user, identity-based group signatures (IBGS) method is proposed in [10]. The identity of the user is used in the generation of public keys, eliminating the certificate management issue. Intending to avoid malicious users accessing privileges or benefits using fake location claims, a block chain-based location proof generation with verification is presented in [11]. In case of dispute in location-based services, there would arise a scenario in which a node needs to prove its location at a specified instant. A prover node, which needs proof of its location, requests the neighboring nodes via short range communication methods such as Bluetooth. Nodes that are willing to respond to the prover node request authenticate the prover node and location proof signed by it is generated and broadcasted on to peer-to-peer. Location verification is employed as a security mechanism in [12] which uses the Global Navigation Satellite Systems (GNSS) for providing location verification securely with global coverage. The advantage of this system is it can be used by nodes that do not have a facility of navigation or GPS. A virtual force model-based location verification based on anchor promotion algorithm for WSN using incremental

refinement is presented in [13]. The mechanism proposed re-locates the drifted nodes and elevates the malicious anchor-based attacks. A two-level verification of location information method is presented in Information Verification cum Security (LIVES) [14]. The first level of location verifications is done using tiles-based verification process where the position of a vehicle node is calculated in terms of latitude and longitude by itself and announced to its neighbors. Next, it is verified based on the signal strength received by the neighboring nodes using the shadowing propagation model. In the second layer, it uses a collective belief-based Transferable Belief list from the neighboring vehicles, for the two-fold verification.

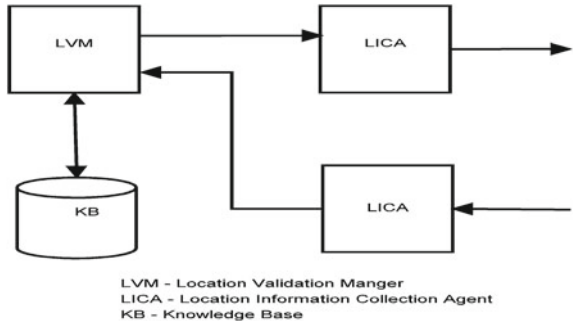
A detailed study for data location verification in the cloud is presented in [15]. Different location verification methods along with vulnerabilities due to malicious CSP attacks are discussed. Further landmark-based approach like a landmark, types of landmarks, landmark distribution scale, landmark selection, Coordination of participant's training, location granularity is presented. In the position-based routing method, participating nodes need to know the destination node's location along with recognizing its geographic locations to route messages. Location service provides current geographic positions of nodes. In [16], a protocol for providing such location service securely using location information which is self-signed by the nodes is presented. But the protocol does not preserve the privacy of the node regarding location information. The location information over a geographic area that is visited by them frequently over a long period is observed and is defined as their Physical Social Location (PSL) which is further used in Mobile phone-based Physical Social Location (MPSL). In the location verification process, the nodes in the PSL act as witnesses to prove the presence of the node [17].

A detailed study of recent trends in cloud architecture is presented in [18]. It defines the cloud as a technology where a cloud user can access the resources from the pool of resources which may be application or storage or infrastructure. This can be achieved by cloud services such as software as a service, platform as a service, and infrastructure as a service which can be deployed in private or public or hybrid cloud structure. The location verification in the cloud is usually the verification of the location where the data is stored. In [19], geo-location-based approach, location verification is achieved using supervised classification algorithms. A literature survey in [20] presents the study of the entire spectrum of agent-based modeling and simulation tools. It also discusses the features, merits, and faults of the software. It also provides guidelines for selecting an appropriate agent-based model and its simulation tool. Public Key Infrastructure (PKI)-based authentication process is presented in [21] which find vast applications such as e-commerce, web services, etc.

3 Proposed Work

We present an agent-based location verification method which consists of a Location validation agency as shown in Fig. 2. The agent comprises of Location Validation

Fig. 2 Location validation agency



Manager (LVM), Location Information Collection Agent (LICA), Knowledge Base (KB).

Knowledge Base (KB): It consists of location-related information about vehicle nodes, which include entry and exit time into the geographical area, service taken or given to the cloud, trust rate, etc. The knowledge base is updated frequently by the information collected by LICA by the LVM.

Location Validation Manager (LVM): The Location Validation Agency is installed in the CA. The complete coordination between all the agents is executed by the LVM. When a query is received by the location validation agent, LVM triggers the LICA to collect location information from all the verifier nodes, which are distributed in the network. The knowledge about the verifier nodes is known only to the CA. The LICA visits these verifier nodes only instead of visiting all the nodes. This further reduces the cost of the verification process. The security aspect also increases because the knowledge about the verifier nodes is known only to CA. So attacks can be elevated which are node targeted.

Location Information Collection agent (LICA): The LICA collects information about the vehicle node location from the verifier nodes. The location details are encrypted using public key encryption.

The encrypted message is received by the CA. Table 1 shows the nomenclature. After decrypting the message, the location information is derived from the message. It consists of the ID of vehicle, location information, timestamp about arrival and exit of the vehicle node into the geographical area. The location of the vehicle is tracked from the location information collected by the TICA which is stored in KB.

Let $V_1, V_2, V_3, \dots, V_N$ be Verifier nodes. The location information of node ‘M’ is sent by node ‘V’ by signing the location information as shown in Eq. 1.

$$LocM_V = SigV (V||PosM||TrustM||TVI||TVC||TMI||TMO) \tag{1}$$

CA sends Query as shown in Eq. 2 to verifier node ‘V’ about Node ‘M’ location. This is encrypted using public key of ‘V’ as shown in Eq. 3. ‘Type’ tells the kind of message it is like, a location query message or a simple encrypted data, etc. ‘seq’ is a number used to keep track of all the query messages.

Table 1 Nomenclature used in proposed work

X	ID of node V
V_{PK}, V_{PR}	Public and Private key of node V
$SigV(LOC)$	Location-related information LOC signed by verifier node V using V_{PR}
$CertV$	Self-signed certificate of A 's public key
TVI	Timestamp representing the time APR was first generated
TVC	Timestamp representing the current time generated by A
$PosV$	V 's geographical location (coordinate) obtained from GPS
TMI	Timestamp representing the entry time vehicle node M into the geographical area of V
TMO	Timestamp representing the exit time vehicle node M into the geographical area of V
$Trust_M$	Trust factor of node M

$$\text{Loc request} = [\text{Type, seq, } M] \quad (2)$$

$$\text{Encrypted message} = V_{PK}[\text{Loc request}] \quad (3)$$

Only 'V' can decrypt Eq. 3 as the message is encrypted using public key of V. The Node 'V' decrypts the message using its private key. If Node 'V' has an encounter with Node 'M,' it responds as shown in Eq. 4, else it does not respond. This message is encrypted using public key of CA as shown in Eq. 5.

$$\text{Loc response} = [\text{Type, Seq, Loc}M_V] \quad (4)$$

$$\text{Encrypted message} = CA_{PK}[\text{Loc response}] \quad (5)$$

The message received by CA can only be decrypted by CA as it has the private key. So the PKI elevates the malicious nodes manipulating the location information at any phase. So the public and private key pair is very important and is crucial. The public and private key pair is generated using RSA algorithm as shown in Table 2. The following steps show the generation of keys. The proposed work has the following advantages:

1. A separate pair of public and private keys can be used for the location verification process. As the number of verifier nodes is low, there is no significant increase in the storage requirement.

Table 2 RSA algorithm

Generation of keys:

1. Choose two large prime numbers 's' and 'r.'
2. Calculate $n = sr$.
3. Calculate $\phi(n)$ using $\phi(n) = (s - 1)(r - 1)$

Here $\phi(n)$ is the *totient function*, also known as *Euler's totient function*. It is the number of integers strictly smaller than 'n' that is relatively prime to 'n,' where 1 is counted as being relatively prime to all numbers

4. Choose a number 'e' which is relatively prime to $\phi(n)$
 5. To decrypt the message using $de \equiv 1 \pmod{\phi(n)}$ 'd' is the private key can be found using the extended Euclidean algorithm
-

Encryption of message:

At the transmitter, the message 'L' is encrypted using the public key

$$E = L^e \pmod{n}$$

Where E is an encrypted message and L is the location details of the node

Decryption of message:

At the receiver, the secret information 'd' regarding the secret key is used for decryption

$$E^d \equiv (L^e)^d \equiv L^{ed} \equiv L^{N\phi(n)+1} \equiv L \pmod{n}$$

2. The CA has a list of all the verifier nodes location, public, and private key details in its database which is further divided, based on geographical area.
3. When there is a need to verify the location of a node, the CA checks the list of verifier nodes in the geographical location. Send query message including the ID of a vehicle about which information is needed. The query message is encrypted using the public key of the verifier node which ensures only the verifier node can read the query. As there is no need to request a public key certificate, there is a considerable reduction in the number of messages transferred between CA and the verifier node.
4. After decrypting the message, the verifier node checks its database and responds only if it has an encounter with that vehicle node in the given period. Due to this, the number of responses will be significantly reduced, and the computation time required to process the responses received by CA will again be greatly reduced. The response message is encrypted using the public key of CA.
5. Malicious nodes cannot manipulate the location information for the following reasons. The query message is visible to only the verifier node. The response message by the verifier node can be decrypted only by CA.

Even if any verifier node or any query or response message is compromised, it does not tamper the location verification process. Suppose, we have verifier nodes 'e,' 'f,' 'g,' 'h' which have an encounter with the vehicle node in a given duration of time and date. If verifier node 'f' or message from it gets compromised, the other three 'e,' 'g,' 'h' verifier nodes respond correctly. So we can trace the location of the vehicle easily.

4 Security Analysis

Privacy: The proposed framework ensures privacy of the verifier nodes identity. Even when the digital location evidence is accessed by a third party, it is unable to learn anything about either the query message or the identity of the verifier node responding to the query message. The completeness and importance of the location information is maintained. The involvement of hidden verifier nodes attains protection against misbehaving nodes trying to submit incorrect location data. The verifier nodes are highly trusted and authenticated nodes, the possibility of them being dishonest is rare.

Confidentiality: The entities involved in location verification are verifier nodes and CA. The messages cannot be decrypted by authorities with highest power such as RSUs. Thus, confidentiality against authorities which possess high power is achieved.

Computationally secure: RSA is computationally secure. It takes considerably very long period to break the RSA. A simple computer will take approximately 300 trillion years to break a RSA with 2048 bit encryption key. Meanwhile, the purpose of the encryption would have been attained.

Avalanche effect: It is defined as the change in number of bits of cipher text corresponding to the change in the number of bits in plain text. A slight change in the plain text must cause significant change in the cipher text. In Fig. 3, it can be observed that RSA produces good avalanche effect when compared to other standard encryption algorithms.

Comparison of different encryption algorithms is presented in Table 3. It can be observed that symmetric encryption method yield good results as compared to asymmetric encryption method. But in the proposed framework, it is crucial that the query messages are accessible only to the intended verifier nodes. In symmetric encryption, for some reason if the key is compromised, then the query message and the

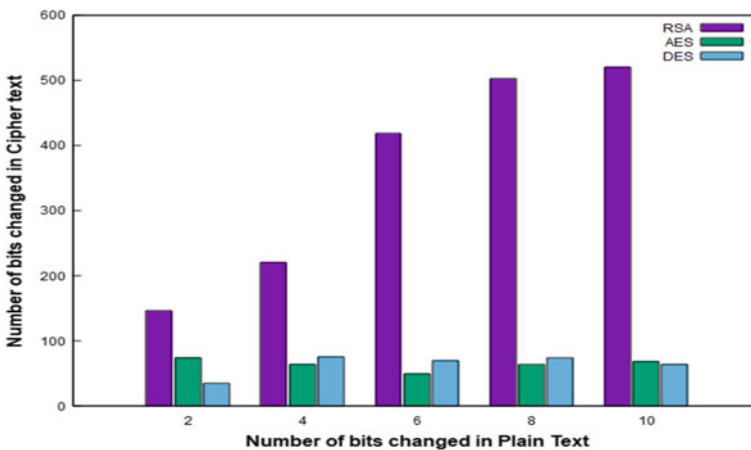


Fig. 3 Avalanche effect

Table 3 Comparison of different encryption methods

Parameters	Symmetric encryption			Asymmetric encryption	
	DES	AES	Blowfish	RSA	Diffie–Hellman
Key size	56 bits	128,192,256 bits	32 to 448 bits	>1024 bits	Key exchange management
Throughput	<AES	<Blowfish	High	Low	<RSA
Power consumption	Low	Low	Low	High	<RSA
Speed	High	High	High	High	Low
Secure against attack	Brute force attack	Chosen—Plain, known plaintext	Dictionary attack	Timings attacks	Eavesdropping
Ciphering and deciphering algorithm	Different	Different	Different	Same	Same
Security	As secret key is shared, risk of key compromise is high	As secret key is shared, risk of key compromise is high	As secret key is shared, risk of key compromise is high	More secure as private key is not shared	More secure as private key is not shared
Block size	64 bits	128 bits	64 bits	Minimum 512 bits	Not applicable

response message both will be compromised. This may lead to several attacks like masquerade, message modification flooding attack, etc. Thus asymmetric encryption method is chosen. RSA and Diffie Hellman are the two popular asymmetric encryption methods. But Diffie Hellman encryption method is used most widely for key exchange. RSA is computationally heavy and requires more power. It is also found that if not implemented correctly, they are easily prone to attacks. But the frequency of dispute arrival and number of location messages transferred are considerably low in number when compared to normal messages communicated. Thus, the extra overhead above factors can be sustained by the CA and verifier nodes.

5 Conclusion

Location-based services in VC face security issues such as location tampering, fake location claims to enjoy more benefits, denying about its location at a particular time and date. In this paper, we have presented a location verification method using special nodes called verifier nodes and trusted CA installed with location verification agent. The agent travels and collects information about nodes location and verifies the same. We use PKI to communicate between the CA and verifier nodes. As the presence of

verifier nodes is unknown to normal nodes, security increases as manipulating the location information targeting the nodes are completely elevated.

References

1. Kopylova, Y., Farkas, C., & Wenyuan, Xu. (2011). Accurate accident reconstruction in VANET. In *IFIP Annual Conference on Data and Applications Security and Privacy* (vol. 26, pp. 271–279). Berlin: Springer.
2. Huang, B., Cheng, X., & Cheng, W. (2018). Enhancing negative messages broadcasting with meet-table and TTL in VANET. *Procedia Computer Science*, 129, 185–187.
3. VinothChakkaravarthy, G., Lavanya, R., & Alli, P. (2012). Communication efficient distributed decentralized key management framework for message authentication in vanet. In *International Conference on Advances in Communication, Network, and Computing* (vol. 3, pp. 405–408).
4. Soyturk, M., Muhammad, K. N., Avcil, M. N., Kantarci, B., & Matthews, J. (2016). From vehicular networks to vehicular clouds in smart cities. *SmartCities and Homes*, 12, 149–171.
5. Lee, E., Lee, E., Gerla, M., & Oh, S. Y. (2014). Vehicular cloud networking: Architecture and design principles. *IEEE Communications Magazine*, 52, 148–155.
6. Boukerche, A., & Robson, E. (2018). Vehicular cloud computing: Architectures, applications, and mobility. *Computer Networks*, 135, 171–189.
7. Xue, X., Lin, N., Ding, J., & Ji, Y. (2010) A trusted neighbor table based location verification for VANET Routing. In *IET 3rd International Conference on Wireless, Mobile and Multimedia Networks (ICWMNN 2010)*, Beijing (pp. 1–5).
8. Ahmad, F., Kazim, M., Adnane, A., & Awad, A. (2015). Vehicular cloud networks: Architecture, applications and security issues. In *2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC)* (vol. 5, pp. 571–576).
9. Wu, H.-T., & Homg, G.-J. (2016). Vehicular cloud network and information security mechanisms. *International Conference on Advanced Materials for Science and Engineering (ICAMSE)*, 7, 196–199.
10. Qin, Bo., Qianhong, Wu., Domingo-Ferrer, J., & Zhang, L. (2011). Preserving security and privacy in large-scale VANETs. *International Conference on Information and Communications Security*, 8, 121–135.
11. Nosouhi, M. R., Yu, S., Zhou, W., Grobler, M., & Keshtiar, H. (2020). Blockchain for secure location verification. *Journal of Parallel and Distributed Computing*, 136, 40–51.
12. Foresti, S., & Jajodia, S. (2010). Data and applications security and privacy XXIV. In *24th Annual IFIP WG 11.3 Working Conference, Rome, Italy, June 21–23, 2010, Proceedings* (vol. 6166, pp. 112–126).
13. Miao, C., et al. (2015). Collaborative localization and location verification in WSNs. *Sensors (Basel, Switzerland)*, 15, 10631–49.
14. Sheet, D. K., Kaiwartya, O., Abdullah, A. H., & Hassan, A. N. (2015). Location information verification cum security using TBM in geocast routing. *Procedia Computer Science*, 70, 219–225.
15. Irain, M., Jorda, J., & Mammeri, Z. (2017). Landmark-based data location verification in the cloud: Review of approaches and challenges. *Journal of Cloud Computing*, 1, 31–34.
16. Lim, J., Kim, S., & Heekuck, Oh. (2007). A secure location service for ad hoc position-based routing using self-signed locations. *International Conference on Cryptology and Network Security*, 2, 121–132.
17. Ni, X., Luo, J., Zhang, B., Teng, J., & Bai, X. (2012). Mpsl: A mobile phone-based physical-social location verification system. In *International Conference on Wireless Algorithms, Systems, and Applications* (vol. 5, pp. 488–499). Berlin: Springer.
18. Odun-Ayo, I., Ananya, M., Agono, F., & Goddy-Worlu, R. (2018). Cloud computing architecture: A critical analysis. In *2018 18th International Conference on Computational Science and Applications (ICCSA)* (vol. 9, pp. 1–7).

19. Ries, T., Fusenig, V., Vilbois, C., & Engel, T. (2011). Verification of data location in cloud networking. In *2011 Fourth IEEE International Conference on Utility and Cloud Computing* (pp. 439–444).
20. Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. (2017). Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review*, *24*, 13–33.
21. Haidar, A. N., & Abdallah, A. E. (2009). Formal modelling of pki based authentication. *Electronic Notes in Theoretical Computer Science*, *235*, 55–70.

An Analytical Approach for Traffic Grooming Problems Using Waiting Probability in WDM Networks



Priyanka Kaushal, Neeraj Mohan, Surbhi Gupta, and Seifedine Kadry

Abstract Traffic grooming is the optimization of resources in a network. An efficient traffic grooming provides better resource utilization, enhanced performance at a lower cost. Traffic grooming has become very significant for all types of computer networks. As WDM (wavelength division multiplexing) networks are providing very high speed for a huge amount of data transfer, so traffic grooming is even more important for these networks. There are so many parameters that may be considered for traffic grooming. The waiting probability is one of the key parameters for traffic grooming. Waiting probability is a measurement of time spent by a call to get the required resources for further communication. We have proposed an efficient traffic grooming technique. This technique is based on waiting probability calculations. Waiting probability is calculated, and then traffic grooming problems are addressed based on the waiting probability calculations. Some of the other key network performance parameters such as the number of servers required, ideal path length for a source–destination pair, number of free wavelengths required, etc., are also analyzed. It is a low complexity technique for handling traffic grooming problems efficiently in telecommunication and call center management.

Keywords Traffic grooming · Waiting probability · WDM networks · Network optimization

P. Kaushal
Chandigarh Engineering College, Landran, Ajitgarh, India

N. Mohan (✉)
IKGPTU, Kapurthala, India

S. Gupta
GRIRT, Hyderabad, India

S. Kadry
Beirut Arab University, Beirut, Lebanon
e-mail: s.kadry@bau.edu.lb

1 Introduction

Resource sharing and load distribution are the basic purposes of networking. WDM networks can easily share resources both within the network and with other networks. The capability to share resources efficiently makes the WDM networks a preferred choice. Earlier, frequency division multiplexing (FDM) and time-division multiplexing (TDM) were largely used for communication. But now the use of WDM and erbium-doped fiber amplifier (EDFA) has changed the entire networking world. WDM has provided the opportunity to use multiple light beams of different wavelengths in a single optical fiber. Whereas, EDFA can amplify different wavelength signals simultaneously [1, 2]. The WDM networks are very popular in modern era networking due to some factors such as exceptionally high bandwidth, extremely low loss rate, and very low bit error rate. A typical single fiber can provide speed in the range of Tera Bits per second. At the same time, a very low loss rate in the range of 0.2 dB/Km is offered by optical networks. The bit error rate is typically in the range of 10^9 to 10^{15} . All optical network transmits data using lightpaths in the optical domain only [3].

Some applications need a large amount of data transfer in the range of Terabytes or Petabytes. At the same time, these applications are deadline-driven. The WDM networks along with advance reservation (AR) have provided a hopeful solution [4]. The major advantage of traffic grooming is maximum resource utilization of all the resources available in a network. If the resource utilization is at optimal level, then it will certainly enhance the overall network performance. Optical networks along with wavelength division multiplexing (WDM) technology has offered a platform for high-speed networks. An ultra-high-speed, huge capacity, and reliability are the key parameters that are expected from WDM networks. So, several traffic grooming techniques are applied to these networks so that performance may be optimized [5].

Traffic grooming is achieved through various multiplexing techniques, algorithms, methods, etc., in different domains of WDM networks. As the performance of a network depends on many parameters, it makes traffic grooming a challenging job. Further, it is categorized as static and dynamic. All of these categories have different attributes associated with them [6, 7].

2 Related Work

The authors have analyzed the blocking probability (BP) along with mean delay for a single-wavelength optical buffer. Poisson arrival process has been assumed for packets, whereas packet size follows the general distributions. A mathematical expression has been presented using the level crossing method [8]. An analytical model has been proposed for optical delay line buffer. It assumed that the packet spans are exponentially distributed random variables. It has been shown that the proposed model can reduce burst loss probability [9]. A large wireless mesh network

was divided into numerous reduced groups. Each group used to be supported by an optical network unit of a passive optical network (PON). The proposed model was able to handle the heterogeneous node traffic load efficiently [10].

Basic Erlang C formula calculations have been used to determine the traffic route for call centers. The number of calls received per unit time and number of calls forwarded per unit time have given a performance parameter for a particular channel or agent. All these calculations have been done using basic Erlang C formula. Further, mean call handling time and the average number of the processed calls per unit time were also calculated [11]. A method was proposed to calculate the service waiting probability. It was based upon approximating the distribution function. Both service waiting probability and average delay time of packets were calculated [12]. A priority-based mode transformation procedure has been proposed to maximize network resource consumption. The proposed algorithm has improved the quality of service significantly. Different load state combinations were considered. Network resources were allocated depending upon the priority of requests [13, 14].

The Erlang C model has been used as it provided clear and accurate results. It was assumed that the skill level of agents in the contact center was not homogenous. Some of the key parameters such as average waiting time, queuing probability has been calculated for a given number of agents [15, 16]. A two-step method has been proposed to regulate the quantity of spectrum essential for the request of vast data allocations. The demands were generated with a target and blocking rate. Markov chain model was used to allocate variable bandwidth over a lightpath. The overall utilization of all the links was also calculated [17]. Approximate network blocking probabilities have been computed using the proposed mathematical model for heterogeneous all-optical WDM networks. A last-fit-first wavelength assignment scheme has also been proposed for dynamic traffic needs. The blocking probability has been minimized significantly [18].

A traffic grooming model has been introduced for dynamic multicast traffic grooming. This model has significantly reduced the blocking probability. A 0/1 knapsack-based multicast traffic grooming model was proposed [19, 20]. An algorithm based on spray and wait has been proposed to address routing problems. Relay node and token numbers were considered based on delivery probability, delivery time, and buffer status. The simulation results have shown significant improvement in the delivery and delay ratio. The planned algorithm has better efficiency for smaller buffer size [21].

The authors have proposed the utilization of Erlang B and Erlang C formulae for estimating and managing network traffic. It is described that the Erlang B formula does not contain the delay parameter. As the delays are omissible in ATM networks, so it makes Erlang B formula more suitable for ATM networks. The Erlang C formula is capable to estimate delay or waiting probability. So this formula is more suitable for IP networks. These formulae had played a vital role to monitor and ensure the Quality of Service in networks [22]. A model of the queueing system has been proposed with a non-full available server. The results of the proposed model were found quite accurate [23].

It may be summarized that researchers have mainly concentrated on blocking probability to address traffic grooming problems [8, 10, 18, 19, 23, 24]. Still, the waiting probability calculation based traffic grooming is addressed by few researchers [11, 12, 15, 25]. In the next section, the proposed technique is presented which is based upon the waiting probability calculation.

3 Proposed Technique

Several parameters such as blocking probability, server utilization, waiting probability, throughput, and response time determine the performance of a network. The optimal value of these parameters enhances network performance significantly. So, these parameters are known as performance measurement parameters. These parameters are defined as:

Blocking Probability: It is defined as ratio of the calls blocked to the total number of calls.

$$\text{Blocking Probability} = \frac{\text{Number of calls blocked}}{\text{Total number of call received}}$$

Server Utilization: The utilization in a network is usually associated with the utilization of the server. It is the fraction of time for which the server is active. If there are multiple servers, their utilization is calculated in a combined manner. It is calculated as:

$$\text{Server Utilization} = \frac{\text{Arrival Rate}}{\text{Service Rate}}$$

Waiting Probability: It is represented as the time spent by a request in a queue before getting resources. So the response time is the sum of waiting time in a queue and service time. It is calculated as:

$$\text{Waiting Probability} = \frac{\text{Number of calls in the waiting queue}}{\text{Total number of call received}}$$

Throughput: It is defined as the average number of handled requests per unit of time.

Response Time: It is the total time spent by a request in the system [26, 27].

We have proposed a model to calculate the waiting probability of a network. The waiting probability gives an estimation of the resources required to achieve a quality of service level. The call waiting probability (P_w) is described using Erlang C formula as [11]:

$$P_w = \frac{\frac{A^N N}{N!(N-A)}}{\sum_{i=0}^{N-1} \frac{A^i}{i!} + \frac{A^N N}{N!(N-A)}} \quad (1)$$

N is the total number of servers or agents and A is traffic load in Erlangs, and it may be defined as the ratio of the average number of calls per unit time (μ) to the average number of requests processed per unit of time (T), so

$$A = \frac{\mu}{T} \tag{2}$$

Further, the traffic load (A) of a network is dependent upon other parameters. So the traffic load (A) can be calculated as [24, 25, 28]:

$$A = \frac{\lambda}{l} \tag{3}$$

where λ is the number of free wavelengths and l is path length for a particular source–destination pair.

Now, substituting Eqs. (2) and (3) in Eq. (1), we get

$$P_w = \frac{\frac{(\frac{\mu}{T})^{*N}}{N!(N-A)}}{\sum_{i=0}^{N-1} \frac{(\frac{\mu}{T})^i}{i!} + \frac{(\frac{\mu}{T})^{*N}}{N!(N-A)}} \tag{4}$$

$$P_w = \frac{\frac{(\frac{\lambda}{l})^{*N}}{N!(N-A)}}{\sum_{i=0}^{N-1} \frac{(\frac{\lambda}{l})^i}{i!} + \frac{(\frac{\lambda}{l})^{*N}}{N!(N-A)}} \tag{5}$$

Now, we have Eqs. (4) and (5) which are used to calculate waiting probability with respect to various network parameters. The results obtained may be further used for traffic grooming of the networks. These results are very useful for telecommunication problems and call center management.

4 Results and Discussions

The outcomes of the planned procedure have been verified on WDM networks. Certain parameters such as numbers of servers, the path length of the route, total network load, and the number of free wavelengths on the network are considered while calculating and analyzing the waiting probability of the network. The waiting probability and the blocking probability needs to be minimum for all networks as it maximizes the network performance.

First, we have fixed the number of servers as 20. The waiting probability is calculated when the path length is 3 and 4. The number of free wavelengths is increased from 25 to 55 with an interval of 5. It is observed that waiting probability is increased

as the number of free wavelengths is increased (3). It is also observed that path length has an important role to play in the waiting probability calculations. If the path length is increased to 4 from 3, there is a significant fall in waiting probability. These results are shown in Table 1 and Fig. 1.

Network load is a significant network parameter. The performance of a network is largely affected by the network load. Usually, the overall performance of a network is decreased as the network load is increased because the network demands more resources to meet the enhanced network load requirements. We have calculated the waiting probability for network load from 30 to 55 Erlangs with a step of 5. The number of servers is fixed as 55 and 60. The waiting probability keeps on increasing as the network load is increased. Moreover, its value is decreased if the number of

Table 1 Effect of path length on waiting probability

No. of servers	No. of free wavelengths	Waiting probability (path length = 3)	Waiting probability (path length = 4)
20	25	0.000442	0.000096
20	30	0.003731	0.000115
20	35	0.018423	0.000801
20	40	0.062317	0.003732
20	45	0.160461	0.012869
20	50	0.338123	0.035259
20	55	0.614581	0.080719

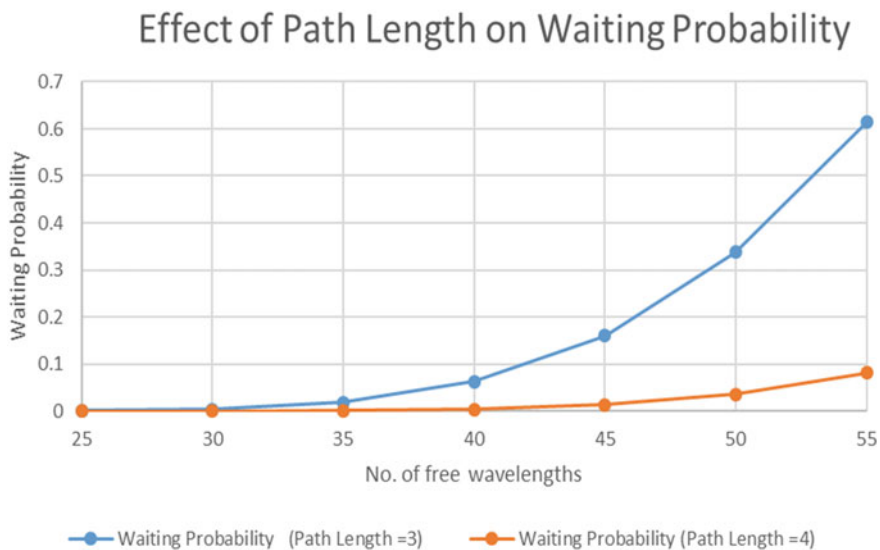


Fig. 1 Effect of path length on waiting probability

Table 2 Effect of network load on waiting probability

Network load (Erlangs)	Waiting probability (no. of servers = 55)	Waiting probability (no. of servers = 60)
30	0.000029	0.0000009
35	0.001146	0.000081
40	0.015896	0.002035
45	0.102372	0.021389
50	0.384547	0.117302

servers is increased. So, it is observed that the required number of servers may be increased to guarantee a minimum level of quality of service. These results are shown in Table 2 and Fig. 2.

The waiting probability is calculated for the different number of servers in the network. The decision about the number of servers is very crucial if the requests demand a minimum level of guaranteed quality of services. The waiting probability is calculated for the number of free wavelengths from 35 to 55 with a step of 5. The number of servers is fixed as 20 and 25. The path length is also fixed as 3. The waiting probability keeps on increasing as the number of free wavelengths is increased. Moreover, its higher value of the number of servers (25) provides better waiting probability. These results are described in Table 3 and Fig. 3.

The results of the proposed technique have been compared with existing techniques [12, 27]. In the existing techniques, the results are based upon approximation methods, whereas the proposed technique has produced calculated results with low complexity.

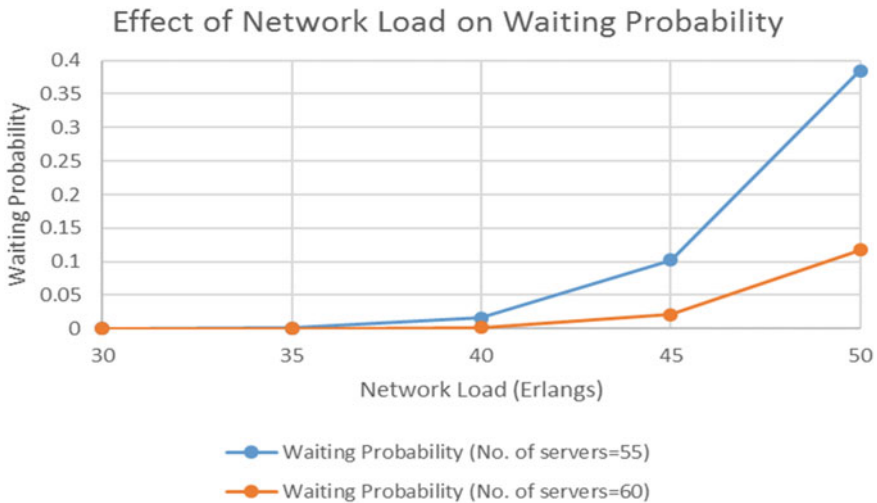


Fig. 2 Effect of network load on waiting probability

Table 3 Effect of number of servers on waiting probability

Path length	No. of free wavelength	Waiting probability (no. of servers = 20)	Waiting probability (no. of servers = 25)
3	35	0.018423	0.000489
3	40	0.062317	0.002973
3	45	0.160461	0.012434
3	50	0.338123	0.039083
3	55	0.614581	0.098504

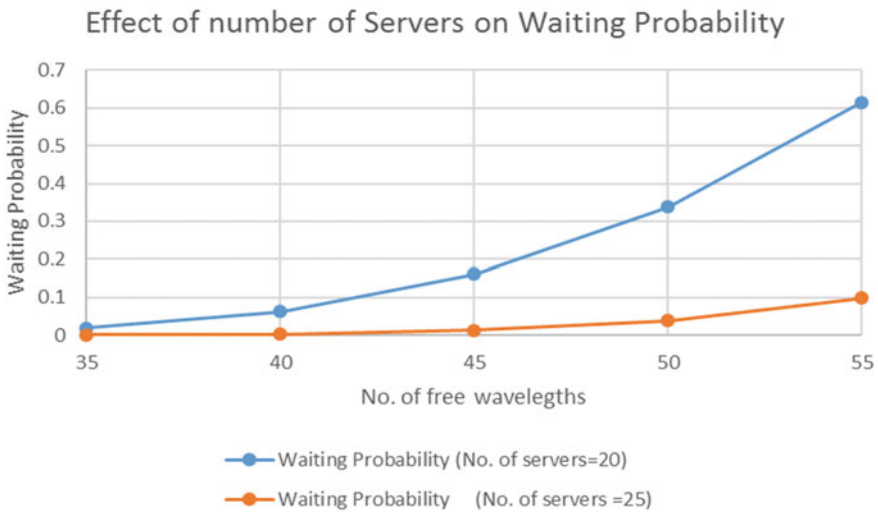


Fig. 3 Effect of number of servers on waiting probability

5 Conclusion

A low complexity traffic grooming model is proposed. This model is based upon the waiting probability calculations. The proposed model is derived using the Erlangs C formula. The waiting probability is a crucial parameter to access the performance of a network. Waiting probability is dependent upon various network parameters. The waiting probability is calculated using the proposed model for parameters such as path length, network load, and the number of servers. Each parameter has an impact on waiting probability calculations. The results have shown that the proposed model has better waiting performance for optical networks. The proposed model is providing calculated values of waiting probability rather than estimated values.

References

1. Mohan, N., Wason, A., & Sandhu, P. S. (2016). ACO based single link failure recovery in all-optical networks. *International Journal for Light and Electron Optics, Optik*, 127(20), 8469–8474.
2. Wason, A., & Kaler, R. S. (2010). Lightpath Rerouting algorithm to enhance blocking performance in all-optical WDM networks without wavelength conversion. *Optical Fiber Technology*, 16, 146–150.
3. Tripathi, T., & Sivarajan, K. N. (2000). Computing approximate probabilities in wavelength routed all-optical networks with limited range wavelength conversio. *IEEE Selected Areas in Communications*, 18(10), 2123–2129.
4. Kim, Y., Atchley, S., Vallée, G. R., Lee, S., & Shipman, G. M. (2017). Optimizing end-to-end bigdata transfers over terabits network infrastructure. *IEEE Transactions on Parallel and Distributed Systems*, 28(1), 188–201.
5. Barr, R. S., Kingsley, M. S., & Patterson, R. A. (2005). Grooming telecommunications networks: Optimization models and methods. In *Technical Report 05-emis-03*, 1–27.
6. Zhu, K., & Mukherjee, B. (2003). A review of traffic grooming in WDM optical networks: Architecture and challenges. *Optical Networks Magazine*, 55–64.
7. Modiano, E., & Lin, P. J. (2001). Traffic grooming in WDM networks. *IEEE Communications Magazine*, 124–129.
8. Tanga, S., & Tana, L. (2018). Single-wavelength optical buffers with general burst size distribution: Blocking probability and mean delay. *Optical Switching and Networking*, 27, 1–6.
9. To, H. L., Lee, S. H., & Hwang, W. J. (2015). A burst loss probability model with impatient customer feature for optical burst switching networks. *International Journal of Communication Systems*, 28(11), 1729–1740.
10. Chen, P. Y., & Reisslein, M. (2015). A simple analytical throughput-delay model for clustered FiWi networks. *Photon Network Communications*, 29, 78–95.
11. Chromy, E., Misuth, T., & Kavacky, M. (2011). Erlang C formula and its use in the call centers. *Information and Communication Technologies and Services*, 9(1), 7–13.
12. Lozhkovskiy, A. (2019). Calculation the service waiting probability with self-similar network traffic. *Journal of Engineering Science*, 16(2), 35–39.
13. Tong, L., Gan, C., Xie, W., & Wang, X. (2019). Priority-based mode transformation algorithm in a multisystem based virtual passive optical network. *Optical Switching and Networking*, 31, 162–167.
14. Xie, W., Gan, C., Xia, W., Zheng, W., & Yu, H. (2017). Mode transformation and united control mechanism supporting wavelength division multiplexing ethernet passive optical network and orthogonal frequency division multiplexing passive optical network in multi subsystem based virtual passive optical network. *IET Communications*, 11, 696–703.
15. Misuth, T., Chromy, E., & Baronak, I. (2010). Performance estimation of contact centre with variable skilled agents. In *Proceedings of the 33rd International Conference on Telecommunications and Signal Processing—TSP 2010* (pp. 391–396). Baden.
16. Misuth, T., Chromy, E., & Baronak, I. (2010). Method for fast estimation of contact centre parameters using Erlang C model. In *Proceedings of the Third International Conference on Communication Theory, Reliability, and Quality of Service, CTRQ 2010* (pp. 181–185). Glyfada: Greece.
17. Feng, D., Sun, W., & Hu, W. (2019). Dimensioning of store and transfer WDM networks with stratified ROADM node. *Optical Switching and Networking*, 31, 100–113.
18. Khan, A. N. (2017). An improved approximate network blocking probability model for all-optical WDM networks with heterogeneous link capacities. *Optical Fiber Technology*, 38, 7–16.
19. Pradhan, A. K., Keshri, S., Das, S., & De, T. (2016). A heuristic approach based on dynamic multicast traffic grooming in WDM mesh networks. *Journal of Optics*, 1–11.

20. Pradhan, A. K., Chatterjee, B. C., Oki, E., & De, T. (2018). Knapsack based multicast traffic grooming for optical networks. *Optical Switching and Networking*, 27, 40–49.
21. Derakhshanfard, N. (2020). Erlang based buffer management and routing in opportunistic networks. *Wireless Personal Communications*, 110, 2165–2177.
22. Chromy, E., Misuth, T., & Weber, A. (2012). Application of Erlang formulae in next generation networks. *International Journal of Computer Network and Information Security*, 1, 59–66.
23. Hanczewski, S., Stasiak, M., & Weissenberg, J. (2019). Non-full-available queueing model of an EON node. *Optical Switching and Networking*, 33, 131–142.
24. Mohan, N., Gupta, O. P., Wason, A., & Sandhu, P. S. (2015). Traffic grooming and blocking optimization in all-optical networks. *International Journal of Scientific & Engineering Research*, 69(1), 409–413.
25. Mohan, N., & Kaushal, P. (2019). Waiting probability based traffic grooming in optical networks. In *Proceedings of the International Conference on Sustainable Computing in Science, Technology & Management, February 26–28, 2019, Jaipur, India*, (pp. 777–781).
26. Chromy, E., & Baronak, I. (2018). Mathematical model of the contact center. *International Journal of Internet of Things and Web Services*, 3, 17–23.
27. Nag, K., & Helal, M. (2017). Evaluating Erlang C and Erlang A models for staff optimization: A case study in an airline call center. In *Proceedings of the IEEE Industrial Engineering and Engineering Management (IEEM), Singapore* (pp. 1–5).
28. Mallika, N. M. (2013). Link failure recovery in WDM networks. *International Journal of Computer Science and Electronics Engineering*, 1(5), 599–602.

Flow-Based Detection and Mitigation of Low-Rate DDOS Attack in SDN Environment Using Machine Learning Techniques



K. Muthamil Sudar and P. Deepalakshmi

Abstract Software Defined Networks (SDN) have become more efficient and popular by having effective controller. Centralized controller makes the decision to handle traffic in data plane by analyzing the entire network. In addition, handling network attack is equivalently complex for controllers. Low-rate Distributed Denial of Service (LR-DDoS) attack restricts legitimate users to access resources by sending unwanted and half-opened request towards the devices in data plane. Hence, it is vital to detect and mitigate LR-DDoS attack in its early stage because nature of attack is very similar to original request. The nature of this attack exhausts the network resources and leads to resource unavailability or delay while processing the legitimate requests. In this paper, we propose a flow-based detection and mitigation framework using machine learning models like Support Vector Machine (SVM), C4.5 Decision tree and Naïve Bayes as classifiers to detect LR-DDoS attack. From every traffic flow samples, we extract the essential features to detect attack. In mitigation phase, we handle the attack flow information and install the mitigation rules to avoid LR-DDoS attack from same source. Our experimental results show that SVM mechanism achieves better accuracy compared to C4.5 and Naïve Bayes techniques.

Keywords SDN · DDoS · Machine learning · SVM · C4.5 decision tree · Naïve Bayes

1 Introduction

Software-Defined Networking (SDN) is an emerging technology which overcomes some drawbacks faced in traditional network environment by providing centralized network controller and programming interface to adopt the dynamic changes in

K. M. Sudar (✉) · P. Deepalakshmi
Kalasalingam Academy of Research and Education, Krishnankoil, Tamilnadu, India
e-mail: k.muthamilsudar@klu.ac.in

P. Deepalakshmi
e-mail: deepa.kumar@klu.ac.in

network [1]. However, providing security is one of the complex tasks in SDN environment. SDN requires proper security mechanisms to protect controllers and other devices in data plane. Distributed Denial of Service (DDoS) attack [2] plays a major role in disrupting the services offered in SDN environment. DDoS attackers generate either high-rate or low-rate traffic towards the target in order to stop processing legitimate requests. Many researchers already addressed about detection of high-rate DDoS attacks in both traditional as well SDN environments. On the other side, only few researchers explored about the impact of low-rate DDoS detection method and proposed detection or mitigation methods. Hence, in this paper, we concentrate on detection of low-rate DDoS attacks in SDN environment by analyzing the essential features.

It is very hard to detect low rate attacks because the features of such attack's traffic are very similar to normal traffic. In recent days, machine learning techniques help to improve the efficiency and accuracy of attack detection system. In this work, we propose a low-rate DDoS detection and mitigation system using machine learning models such as Support Vector Machine (SVM), C4.5 Decision tree and Naïve Bayes with the help of essential features extracted from SDN traffic flows. Since we update every detected attack flow information in blacklist table, we could avoid the attack from same source in the beginning itself.

The following contributions are done in this paper:

- (i) We introduce flow management module to extract the essential features like relative distribution of packet interval, number of packets, duration of the flow and relative distribution of match bytes.
- (ii) We design and implement machine learning models like SVM, C4.5 and Naïve Bayes techniques to detect the LR-DDoS attacks.
- (iii) We introduce blacklist table to gather the information about the attack and malicious traffic flows.
- (iv) We discuss and compare the performance of our proposed work in terms of accuracy, precision and recall.

2 Background and Related Work

2.1 *Software Defined Networking*

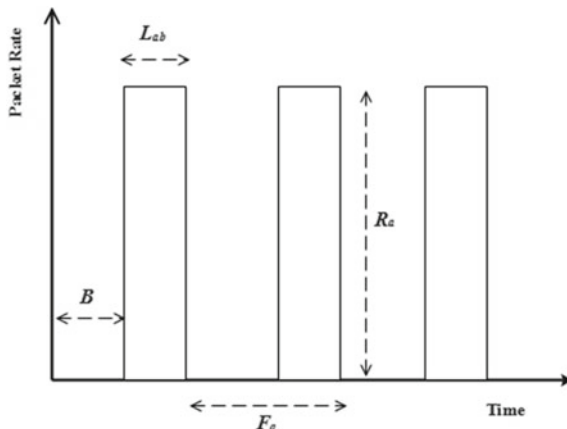
SDN is a promising paradigm which is powerful, dynamic, directly programmable and supports high-bandwidth characteristics of today's applications. SDN architecture isolates control from data plane devices and facilitates the control using programmable interface. Abstract control from data plane devices helps network administrators to adjust traffic flows dynamically based on the needs of changes in network topology and functions. SDN architecture allows the administrators to manage and optimize network resources with the help of automated SDN logical programs. Unlike traditional network architecture, SDN also abridges the design and

network operation because logics are provided by SDN controllers. Network Function Virtualization (NFV) [3] takes apart network functions from hardware devices in the form of virtualized functions. SDN architecture consists of three planes such as application, data and control plane. Southbound API (Application programming Interface) and northbound API are used to enable communication between these planes. Application plane contains all the applications and services running on the network. Control plane contains logically centralized controller which acts as core part of network and decides how to handle traffic based on the information from applications and devices in data plane. Data plane contains forwarding devices like switches and routers which forwards the traffic upon receiving routing information from control plane.

2.2 Low-Rate DDoS Attack

Many researchers and studies reported DDoS attacks as one of the prime threats created by cyber criminals to make massive destruction and have impact in all kinds of network platform. Now a days, the intensity of attack generate by attacker varies from high-level traffic to low-level traffic. DDoS attack with smart, low-level traffic is called low-rate DDoS attack. In low-rate DDoS attack, amount of attack traffic generate towards the target network accounts for 10–20% only. Identifying the behaviour of LR-DDoS attack is quite difficult. Even though the nature of attack is small, it will cause heavy destruction to target system or network. This type of low-rate attack is referred to shrew attack. The nature of shrew attack is extremely opposite to the DDoS flood attack. This attack traffic accounts 10–20% of legitimate network traffic and has strong concealment. L-DDoS attack shows traces of a periodic pulse shape which mainly concentrates on attack energy and leads to small average attack traffic. L-DDoS attack pulses generated from various attack sources with small average flows is associated together to form attack pulse for certain period of time. This leads to bottleneck in the target terminal and degrades the quality of service. L-DDoS attack traffic mainly concentrates on pulse period (i.e.) rectangular pulse in which several packets are sent towards the target. So, power of the attack mainly depends on duration and amplitude of the pulse period. Currently, L-DDoS attack mainly focuses on high speed and centralized services like cloud, big data. The reason is that low rate attack traffic can easily conceal into large traffic generated by cloud and big data environments. The effect of this attack results in huge traffic loss and economical loss to high speed computing environments. The mechanism used to detect general DDoS attack is not sufficient to detect L-DDoS attack because number of attack packets, and average attack rate is very low which looks like legitimate traffic. Figure 1 represents the L-DDoS attack in which L_{ab} indicates length of attack burst, R_a indicates rate of attack, B indicates beginning time of attack and F_a indicates frequency of attack. X and Y axis represents packet rate and time, respectively.

Fig. 1 L-DDoS attack representation



Attackers generally use sockstress attack, slowloris tool, and RUDY tool to generate low-rate attack [20]. Slowloris tool can generate partial HTTP headers after connecting to server and makes the connection open for long period of time. RUDY tool can generate HTTP POST requests and makes the server connection open for long time by sending the data in very slow manner. Sockstress attack mainly identifies the vulnerability in Transmission control protocol (TCP) handshake process and creates a numerous connection. Table 1 indicates the different types of LR-DDoS attack with their effect.

2.3 Related Work

Many researchers proposed different techniques to detect the L-DDoS attack by analyzing diverse parameters in both traditional and SDN environment. In [4], the authors attempted in SDN environment to identify L-DDoS attacks using machine learning models like Support Vector Machine (SVM), J48, Multi-layer Perceptron, REP tree, Random forest and Random tree. They claimed that Multi-Layer Perceptron model achieved an accuracy of 95% against Canadian Institute of Cyber security DoS Dataset. The authors of [5] proposed packet size measurement technique by analyzing the distribution difference in packet size to detect low rate DDoS attack. They demonstrated with real time datasets for various tolerance factors. They claimed that proposed work can detect L-DDoS attacks in short span of time with low false alarm rate.

In [6], the authors used entropy-based optimal technique to classify L-DDoS attacks in traditional environment. The authors of [7] used metric as information distance using generalized SDN control plane. They stated that proposed technique could achieve high accuracy compared to traditional entropy technique like Shannon entropy. In [8], the authors proposed correlation based detection technique to discover

L-DoS attack. By applying this technique they analyzed the traffic characteristics of attack and normal traffic by calculating linear relationship between the variables. They claimed that PRCD technique works better than relative entropy-based scheme.

The authors of [9] proposed statistical-based detection mechanism to detect L-DDoS attack in application server. They used t-statistic measure based on hypothesis test to identify L-DDoS flows. They claimed that the proposed distribution function technique detects the attack in efficient manner. In [10], the authors employed ML techniques to detect low-rate DDoS attack in SDN enabled IoT networks. They demonstrated with different machine learning techniques. They concluded that SVM achieve similar accuracy of 91% in both switch and controller with low false alarm rate.

From the literature work, we have identified that low-rate DDoS attack is one of the notable threat in all kinds of networking environment. Since the nature of attack is looking like original traffic instances, it is very difficult to detect and mitigate. It tries to leave many connections in open state as long as possible. The impact of this attack is very high in terms of bringing down the target but in slow manner. Few researchers have proposed statistical and ML techniques to detect the attack. Statistical-based technique like entropy, probability distribution, t-statistic measure achieved good accuracy but at the cost of high computation overload. Machine learning-based techniques like SVM, RF, KNN, Naïve Bayes, and J48 achieved better accuracy with low false alarm rate and at the same time with low computation overhead also.

3 Architectural Design

To detect and mitigate the L-DDoS attack, we proposed a framework which separates the process of detection and mitigation from network application. This reduces the processing overhead in SDN control plane. The framework consists of two independent phases such as detection phase and mitigation phase. Detection phase consists of flow management for extracting the essential flow information from every flow and ML models to identify the flow as attack or normal flow. Mitigation phase consists of malicious attack management and mitigation rules which will define the flow drop rules for the malicious flow.

3.1 Detection Phase

Low-rate DDoS attack mainly focuses on saturating the data plane devices like switches. Open Flow (OF) switch forwards data packets based on rules specified by controller. Every switch contains flow table listed with flow rules that are installed by the centralized controller after analysing the entire network. General OF switches use Ternary Content-Addressable Memory (TCAM) which can store 1500–3000 flow information [11]. So, data plane devices like switch become the target for attackers

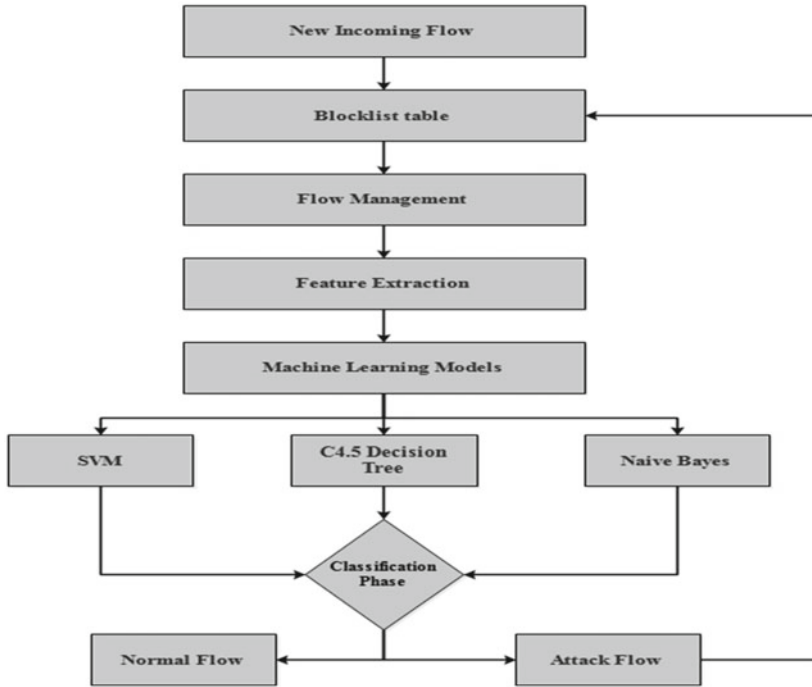


Fig. 2 Proposed architecture

to generate LR-DDoS attacks. This attack is also referred to low TCAM saturation attack. Figure 2 illustrates about architecture diagram of our proposed work.

The attackers generate low-rate attack in following conditions.

- (i) The number of bots controlled by the attacker should be very less to generate the attack.
- (ii) Every bot will send similar set of packets to target switch using real IP addresses. Upon receiving these packets, flow rules get installed in switch's flow table.
- (iii) The rate of packets generating from bots are not too high. So, the attackers do not use spoofed IP address to generate the LR-DDoS attack.
- (iv) Attackers used to generate the data packets from bots in less time interval compared to flow entry idle time out. The idle time out interval is the duration of entry available in the flow table in no-match state.

3.1.1 Flow Management

Flow management in the detection module will extract the essential features from flow samples. The four essential features used in this work are relative distribution of packet interval, number of packets, flow duration and relative distribution of match

bytes. Duration of a flow indicates the time period of corresponding flow available in a flow table. The ultimate aim of LR-DDoS attack in SDN data plane is to exhaust the resources of flow table. So, attackers will send new data packets and makes the connection as open, which occupies flow table for long time. By calculating relative distribution match bytes, we can identify the difference between normal flow and attack flow. Because normal flow matches large data compared to attack flow. Relative distribution match bytes can be calculated using Eq. 1.

$$RDB = \frac{\sum_{i=1}^N (P_i - \mu)^2}{N} \quad (1)$$

Here, P_i indicates each packet size in bytes matched by flow rule; N indicates total number of packets matching with flow rules in a particular time and μ indicates average value of matched packet sizes again in a particular time. Original flow will follow the random and different packet interval based on the size of data packets. But, attack flow will follow periodic and smaller interval in order to act like original flow. So, packet interval also plays a vital role in detection of LR-DDoS attack. Relative distribution packet interval can be calculated using Eq. 2.

$$RDP = \frac{\sum_{i=1}^N |D_i - \lambda|}{N} \quad (2)$$

Since attack flows are very similar to legitimate flows, we have considered all flow samples to detect the attack by using the above mentioned features.

3.1.2 Machine Learning Models

In this module, we have implemented ML models like support vector machine (SVM), C4.5 Decision tree and Naïve Bayes to detect the attack by analyzing the features extracted from flow management module. SVM is one of the commonly used classification technique by placing the best hyper plane between two classes such as attack and normal [12]. C4.5 is a variant of decision tree technique which helps to classify the classes by constructing tree using if-then rules [13]. Naïve Bayes classifier is the simplest classification method based on the probability of each feature [14]. To implement these machine learning models, we used Scikit-learn 0.23 library provided by python. The output of flow management module is passed as input to machine learning models. Once the attack is detected, it will alert the mitigation phase for further processing of attack packets.

3.2 *Mitigation Phase*

3.2.1 **Malicious Attack Management and Mitigation Rules**

Machine learning models will return 0 for normal flow and 1 for attack flow after analyzing the traffic features. Once the flow is identified as malicious, assuming that it could be an attacker, detection module will forward the corresponding source IP address to blocklist table. Mitigation module will alert the switch to block the particular flow by creating blocking port rule immediately. Source IP address of every flow will be matched with the entries in blocklist table before forwarding for further process. The installed flow mitigation rules are set to get removed from flow table after seven days. This helps to analyze and update the flow in regular manner. Since flow management and analysis is already part of SDN, this module will not increase the complexity of SDN process.

3.3 *Algorithm*

Input: Incoming Traffic Flows
Output: Attack flow or Normal Flow
Begin
For each P_i
If P_i already on BT **then**
 Drop the flow
Else
 Forwards to Flow management module
 Extract Features from the traffic flow
 Send the selected features value to ML models
If $flow == malicious$ **then**
 Add source ip address to BT
 Update the mitigation rules
Else
 Label as Normal flow
End
End
End

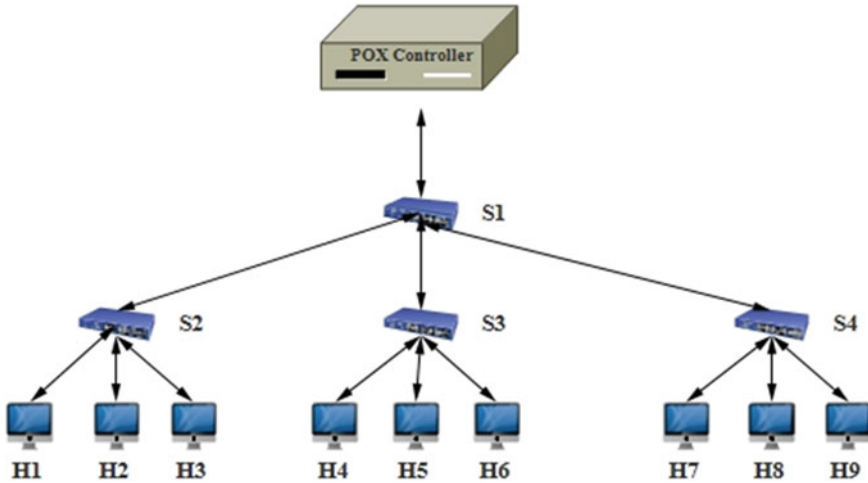


Fig. 3 Network topology

4 Experimental Setup

4.1 Environment Setup

We use Mininet 2.3.2 [15] tool, which is a commonly used tool to simulate the SDN environment by creating virtual network topologies. We configure hosts in data plane as both legitimate and attack devices. OpenFlow switch is another standard virtual component available in Mininet. We use POX—a python-based controller [16, 22] to simulate the functionalities of controller in control plane. The experimental network topology consists of nine host machines (H1–H9) and four switches (S1–S4). Figure 3 portrays the network topology of proposed work. All hosts connect to switch with the data rate of 10 Mbps. We use Scapy tool [17, 18] to generate the low rate attack by forging the fields of packets and rate of transmission. We have considered average rate of 1.6 to 2.2 Mbps to generate the low-rate DDoS attacks from different hosts.

4.2 Dataset

To train our machine learning models, we have used CIC DoS 2019 dataset (Canadian Institute of Cyber security) [19]. CIC dataset contains both benign and common DDoS attacks. CIC dataset contains LR-DDoS attacks compared to commonly used datasets like KDD-CUP’99 and NSL-KDD dataset. CIC dataset consists of eight different low rate attacks and normal instances. Totally, it contains 80 features to provide statistical information about the flows. Table II specifies the distribution of

attack and normal flows in CIC dataset. We use Flowtbag method [4] to convert conventional traffic instances into flow instances with same set of features. This works well in SDN environment too.

4.3 Evaluation Metrics

The performance of proposed work is evaluated using standard metrics like accuracy, recall, precision and F-Score [21]. Accuracy (AC) determines rightly classified data (both attack and normal) over all the classified data. Precision (P) determines rightly classified positive data over all the positive classified data. Recall (R) determines how correctly the model classifies true positives. F-score is a mean of both recall and precision. Accuracy, Recall, Precision and F-score can be computed by Eqs. 3, 4, 5 and 6, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F - Score = 2 * \frac{P * R}{P + R} \quad (6)$$

Here, TP indicates correctly classified normal flows, TN indicates correctly classified attack flows, FP indicates wrongly classified normal flows, and FN indicates wrongly classified attack flows.

Table 2 depicts the performance of our proposed work. SVM achieves better accuracy to detect low rate DDoS attack compared to C4.5 and Naïve Bayes techniques. Figure 4 indicates the confusion matrix for C4.5, SVM and Naïve Bayes models. Figure 5 indicates the accuracy comparison of different machine learning models for our proposed work.

Table 1 LR-DDoS Attacks and Effects

LR-DDoS Types	Effect of Attack
Shrew DDoS	Reduce Quality
Fake Session	Delay in service
Session attack	Delay in service
HTTP Fragmentation	Delay in service e
Bandwidth Congestion	Reduce Quality

Table 2 CIC dataset traffic flow distribution

Type of Traffic	Number of Flows
DDoSSIM	2185
Goldeneye	661
H.U.L.K	2088
R.U.D.Y	2070
Slowbody2	4391
Slowheaders	3185
Slowloris	2912
Slowread	1464

Table 3 Performance evaluation

Algorithm	Accuracy	Precision	Recall	F-Score
SVM	0.93	0.93	0.99	0.95
C4.5 Decision Tree	0.91	0.91	0.98	0.94
Naïve Bayes	0.87	0.87	0.97	0.92

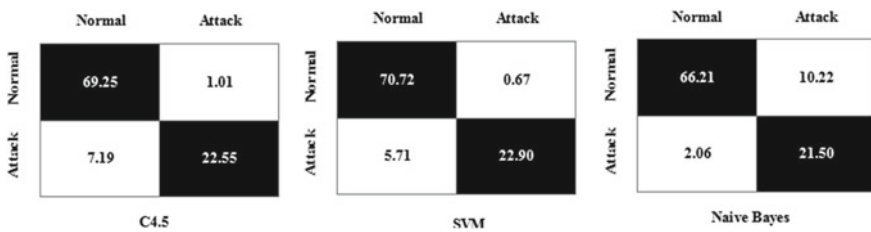
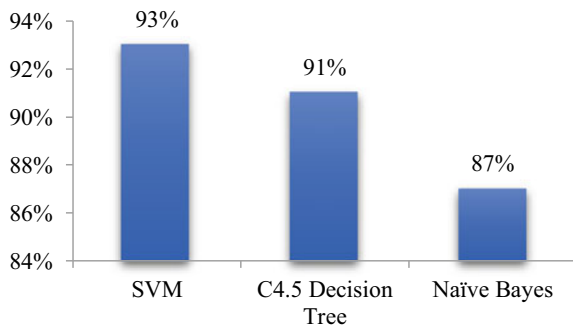


Fig. 4 Confusion matrix

Fig. 5 Accuracy comparison of our proposed work



5 Conclusion

In this paper, we propose mechanism to detect and mitigate LR-DDoS attack in SDN environment using machine learning techniques. Since normal features like port number, source and destination address are not sufficient to detect attack, we have used four essential features viz. relative distribution of packet interval, number of packets, duration of the flow and relative distribution of match bytes from the SDN flow. In this paper, we used machine learning techniques like SVM, C4.5 Decision tree and Naïve Bayes as classifiers to detect attacks. We trained machine learning models with the help of CIC DoS dataset, which contains special types of low rate DDoS attacks. Our experimental results indicate that SVM aids to detect attack with better accuracy compared to C4.5 and Naïve Bayes. Once attack is detected, mitigation module adds particular attack flow details to blacklist table and alerts the controller to drop particular flow from flow table by entering mitigation rules. Our experimental results illustrate the efficacy of our proposed work. In future, we plan to implement multi-layer machine learning-based techniques to distinguish low rate and high rate DDoS attacks using SDN flow information.

References

1. Sudar, K. M., & Deepalakshmi, P. (2020). Comparative study on IDS using machine learning approaches for software defined networks. *International Journal of Intelligent Enterprise*, 7(1–3), 15–27.
2. Muthamil Sudar, K., & Deepalakshmi, P. (2020). A two level security mechanism to detect a DDoS flooding attack in software-defined networks using entropy-based and C4.5 technique. *Journal of High Speed Networks*, (Preprint), 1–22.
3. Deepa, V., Sudar, K. M., & Deepalakshmi, P. (2018, December). Detection of DDoS attack on SDN control plane using Hybrid Machine Learning Techniques. In 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 299–303). IEEE.
4. Pérez-Díaz, J. A., Valdovinos, I. A., Choo, K. K. R., & Zhu, D. (2020). A flexible SDN-based architecture for identifying and mitigating low-rate DDoS attacks using machine learning. *IEEE Access*, 8, 155859–155872.
5. Zhou, L., Liao, M., Yuan, C., & Zhang, H. (2017). Low-rate DDoS attack detection using expectation of packet size. *Security and Communication Networks*, 2017.
6. Jadhav, P. N., & Patil, B. M. (2013). Low-rate DDOS attack detection using optimal objective entropy method. *International Journal of Computer Applications*, 78(3).
7. Sahoo, K. S., Puthal, D., Tiwary, M., Rodrigues, J. J., Sahoo, B., & Dash, R. (2018). An early detection of low rate DDoS attack to SDN based data center networks using information distance metrics. *Future Generation Computer Systems*, 89, 685–697.
8. Bhuyan, M. H., Kalwar, A., Goswami, A., Bhattacharyya, D. K., & Kalita, J. K. (2015, April). Low-rate and high-rate distributed DoS attack detection using partial rank correlation. In 2015 Fifth International Conference on Communication Systems and Network Technologies (pp. 706–710). IEEE.
9. Bhushan, K., & Gupta, B. B. (2018). Hypothesis test for low-rate DDoS attack detection in cloud computing environment. *Procedia computer science*, 132, 947–955.
10. Cheng, H., Liu, J., Xu, T., Ren, B., Mao, J., & Zhang, W. (2020). Machine learning based low-rate DDoS attack detection for SDN enabled IoT networks. *International Journal of Sensor Networks*, 34(1), 56–69.

11. Deepa, V., Sudar, K. M., & Deepalakshmi, P. (2019, March). Design of Ensemble Learning Methods for DDoS Detection in SDN Environment. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1–6). IEEE.
12. Sudar, K. M., Deepalakshmi, P., Nagaraj, P., & Muneeswaran, V. (2020, November). Analysis of Cyberattacks and its Detection Mechanisms. In 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) (pp. 12–16). IEEE.
13. Singh, S., & Gupta, P. (2014). Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey. International Journal of Advanced Information Science and Technology (IJAIST), 27(27), 97–103.
14. Jiang, L., Wang, D., Cai, Z., & Yan, X. (2007, August). Survey of improving naive bayes for classification. In International Conference on Advanced Data Mining and Applications (pp. 134–145). Springer, Berlin, Heidelberg.
15. Mininet Team Mininet: An Instant Virtual Network on your Laptop (or other PC). Available online: <http://mininet.org/> (accessed on 3 January 2021).
16. S. Wilson Prakash and P. Deepalakshmi, Flow-based Dynamic Load balancing algorithm for the Cloud networks using Software Defined Networks, International Journal of Cloud Computing, 8(4) (2019), 299–318.
17. Scapy-Packet Crafting Tool. Available online: <https://scapy.net/> (accessed on 3 January 2021)
18. S. Wilson Prakash and P. Deepalakshmi, DServ-LB: Dynamic server load balancing algorithm, International Journal of Communication Systems, 32 (1) (2019), 1–11.
19. Leevy, J. L., & Khoshgoftaar, T. M. (2020). A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data. *Journal of Big Data*, 7(1), 1–19.
20. Zhijun, W., Wenjing, L., Liang, L., & Meng, Y. (2020). Low-Rate DoS Attacks, Detection, Defense, and Challenges: A Survey. *IEEE Access*, 8, 43920–43943.
21. Muthamil Sudar, K., & Deepalakshmi, P. An intelligent flow-based and signature-based IDS for SDNs using ensemble feature selection and a multi-layer machine learning-based classifier. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1–20.
22. Balakiruthiga, B., Deepalakshmi, P., Mohanty, S. N., Gupta, D., Kumar, P. P., & Shankar, K. (2020). Segment routing based energy aware routing for software defined data center. *Cognitive Systems Research*, 64, 146–163.

Design and Simulation of MEMS Based Capacitive Accelerometer



S. Veena, Newton Rai, Amogh Manjunath Rao Morey, H. L. Suresh, and Habibuddin Shaik

Abstract Accelerometer is an electromechanical device, which is used for physical measurement along the orthogonal coordinates. Micro Electro Mechanical Systems (MEMS) based capacitive accelerometers are embedded in many modern technological applications. This paper presents the comparison between two single axis MEMS based capacitive accelerometers, which have the natural frequencies of 7 and 2.2 kHz. This work includes design, simulation, analytical modelling, and finite element modelling of each MEMS comb type capacitive accelerometer with different operating frequencies. The accelerometer was designed using COMSOL Multiphysics and MATLAB simulator tool.

Keywords Accelerometer · MATLAB · COMSOL

1 Introduction

Accelerometer is an electromechanical device, which is used for physical measurement like vibration, acceleration or force of a moving solid. Most accelerometers are based on the principle of mechanical vibration. The basic structure of mems accelerometer contains the seismic mass supported by beams. The mass is frequently appended to a dashpot that gives the essential damping impacts [1, 2]. The spring and the dashpot are in turn connected to frame as shown in Fig. 1 [3]. An accelerometer, which is stored at rest on the surface of earth, will measure acceleration governed by gravity ($g \approx 9.81 \text{ ms}^{-2}$) and in contrast, accelerometers, which are in free descend state, will measure to zero [3]. The mathematical analysis was executed with accordance of damping [5] in order to analyse the characterization of the device. The

S. Veena · N. Rai (✉)

Nitte Meenakshi Institute of Technology, Bengaluru, India

A. M. R. Morey · H. L. Suresh · H. Shaik

Sir. M Visvesvaraya Institute of Technology, Bengaluru, India

e-mail: hl_suresh@rediffmail.com

H. Shaik

e-mail: habibuddin.shaik@nmit.ac.in

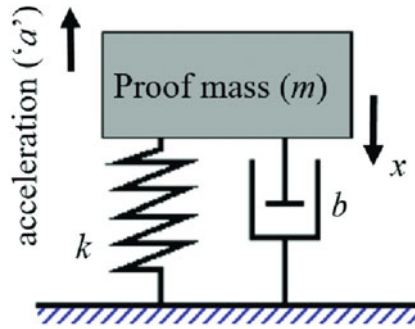


Fig. 1 Typical spring mass damper model of accelerometer (Source MDoF [3], p. 3)

capacitive based MEMS accelerometers measures capacitance fluctuation between a proof mass and an anchored conductive electrode disembodied by a small gap [6]. Here, the sensitivity of the accelerometer is a measure of displacement with respect to acceleration [8]. Although commercially different accelerometers are available, the main aim for selecting capacitive based MEMS accelerometers are high sensitivity with respect to resonant frequency [8] and reliability. This work presents a comparison between two novel single folded beam type capacitive MEMS Accelerometers. There are a wide scope of utilizations that require acceleration measurement such as automotive industry, biomedical applications [11], vibration analysis, navigation system, robotics, and this measurement is also achievable with different accelerometers fabricated through the CMOS micromachining process [9]. Hence, the need of exploration with respect to this field is rampant, as the configurations are getting more diverse. Air was elected as the di-electric medium between the combs [13]. This work was carried out, starting with the structural design and mathematical analysis along with the subsequent simulations to verify the same and an overview about the comparison between the two accelerometers.

2 Structure Design and Working Phenomenon

The structure design of proposed capacitive based MEMS accelerometer is as shown in Fig. 2 [1]. The proposed design of MEMS accelerometer has unique structure in comparison to former design, where two proof masses is suspended on either side by a single folded beam. The structure of mems comb accelerometer is symmetry in shape. Movable parts comprises two proof mass, single folded beam, movable fingers and Fixed parts comprises anchors, pad metal, fixed fingers attached to pad metal [9]. The two proof mass are symmetrically suspended to central anchor through single folded beam. The fixed fingers are connected to the left and the right anchors.

The two proof masses are responsible for vibrating in response to the incident inertial force. The portable fingers are attached on one side of the each proof mass,

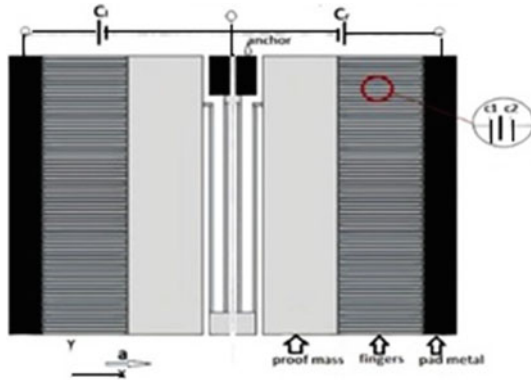


Fig. 2 Design of proposed MEMS comb accelerometer

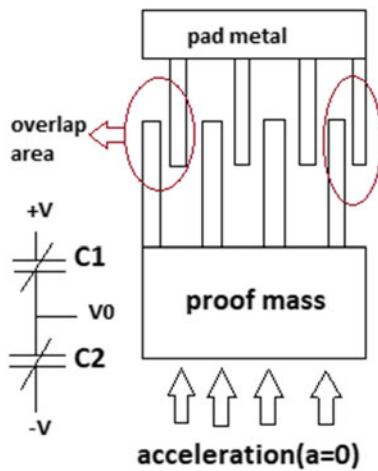


Fig. 3 Resting position of the Accelerometer

and the other side is attached to the beam. The portable fingers (Number of portable fingers $[N_f]$ for Model X-66 and Model Y-13) are surrounded by anchored fingers which establishes the differential capacitance couple C_1 and C_2 with anchored fingers on either side [10]. The structure consists of two positive electrodes and one ground electrode [11].

Keeping this in check, the models were reconstructed in Comsol Mutliphysics 5.1a. Silicon material was chosen from the designated Comsol library function.

In order to draw comparisons between the two different frequency oriented accelerometers, the working phenomenon should be utilized. At steady state (when acceleration = 0), the proof mass does not experience any inertial force so the portable fingers are in rest. As there is no significant change in area between fixed and portable

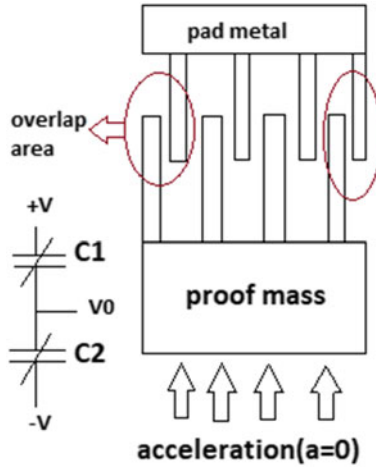


Fig. 4 Accelerometer in movement

fingers, the capacitance couple C1, C2 are equal. Under steady state an external driving voltage is applied through positive and negative electrode, which develops an electrostatic force and this force will try to displace the sensing electrodes and the corresponding capacitance.

3 Mathematical Analysis

3.1 To Find the Natural Frequency of the Accelerometers

The natural frequency of a general spring mass damper system [1] is

$$f_2 = \frac{1}{2\pi} \sqrt{\frac{k}{m}}, \tag{1}$$

where

the total sensing mass of accelerometer is given by [12], $m = \text{mass}$

$K = \text{stiffness of spring}$ is given by

$$K = \frac{EtW^3}{4L^3} \tag{2}$$

where L , t , and W are the length, thickness, and width of the beam respectively, $E = \text{Young's modulus of elasticity (silicon)} = 170 \times 10^9 \text{ N/m}^2$.

3.2 To Find the Damping Ratio of the Accelerometers

1. Model X Accelerometer

Squeeze film damping: The pressure is developed due to the compression of air between moving fingers and fixed fingers, and this effect is called damping effect. It acts as a dragging force and opposes the movement of sensing fingers [13]. An analytical formula [6] for the air squeeze film damping is given as

$$b = N_f n_{eff} l_f \left(\frac{t}{d_o} \right)^3, \tag{3}$$

where

n_{eff} is the effective viscosity of air 18.5×10^{-6} Ns/m², number of portable fingers (N_f) and l_f, t, d_o -Table [1], after substituting all values in Eq. 3 the damping coefficients is $b = 3.815625 \times 10^{-5}$ Ns/m.

At Steady state:

$$S = \frac{1}{w_n^2}, \tag{4}$$

where, S = sensitivity.

At dynamic state: The dynamic behaviours of spring mass damper system is defined by parameters like angular frequency (w_n) and damping ratio (ζ). The dynamic analysis of this system is done in Laplace domain of equation [2, 9]

when x → F

when x → a

$$\frac{X(s)}{F(s)} = \frac{1}{ms^2 + bs + k}, \tag{5}$$

$$\frac{X(s)}{a(s)} = \frac{1}{s^2 + \frac{b}{m}s + \frac{k}{m}}, \tag{6}$$

where

$$w_n = \sqrt{\frac{k}{m}}, \tag{7}$$

$$Damping\ ratio\ (\zeta) = \frac{b}{2mw_n}. \tag{8}$$

The value of damping ratio (ζ) directly controls the nature of the system response.

For $0 \leq \zeta < 1$, the system is under damped system

For $\zeta = 1$, the system remains critically damped

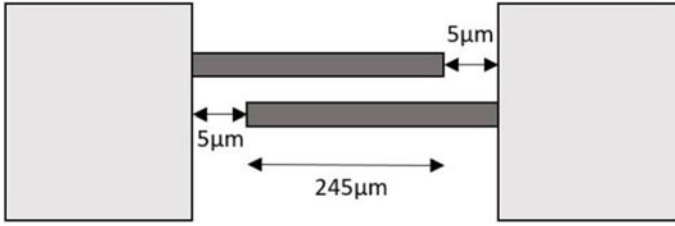


Fig. 5 Positioning of the combs in the accelerometer structure

For $\zeta \geq 1$, the system remains overdamped

$K, m =$ Table [2]

On substituting in above equations, we get damping ratio as = **0.03208**.

2. Model Y Accelerometer

The length of the comb and the distance between the comb and the proof mass is shown in Fig. 5.

Here the value of damping coefficient is given by [13]

$$b = \frac{64\sigma p_a L w}{g \pi^6 w_n} \sum_{min} \frac{m^2 + \left(\frac{n}{B}\right)^2}{(mn) \left[\left(m^2 + \frac{n^2}{B^2}\right)^2 + \frac{\sigma^2}{\pi^4} \right]} \pi^6, \tag{9}$$

where

$L = 245\mu m, W = 10\mu m, \sigma =$ compression ratio of air, $P_a =$ Ambient pressure (1.01×10^{-5} Pa), and $W_n =$ natural frequency.

And natural frequency is given as

$$W_n = 2\pi f \tag{10}$$

For $f = 7$ kHz, $W_n = 14000\pi$ Hz.

And this compression ratio is given as

$$\sigma = \frac{12uw^2w_n}{p_a g^2}. \tag{11}$$

Here,

$u =$ coefficient of dynamic viscosity (1.85×10^{-5}), $\sigma = 0.19576$.

The variable B is given by

$$B = L/W \tag{12}$$

$B = 245 \mu m / 10 \mu m, B = 24.5$.

Table 1 Dimensions of the capacitive MEMS Accelerometers

Accelerometer parameters	Dimensions (μm) of Model X accelerometer	Dimensions(μm) of Model Y accelerometer
Right proof mass	225X1000X25	225X400X25
Left proof mass	225X1000X25	225X400X25
Gap between two fingers (d_0)	5	5
Pad metal size	100X1000X25	100X400X25
Thickness of device (t)	25	25
Beam length (L)	550	250
Beam width (W)	10	10
Length of finger (l_f)	245	245
Breadth of finger (b_f)	10	10

Substituting this in the main damping coefficient equation we get $b = 7.304 * 10^{-6}$ Ns/m; on substituting in the equations [7, 8] we get damping ratio as = **0.0549**. The damping ratio for both the systems is less than one (< 1), hence the proposed accelerometers are underdamped systems. The above calculations for both the accelerometers was carried out to execute the simulations in MATLAB which is discussed further (Sect. 4) in this paper.

3.3 To Find the Static Capacitance of the Accelerometers

The static capacitance is expressed as

$$C_0 = \frac{EN_f l_f t}{d_o} \quad (13)$$

where

C_0 is the static capacitance, Length of finger (l_f), Number of movable fingers (N_f), Thickness (t), distance between the plates (d_o)

Referring Table 1, the static capacitance obtained for Model X Accelerometer is $C_0 = 0.730455\text{pf}$ and Model Y Accelerometer $C_0 = 0.28199\text{pf}$.

3.4 To Find the Displacements Sensitivity of the Accelerometers

$$F_{inertia} = M * a \quad (14)$$

$$F_{damping} = K * x \quad (15)$$

The displacement sensitivity of proof mass is given by

$$S_d = x = \frac{M}{K} * a, \tag{16}$$

where

M = mass, K = stiffness of the spring.

It is found that the displacement sensitivity of the proof mass of Model X Accelerometer is $S_d = 3.53625 * 10^{-8}m/g$ and of model Y Accelerometer is $S_d = 4.5096 * 10^{-9}m/g$.

4 Simulation and Results

4.1 Model X

1. **Frequency simulation:** Silicon (Single crystal) as the base material, as shown in Fig. 6, the simulated frequency obtained was 2.1084 kHz. The simulated frequency values when compared to the calculated values referred from Table 2 gives us an error difference of about 0.492 kHz for the model X accelerometer.

Fig. 6 Frequency simulation of Model X comb

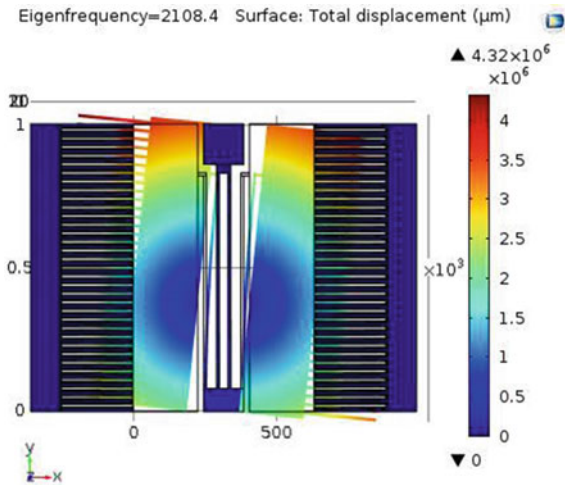


Table 2 Natural frequencies of the accelerometers

Model X Accelerometer	Model Y Accelerometer
K = 10 N/m and m = $3.53625 * 10^{-8}$ Kg	K = 31 N/m and m = $1.42651 * 10^{-8}$ Kg
$f_n = 2.6$ kHz	$f_n \text{ } f7 = 7.41$ kHz

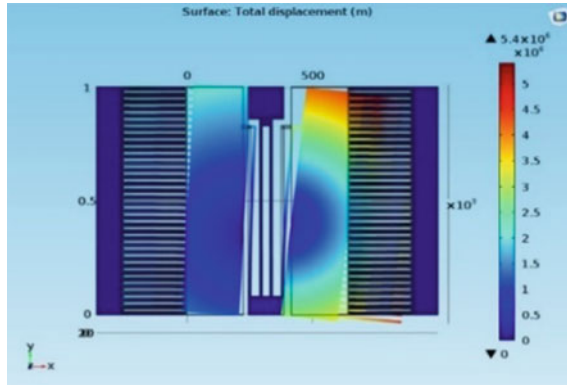


Fig. 7 Total Displacement for the Model X comb

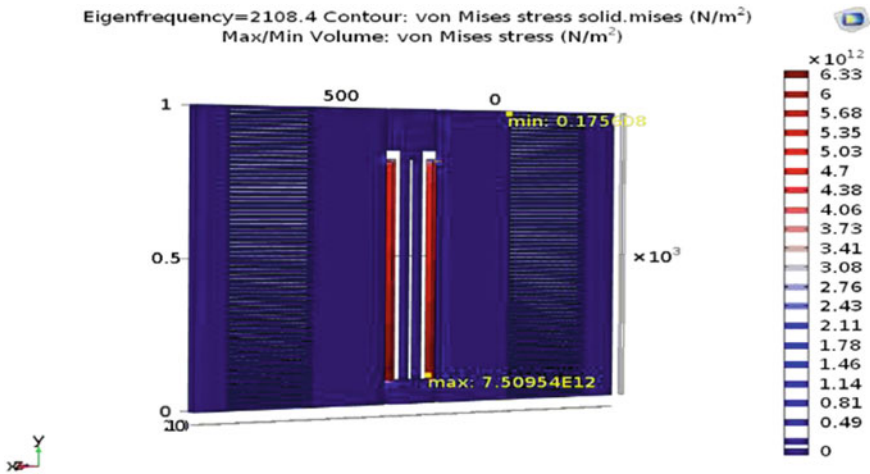


Fig. 8 Stress evaluation in Model X com

2. **Displacement simulation:** Silicon (Single crystal) as the base material, as shown in Fig. 7, the total displacement was observed to be 5.4×10^{-8} .
3. **Stress simulation:** Here the yield point is up to $7.509 \times 10^{-12} \text{ N/m}^2$, beyond which failure could occur in the bottom region of Model X. The module can toggle with a stress range of $7.33 \times 10^{-12} \text{ N/m}^2$ given the min value is $0.175 \times 10^{-12} \text{ N/m}^2$.

4.2 Model Y

1. **Frequency simulation:** Silicon (Single crystal) as the base material, as shown in Fig. 9, the simulated frequency obtained was 7.074 kHz.
2. **Displacement simulation:** Silicon (Single crystal) as the base material, as shown in Fig. 10, the total displacement was observed to be 6.03×10^{-9} m.
3. **Stress simulation:** As shown in Fig. 11, here the yield point is up to 6.361×10^{-12} N/m², beyond which failure could occur in the upper region of Model Y. The module can toggle with a stress range of 7.33×10^{-12} N/m² given the min value is 0.175×10^{-12} N/m² around the corners.

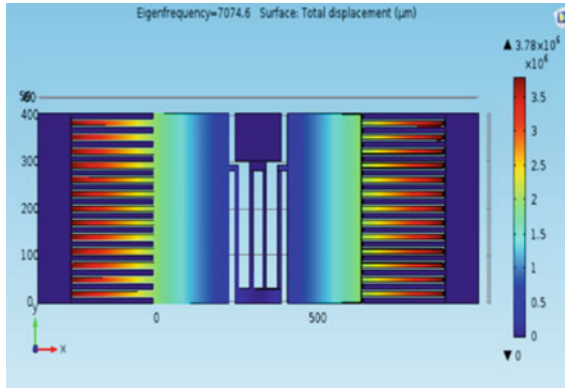


Fig. 9 Frequency simulation of Model Y accelerometer

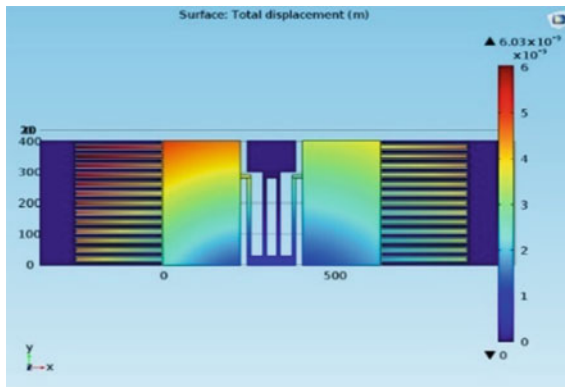


Fig. 10 Total Displacement of Model Y comb

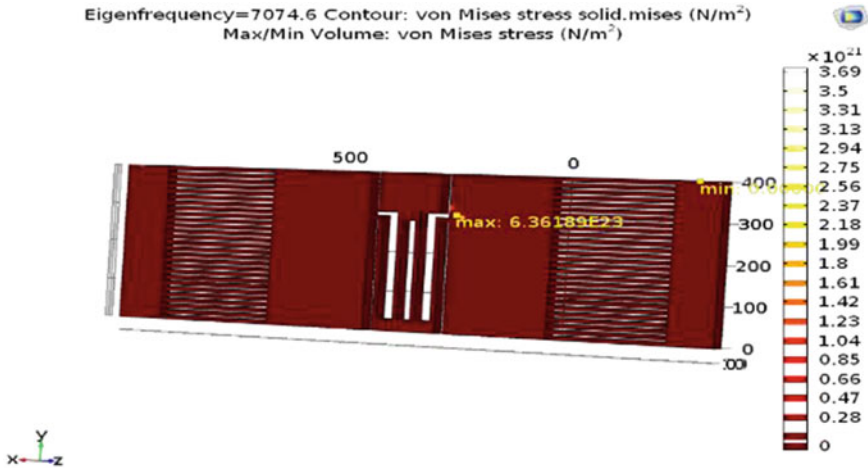


Fig. 11 Stress evaluation in Model Y comb

4.3 Simulations to Check Variation of Mass and Dimensions of the Proof Mass Over Resonant Frequency

1. Width of proof mass is reduced to half keeping the beam length unaltered: Given that both the accelerometers are underdamped systems, one of them (Model Y) was considered to check variation of mass and dimensions of the proof mass over resonant frequency.

2. Both Width of Proof Mass and Beam Length Reduced to Half

As shown in Figs. 12, 13, 14, and 15, it was observed that the resonant frequency of accelerometer depends on its design parameter like beam length and sensing mass.

Fig. 12 Resonance frequency is 10.8 kHz

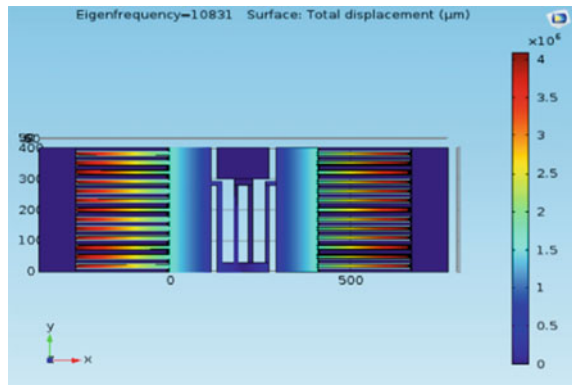


Fig. 13 Max displacement at resonance frequency 10.8 kHz

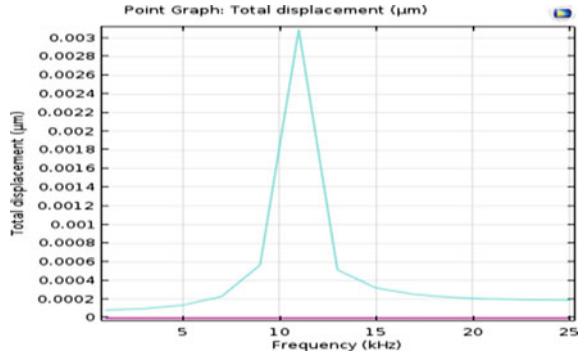


Fig. 14 Resonance frequency is 15.8 kHz

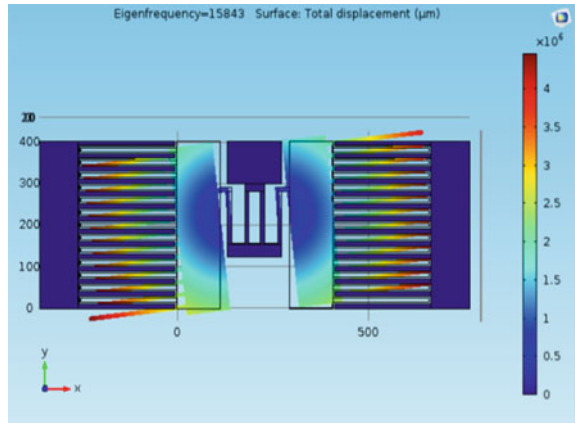
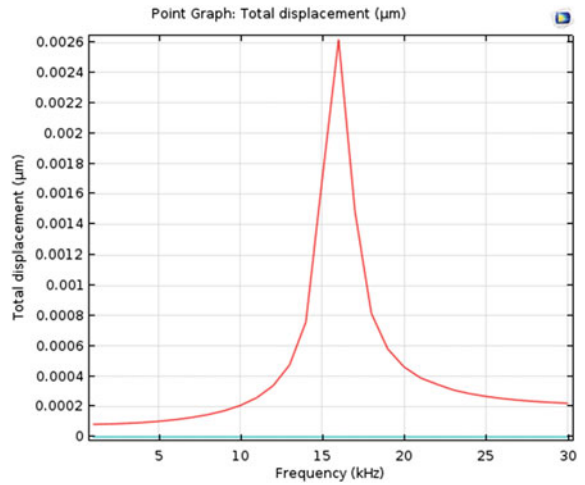


Fig. 15 Max displacement at resonance frequency



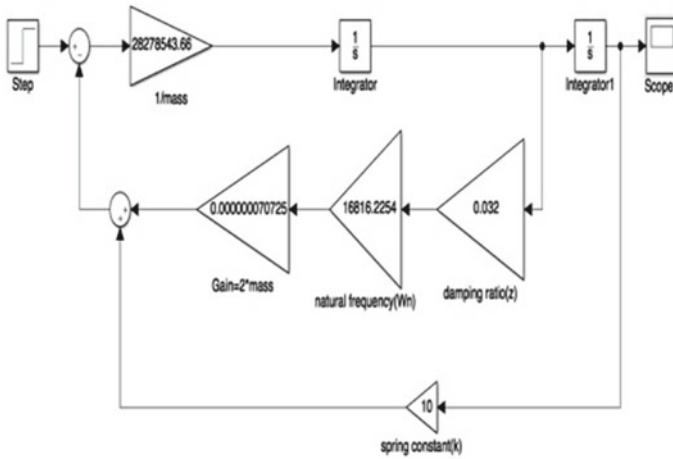


Fig. 16 Simulink model of a

Fig. 17 Step response for $\zeta = 0.5$ (Model X)



4.4 MATLAB Simulations

1. Open loop model of 2 KHZ accelerometer:

In an open-loop system no additional control circuits are required, the electrical output signal is directly interfaced to analog front end circuit (AFE) for capacitance to voltage conversion and thereafter output voltage is used for measurement analysis [5]. The Eq. (7) can be rewritten into a standard form with z and w_n as

$$\frac{X(s)}{a(s)} = \frac{1}{s^2 + 2\zeta s w_n + w_n^2} \tag{17}$$

The open loop Simulink model of two accelerometers are as shown in Figs. 16 and 18. The step responses of both the accelerometers as shown in Figs. 17 and 19 are analogous to that of an underdamped system, where increasing the damping ratio stabilized the Model X more.

2. Open loop model of 7 kHz accelerometer

See Figs. 18 and 19.

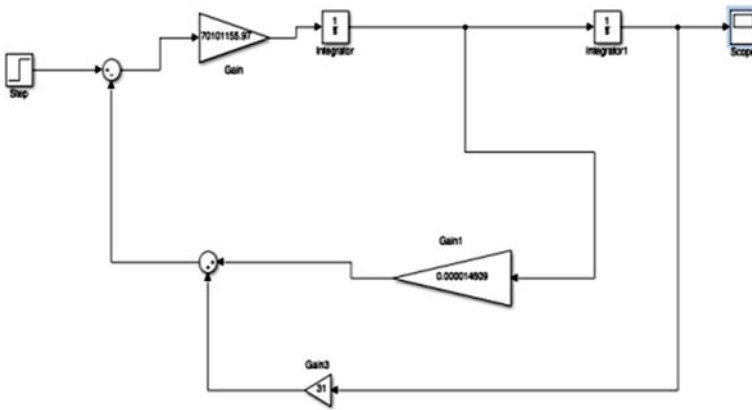


Fig. 18 Simulink model of an open-loop. 7 kHz accelerometer

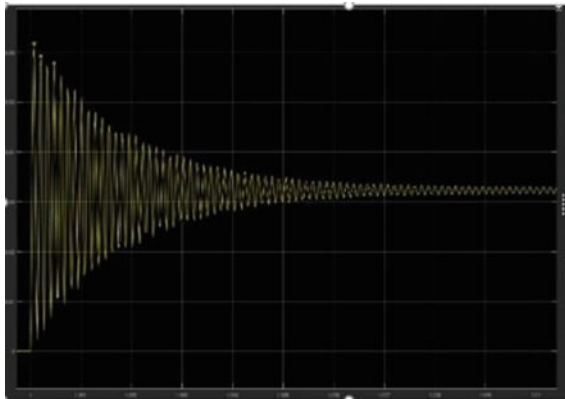


Fig. 19 Step response for $\zeta = 0.5$ (Model Y)

5 Results and Discussions

In this work an attempt is made to employ the detailed design, simulation, mathematical modelling and sensitivity analysis of single axis mems based capacitive accelerometer of different operating frequencies (2 kHz, 7 kHz). Mathematical modelling of frequency, displacement and capacitance is done and found that it is almost closer to the values obtained from simulation. The frequency, displacement and capacitance values are gained by simulating the desired accelerometer in COMSOL Multi-physics software. Both the accelerometers were analysed to be underdamped systems. The error difference between the simulated and calculated value is on an average of 0.045 kHz. Considering the effect of the physical dimensions on sensitivity, it is concluded that though the total displacement remained more or less the same (0.03 + um for 10.8 kHz and 0.028um for 15.8 kHz), the resonant frequency changed in regard to multiple configurations of physical parameters. Moreover, the displacement was found to be more in regard to 2 kHz than 7 kHz; it could be concluded that lower the resonant frequency, greater will be the sensitivity of the accelerometer. In addition, this is cross-verified by comparing the displacement sensitivity factor which is relatively more with respect to the Model X accelerometer making it more susceptible to displacement. Model X has more stable capacitive range as its static capacitance is more, which allows it to be used to detect multiple voltage surges caused by various displacement factors. Moreover, taking the von mess stress into consideration, the Model Y accelerometer can withstand considerably lower stress, where the maximum stress hold point is quite opposite in Model X (lower with respect to the beam). Analytically, the damping ratio for 2 kHz accelerometer is calculated and found to be 0.03208 and concluded this device to be underdamped. It was also observed that by improving to 0.5, the system becomes more stable. The comparison between the simulated and calculated values of the two accelerometers is displayed in Tables 3 and 4.

Table 3 Comparison between simulated and calculated values

Model X		Model Y	
Simulated value	Calculated value	Simulated value	Calculated value
$f_n = 2.1$ kHz	$f_n = 2.6$ kHz	$f_n = 7.0$ kHz	$f_n f7 = 7.4$ kHz

Table 4 Comparison of the two accelerometers

Parameters	Model X	Model Y
Natural frequency	$f_n = 2.6$ kHz	$f_n f7 = 7.41$ kHz
Static capacitance	$C_o = 0.730455$ pf	$C_o = 0.28199$ pf
Displacement sensitivity	$S_d = 3.53625 * 10^{-8}$ m/g	$S_d = 4.5096 * 10^{-9}$ m/g
Stress range	Min:0.0175 N/m ² to max-7.509E12 N/m ²	Min-0.00 N/m ² to max -6.361E23 N/m ²

6 Conclusion and Future Scope

The Model X accelerometer surpasses the Model Y in many parameters, making it a viable selection over the other; it can be concluded to extend this work towards improving the of the model X accelerometer. Although, great number of research papers was available on improvement of sensitivity of capacitive MEMS accelerometers, still there is a lot of scope to carry out the research on capacitive accelerometers to improve its sensitivity. Therefore, the future development of this work will be to enhance performance and sensitivity of capacitive accelerometers that includes change of finger width, shape of proof mass, applied load and by using a complex pair of folded beams structure and produce the more optimized structure.

Acknowledgement We take immense pleasure in expressing our gratitude to Dr.Veda Sandeep Nagaraja, Associate Professor, ECE department, NMIT, and all the staff members of EEE Dept, NMIT, and Bangalore for the assisting in our learning. We also thank the staff members of Nano materials and MEMS COE for the support extended. We further acknowledge the guidance provided by the management of NMIT, Bengaluru.

References

1. Padmanabhan Y (2017) Mems based capacitive accelerometer for navigation. Dissertation University of Windsor. <https://doi.org/10.13140/rg.2.2.35625.49769>.
2. Kei-Ming Kwong (2017) MEMS Accelerometer Specifications and Their Impact in Inertial Applications Dissertation, Master of Applied Science, Department of Computer Engineering, the University of Toronto
3. Zakriya M, Ibrahim (Abe) ME, Mahmoud R (2018) Monolithic Multi Degree of Freedom (MDoF) Capacitive based MEMS Accelerometers. *J Micromachines* (Basel). 10.3390/mi9110602
4. Senturia (2001) *Microsystem Desig.* ISBN 0-7923-7246-8 Publisher: Kluwer Academic Publishers
5. Akila Kannan BE (2005) Design and modelling of a mems-based accelerometer with pull in analysis by (Electrical Engineering), the Anna University
6. Albarbar A, Mekid S, Starr A. (2008) Suitability of MEMS Accelerometers for Condition Monitoring: An experimental study. *Journal of Sensors*. <https://doi.org/10.3390/s8020784>
7. Amini BV (2006) A mixed-signal low-noise sigma-delta interface IC for integrated sub-micro-gravity capacitive SOI accelerometers. Dissertation, School of Electrical Engineering, Georgia Institute of Technology
8. Wong WC, Azid IA, Majlis BY (2010) Theoretical analysis of stiffness constant and effective mass for a round-folded beam in MEMS Accelerometer. *J Mech Eng* 57(2011)6, 517–525 Paper received: 03.11.2009 <https://doi.org/10.5545/sv-jme.2009.151>.
9. Benmessaoud, M., & Nasreddine, M. M. (2013). Optimization of MEMS capacitive accelerometer. *Microsyst Technology*, 19, 713–720. <https://doi.org/10.1007/s00542-013-1741-z>.
10. Gupta V, Mukherjee T (2012) Layout synthesis of CMOS MEMS accelerometers. Department of ECE, Carnegie Mellon University, Pittsburgh, PA-15213.
11. Nagaraja VS, Rudresha KJ, Pinjare, S. L. (2018). Design, Fabrication and Characterization of a Biologically Inspired MEMS Directional Microphone. *IEEE SENSORS*, New Delhi, 1–4

12. Sharma, K., Macwan, I. G., Zhang, L., Hmurcik, L., & Xiong, X. (2012). *Design optimization of a MEMS comb Accelerometer*. Department of electrical engineering: University of Bridgeport, Bridgeport.
13. Zhou W., Chen Y., Peng B., Yang H., Yu H., Liu H., and He X. (2014). Air Damping Analysis in Comb Microaccelerometer. *Journal of Advances in Mechanical Eng*
14. Kumar D. P. and Karat (2006). System Level Simulation of Servo Accelerometer in Simulink. Department Of Earthquake Engineering, I.I.T.Roorkee, India. *Journal of Physical Sciences*, Vol. 10, 2006, 143 – 150.

Transport Tracking Using RFID and GSM Based Technique



N. Subbulakshmi, R. Chandru, and R. Manimegalai

Abstract Nowadays, the crime has increased and the occurrence of the accidents are more in the cities. People are afraid about the current scenario. Moreover parents are worried on children when they travel to schools. In order to provide protection and safety to the children, there is a necessity for a technology based security and alert. In this paper, an idea to identify the status of the children when they are out to school so that they can identify the location of their children. Using GSM and RFID the tracking of transport is done. It has two units; one is bus unit for tracking the activity and alerting parents during the school travel. Another is school unit for tracking child inside the school. The results show that our approach will provide safety and security to school children.

Keywords RFID · GSM module · Sensors · GPS module · PIC controller · Tracking system

1 Introduction

At present accidents become predominant, and people are conscious about the safety and life. Parents are very much concerned about their children, and they admit their children in highly reputed schools which has hi-tech amenities. Recently, most of the schools have transport services in around that city peripheral. Though children go by school bus, parents are worried whether the child reaches the school safely or is in some danger situation, i.e., met with any accidents [1]. This paper identifies such accidents and alert the parents. This method conveys the alert messages which are used to send pick up and drop information to the parents. As discussed above, bus unit

N. Subbulakshmi (✉)

Francis Xavier Engineering College, Tirunelveli, Tamilnadu, India

e-mail: subbulakshminammalwar@francisxavier.ac.in

R. Chandru

Sri Ramakrishna Engineering College, Coimbatore, Tamilnadu, India

R. Manimegalai

PSG Institute of Technology and Applied Research, Coimbatore, Tamilnadu, India

consists of RFID (Radio-Frequency Identification) reader that uses electromagnetic fields to automatically track and identify the tags that is being attached in ID cards of the children so that once the card is read it means child has entered the bus. It has a local power source like battery.

2 Literature Review

Various types of sensors like temperature, gas sensor, and accelerated tilt sensors are also placed in the transport any accidents during driving. When there is such a danger, the alert message will reach the parents with help of GSM and GPS modules. GSM (Global System for Mobile Communications) is used for mobile communication which can operate over a wide range of network. GPS (Global Positioning System) is used to identify the geographical location of the child. This ensures that each movement of the child is being identified by their parents. Next the school entity comprises of RFID reader and GSM module. Both the units are being processed by the PIC Controller 16F877A. RFID tag is worn by the child and when it is placed [2]. The sensors and the RFID tag are interfaced through the controller. There will be a LCD display for drivers to view where the child is getting down. If a child gets down somewhere else other than school the message of the location will be displayed to the driver. The GSM modem send message to authorized person according to the data received [3].

Parents will get message when child enters the bus, leaves the bus, when enters the school premises and leaves the school. The RFID reader reads the ode in all these cases. Each activity and location of the child is monitored by their parents through the GPS. If any accident on the way to school, that time also alert message will reach the parents. This system will be efficient and ensure security and safety of the child. The paper is proposed to reduce the risk of accidents and loss of life of school children without any reason. There were so many school bus accidents that occurred which has taken life of so many children. There were so many safety measures taken for school children to prevent accidents like speed control and checking devices to avoid over speeding of bus, limitation of children count in bus, etc. But there were no alert systems and tracking systems. The RFID based this system will track the activity of the child and will report if there is any danger [2]. The live location is also tracked for quicker recovery. It helps in intimating school management and parents regarding the danger in bus. And also when there is tilt in the bus driving, that is also alerted. SMS based kids' tracking and safety system is introduced in [4]. RFID GSM imparted school children security system is described in [5]. Efficient real time tracking is designed in [6]. A robust tactic-based real time framework for public transport synchronization is discussed in [7].

3 Design and Implementation

The block diagram of the proposed system is shown below as depicted in Fig. 1.

The main components used in the system is shown in Fig. 1. The main component is PIC 16F877A. It acts as a main control unit of the whole proposed system. All the other components like sensors and GPS modules are fed as input to the controller. The alert messages can be carried to parents along with GSM modem. The details of the components in the system is described below.

3.1 Pic 16f877a

The cost of this controller is low, and its handling is also easy. Because of these features it is used in the system. It processes the data and then sends output data to GSM modem and LCD display. Figure 2 shows PIC 16F877A.

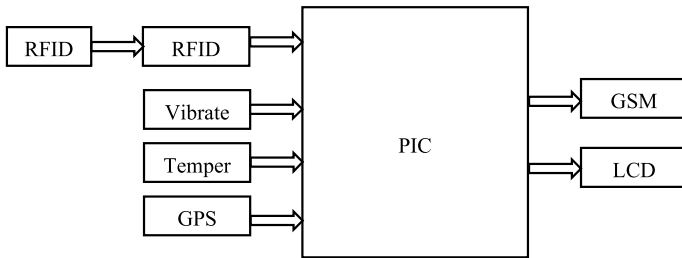


Fig. 1 Block diagram of the system

Fig. 2 PIC 16F877A



3.2 RFID Tag

The RFID tag is worn by child. RFID tag is active tag, and another is passive tag. The active tags have battery, and they provide longer read range than the latter and moreover it broadcast its own signals [3]. Whereas passive tag is opposite of active tags. The passive tag is shown in Fig. 3.

The process of RFID is shown in Fig. 4. The passive tag consists of a chip and an antenna. The antenna collects the signal from the reader and controls the chip. Then the signal will be adjusted based on unique information which is available the block. This signal is again retransmitted to the antenna. Passive tag will be appropriate for this system because it has short reading range which is necessary for this application. So that it can identify the person who comes near to tag reader.

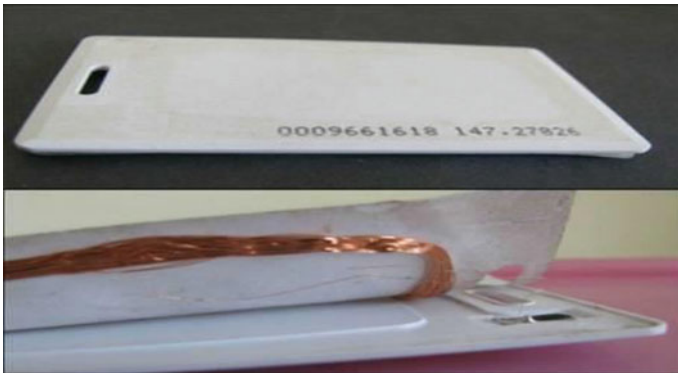


Fig. 3 RFID TAG

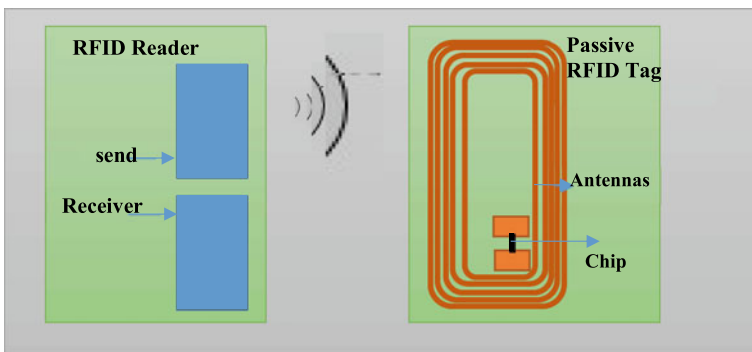


Fig. 4 Process of RFID

3.3 RFID Reader

The RFID reader is used to read the signal that is being sent from tag. It is of three types that is classified based on their frequency range. The types are low frequency, high frequency and ultrahigh frequency readers. The reader that is suitable for this application is low frequency reader. EM-18 is being used which has a frequency range of 125 KHz. It has serial output with range of distance 8–12 cm. Another feature is that it has an in built antenna which can be directly connected to the PC by RS232 protocol [4] and [3]. The reader also has on-chip antenna and can be powered up with help of a 5 V power supply, and moreover it is inexpensive [5].

3.4 Temperature Sensor

LM35 is a precision temperature sensor that is used for this proposed system. It monitors the temperature exactly when compared to thermistor. The operating voltage is from 4 to 30 V. Its accuracy is approximately about 0.5 Celsius. The range of the system is from –55 to 150 C. It has three pins in which 1st is power supply pin, 2nd pin is output that is being connected to PIC, and 3rd pin is ground pin.

3.5 Accelerated Tilt Sensor

Accelerometer is a device that measures the proper acceleration. The ADXL330 is a smallest triaxle accelerometer. It is thin, and it is low power device. It measures the dynamic and static acceleration. Dynamic acceleration results from vibration, shock or motion. Whereas static acceleration is gravity in tilt-sensing. This sensor is being interfaced with the PIC. The sensed data is processed by the PIC16F877A and the output signal, i.e., the alert message is being transmitted to parents using GSM.

3.6 Gas Sensor

Gas sensor is used to detect the presence of gases. It will detect toxic gases, flammable, combustible and oxygen depletion. It is widely used for safety purpose. MQ6 is being used here which has high sensitivity to gases like Butane, LPG and Propane. It can detect concentration of gases from 200 to 1000ppm. This is also interfaced with PIC and the sensed output information will be transmitted to parents through GSM.

3.7 *GPS Module*

GPS stands for Global Positioning System. It helps in tracking and identifying the location of the child. It has coordinates, and it is captured. The GSM modem sends the message to parents and school management. The location is accurately determined by it. GPS is a satellite navigation system that accurately identifies location and time [4]. The GPS receiver identifies the location of the students and also there are many other satellites and ground stations to support this.

3.8 *GSM Module*

GSM clearly explains about the 2G networks (cellular) which are used in mobile communication. It has a wide range of usage in around 219 countries and union territories. This technology was implemented using one of the narrow band techniques named as Time Division Multiple Access (TDMA). Its process is as follows. The data is digitized, and then it is compressed and then is passed through the transmitting channel. The frequency band is in the range 900 MHz or 1800 MHz. Since it is present globally it allows the users to use their same mobile phone all over the world but by changing the SIM (Subscriber Identity Module) card [4].

4 **Hardware Implementation**

The GPS module, GSM module, the various types of sensors and the RFID reader are connected to the PIC controller. The gas sensor and accelerated tilt sensor detect the gases entering into the bus and the tilting of the bus, respectively, and send the received signal to the PIC controller. Figure 5 represents the hardware implementation of the model.

The controller processes the received information from all those inputs and gives relevant output to the GSM module and the LCD display. When the RFID tag is placed near the reader it reads the code and it will interface it to PIC.

At the same time if there is any fire accidents, road accidents or any gas leakage into the bus, the system has different types of sensors to detect the accidents. As discussed previously all these are interfaced with controller. This controller issues a command to the GSM modem in order to deliver the received messages for respective persons [2]. The GPS plays a major role here. It along with the messages sends the location. At last the parents and the school management receives the messages from the GSM modem [4]. In paper [3] also this block can be implemented for the automatic design.



Fig. 5 Hardware implementation

5 Results and Discussions

Some of the results of the system is shown below. When the child safely boarded and de-boards the bus the following message is received by the parents [6]. The current situation of the child is being mentioned, and it is done through GSM modem. This controller sends some instruction to the GSM modem. According to this the alert message will be received by the school authorities and parents. References [8, 9] describes the survey of Filter bank algorithms. Low power techniques have been discussed in [10] and [11].

This shows the fire accident alert message. When the fire is being detected in the transport is identified by the sensor present, and it will send a message to the GSM modem. The temperature sensor that is used here will identify the changes in temperature will sense that and gives information to the controller. From the information of the controller, the information will reach the authorized persons through GSM modem. Similarly, the gas sensor detects the presence of gas in the transport and accelerated tilt sensor detect the change in acceleration in the traveling bus and sends the information in a similar way to the controller which in turn sends alert message to parents [7]. All three sensors work in a similar way in transmitting the messages. Many algorithms were discussed for the gearing aid applications in [9].

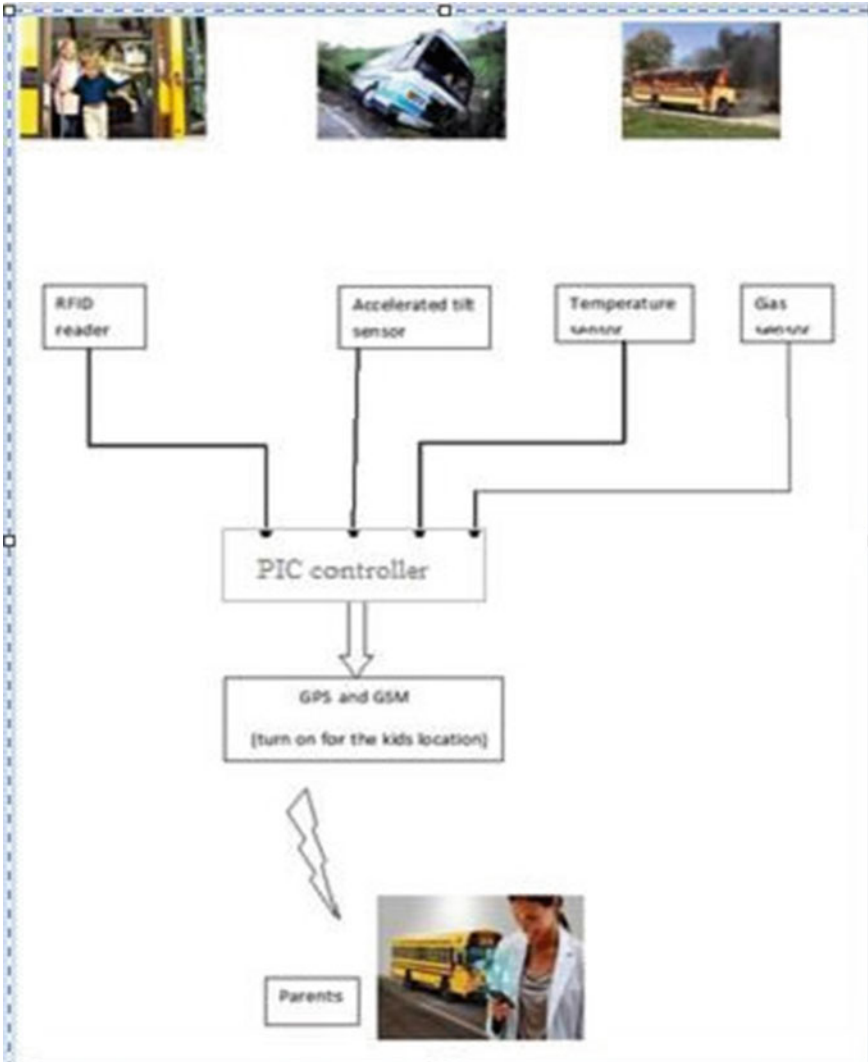


Fig. 6 PIC controller

6 Conclusion

This paper focuses on safety of the school children during their school travel. GSM module is responsible for issuing messages to the parents during boarding and de-boarding of the bus. Similarly, the gas sensor detects the presence of gas in the transport and accelerated tilt sensor detect the change in acceleration in the traveling bus and sends the information in a similar way to the controller which in turn sends alert message to parents. The types of sensors identify the accidents. It has two units

Fig. 7 Alert message of Boarding and Deboarding

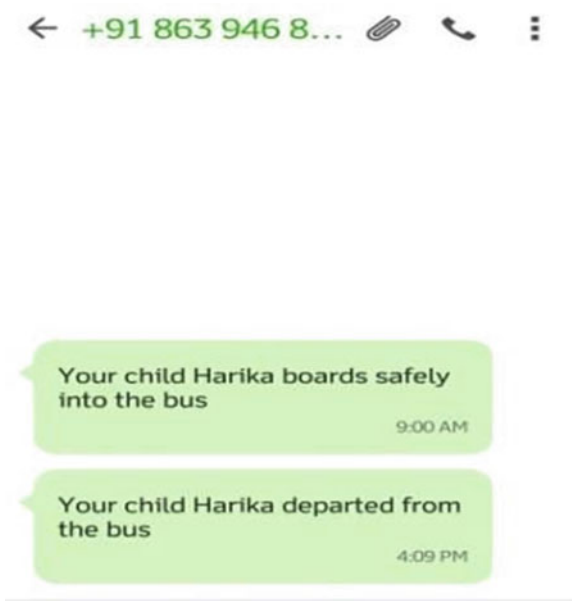
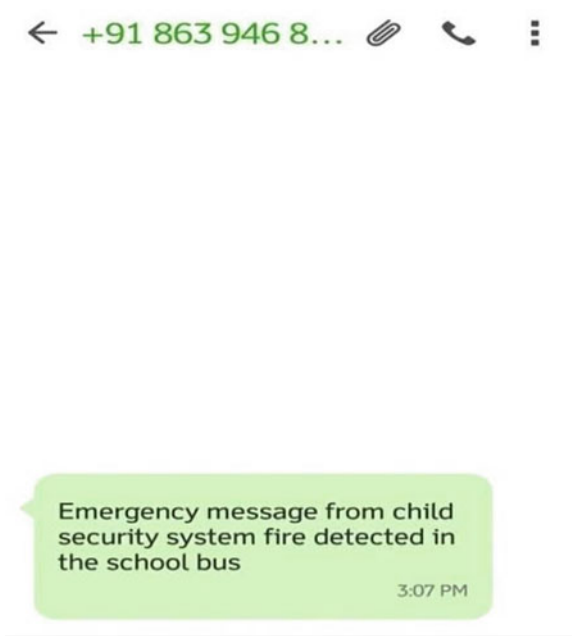


Fig. 8 Accident message



one is Bus unit for tracking the activity and alerting parents during the school travel. Another is school unit for tracking child inside the school. The results show that our approach will provide safety and security to school children. The location of the accident or the status of the child is detected using GPS module. This will enhance the safety of the children.

References

1. Anusha, R., & China Appala Naidu, R. (2016). GPS and RFID Based School Children Tracking System. *International Journal of Advanced Research in Computer Engineering & Technology (JARCET)*, 5(6).
2. Malliga, R., & Narmatha, T. (2016). RFID-based system for school children transportation safety enhancement. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(4).
3. Subbulakshmi, N., & Manimegalai, R. (2020). Novel lifting filter bank for bird call analysis. In L. Ashok Kumar, L. S. Jayashree, & R. Manimegalai (Eds.), *Proceedings of international conference on artificial intelligence, smart grid and smart city applications* (pp. 871–880). Springer, Cham. https://doi.org/10.1007/978-3-030-24051-6_82.
4. Shyam, N., Kumar, N., Shashi, M., & Kumar, D. (2015). SMS based kids tracking and safety system by using RFID and SGSM. *International Journal of Innovative Science, Engineering & Technology*, 2(5).
5. Vidyasagar, K., Balaji, G., & Narendra Reddy, K. (2015). RFID- GSM imparted school children security system. *Communications on Applied Electronics (CAE)*.
6. James, J. G., & Nair, S. (2017). Efficient, real-time tracking of public transport, using LoRaWAN and RF transceivers. In *Proceedings of the IEEE region 10 conference TENCON 2017–2017*, Penang (pp. 2258–2261). <https://doi.org/10.1109/TENCON.2017.8228237>.
7. Nesheli, M. M., (Avi) Ceder, A., & Liu, T. (2015). A robust, tactic-based, real-time framework for public-transport transfer synchronization. *Part of special issue: Papers selected for Poster Sessions, The Transportation Research Procedia*, 9(2015), 246–268.
8. Elleuch, W. (2014). Mining road map from bug database of GPS data. In *14th international conference on hybrid intelligent systems* (pp. 193–198). IEEE Press, Kuwait.
9. Subbulakshmi, N., & Manimegalai, R. (2014). A survey of filter bank algorithms for biomedical applications. In *2014 international conference on computer communication and informatics, Coimbatore* (pp. 1–5). <https://doi.org/10.1109/ICCCI.2014.6921778>.
10. Rooban, S., Subbulakshmi, N., & Poorna Vamsi, Y. (2021). Low power circuit design for dynamic comparator. *International Journal of Performability Engineering*, 17(5), 444.
11. Subbulakshmi, N., & Manimegalai, R. (2017). MACGDI: low power MAC based filter bank using GDI logic for hearing aid applications. *International Journal of Electronics and Electrical Engineering*, 5(3), 235–239.

An Ultra-Wide Band Patch Antenna for Commercial Communication Applications



L. Diana Evangeline, G. Shine Let, and C. Benin Pratap

Abstract Modern wireless systems require a single antenna to perform multi-band functionality considering different applications. As per the FCC report, the designed antenna works in the ultra-wide band from 0.45 to 14.37 GHz with an impedance bandwidth of 187.8%. For the designed antenna, efficiency is greater than 50% and gain is greater than 1 dBi for the entire UWB band. By tuning the feed line width, the same antenna can operate in eight narrow bands covering various mobile communication bands, military applications, global positioning system, satellite communication, and some part of Ku band that is from 0.45 to 14.37 GHz. The designed antenna provides a reflection coefficient of less than -10 dB for all eight narrow bands and has good impedance matching. Moreover, the proposed structure can also support many mobile applications like LTE, Wi-Fi, Wi-Max, and 5G.

Keywords Ultra-wide band · Narrow band · Antenna · Communication · Microstrip patch

1 Introduction

Federal Communication Commission (FCC) in 2002 reported, frequency range from 3.1 to 10.6 GHz is called Ultra-wide band (UWB) and can be used for various commercial communication applications. Within the UWB band, ITU-T specifies various narrow bands for different applications [1]. Table 1 shows the different frequency ranges and their applications.

Many UWB antennas have been researched and developed for industry and academic purposes [2]. The system integrated with cognitive radio capability should have a UWB antenna and a narrowband antenna for proper communication. For efficient spectrum sensing purposes, a UWB antenna is preferred. To have data communication, a narrowband functionality antenna is used. Due to the rapid advancements in the standards of technology, it would be easier to have a single antenna covering

L. D. Evangeline · G. S. Let (✉) · C. B. Pratap
School of Engineering and Technology, Karunya Institute of Technology and Sciences,
Coimbatore, India

Table 1 Frequency range and its applications

Bands	Frequency range (GHz)	Applications
L band	1–2	RADAR, GPS, GSM-1800 band
S band	2–4	LTE, 5G, Bluetooth, Wi-Fi, ISM band
C band	4–8	Satellite communications, WiMax, Wi-Fi, ISM band
X band	8–12	Military applications
Ku band	12–18	Broadcasting, satellite communication

multiple applications [3]. Researchers have carried out various wideband designs with strips, slots, metamaterial structures, and meander designs to encounter several operating bands. For high-speed communication, a coplanar waveguide (CPW) fed UWB antenna is proposed in the literature [4]. The author used an elliptical radiating structure with a partially bent ground plane. Wideband antenna structures are required for cognitive radio spectrum sensing applications. Various shapes such as circular, elliptical, rhombic, semi-circular crown-shaped radiating structures are designed for wideband spectrum sensing application [5]. Rectangular shape low profile dielectric resonator antenna is proposed by authors for spectrum sensing in cognitive radio applications [6]. A U-shaped tunable radiating stub antenna is designed for cognitive radio applications [7].

Half circular ring and half square ring 2-port monopole UWB antenna are in literature for automotive communications [8]. Electromagnetic bandgap structures near the feed point are used to make the antenna operate in the UWB band [9]. Square shape metamaterial structure is reported in the literature to make the antenna operate in UWB for wireless body area network (WBAN) application [10]. 4-port rectangular shape CPW fed monopole antenna is proposed for multiple-input multiple-output (MIMO) applications [11]. A defective ground structure UWB antenna is proposed by authors for 4G applications [12]. Generally, the slot is a desirable technique to make an antenna operate over a larger bandwidth. The slots can be of any interesting shape. A circular slot structure is used in radiating patch to operate in the UWB band [13–15].

A lotus-shaped printed monopole antenna is designed to operate in the ultra-wide band and is proposed in this article. By varying the feed line width, the same antenna can function in multiple narrow bands within the UWB band. The design considerations, various analyses carried out on the deigned lotus-shaped antenna are discussed in the resulting sections.

2 Design of Lotus-Shaped UWB Antenna

The deigned lotus-shaped UWB antenna structure is shown in Fig. 1. The radiating patch exploits a lotus-shaped geometry, and a partial ground plane is considered. Copper material of thickness 35 μm is used for radiating patch and ground plane.

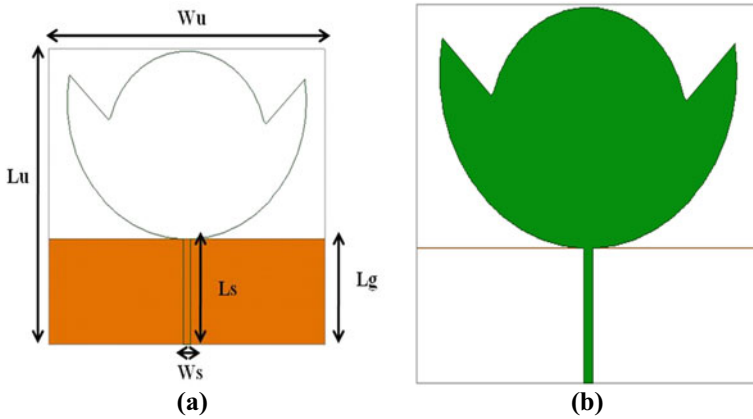


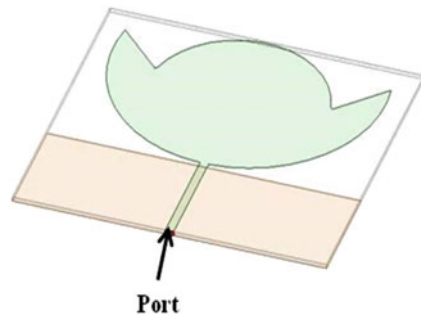
Fig. 1 Lotus-shaped UWB antenna structure **a** ground plane **b** radiating structure

Table 2 Design specifications

Parameter	Dimensions (mm)
Lu	140
Wu	150
Ls	50
Ws	4
Lg	50

Glass epoxy-based commonly available low profile flame retardant-4 (FR4) is used as a substrate. FR4 substrate has a thickness of 1.6 mm, dielectric constant 4.4, and dielectric loss tangent 0.02. $Lu \times Wu$ (mm^2) is the total area of the proposed lotus-shaped antenna. Table 2 provides the design specification considered for the antenna design. The trimetric view and the microstrip feed provided to the antenna are shown in Fig. 2.

Fig. 2 Lotus-shaped UWB antenna—feed point



3 Results and Discussions

The lotus-shaped UWB antenna is simulated using ANSYS HFSS software. The optimum dimensions of the antenna are obtained by using parametric analysis. Using parametric analysis, the length and width of the antenna are varied within a particular range. The simulated performance of the antenna is verified at every step size which is within the given range. From the literature, the operating band of the antenna is decided when the voltage standing wave ratio (VSWR) is less than two or the reflection coefficient is less than -10 dB. The reflection coefficient -10 dB indicates that out of the total transmitted power, 90% power is effectively radiated by the antenna, and 10% power is reflected. Hence, if $S_{11} < -10$ dB, the reflected power by the antenna is less than 10% and transmitted power is greater than 90%. The reflection coefficient is based on the antenna terminal impedance (Z_t) and characteristic impedance (Z_0) and is shown in Eq. (1). The relation between VSWR and S_{11} is given by Eq. (2). If the above-said criterion is met for the frequency range between 3.1 and 10.6 GHz, the antenna is accurately operating at the UWB band.

$$S_{11} = \frac{Z_t - Z_0}{Z_t + Z_0} \tag{1}$$

$$VSWR = \frac{1 + S_{11}}{1 - S_{11}} \tag{2}$$

As the width of the feed line (W_s) is varied, there is a vital change in the operating frequency range. Figure 3 shows the reflection coefficient (S_{11}) characteristics of the lotus-shaped UWB antenna for the width $W_s = 4$ mm. From Fig. 3, the operating band of the designed antenna is from 0.45 to 14.32 GHz and the obtained impedance bandwidth is 187.8%. Figure 4 shows the VSWR characteristics plot.

Microstrip feed is provided to the designed antenna through a feedline of width (W_s) 4 mm and length (L_s) 50 mm. Figure 5 shows the current flow through the

Fig. 3 Reflection coefficient characteristics for $W_s = 4$ mm

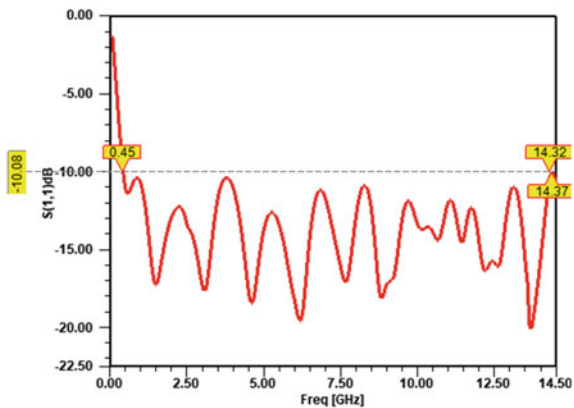


Fig. 4 VSWR characteristics for $W_s = 4$ mm

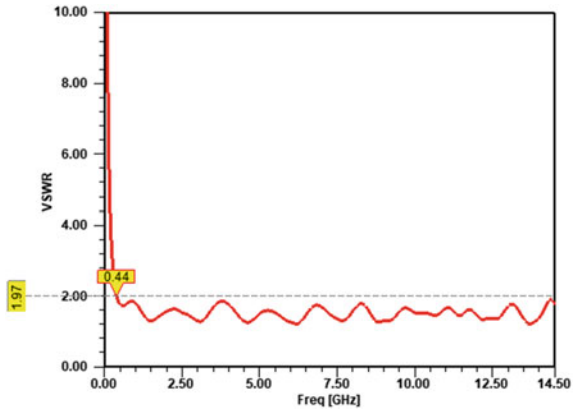
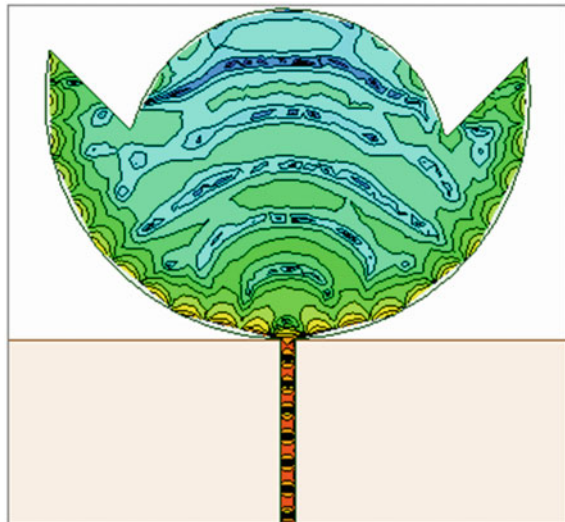


Fig. 5 Current distribution



radiating patch as the feed is provided. Figure 6 depicts the far-field radiation pattern of the lotus-shaped UWB antenna at $\phi = 90^\circ$ and $\phi = 0^\circ$. When the power is fed to the antenna more than 90% of the input power is effectively radiated and the direction of radiation is shown in Fig. 6.

The shape and area of the antenna determine the directivity of the antenna for the entire frequency of operation. Gain (G), efficiency (η) and directivity (D) are closely related to each other and are shown in Eq. (3). The figure of merit of the antenna is represented by the gain of the antenna. Figs. 7 and 8 shows the gain and efficiency of the antenna, respectively, for the operating UWB frequency range.

$$\eta = \frac{G}{D} \tag{3}$$

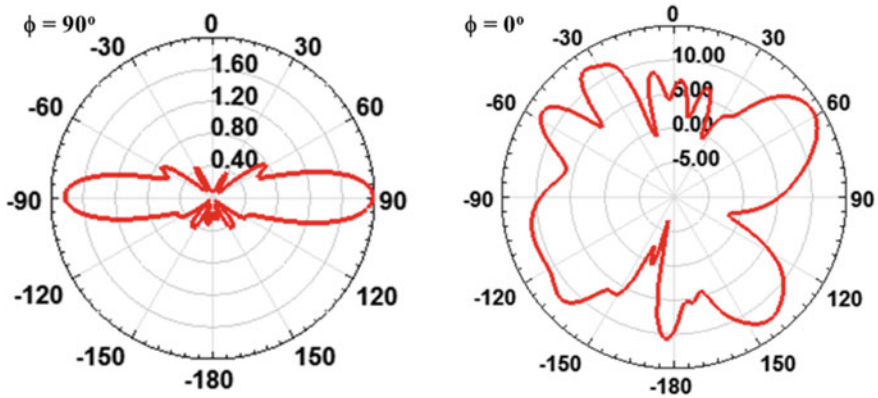
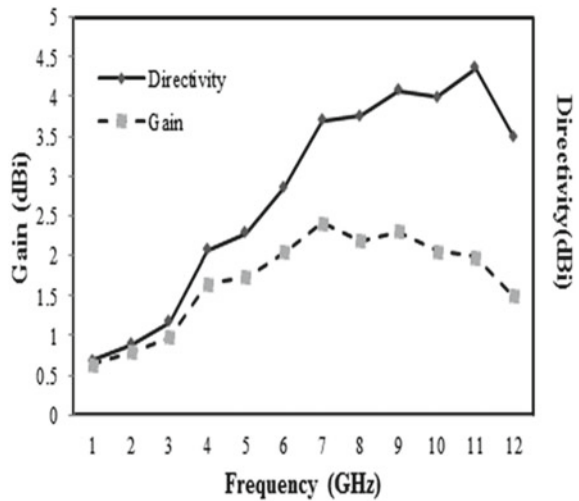


Fig. 6 Radiation pattern at $\phi = 90^\circ$ and $\phi = 0^\circ$

Fig. 7 Lotus-shaped UWB antenna gain versus frequency



The same UWB antenna can be used to operate in multiple narrow bands if the feed width (W_s) is increased to 6 mm. In eight narrow bands, the antenna can operate with a good reflection coefficient. The narrow bands are working in different communication applications mentioned in Table 1. Figure 9 shows the reflection coefficient characteristics for $W_s = 6$ mm.

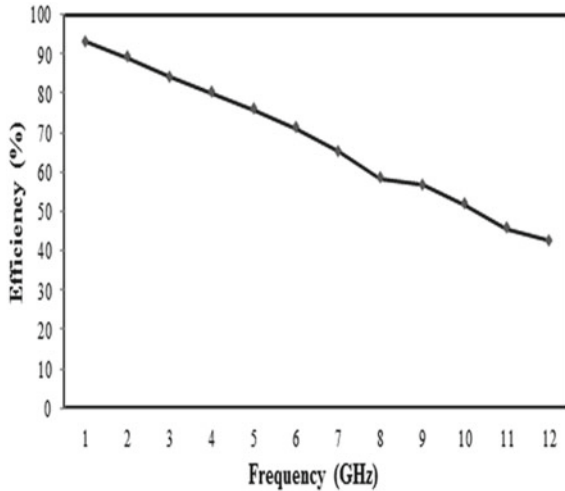


Fig. 8 Lotus-shaped UWB antenna efficiency versus frequency

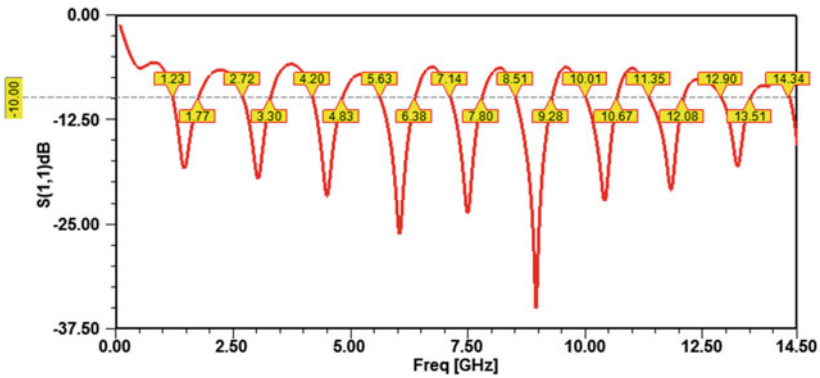


Fig. 9 S₁₁ versus frequency for width W_s = 6 mm

4 Conclusion

In this paper, a simple lotus-shaped ultra-wide band antenna operating at the frequency range of 0.45–14.32 GHz is proposed. The antenna is imprinted on a low-cost, low-profile FR4 substrate. The performance of the Lotus-shaped UWB antenna is validated by showing the reflection coefficient characteristics, VSWR characteristics, current distribution, radiation pattern, gain, directivity, and efficiency. For the entire operating impedance bandwidth of 187.8%, S₁₁ is less than –10 dB, efficiency is greater than 50%, and gain is greater than 1 dBi. The proposed antenna can be used for UWB commercial communication applications.

References

1. I.T.U. (ITU-R). (2014). Fixed service use and future trends (F.2323-0).
2. Li, Y., Li, W., Ye, Q., & Mittra, R. (2014). A survey of planar ultra-wideband antenna designs and their applications. *Forum for Electromagnetic Research Methods and Application Technologies (FERMAT), 1*, 1–16.
3. Porcino, D., & Hirt, W. (2003). Ultra-wideband radio technology: Potential and challenges ahead. *IEEE Communications Magazine, 41*(7), 66–74.
4. Ibrahim, A.A. (2019). Compact planer UWB antenna for high speed communications. In *2019 International Conference on Innovative Trends in Computer Engineering (ITCE)* (pp. 266–269).
5. Durbhakula, M.K.C., & Nalam, V.K.R. (2017). Wide band patch antenna structures for cognitive radio applications. In *2017 IEEE 7th International Advance Computing Conference (IACC)* (pp. 434–439).
6. Atallah, H.A., Abdel-Rahman, A.B., Yoshitomi, K., & Pokharel, R.K. (2014). Novel compact UWB monopole RDRA for cognitive radio spectrum sensing applications. In *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)* (pp. 462–466).
7. Bharathi, A., & Pandharipande, V.M. (2016). Design and development of antenna with U-shaped tuning stub for UWB applications. In *2016 IEEE Annual India Conference (INDICON)* (pp. 1–4).
8. Alsath, M. G. N., & Kanagasabai, M. (2015). Compact UWB monopole antenna for automotive communications. *IEEE Transactions on Antennas and Propagation, 63*(9), 4204–4208.
9. Dalal, P., & Dhull, S. K. (2020). Upper WLAN band notched UWB monopole antenna using compact two via slot electromagnetic band gap structure. *Progress in Electromagnetics Research, 100*, 161–171.
10. Yalduz, H., Koç, B., Kuzu, L., & Turkmen, M. (2019). An ultra-wide band low-SAR flexible metasurface-enabled antenna for WBAN applications. *Applied Physics A, 125*(9), 609.
11. Naktong, W., & Ruengwaree, A. (2020). Four-port rectangular monopole antenna for UWB-MIMO applications. *Progress in Electromagnetics Research, 87*, 19–38.
12. Ibrahim, M.A., Swelam, W., Abd El-Azeem, M.H., & El Hennawy, H. (2019). Ultra-wide band microstrip antenna for 4G applications. In *IOP Conference Series: Materials Science and Engineering* (Vol. 610, No. 1, p. 012025). IOP Publishing.
13. Florencio, R., Boix, R. R., Carrasco, E., Encinar, J. A., Barba, M., & Pérez-Palomino, G. (2014). Broadband reflectarrays made of cells with three coplanar parallel dipoles. *Microwave and Optical Technology Letters, 56*(3), 748–753.
14. Jang, J. W., & Hwang, H. Y. (2009). An improved band-rejection UWB antenna with resonant patches and a slot. *IEEE Antennas and Wireless Propagation Letters, 8*, 299–302.
15. Guo, Z., Tian, H., Wang, X., Luo, Q., & Ji, Y. (2013). Bandwidth enhancement of monopole UWB antenna with new slots and EBG structures. *IEEE Antennas and Wireless Propagation Letters, 12*, 1550–1553.

8-Bit Carry Look Ahead Adder Using MGDI Technique



P. Ashok Babu, V. Siva Nagaraju, and Rajeev Ratna Vallabhuni

Abstract High-performance and low power consumption are major factors that describe the significance of a design in VLSI. At low and ultra-low power applications, power consumption and logic delays are a fundamental problem. Nowadays, higher performance designs are built on the concept of computation units like ALU, where adders and multipliers are the essential components. To optimize low-power and high-performance applications, adders and multipliers need to be engineered to meet specifications such as speed and power dissipation. Using Modified Gate Diffusion Technique, this paper proposes a carry look-ahead adder (CLA). Compared with other adder models, the carry look ahead adder has much less propagation delay. The proposed adder is implemented on 90 nm technologies using the cadence environment. This paper's main objective is to use Modified Gate Diffusion Input (MGDI) Technology to develop and implement a CLA.

Keywords ALU · EDP · MGDI · Carry look ahead adder · PDP

1 Introduction

Nowadays, in our day-to-day life, we are using many electronic devices of low power and high-performance VLSI circuits. Due to modernization, there have been improvements in designs, and the need to make various transistors in one single chip has increased. This also affected the increase in operating frequency. Even the size of the construction chip is gradually reduced. As a result of all these factors, it has caused an increase in energy dissipation [1–5].

P. A. Babu (✉) · V. S. Nagaraju
Institute of Aeronautical Engineering, Dundigal, Hyderabad 500043, India
e-mail: p.ashokbabu@iare.ac.in

V. S. Nagaraju
e-mail: v.sivanagaraju@iare.ac.in

R. R. Vallabhuni
Bayview Asset Management, LLC, Coral Gables, FL 33146, USA
e-mail: rajeevratna@ieee.org

Carry look-ahead adder is a fast and resilient among various adders with very less propagation delay and low power dissipation. The main objective of this design is to reduce the propagation delay. Other adders more than 1-bit are designed using serial connection of 1-bit full adders. When we compare it with the existing adder designs the output carry of 1-bit adder is given as input carry to the next adder and so on [6–9].

2 Literature Review

P. Balasubramanian and J. John have proposed a low-power design of 2-input *OR* gate and 2-input *XOR* gate in MGDI technology, which are to be used to design the various complex circuits like arithmetic and logical circuits. The focus of this paper is on power efficiency. They created the logic gates using the MGDI to obtain a low power than the existing gate diffusion input technology (GDI). They have worked on the 0.35 μm CMOS technology and compared the results with the static CMOS logic. They have calculated power consumption, propagation delay, transistor count and have done the area comparison and the PDP comparison for different schemes. Power efficiency is a significant problem in designing digital devices, and high power dissipation can lead to device failure. Hence to overcome those issues, the design of the systems is developed at the transistor level [10–15].

The source station of the PMOS is always connected towards the VDD in the MGDI cell, and the NMOS remains forever connected towards the ground to reduce the consumption of electricity. The 2-input or and *XOR* gate design is carried out in design architect on the Linux platform [3]. S. Sapana and A. Smitha (2015) have proposed low power NAND and NOR gates using 16 and 32 nm MGDI Technology. The designed logic gate performance is compared Conventional Metal Oxide Semiconductor (CMOS) technology. In this paper, the dynamic and static power remain calculated using the tanner EDA tool and showed that technology requires less power for electronic devices' applications. This paper also showed different design considerations for various logic gates and observed the static, dynamic power, and the area of the 16 and 32 nm Technology [16–21].

Ebrahim Abiri et al. (2014) have proposed Logic Gates' design using the 32 nm CNTFET technology in Modified Gate Diffusion Input (MGDI) Technology. In this paper, the design of the MGDI cell is carried out using a nano process. It has shown that the area of the networks has reduced to 80% for the pull up section and up to 50% for the pull down section when compared to the GDI technology. In this paper, the calculations of the power dissipation, delay, and PDP are carried out using the H-SPICE software at 0.9 V supply. It proved that these transistors have less power dissipation than that of the conventional Si-MOSFET, and the area of the networks also decreased drastically. The PDP of the multiplexer is very low. The designs of the GDI, conventional CMOS and the MGDI technology are compared. The power dissipation, propagation delay, and the PDP of the logic gates and the multiplexer

are calculated and showed that the MGDI cell best solution of the VLSI designers [5, 22].

2.1 Working of CLA

CLA is fast and resilient among various adders with significantly less propagation delay and low power dissipation. The addition of two parallel binary numbers means that all levels and additive effects are obtained simultaneously. The signal must propagate through the gates as with any merging period until the output edges' correct output value is obtained. The total stream time, periods the amount of gate stages now the round, is proportional to the standard gate propagation delay. The most prolonged delay in additions is the time it takes to refine complete add-ons [23, 24]. As the total output bit depends on the number of input entries, it is only after the entries have been added to that category that the S_i sum in any additional category in the addendum will be at its final provincial value [25].

Using the subscriptions, the input and output variables define a standard add-on group. Signs on the P_i and G_i after spreading through their gates are still up to their influential state principles. These two characteristics remain corporate towards altogether half-adders also are based exclusively on adding and adding parts to the auction. At the exit, the signal after the input carries the C_i carrying the C_{i+1} propagating finished the OR gate and the OR gate, producing two stages of the gate [26].

The carrier transmission period remains a significant feature of the additive since it decreases the rapidity when two digits remain added. Though the add-on-or, designed aimed at that matter, somewhat integration circuit-will still consume approximately value in its output workstations, unless signals remain assumed extended period towards transmit complete the output gateways, the results will not be nice. Because with consecutive addition, all other arithmetic operations are used, the time spent during the process is significant. Using a faster gate with reduced delay is the apparent option for decreasing the carrier delay time [27]. Duration of a transmitted stream in the same add-on can be shortened in several respects. The principle of moving on is the most commonly used technique. If we define two new binary variables, consider the complete adder's circuit [28].

$$P_i = A_i \oplus B_i$$

$$G_i = A_i B_i.$$

The production sum and carry can remain conveyed as, respectively,

$$S_i = P_i \oplus C_i$$

$$C_{i+1} = G_i + P_i C_i.$$

G_i remains called a generate carry, and when both A_i and B_i are 1 regardless of the input carry C_i , it generates a carry of 1.

P_i is known as a carry propagate because it decides whether to propagate a carry into stage $i + 1$.

Boolean functions hold outputs for each step and replace the value of each C_i with the following equations:

$$\begin{aligned}
 C_1 &= G_0 + P_0 C_0 \\
 C_2 &= G_1 + P_1 C_1 = G_1 + P_1(G_0 + P_0 C_0) = G_1 + P_1 G_0 + P_1 P_0 C_0 \\
 C_3 &= G_2 + P_2 C_2 = G_2 + P_2 G_1 + P_2 P_1 G_0 + P_2 P_1 P_0 C_0 \\
 C_4 &= G_3 + P_3 C_3 = G_3 + P_3 G_2 + P_3 P_2 G_1 + P_3 P_2 P_1 G_0 + P_3 P_2 P_1 P_0 C_0 \\
 C_5 &= G_4 + P_4 C_4 = G_4 + P_4 G_3 + P_4 P_3 G_2 + P_4 P_3 P_2 G_1 + P_4 P_3 P_2 P_1 G_0 \\
 &\quad + P_4 P_3 P_2 P_1 P_0 C_0 \\
 C_6 &= G_5 + P_5 C_5 = G_5 + P_5 G_4 + P_5 P_4 G_3 + P_5 P_4 P_3 G_2 \\
 &\quad + P_5 P_4 P_3 P_2 G_1 + P_5 P_4 P_3 P_2 P_1 G_0 + P_5 P_4 P_3 P_2 P_1 P_0 C_0 \\
 C_7 &= G_7 + P_7 C_6 = G_7 + P_7 G_6 + P_7 P_6 G_5 + P_7 P_6 P_5 G_4 + P_7 P_6 P_5 P_4 G_3 \\
 &\quad + P_7 P_6 P_5 P_4 P_3 G_2 + P_7 P_6 P_5 P_4 P_3 P_2 G_1 + P_7 P_6 P_5 P_4 P_3 P_2 P_1 G_0 \\
 &\quad + P_7 P_6 P_5 P_4 P_3 P_2 P_1 P_0 C_0 \\
 C_8 &= G_8 + P_8 C_7 = G_8 + P_8 G_7 + P_8 P_7 G_6 + P_8 P_7 P_6 G_5 + P_8 P_7 P_6 P_5 G_4 \\
 &\quad + P_8 P_7 P_6 P_5 P_4 G_3 + P_8 P_7 P_6 P_5 P_4 P_3 G_2 + P_8 P_7 P_6 P_5 P_4 P_3 P_2 G_1 \\
 &\quad + P_8 P_7 P_6 P_5 P_4 P_3 P_2 P_1 G_0 + P_8 P_7 P_6 P_5 P_4 P_3 P_2 P_1 P_0 C_0.
 \end{aligned}$$

As the Boolean function is represented in the sum-of-product form for each output carrier, each purpose can be executed through one level of AND gates shadowed by an OR gate or a two-level NAND gate. The three Boolean occupations are implemented in the carry look ahead generator for C_1 , C_2 , and C_3 . The circuit can be introduced in less time as C_3 does not have to wait on behalf of propagation by C_2 and C_1 ; C_3 is also transmitted on the similar period as C_1 also C_2 . At the cost of extra complexity, this increase in operating speed is achieved.

In the designing of 8-bit CLA, an 8-bit input bit stream is required A_0 – A_7 and B_0 – B_7 and a carry input signal (C_{in}). In this 8- P is and 8- G is are generated as per the equations. The schematic is implemented by the 8-bit Carry Look Ahead Adder. Block diagrams of 4-bit CLA and 8-bit CLA are represented as the basic architecture is common in CMOS, GDI and MGDI technique; common block diagrams are used to represent the three methods that are to avoid the redundancy.

3 Methodology

3.1 Mod-GDI Technology

The simple equation of the improved gate input method is given by

$$D = P\bar{G} + NG \tag{1}$$

Due to its special configuration, power dissipation is becoming a major concern in complex VLSI circuits that can be overcome using the MGDI technique. The low voltage terminal [SP] is still associated with the V_{DD} in MGDI, and the high voltage terminal [SN] is always associated towards the ground, as shown in Fig. 1. The three terminals left as they are and can be used as input to the transistor. Thus, as the number of inputs has increased, the number of transistors required to perform a particular operation are decreased drastically [8, 9]. Hence, using the MGDI method, the number of transistors required is reduced to a greater extent, which reduces the power dissipation and propagation delay. So, it is the best suitable solution for the VLSI designers in this rapid growth of technology.

The various operations that can be performed using MGDI cell are shown in the Table 1 given above. All these operations are obtained using the Eq. (1).

Fig. 1 Basic MGDI cell

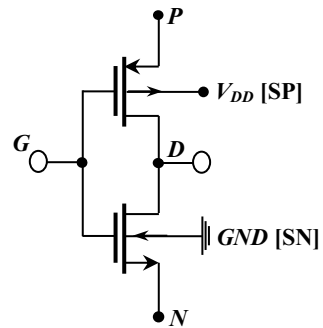
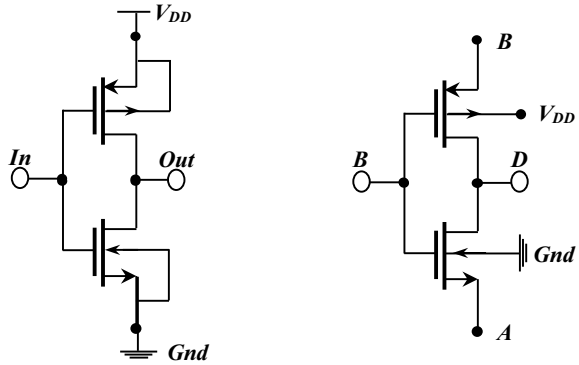


Table 1 Operations of MGDI cell

N	SN	P	SP	G	Result	Function
0	0	1	1	A	A'	Inverter
A	0	0	1	B	AB	AND
A	0	B	1	A	A + B	OR
B'	0	1	1	A	A'B'	NOR
A'	0	A	1	B	A'B + AB'	XOR
B	0	A	1	S	AS' + BS	2X1 MUX

Fig. 2 a Schematic of MGDI using FinFET based Inverter. **b** Schematic of MGDI based AND gate



3.2 Design of 8-bit CLA

The design of inverter gate using improved gate diffusion input method using FinFET is depicted in Fig. 2a. It consists of two transistors, one pMOS and one nMOS, which are connected in series where the nMOS source terminal is associated towards the ground also the drain of the pMOS is associated to the V_{DD} . When the input of logic 0 is applied, then the nMOS acts as an off-state connecting the output to the V_{DD} through pMOS, providing the output of logic 1. When the input is of logic 1 is applied, then the pMOS acts as an off state connecting the output to the ground through nMOS, providing the output of logic 0. The implementation is done using cadence virtuoso tool at 18 nm technology for 1 V as supply voltage, as shown in Fig. 2.

3.3 Design of AND Gate in MGDI Technique

The design of AND gate using modified gate diffusion input technique using FinFET is depicted in Fig. 2b. It consists of two transistors, one pMOS and one nMOS, which are connected in series where the nMOS source terminal is connected to the input signal A and the drain of the pMOS is connected to the input signal B. The pMOS' SP terminal is associated with the V_{DD} , and the SN terminal of the nMOS is related to the ground. The common gate terminal is connected to the input signal B. If any one of the input is of logic 0 then the nMOS acts on connecting the output to ground hence providing the output as logic 0. When both the inputs are of logic 1, then the pMOS connects the output to the V_{DD} hence providing the output as logic one. The implementation is done using cadence virtuoso tool at 18 nm technology for 1 V as supply voltage.

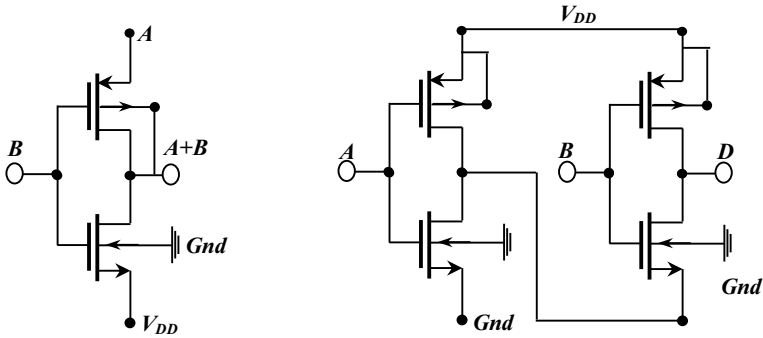


Fig. 3 a Schematic of MGDI based OR gate. b Schematic of MGDI based XOR gate

3.4 Design of OR Gate in MGDI Technique

The OR gate design using modified gate diffusion input technique using fin field-effect transistor is depicted in Fig. 3a. It consists of two transistors, one pMOS and one nMOS, linked in sequence, connecting the nMOS source terminal to V_{DD} and connecting the pMOS drain to the A input signal. The pMOS SP-terminal is associated with the output; the nMOS SN-terminal remains associated towards the ground. It is connected to the input signal B by the common gate terminal. The implementation is carried out as a supply voltage using cadence virtuoso tool at 18 nm technology for 1 V.

3.5 Design of XOR Gate in MGDI Technique

The inverter is used to get the invert of the input signal A that is provided to the nMOS source; also the drain station of the pMOS is associated to the input signal A. Figure 3b shows the XOR gate configuration. The common gate terminal is connected to the input signal B. The pMOS SP terminal is connected to the V_{DD} ; also the nMOS SN terminal is associated with the deck. Hence based upon the fundamental equation of the MGDI Eq. (1), the output is obtained as below.

$$D = (A.B') + (A'.B) \tag{2}$$

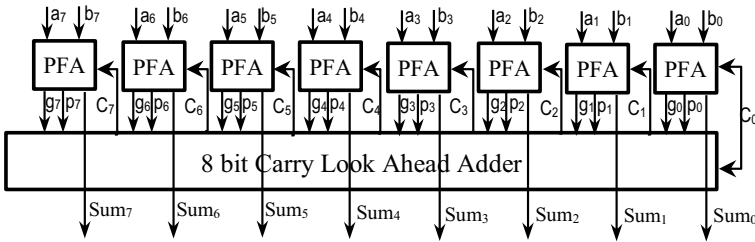


Fig. 4 Schematic of 8 bit carry look ahead adder

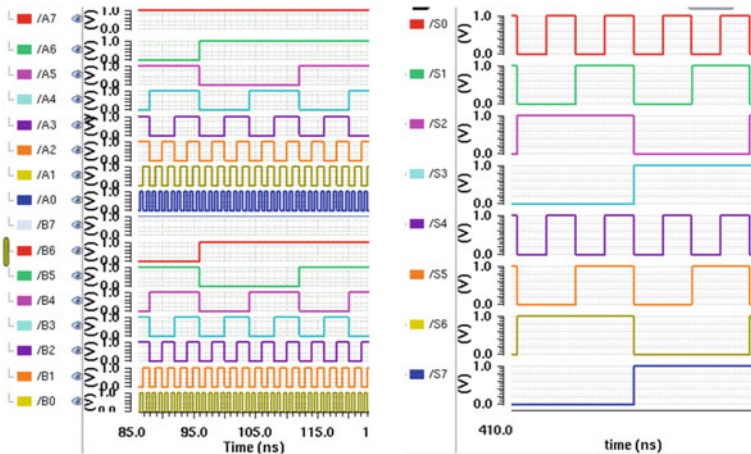


Fig. 5 a Inputs for 8 bit CLA. b Outputs of 8 bit CLA

3.6 Design of 8-bit CLA in MGDI Technique

Schematic of 8 bit CLA and internal circuit is presented in Fig. 4b. Inputs for 8 bit CLA are provided in Fig. 5a, b shows the outputs of 8 bit CLA.

4 Results and Discussions

This chapter tabulates the total power dissipation, propagation delay, and transistor count of each circuit implemented by the three techniques, i.e., CMOS, GDI, and MGDI technology. The calculations are performed directly using the cadence virtuoso method. Table 2 offers a comparison of the total power dissipation, propagation delay, and a number of circuit transistors using the FinFET with and without the Modified gate diffusion technology. We can observe from these tables that there is a drastic difference in the values of the circuits using different technologies; hence, we can say that MGDI is best solution for designing complex circuits with low power dissipation and propagation delay. The obtained results below prove that MOD-GDI is the efficient technique to implement the CLA with low power consumption.

Table 2 Comparison of 8 bit CLA using different technologies

Parameter	CMOS	GDI	MGDI
Power dissipation in nW	250×10^3	248.85	190.42
Propagation Delay in ns	155	95	72.35
Transistor count	294	246	152

Table 2 shows that the power dissipation of the circuits designed using MGDI technique is very low compared to that of the other technologies. This is mainly due to the reduction of the leakage power and the area of the circuits. Hence the power dissipation is significantly less in circuits using MGDI technique.

5 Conclusion

In this paper, the high-performance adder called Synchronous Carry Generate Adder/Carry look-ahead adder remains calculated by exhausting existing techniques (CMOS and GDI) and proposed MGDI technique. The functionalities of MGDI based designs have been verified. It is observed from Table 2 that MGDI based designs offer less power dissipation, high speed, and require less no. of transistors. Hence, it is concluded that MGDI based adders can be used in high-performance applications and can be used in the multipliers. The focus was on reducing the size of the transistor and speed, which later increased the problem of power intemperance in the circuits. It remains becoming the major problem as the device's lifetime and performance are getting significantly affected. The future work can be implemented using the Tunnel FET (TFET), which is used to decrease the power dissipation further. The implementation of the CLA can be done using the TFET.

References

1. Manchala, S., & Vallabhuni, V. (2020). A unique approach to provide security for women by using smart device. *European Journal of Molecular and Clinical Medicine*, 7(1), 3669–3683.
2. Siva, N.V., et al. (2020). Design and implementation of low power 32-bit comparator. In *ICICNIS 2020*, 10–11 Dec 2010 (pp. 1–8). Palai, India.
3. Rani, B.M.S, et al. (2020). *Retinal vascular disease detection from retinal fundus images using machine learning*. Australia patent 2020101450.
4. Venkateswarlu, S. C., et al. (2019). Implementation of area optimized low power multiplication and accumulation. *International Journal of Innovative Technology and Exploring Engineering*, 9(9), 2278–3075.
5. Vallabhuni, R.R., et al. (2020). High speed energy efficient multiplier using 20nm FinFET technology. In *Proceedings of the International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS 2020)*, 10–11 Dec 2020 (pp. 1–8). Palai, India.
6. Rajeev, R.V., et al. (2020). Comparative validation of SRAM cells designed using 18nm FinFET for memory storing applications. In *Proceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics and Cloud in Computational Vision and Bio-Engineering* (pp. 1–10).

7. Vallabhuni, V., & Avireni, S. (2019). A low power waveform generator using DCCII with grounded capacitor. *International Journal of Public Sector Performance Management*, 5, 134–145.
8. Vallabhuni, V., et al. (2019). System and method to improve performance of amplifiers using bias current. The Patent Office Journal No. 43/2019, India. International classification: C12Q1/6869. Application No. 201941042648 A.
9. Babu, P.A., et al. (2021). Realization of 8×4 barrel shifter with 4-bit binary to gray converter using FinFET for low power digital applications. *Journal of Physics: Conference Series*, 1714(1). IOP Publishing.
10. Rajeev, R.V., et al. (2020). An advanced computing architecture for binary to thermometer decoder using 18nm FinFET. In *2020 Third International Conference Smart Systems and Inventive Technology (ICSSIT)*, 20–22 Aug 2020 (pp. 510–515). Tirunelveli, India.
11. Vallabhuni, V., et al. (2019). A simple and enhanced low-light image enhancement process using effective illumination mapping approach. *Lecture Notes in Computational Vision and Biomechanics*, Cham, Switzerland (pp. 975–984).
12. Vallabhuni, V. (2017). *Second generation differential current conveyor (DCCII) and its applications*. Vignana's Foun. Sci., Tech. & Res. (Deemed to be University).
13. Rajeev, R.V., et al. (2020). Performance analysis: D-Latch modules designed using 18nm FinFET Technology. In *2020 International Conference on Smart Electronics and Communication (ICOSEC), September 10–12* (pp. 1171–1176). Tholurpatti, India.
14. Vallabhuni, V., & Avireni, S. (2017). A novel square wave generator using second generation differential current conveyor. *Arabian Journal for Science and Engineering*, 42(12), 4983–4990.
15. Vijay, V., et al. (2012). Performance evaluation of the CMOS full adders in TDK 90 nm technology. *International Journal of Systems, Algorithms and Applications*, 2(1), 711.
16. Vallabhuni, V., et al. (2020). High performance 2:1, 4:1 and 8:1 binary and ternary multiplexer realization using CNTFET technology. *Journal of Critical Reviews*, 7(6), 1159–1163.
17. Ratna, V.R., et al. (2021). High speed energy efficient multiplier using 20nm FinFET technology. In *ICICNIS 2020*. Retrieved from SSRN <https://ssrn.com/abstract=3769235>, <https://doi.org/10.2139/ssrn.3769235>
18. Rajeev, R.V., et al. (2021). Universal shift register designed at low supply voltages in 20nm FinFET using multiplexer. *Lecture Notes in Networks and Systems*.
19. Siva, N.V., et al. (2020). Design and implementation of low power 32-bit comparator. In *ICICNIS 2020*. Retrieved from SSRN <https://ssrn.com/abstract=3769748>
20. Saritha, P., et al. (2020). 4-Bit Vedic multiplier with 18nm FinFET technology. In *2020 International conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1079–1084). Coimbatore, India.
21. Vijay, V., et al. (2012). A review of the 0.09 μm standard full adders. *International Journal of VLSI Design & Communication Systems*, 3(3), 119.
22. Shaker, P.C., et al. (2020). Realization and comparative analysis of thermometer code based 4-bit encoder using 18 nm FinFET technology for analog to digital converters. *Advanced Intelligent Systems and Computing (AISC)*.
23. Mohammad, K., et al. (2020). Design of carry select adder based on a compact carry look ahead unit using 18 nm FinFET technology. *Journal of Critical Reviews*, 7(6), 1164–1171.
24. Rajeev, R.V., et al. (2020). Design of comparator using 18 nm FinFET technology for analog to digital converters. In *2020 7th International Conference on Smart Structures and Systems (ICSSS)* 23–24 July 2020 (pp. 318–323). Chennai, India.
25. Rajeev, R.V., et al. (2020). Smart cart shopping system with an RFID interface for human assistance. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 4–5 Dec 2020 (pp. 165–169). Palladam, India.
26. Vallabhuni, V., et al. (2019). System for minimizing crosstalk effects of shells and designing multiwalled carbon nanotube models. The Patent Office Journal No. 43/2019, India. Int. classification: B82Y10/00. Application No. 201941042460 A.

27. Rajeev, R.V., et al. (2020). Comparative analysis of 8-bit manchester carry chain adder using FinFET at 18 nm technology. In *2020 3rd International Conference Intelligent Sustainability Systems (ICISS)*, 4–5 Dec 2020 (pp. 1579–1583). Palladam, India.
28. Rajeev R.V., et al. (2020). Transistor SRAM cell designed using 18 nm FinFET technology. In *2020 3rd International Conference Intelligent Sustainability System* (pp. 1584–1589). Palladam, India.

Improved Scientific Workflow Scheduling Algorithm with Distributed Heft Ranking and TBW Scheduling Method



Ramandeep Sandhu and Kamlesh Lakhwani

Abstract Scheduling is a process that manages the workflow tasks during execution on different resources. Virtual infrastructure is a dynamic mapping of system resources to applications in order to maximize its utilization. In today's technological world, cloud has taken a long stride on the success towards maximum throughput as well as highest qualitative services to its consumers. Yet, approaches for maximizing the utilization of cloud resources are at peak demand. Each cloud service provider focuses on maximum utilization with minimum consumption of cloud resources, although managing and providing computational resources to maximum number of users and to execute such huge applications is a challenging one. In this paper, a scheduling algorithm with name TBW (Tabu Bayesian Whale Optimization) has been proposed. Basically, the algorithm is used to target the improvement in scheduling of scientific workflows. The complete framework has firstly used a ranking algorithm named distributed HEFT ranking and then applied TBW algorithm on ranked tasks of input workflows. The work has been executed for five scientific workflows LIGO, MONTAGE, Epigenomics, SIPHT and Cybershake. TBW is using tabu method on workflow tasks for fast local search in cloud system, and Bayesian Optimization is used to find out best possible combinations of resources where tasks are mapped and then whale optimization maps the tasks on the resources in a smart way. In the whole process, total execution time and cost parameters are minimized under deadline constraints.

Keywords Cloud VMs · Task mapping · Tabu search · Bayesian optimization · Whale optimization

R. Sandhu (✉) · K. Lakhwani
Lovely Professional University, Phagwara, India

1 Introduction

As cloud is a new trend with myriad satisfactory features for consumer as well as for provider, the best is on demand service with pay-as-per-use. Workflow scheduling under deadline constraints is like a fast medium for easy and efficient management of cloud environment with big data [1]. Moreover, parameters like time, cost, response time and energy (TCRE) are highly constrained to implement a cloud in capable as well as proficient way. Cloud computing provides features like elasticity, most favorable resource consumption, cost lessening and better-quality flexibility. Cloud computing is used factually everywhere in the world like big data analysis, file storage, backup, development of projects, testing, etc. [2–4]. Effective resource utilization makes it a prior requirement in industrial world. Albeit, the world today is fast-moving towards success in each and every era, yet without cloud, the technological world is not near satisfaction. Almost each business is feeling smooth and flexible execution in cloud environment. On the other side of cloud users, cloud-based services are if less complex and with genuine expenditure then each attached user will become a pleased member of cloud. So, visibility is an important fact [5].

To attain it, convex optimization of cloud is highly supportive. It will automatically thrust high the delivery on time as well as best utilization of resources. Furthermore, task mapping across means movement of data across machines. It should be complete secure, with zero response time as well as time and cost in movement must be negligible. At each glance, consistency is major concern [6].

Disregards confronted by researchers in Scientific Applications

Researchers in their studies have given many openings that the cloud has brought in, such as superior consumption of resources, improved responsiveness thereby improving user practice, enabling a generation of collaborative scientific workflows and reducing the cost in challenges and opportunities in running scientific workflows on cloud. The challenges faced by the applications are architectural challenges, service challenges, challenge in compute intensive applications, challenge for data management, and service management challenge. Even, Qi Zhang et al. [7] have discussed a variety of cloud computing skills and commercial products in detail. The profitable products have been evaluated on parameters like cloud provider, computing classes, target application, computation, auto scaling and storage. The research challenges expressed in this work are managing energy and providing. Furthermore, as noticeable concern of high cloud efficiency, Christian Vecchiola et al. [8] have given a judgment of computing enlightenment such as Amazon EC2, Google App Engine, and Microsoft Azure. This evaluation is completed on the basis of parameters like type of tune, value added provider, if PaaS, ability to organize on third party IaaS, platform, Virtualization, consumption Model and interface for user access. Even, the aneka architecture, deployment model and application model are discussed in detail. Programming models like task model, map reduce model, thread model, parameter sweep model, and workflow have been compared on the basis of execution services,

applications and execution unit. Apart from this, Suraj Pandey et al. [9] have worked on reducing the computing cost of the function by using particle swarm optimization (PSO) algorithm. Researchers in [10] have discussed dependency of cost on execution models in this work. In this work, the cost is calculated as a function of number of processors. The cost to execute montage workflow has been anticipated by running simulation using GridSim tool. The cost of running each of the data management models have been compared graphically. The Whale Optimization Algorithm (WOA) is a recently developed meta-heuristic swarm-based algorithm based on the bubble-net hunting manoeuvre technique of humpback whales to solve the complex problems of optimization [11].

2 Literature Review

Alkhanak et al. [12] proposed a cost optimization approach for scientific workflow scheduling in cloud computing. The proposed approach employs the four meta-heuristic algorithms which are based on the population. The approach helps in reducing cost and time of the service providers. The execution cost and time are reduced as compared to baseline approaches (Rimal et al. [13]). The result of the proposed work is compared with existing approaches and algorithms. The recreation result of the proposed approach shows more effective results than the existing approaches. Casas and Israel et al. [14] focused on Parallelization and proposed a scheduling approach called Balanced and file Reuse–Replication scheduling. This approach is used to schedule the scientific application workflows. It splits the workflows into sub-workflows which help in proper utilization via parallelization process. This approach provides the facility if data reuse and replication which helps in optimization of data and transfer it at run time. The optimization process is based on execution time and monetary cost of workflows. The work in [15–19] reduced time and cost and introduced a gravitational search algorithm for workflow scheduling in the cloud environment. The optimizations in workflow reduce the cost and make span. Two algorithms are hybridized GSA and HEFT for workflow scheduling. The performance evaluation is done on the basis of two metrics that are monetary cost ratio and schedule length ratio. The justification of result is also tested by ANOVA test, and it shows that the proposed approach does better. Sagnika et al. [20] proposed BAT algorithm for workflow scheduling in cloud computing which helps to handle the large size of data. The scheduling process decides that which task is executed first and which is last according to their requirement of the resources. It manages the resources according to the task size and execution time. The result of the proposed algorithm is compared with particle swarm optimization algorithm and Cat swarm optimization algorithm. The convergence of current study is better than the existing algorithms. Thiago Genez et al. [21] have worked with selection of CPU frequency configuration for resources carefully to reduce the total make span. It is combination of PSO and HEFT schedulers to make it better in case of time. A fitness function without any parameter value is imposed to measure performance of various particles.

Less is fitness value, better is solution. So, particles of swarm are moved towards less fitness value region the workflows with different sizes are accepted for simulation. Cybershake, SIPHT and LIGO are the workflows on which experiments are performed.

3 Proposed Ideology

In today's advance world of technology, cloud and its applications have provided various techniques. In the literature survey, many papers have been studied related to workflow scheduling in which cloud environment depends on the static configuration of virtual machine, which is not a real condition. Rather than optimizing and searching virtual machine utilization criteria, it is formulated that scanning and maximizing migration of tasks among VMs should be major concern. Also, distribution of tasks should not be random. Even, optimization initialization in a mannered way is more important which will enhance the resources utilization. Although, both consumers and providers of cloud are highly linked comfortably through the best way utilization of cloud resources yet the issue of underutilized virtual machines (called cloud resources) cannot be ignored. By considering it, few limitations have been mentioned below which are basic pillar for proposed study. Random distribution of tasks [22] has been done in existing works. Also, optimization of the task depends on single objective, which is sometimes conflict like time and cost [23]. Furthermore, only workflow dependency is main aspect on which deadlines are totally dependent. In existing approaches, optimization use local VM or global data center [14, 24–36].

3.1 Proposed Representation

The TBW approach has been proposed for mappings tasks to VMs in cloud. As depicted in above Fig. 1, task scheduling using Tabu Search, Bayesian Optimization and Whale Optimization algorithm has been done.

3.2 Proposed Flowchart and Phases of Proposed Methodology

Various phases of proposed methodology are listed below: Phase 1: Total five workflows are accepted as input named CYBERSHAKE, MONTAGE, LIGO, SIPHT and EPIGENOMICS.

Phase 2: Parsing is a step of analyzing input. In proposed cloud workflow framework, parsing will occur at initial stage. It will show tasks with dependencies.

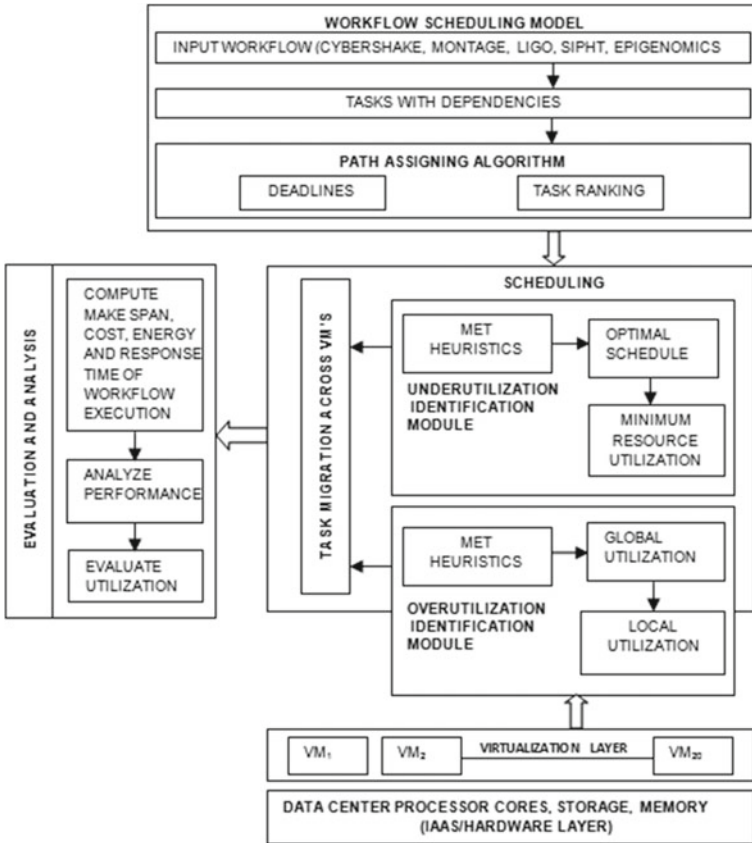


Fig. 1 Graphical representation of proposed methodology

Phase 3: A path assignment algorithm will be used for ranking various tasks of workflow. Here, deadlines are prior concern. Moreover, budget and heuristics for initializations are also important factors.

Phase 4: Scheduling is performed in two scenarios, and the outcome in both cases is mapping on Virtual Machines (VMs). It means tasks migration across VMs should be optimized so that under-utilized VMs can be scanned.

Phase 5: It is basically evaluation and analysis. After performing VM migration in optimized manner, Makespan, Cost, energy and response time of Workflow Execution will be calculated. So, these four parameters will help to analyze performance of proposed framework. In whole system, utilization of resources will be increased to much high level.

So, proposed framework improves the utilization of resources which reduce the cost and time of execution. Also, improves the virtual machine migration because of optimization schedule. Apart from this, it reduces the decision time of Virtual

machine mapping. The proposed study has inclined the research towards energy-efficient scheduling for optimum resource utilization in cloud environment by designing and deploying an effective task ranking algorithm. By creating correlation between time and deadline of input tasks, it has been evaluated that results are better than HEFT method. In HEFT, TET (Total execution Time) parameter of all input tasks is also calculated on cloudsim [25].

As shown in Tables 1 and 2, algorithms for finding critical path have been used.

In the proposed ideology, three algorithms are used: Use of Tabu Search, Use of Bayesian Optimization and Use of Whale Optimization.

Then all three steps have been combined to make it TBW Scheduling Approach.

Tabu Search: it takes Input: ranked tasks mapped to VMs. Then apply tabu search algorithm to find neighbours of current VM. It works iteratively. If neighbour is less utilized than current VM then adds such neighbour in tabu list. Result after applying tabu search—we get a final list of VMs which are not efficiently utilized.

After applying tabu search method, next step targets towards try to shift tasks from these VMs on other VMs (efficiently utilized) but without making Queue. Here

Table 1 Initialization algorithm (W)

```

Begin
Initialize number of fog nodes and workflows
N ← VM
W ← Number of workflows
Parse workflows
while (W)
start
Parse  $W_i$ 
Extract task
 $T_i \leftarrow \text{critical path}(W_i)$ 

End
 $X \leftarrow \sum_{i=1}^n T_i + W_i \dots \dots \dots (1)$ 
Return X
End
    
```

Table 2 Algorithm to find critical path (W_i)

```

Begin
Cp = ∅, taski = taskexit, Critical path = ∅
While(taski != taskentry) do
Calculate critical parent taskp of taski
Criticalparent(taski) =
{taskp | max taskp {finishtime+totaltime}} .....(2)
Cp = Cp U taskp
taski = taskp
End While
Return Cp
End
    
```

the meaning of queue is that increase in total execution time. So, use of Bayesian Optimization is helpful here. Basically Bayesian optimization works for choosing best combination for mapping tasks of not properly utilized VMs on other utilized VMs. It means, it targets tasks of the VMs which are not efficiently utilized and targets those VMs which are in the underutilized category. It provides all combinations where tasks can be shifted. The next part of the TBW algorithm is whale optimization. Whale works in an intelligent way. It takes input from Bayesian Optimization. Based on its objective functions, it chooses best combinations and shifts tasks on VMs efficiently without increasing TEC and TET. The overall results are better than GA-PSO for different scientific workflows.

4 Conclusion

In this paper, we have introduced our ranking algorithm named distributed HEFT ranking also presented a new approach for better scheduling in cloud system (named as TBW—Tabu Bayesian Whale Algorithm) included three optimization approaches (Tabu Optimization, Bayesian Optimization and Whale Optimization). As researchers are already using various techniques for cloud optimization and for minimizing the time and cost parameters, here, in this study we have provided the working methodology of TBW. Also study of various algorithms which have used the smart optimization techniques has been done. In further work, results with TBW advance approach will be calculated and also will be compared with existing optimization approaches like GA, PSO, GA-PSO and WOA.

References

1. Deelman, E., Singh, G., Livny, M., Berriman, B., & Good, J. (2008). The cost of doing science on the cloud: The montage example. In *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, Austin, TX* (pp. 1–12).
2. Bölöni, L., & Damla, T. (2017). Value of information based scheduling of cloud computing resources. *Future Generation Computer Systems, 71*, 212–220.
3. Chirkin, M., Adam, B., Sergey, K., Marc, M., Mikhail, M., Alexander, V., & Denis, A. (2017). Execution time estimation for workflow scheduling. *Future Generation Computer Systems, 75*, 376–387.
4. Zhang, L., Kenli, L., Changyun, L., & Keqin, L. (2017). Bi-objective workflow scheduling of the energy consumption and reliability in heterogeneous computing systems. *Information Sciences, 379*, 241–256.
5. Vöckler, J., Juve, G., Deelman, E., Rynge, M., & Berriman, B. (2011). Experiences using cloud computing for a scientific workflow application. In *Proceedings of the 2nd International Workshop on Scientific Cloud Computing* (pp. 1–10). ACM.
6. Wang, Y., Jiajia, J., Yunni, X., Quanwang, W., Xin, L., & Qingsheng, Z. (2018). A multi-stage dynamic game-theoretic approach for multi-workflow scheduling on heterogeneous virtual machines from multiple infrastructure-as-a-service clouds. *International Conference on Services Computing* (pp. 137–152). Cham: Springer.

7. Zhang, C., & Raouf, B. (2018). Cloud computing: State-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1), 7–18.
8. Vecchiola, C., Suraj, P., & Buyya, R. (2009). High-performance cloud computing: A view of scientific applications. In *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN)* (pp. 1–9). IEEE.
9. Pandey, S., Wu, L., Guru, S.M., & Buyya, R. (2010). A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In *2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA)* (pp. 1–10). IEEE.
10. Goyal, M., & Mehak, A. (2017). Optimize workflow scheduling using hybrid ant colony optimization (ACO) & particle swarm optimization (PSO) algorithm in cloud environment. *International Journal of Advance Research, Ideas and Innovations in Technology*, 3(2).
11. Alameer, Z., Elaziz, M., Ewees, A., Ye, H., & Jianhua, Z. (2019). Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resources Policy*, 61, 250–260.
12. Alkhanak, E., Lee, S., Rezaei, R., & Parizi, R. (2018). Cost optimization approaches for scientific workflow scheduling in cloud and grid computing: A review, classifications, and open issues. *Journal of Systems and Software*, 113, 1–26.
13. Israel, C., Javid, T., Rajiv, R., Lizhe, W., & Albert, Z. (2017). A balanced scheduler with data reuse and replication for scientific workflows in cloud computing systems. *Future Generation Computer Systems*, 74, 168–178.
14. Mao, M., & Humprey, M. (2011). Auto-scaling to minimize cost and meet application deadlines in cloud workflows. In *SC'11: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, Seattle, WA* (pp. 1–12).
15. Kapoor, N., & Kumar, Y. (2020). The efficient management of renewable energy resources for Vanet-cloud communication. In *Nature-Inspired Computing Applications in Advanced Communication Networks* (pp. 228–253).
16. Kumar, Y., & Kaul, S. (2019). Effective use of the machine learning approaches on different clouds. In *International Conference on Sustainable Computing in Science, Technology and Management* (pp. 892–897).
17. Kumar, Y., & Mahajan, M. (2019). Intelligent behavior of fog computing with IOT for healthcare system. *International Journal of Scientific and Technology Research*, 8, 674–679.
18. Liu, L., Zhang, M., Buyya, R., & Fan, Q. (2016). Deadline-constrained co evolutionary genetic algorithm for scientific workflow scheduling in cloud computing. *Concurrency and Computation Practice and Experience*, 1–9.
19. Masadari, M. (2017). Towards workflow scheduling in cloud computing: A comprehensive analysis. *Journal of Network and Computer Applications*, 1–27.
20. Sagnika, S., Saurabh, B., & Bhabani, S. (2018). Workflow scheduling in cloud computing environment using bat algorithm. In *Proceedings of First International Conference on Smart System, Innovations and Computing* (pp. 1–13). Springer, Singapore.
21. Vinothina, V., & Sridaran, R. (2018). An approach for workflow scheduling in cloud using ACO. In *Big data analytics* (pp. 525–531). Springer.
22. Genez, T., Pietri, I., Sakellariou, R., Bittencourt, F., & Madeira, E. (2015). A particle swarm optimization approach for workflow scheduling on cloud resources priced by CPU frequency. In *Data Science and Symptoms* (pp. 1–9).
23. Kumar, B., Mala, K., & Poonam, S. (2017). Discrete binary cat swarm optimization for scheduling workflow applications in cloud systems. In *2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT)* (pp. 1–6).
24. Verma, A., & Sakshi, K. (2010). A hybrid multi-objective particle swarm optimization for scientific workflow scheduling. *Parallel Computing*, 62, 1–19.
25. Sandhu, R., & Lakhwani, K. (2020). Optimal cloud system enhancement using improved workflow task ranking system. *Test Engineering and Management*, 83, 10092–10101.
26. Choudhary, A., Gupta, I., Singh, V., & Jana, P.K. (2018). A GSA based hybrid algorithm for bi-objective workflow scheduling in cloud computing. *Future Generation Computer Systems*, 1–10.

27. Dillon, T., Elizabeth, C., & Chen, W. (2010). Cloud computing: Issues and challenge. In *24th IEEE International Conference on Advanced Information Networking and Applications*.
28. Garg, J., & Gurjit, B. (2017). Research paper on genetic based workflow scheduling algorithm in cloud computing. *International Journal of Advanced Research in Computer Science*, 8(5), 1–7.
29. Ghose, M., Verma, P., Karmakar, S., & Sahu, A. (2017). Energy efficient scheduling of scientific workflows in cloud environment. In *2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Bangkok* (pp. 170–177).
30. Iosup, A., Ostermann, S., Yigitbasi, M., Prodan, R., Fahringer, T., & Epema, D. (2011). Performance analysis of cloud computing services for many-tasks scientific computing. *IEEE Transactions on Parallel and Distributed Systems*, 22(6), 931–945.
31. Jiang, J., Yaping, L., GuoqiXie, L., & Junfeng, Y. (2017). Time and energy optimization algorithms for the static scheduling of multiple workflows in heterogeneous computing system. *Journal of Grid Computing*, 15(4), 435–456.
32. Quang, H., Nguyen, S., & Nam, T. (2017). Energy-saving virtual machine scheduling in cloud computing with fixed interval constraints. In *Transactions on large-scale data-and knowledge-centered systems XXXI* (pp. 124–145). Springer.
33. Scott, C., Ewa, D., Dan, G., Gideon, J., Philip, M., Christopher, B., Karan, V., Kevin, M., Robert, G., Edward, F., David, O., & Thomas, J. (2010). Scaling up workflow-based applications. *Journal of Computer and System Sciences*, 76, 428–446.
34. Zhao, Y., et al. (2011). Opportunities and challenges in running scientific workflows on the cloud. In *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE.
35. Ahmed, M., Houssein, E., Hassanien, A., Taha, A., & Hassanien, E. (2019). Maximizing lifetime of large-scale wireless sensor networks using multi-objective whale optimization algorithm. *Telecommunication Systems*, 72(2), 243–259.
36. Ben Oualid Medani, K., Sayah, S., & Bekrar A. (2017). Whale optimization algorithm based optimal reactive power dispatch: A case study of the Algerian power system. *Electric Power Systems Research*, 163, 696–705.

Selection of OLAP Materialized Cube by Using a Fruit Fly Optimization (FFO) Approach: A Multidimensional Data Model



Anjana Yadav and Anand Tripathi

Abstract The Online Analytical Processing (OLAP) based multidimensional examination hassles for several stockpiling magnificence over huge data. For as much to recognize queries answering time companionable by OLAP framework users and understanding entire business perceive mandatory, OLAP data is structured as a data cube (a multidimensional model). The OLAP queries are responded in speedy and steady time by utilizing the cube materialization for assessments takers. But, this also involves unendurable expenses, regarding to stockpile memory and period, and as a data depot, OLAP has an average dimension and dimensionality which is to be significant on query processing. Consequently, cube assortment has got to be finished motivating to diminish inquiry management expenses, maintaining as a restraint the materializing gap. Several techniques and heuristics like deviationist and insatiable algorithms have been utilized to offer an estimated result. In this work, a Fruit Fly Optimization (FFO) approach is implemented in a lattice structure to obtain an optimal materialized data cube for reducing the query processing expenses. The results illustrate that FFO generates better performance than Particle Swarm Optimization (PSO) in terms of frequency and number of dimensions.

Keywords Cube materialization · Data cube · Fruit fly optimization · OLAP · Multidimensional model

1 Introduction

Data warehouse [1] is a collection of huge information of assorted data about several organizations utilizing for assessments. These assessments are taken by complex queries applied to data warehouse [2] to reduce the answer time. The materialization in data cubes is well utilized for query processing in an efficient manner with lesser time consuming. Entire views of data cubes can be materialized to access the data quickly with minimizing the answering time of queries. The Online Analytical Processing (OLAP) [3] is used for query processing over data warehouse and

A. Yadav (✉) · A. Tripathi
P. K. University, Shivpuri, India

extracting the useful information for analysis work. It is also helpful for assessment support to provide the users overviews. OLAP is combined with Key Performance Indicators (KPIs) [4, 5] to provide a dashboard application over historical data. The production efficiency is explained with numerical examples and developed in the Java programming language.

OLAP is also utilized for energy cost analysis in commercial areas to reduce the expenditure with improving the power efficiency [6]. A multidimensional cube model is utilized for evaluating the power consumption at several stages of generalization. This time OLAP [7] is used with association rules to generate feasible solutions for huge data about buildings in commercial sectors. Unified Modeling Language (UML) [8] and Structured Query Language (SQL) are also introduced for query processing in OLAP. The web based software is implemented for analyzing the students' results based on the object oriented methodology. The analytical comparison of huge student data is easily performed by OLAP to minimize the stress and workload of schools. Lectures are easily delivered to students through this object oriented platform [9]. Another application of OLAP is a police intellect, a decision scheme to catch the criminals and take an efficient decision about crime [10].

The analytical process is also utilized for intellectual study over multidimensional data for business purposes [11] in Social Business Intelligence (SBI). The social, private and public data can be analyzed easily using OLAP semantic analysis to model the huge information of companies for business point of views [12, 13]. The health care data [14] is also formulated and structured by mining in OLAP. The Internet of Medical Things (IoMT) systems are well suited for providing health information of patients and also predict the treatment of diseases by using medical data analysis [15]. The extraction of solar radiation data from huge amount of information is performed by query processing in several geographical positions and structures. The sensor data of this system accesses by user to initiating the query on solar data and analyzing the data at several stages [16].

In the above analysis of works, materialized data cubes are generated from important data extraction from the huge amount of data. These data cubes are selected by using optimization approaches in large space to provide efficient query processing in an OLAP multidimensional data model. The PSO is one of the nature inspired optimization approach initiating on data cubes to generate optimal materialized data cubes. Here, we proposed another optimization technique FFO, which generates better results than PSO over OLAP multidimensional model in terms of frequency and number of dimensions to reduce the query processing expenditure in several constraint spaces.

2 The Fruit Fly Optimization (FFO) Approach for Selection of OLAP Materialized Cube (A Multidimensional Data Model)

2.1 FFO Approach

A bio-inspired Fruit Fly Optimization (FFO) approach is utilized for generating the global optimal solutions on the basis of foraging inspired by fruit flies. Various realistic explanations to optimization quandary are illustrated by foraging of fruit flies in FFO. The fruit fly gaits by plunging to the food, exploits its vigilant spirit to realize food and where it correlates flock and then it gaits by plunging into a route.

Start

Step 1. Put primary standards of position, fitness function, smell, generation and population randomly of entire fruit flies (Eq. (1)).

$$\begin{aligned} X_{\emptyset} &= X_{axis_value} + random_value \\ Y_{\emptyset} &= Y_{axis_value} + random_value \end{aligned} \tag{1}$$

Step 2. Compute the fitness standards on the basis of distance (*Dist*) and smell (*Sml*) for entire fruit flies. Hence, optimized solutions are obtained with fitness of individual and population (Eqs. (2), (3) and (4)).

$$Dist = \sqrt{X_{\Phi}^2 + Y_{\Phi}^2} \tag{2}$$

$$Sml = \frac{1}{Dist} \tag{3}$$

$$Fitness_Function = Function(sml) \tag{4}$$

Step 3. Change the standards of best index and position for entire fruit flies (Eqs. (5) and (6)). Hence, update the fruit fly's position (Eqs. (7) and (8)).

$$\begin{aligned} BestIndex(t + 1) &= \mu \times BestIndex(t) + \eta_1 \times Rand \\ &\times (Dist - X_{\Phi}(t)) + \eta_2 \times Rand \times (Sml - X_{\Phi}(t)), \end{aligned} \tag{5}$$

where μ = indolence coefficient.

η_1 and η_2 = Constant.

$$[BestSmell \ BestIndex] = minimum(Sml) \tag{6}$$

$$X_{axis_value} = X(BestIndex) \tag{7}$$

$$X_{\Phi} = Round \left(X_{\Phi} \times \left(\frac{EntireNodes - EntireSourceNodes}{-EntireDestinationNodes} \right) \right) + EntireSourceNodes + 1. \tag{8}$$

Step 4. Establish optimized result, if not, go over 2.

If established the optimized result, found best node.

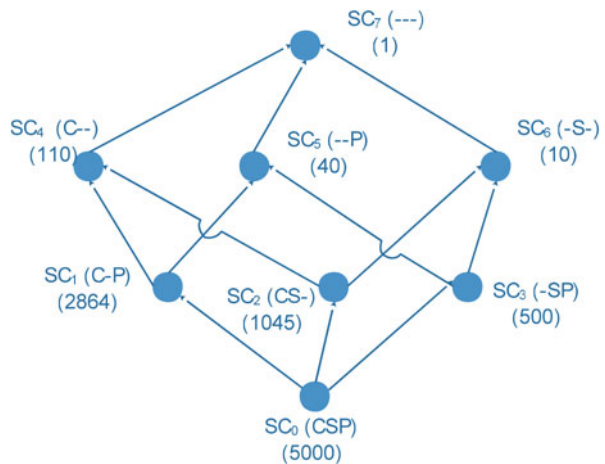
End

2.2 Lattice Structure

The lattice structure combines entire probable data cubes at dissimilar stages of concentricity by describing cubes on their reliance. Two cubes C_x and C_y are connected by a route for generating reliance association ($C_x \leq C_y$). This association illustrates that an answer to the OLAP query given by C_x , can also be provided by C_y . In lattice structure, 2^N data cubes are possible for N dimensions for an information association. Here, $2^3 = 8$ data cubes are obtained from three dimensional data “Sales” (Customer (C), Supplier (S), Part (P)) (Fig. 1).

The minimum concentricity represents by bottom cube (e.g., CSP) and maximum concentricity represented by the top cube. The structure illustrates that an answer

Fig. 1 Lattice structure



to the OLAP query given by child cube (*, S, *) can also be provided by any parent cubes (*, S, P), (C, S, *) or (C, S, P) by concluding data beside few dimensions.

2.3 Cube Selection Using FFO Approach

The objective of an OLAP model is to reduce expenditure of query and preservation with fulfilling a constriction like materialized space. The Entire Query Expenditure (EQE) for an OLAP query is evaluated by utilizing answering cost and frequency of query using Eq. (9).

$$EQE = \sum_{x=1}^N fq_x \times E(q_x, C_M) \quad (9)$$

Here,

fq_x = query q_x frequency.

$E(q_x, C_M)$ = query q_x answering expenditure (cost) with C_M (Materialized Cube) which is calculated by using Eq. (10).

$$E(q_x, C_M) = \text{Minimum}(|SC_x|, Lpre(q_x, C_M)) \quad (10)$$

Here,

$|SC_x|$ = Number of sub-cubes of query q_x .

$Lpre(q_x, C_M)$ = Least Predecessor (Lpre) conception of query q_x with C_M .

The Preservation Expenditure (PE) of a materialized cube $C_x \in C_M$ is evaluated in terms of least predecessor (Lpre) of C_x using Eq. (11).

$$PE(C_x, C_M) = |Lpre(C_x, C_M)| \quad (11)$$

After that Entire Preservation Expenditure (EPE) is evaluated by Eq. (12).

$$EPE = f_{PE} \sum_{C \in C_M} PE(C, C_M) \quad (12)$$

Here,

f_{PE} = frequency of inclusion in support association.

Hence, the Entire Cost Function (ECF) is calculated by combining the EQE (Entire Query Expenditure) and EPE (Entire Preservation Expenditure) using Eq. (13).

$$\text{Minimum} \quad ECF = \sum_{x=1}^N fq_x \times E(q_x, C_M) + f_{PE} \sum_{C \in C_M} PE(C, C_M) \quad (13)$$

The optimal value of ECF is obtained by applying FFO approach on this objective function ECF. The initial values of expenditure can be evaluated from pedestal association and the value of C_M is empty initially. After that optimal values are evaluated using FFO. The ECF function is utilized as variation function, as it discovers the solution fitness throughout the following of the several targets. The ECF function is minimized to obtain least expenditure of query with highest fitness.

3 Result and Analysis

The FFO and PSO approaches have implemented in MATLAB 2019a environment and analyzed in terms of frequency and number of dimensions. The results illustrate the best quality performance of FFO over PSO with minimum query processing expenditure. Here, we measured several belongings of space constrictions as 10, 20, 30, 40, 50 and 60% to analyze the FFO and PSO approaches.

We performed the implementation of FFO and PSO on the multiple dimensional data like three dimensions ((Customer (C), Supplier (S), Part (P)), and four dimensions (Customer (C), Supplier (S), Part (P), Time (T)) evaluate the results.

Figures 2 and 3 illustrate that the FFO generates better results for selecting optimal data cubes with minimum OLAP query processing expenditure as compared to PSO for all dimensional data.

We also evaluated the results of FFO and PSO in terms of identical and arbitrary frequencies (range 0–1).

Figures 4 and 5 illustrate that the FFO obtains more efficient performance for choosing optimal data cubes as compared to PSO in all frequencies with least OLAP query processing expenditure.

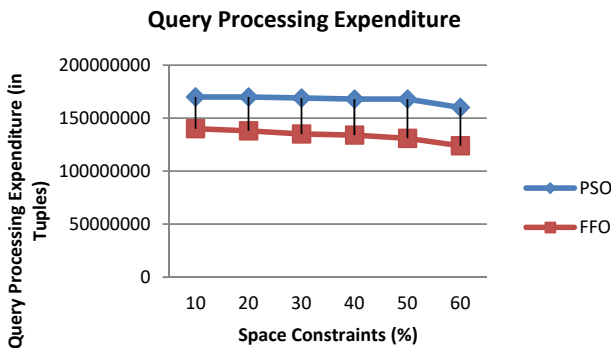


Fig. 2 Query processing expenditure for PSO and FFO approaches (3 Dimensions)

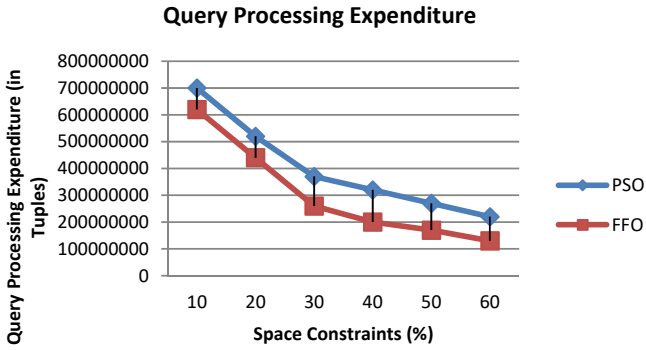


Fig. 3 Query processing expenditure for PSO and FFO approaches (4 Dimensions)

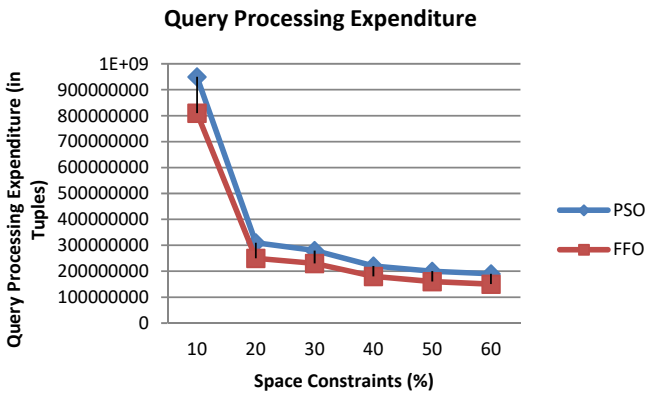


Fig. 4 Query processing expenditure for PSO and FFO approaches (Identical frequencies)

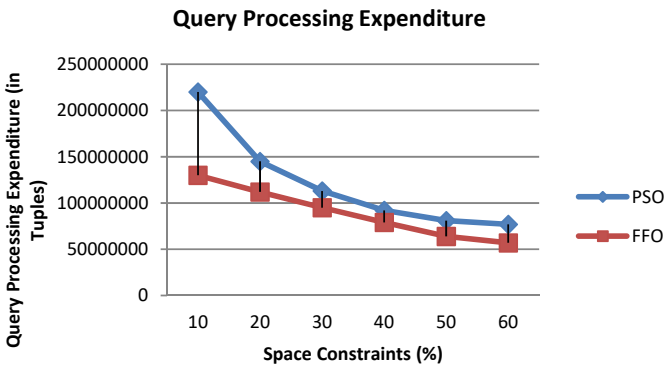


Fig. 5 Query processing expenditure for PSO and FFO approaches (Arbitrary frequencies)

4 Conclusion

In this work, FFO based optimal materialized cube selection is performed on lattice structure with multidimensional data over OLAP framework. The results are evaluated on multidimensional data in terms of frequency and number of dimensions. The analysis of performance of FFO illustrates the improved quality, efficiency of FFO to reduce the OLAP query processing expenditure as compared to PSO. In the future, several optimization approaches will be implemented for optimized cube selection by considering the time complexity as a factor.

References

1. Vaisman, A., & Zimanyi, E. (2019). Mobility data warehouses. *International Journal of Geo-Information, MDPI*, 8(170), 1–22.
2. Agapito, G., Zucco, C., & Cannataro, M. (2020). COVID-warehouse: A data warehouse of Italian COVID-19, pollution, and climate data. *International Journal of Environment Research and Public Health, MDPI*, 17(5596), 1–22.
3. Jukic, N., Jukic, B., & Malliaris, M. (2008). Online analytical processing (OLAP) for decision support (pp. 1–25).
4. Papacharalampopoulos, A., Giannoulis, C., Stavropoulos, P., & Mourtzis, D. (2020). A Digital twin for automated root-cause search of production alarms based on KPIs aggregated from IoT. *Applied Science, MDPI*, 10(2377), 1–16.
5. Stefanovic, N. (2014). Proactive supply chain performance management with predictive analytics. *The Scientific World Journal, Hindawi*, 1–18.
6. Noh, B., Son, J., Park, H., & Chang, S. (2017). In-depth analysis of energy efficiency related factors in commercial buildings using data cube and association rule mining. *Sustainability, MDPI*, 9(2119), 1–20.
7. Almeida, D. R. D., Baptista, C. D. S., Andrade, F. G. D., & Soares, A. (2020). A survey on big data for trajectory analytics. *International Journal of Geo-Information, MDPI*, 9(88), 1–24.
8. Ciferri, C., Ciferri, R., Gomez, L., Schneider, M., Vaisman, A., & Zimanyi, E. (2012). Cube algebra: A generic user-centric model and query language for OLAP cubes. *International Journal of Data Warehousing and Mining*, 1–23.
9. Emmanuel, N. E., Obiageli, J. A., & Osinachi, V. (2019). Design and implementation of multi-dimensional students result analytical processing of tertiary institutions. *International Journal of Engineering and Computer Science*, 8(8), 24814–24828.
10. Shen, L., Liu, S., Chen, S., & Wang, X. (2012). The application research of OLAP in police intelligence decision system. In *International Workshop on Information and Electronics Engineering (IWIEE)* (Vol. 29, pp. 1–6). Elsevier.
11. Venkatraman, S. (2017). A proposed business intelligent framework for recommender systems. *Informatics, MDPI*, 4(40), 1–12.
12. Cruz, I. L., Berlanga, R., & Aramburu, M. J. (2018). Modelling analytical streams for social business intelligence. *Informatics, MDPI*, 5(53), 1–17.
13. Fuertes, W., Reyes, F., Valladares, P., Tapia, F., Toulkeridis, T., & Perez, E. (2017). An integral model to provide reactive and proactive services in an academic CSIRT based on business intelligence. *Systems, MDPI*, 5(52), 1–20.
14. Qwaider, W. Q. (2012). Apply on-line analytical processing (OLAP) with data mining for clinical decision support. *International Journal of Managing Information Technology (IJMIT)*, 4(1), 1–13.

15. Rubi, J. N. S., & Gondim, P. R. L. (2019). IoMT platform for pervasive healthcare data aggregation, processing, and sharing based on OneM2M and OpenEHR. *Sensors, MDPI, 19*(4283), 1–25.
16. Cervantes, J. L. S., Radzimski, M., Enriquez, C. A. R., Hernandez, G. A., Mazahua, L. R., Ramirez, C. S., & Gonzalez, A. R. (2016). SREQP: A solar radiation extraction and query platform for the production and consumption of linked data from weather stations sensors. *Journal of Sensors, Hindawi*, 1–19.

Fault Tolerant Multimedia Caching Strategy for Information-Centric Networking



Dharamendra Chouhan, Sachinkumar Hegde, N. N. Srinidhi, J. Shreyas, and S. M. Dilip Kumar

Abstract Extensive usage of mobile multimedia-based applications is increasing the user's demand on the internet. Edge computing makes it possible to bring storage of data and computation in close vicinity to mobile users. To improve the experience of the users, the proposed method uses edge Internet of Things equipment-assisted Fault Tolerant Multimedia Caching Strategy (FTMCS) for information-centric networking. To accomplish this, we propose a new technique composed of optimized task distribution and fault tolerant network services. The task distribution method efficiently distributes the downloading task among the edge nodes having required data. Fault tolerant service makes it possible for the network to withstand node failures that may occur during the file downloading, and by using this service, fast and reliable transfer of content happens seamlessly according to user preferences. The experimental results show that the FTMCS provides better network performance compared to the existing solution in terms of download speed and throughput.

Keywords Caching · Edge computing · Task distribution · Fault tolerant · ICN · Internet of things · Multimedia

1 Introduction

The demand for multimedia applications is continuously increasing on the internet, which increases the data volume on the internet [1]. Nevertheless, in conventional Transmission Control Protocol/Internet Protocol (TCP/IP), repeated transmission of data occurs during transmissions involving massive information. Recognizing a method that empowers various applications to become smart during usage is an essential topic focused on future networking administration.

In this era, Information-Centric Networking (ICN) is a type of internet architecture, which has gained significant attention. In ICN, the location of the content is not

D. Chouhan · S. Hegde · J. Shreyas (✉) · S. M. D. Kumar
University Visvesvaraya College of Engineering, Bangalore, India

N. N. Srinidhi
Sri Krishna Institute of Technology, Bangalore, India

considered by the user, but only the content. Based on the content, ICN will be able to route, position, and transmit the data. In general, ICN offers caching with seamless and built-in functionality to speed up the delivery of data. It also maximizes the network resource usage to avoid congestion [2] generated by the growing network traffic.

With the advancement in Information-Centric Networking with respect to mobility, the need for a technique to get superior user experience through caching has gained significant importance. Scholars have done considerable work to provide better transmission by modifying the cache entities. Nevertheless, it still remains challenging to obtain a good caching strategy for mobile users because of the movement of the users and limited device resources.

The development of edge computing gives a new perspective on these issues. Edge computing [3] is a modern computing technique that relies upon cloud and IoT services [4]. These services are required to be deployed at the edge of networks. This offers advantages like better response time, improved battery life, and minimized usage of bandwidth. Edge computing is effective in offering a possible improvement in network throughput and enhanced experience of the users. In addition, edge computing plays an important role in fostering new and popular solutions, like smart homes and smart cities, based on a smart environment [5]. Caching focuses on learning how to efficiently reduce the time needed for repetitive traffic to access content to achieve efficient content transmission. Caching decisions, therefore, play an important role in enhancing the user experience.

Caching mechanism is an innovation that exchanges time for space, and it concentrates on adequately diminishing the time needed to acquire information and accomplish the proficient transmission of information. In the proposed work, research is done on the issues related to the caching methodology for portable multimedia in ICN with edge computing. The outline of overall contributions of this paper can be summed up as follows:

- FTMCS for Edge IoT, a standard framework that enables task distribution is proposed. It can effectively distribute the downloading task among the edge nodes having required data.
- An optimal solution to deal with the failure of IoT nodes is proposed. It allows smooth relocation of existing networks, such as peer-to-peer, and thoroughly explores their potential.
- The performance of FTMCS is evaluated by conducting the experiments. The obtained results clearly show that the proposed FTMCS provides clients with higher download speed and throughput compared to the existing system.

The rest part of the paper is summarized as follows. Related research work and problem statement have been explained in Sects. 2 and 3, respectively. Sections 4 and 5 provide the system model and proposed method along with evaluation in Sect. 6. The conclusion part has been provided in Sect. 7.

2 Related Work

There has been much research works done on caching technology in ICN. Among them, Ahgren et al. [6] conducted a survey of Information-Centric networking. There are two main functions in ICN which must be done by resolution of name and routing when a particular NDO is requested. The first step is to search for a node that contains a copy of NDO and deliver the request to that node. The next step is to find a route back to the requestor from that node from which the NDO can be delivered. To offer better performing and more efficient transport services, ICN typically uses in-network storage. NDOs can be cached on the path, but by notifying them in a routing protocol or registering them in a name-resolution service, ICN can also make cached objects accessible for off-path requests. Research also reveals that there are no end-to-end linkages to the ICN approach. Xylomenos et al. [7] established a set of problems that motivates a reimagining of how the internet can function to meet the continuously evolving requirements. ICN is convenient for the infrastructure of the future Internet. ICN enables the proper and on-time transmission of information to the end-users. In this survey, the main targets are: (i) Identifying the important features of information-centric networking designs, (ii) To explain the main objectives of the information-centric networking in the form of tutorials, representing the differences and common factors between them with reference to those key features, (iii) Shortcomings of the information-centric networking proposals should be detected, and the unsolved research problems in this area of network research are identified. Wu et al. [8] proposed a Grey Relational Analysis (GRA)-based cross-layer caching strategy and a cross-layer cooperating caching technique for Content-Centric Networking. User priority is enforced in the form of the application layer, physical layer, and network layer metrics. Then, among all nodes along the distribution road, a caching probability of caching function is developed based on the grey relational analysis. GRA-based cooperating caching technique is developed based on the above-mentioned three-layered caching parameters. Grey relational analysis is a part of the grey method theory, which is best suited for resolving problems of complex connections between multiple variables and factors. This strategy does not provide scalability. Quan et al. [9] conducted a study that focuses on finding the possibility for social collaboration between neighboring highway vehicles to facilitate the use of resources for caching. In particular, the authors propose a cooperative caching mechanism based on the highway customized ICN, namely, ICoC, aimed at optimizing the quality of video streaming without excessive startup time and reducing the playback buffering ratio. In order to allow advanced QoE optimization, this study primarily characterizes information-centric delivery and highway traffic. Partner assist cooperation and courier-assist cooperation are being established in order to improve the efficiency of caching in highway VANETs. The performance of the caching mechanism proposed in this study is poor.

3 Problem Statement and Objectives

The problem considered in the proposed work is to formulate an advanced caching mechanism for information-centric networking by developing fault tolerance and task distribution features to achieve better performance. To accomplish the research goal, the following are the objectives set forth:

- To design an efficient caching strategy with task distribution to maximize the download speed.
- To develop a fault tolerance feature to the caching strategy to withstand the node failures in the network.
- To achieve reduced download time and increased throughput.

4 System Model

The architecture of the proposed system is shown in Fig. 1. The proposed system has three parts. (i) User storage, (ii) Edge node, and (iii) Server.

The user storage constitutes the private space of a client, which can store something from the cache system that they like. In addition, all the user instructions are transferred through the user configuration to the system.

The edge node is the second element. Where, each edge node, updates the cached multimedia content, can be composed of customized storage space and computation power.

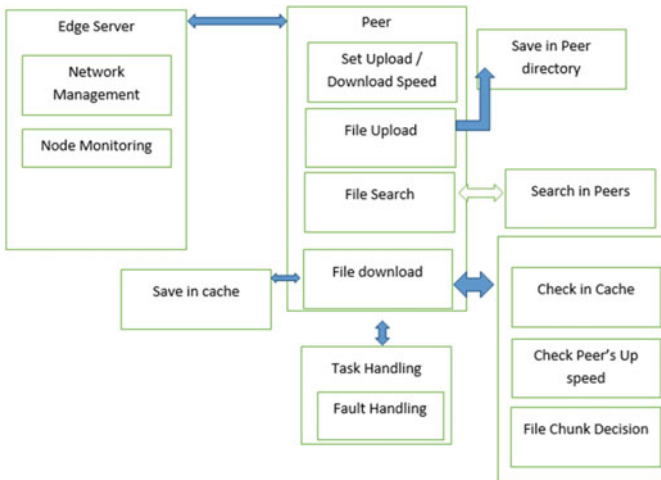


Fig. 1 System architecture

The server is the last part of the device that involves the computation of data, cache, and storage. It stores the information resources, and also archives all clients' interest content and mobile traces. It monitors the nodes and manages the network.

The multimedia data is cached in the peers to respond to user queries faster, and implemented fault tolerant ICN will save network bandwidth. The fault tolerance feature equips the network with the ability to withstand the unexpected node failures that may occur in the network. It essentially monitors the ongoing network operation and can assign the task to the next available peer in case of unexpected node failures. The proposed system also has a task distribution feature. A data request is divided among nodes, and parallel processing of data collection and caching mechanism makes the overall network efficient. The task is divided among the selected nodes equally, and all the tasks are processed simultaneously so that the data chunks from different nodes are collected and merged together in less span of time. The proposed system provides better response time, battery life, transmission capacity, throughput, and enhanced client experience.

5 Implementation

In this section, we discuss the details of the proposed algorithm. The size of the chunk file which is to be downloaded from available peers is calculated, and then data will be downloaded from them through parallel processing.

5.1 Calculation of Chunk Size

Algorithm 1 shows the way in which chunk size is calculated. While downloading the content, all the peers having the required cached data will be involved. The file will be downloaded in the form of chunks from each of the peers, and the size of the chunk that needs to be downloaded from each of the peers has to be calculated. This is done on the basis of the upload speed of the respective peer. Upload speed depends upon the configuration done by the user. Total upload speed can be calculated by using the formula (1).

$$Total_{Up_Speed} = \sum_{i=0}^n (Up_speed(node_i)) \quad (1)$$

The size of chunks from each peer can be calculated by using the formula (2).

$$Chunk_{size} = \frac{file_{size}}{Up_{speed}(node)} \quad (2)$$

A chunk of higher size will be downloaded from the peer with the highest upload speed. Likewise, chunk size will be calculated for all of the available peers based on their upload speed limit set by the user. The calculated chunk size information will be used later while downloading the file.

Algorithm 1: Chunk Size Calculation Algorithm

Input: num_peers, upload_speed, size_file

Output: sorted peers, chunk size

1. Assign arr \leftarrow sorted_selected_list
2. $t \leftarrow 0$
3. for $i=0$ to length (node) do
4. total_up_speed = up_speed (node_{*i*})
5. end for
6. $t \leftarrow (\text{size_file} / t_up_speed)$
7. $st \leftarrow 0$
8. End $\leftarrow 0$
9. for each peer
10. Bytes \leftarrow up_speed * t
11. end \leftarrow st + Bytes
12. st \leftarrow end + 1
13. end for
14. save peers_name and chunk_size
15. Stop

5.2 Download File

Based on the previous chunk size calculation, the file will be downloaded as shown in Algorithm 2. The requester sends the request for a chunk from the assigned peer. When the assigned peer receives the chunk request, extracts the corresponding chunk from the file, and sends it back to the requester. The requester on receiving the chunk assigns the chunk id and then, it saves the chunk in the buffer. Later, it sorts all the received chunks based on the chunk id. After it receives a complete set of chunks, merges them in sorted order to get the complete file, and saves them in the application directory as well as in the cache memory.

The network is able to address the issues arising due to node failures. After assigning the task to the selected peers, if any network failure occurs with any of the peers, the requester assigns the task to get the chunk to the next available peer and follows the previously mentioned procedures. The requester continuously monitors the progress of the download of chunk and is able to detect the node failures. Thus, it ensures the integrity of data.

Algorithm 2: Download file algorithm with task distribution and fault tolerance

Input: filename, peer_name , ip , port, chunk_size

Output: downloaded file

1. Send chunk from source node to destination node
 - a. Create object soc
 - b. $\text{soc} \leftarrow \text{peer_ip}, \text{peer_port}$
 - c. Create ObjectOutputStream object
 - d. write chunk_details, chunk_id, chunk_size to object
 2. Read file
 - a. Load the file f from dir
 - b. Read file f using BufferedReader
 - c. for index = 0 to length (index) do
 - d. read f[index]
 - e. end for
 3. transfer chunk to requested peer
 - a. $\text{soc} \leftarrow \text{peer_ip}, \text{peer_port}$
 - b. Create ObjectOutputStream object
 - c. write chunk_data to object
 4. monitor peer transfer
 5. if connection_lost
 6. Search(next_peer)
 7. if found
 8. Send_chunk_request(fname,chunk_id)
 9. end if
 10. end if
 11. Sort chunks using chunk id
 12. Merge chunks
-

6 Performance Evaluation

Performance of the Proposed FTMCS is assessed by comparing it with the existing SCM [10]. The following performance metrics have been used to evaluate the performance: Download time with cache and throughput. The results obtained are depicted in the form of graphs.

6.1 Download Time

In Fig. 2, a comparison of the proposed system and the existing system is done using files of different sizes. Here, it can be observed that the time taken to download the same size file is less in the proposed system. In the proposed system, the download task will be distributed among all the peers having the required data, and while downloading, data will be downloaded from all those peers in the form of chunks. In the existing system, download can happen from a single peer at a time. This makes it possible for the proposed system to achieve a faster download speed.

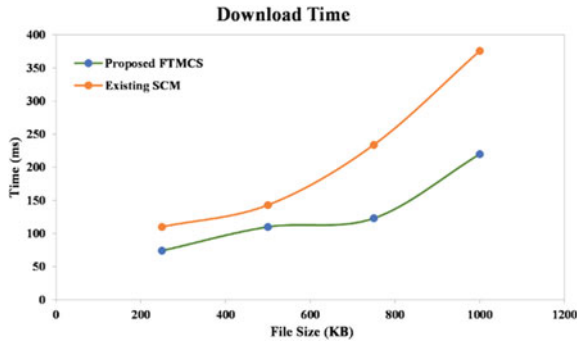


Fig. 2 Download time

6.2 Throughput

In Fig. 3, it can be observed that the throughput in the proposed system is improvised. The proposed system is tolerant to node failures in the network. If any node fails while transmitting the data, the proposed system will continue to download from the next available node. The existing system is not tolerant of node failure. This enables the proposed system to achieve increased throughput than the existing system.

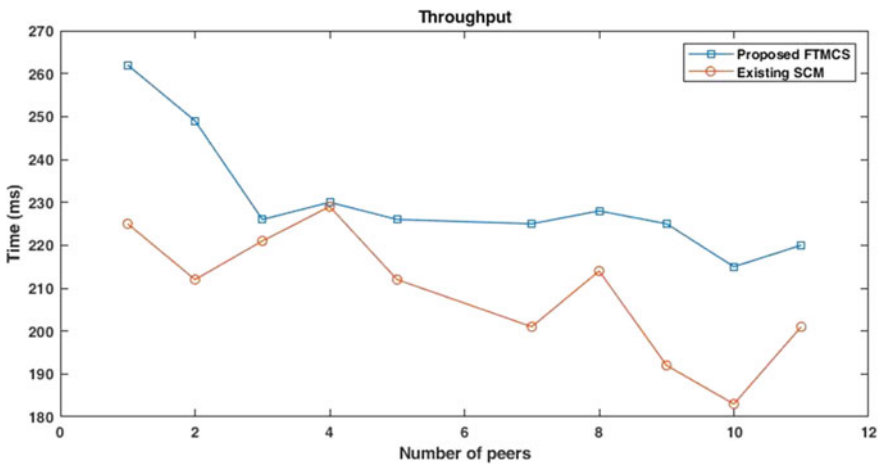


Fig. 3 Throughput

7 Conclusion

In this paper, an optimized solution for caching multimedia data in information-centric networking has been proposed. The caching strategy in the proposed work is inclusive of task distribution and fault tolerance. Task distribution essentially divides the task among the peers having the required file. Fault tolerance ensures the proper transmission of data even if there is a node failure in the network. A major issue is being addressed by the algorithm presented in this paper. Multimedia data is cached in the peers for a faster response to the queries. The proposed algorithm will increase performance dramatically, optimizes the time for data collection, and improves efficiency. A comparison between the existing system and the proposed system shows that the proposed system provides better download speed and throughput. In-network caching for ICN can be used as a further extension to the proposed method.

Acknowledgements This research work has been funded by SERB-DST Project File No: EEQ/2017/000681.

References

1. Shreyas, J., Jumnal, A., Kumar, S. M. D., & Venugopal, K. R. (2020). Application of computational intelligence techniques for internet of things: An extensive survey. *International Journal of Computational Intelligence Studies*, 9(3), 234–288.
2. Shreyas, J., Singh, H., Bhutani, J., Pandit, S., Srinidhi, N. N., & Kumar, S. M. D. (2019). Congestion aware algorithm using fuzzy logic to find an optimal routing path for IoT networks. In *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, December 11–12 (pp.141–145). Dubai, United Arab Emirates.
3. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things*, 3(5), 637–646.
4. Srinidhi, N. N., Sagar, C. S., Chethan, S. D., Shreyas, J., & Kumar, S. M. D. (2019). Machine learning based efficient multi-copy routing for opriot networks. In *International Conference on Computational Intelligence, Security and Internet of Things*, December 13–14 (pp. 288–302). Agartala, India.
5. Srinidhi, N. N., Sunitha, G. P., Nagarjun, E., Shreyas, J., & Kumar, S. M. D. (2019). Lifetime maximization of IoT network by optimizing routing energy. In *IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, November 15–16 (pp. 1–4). Bangalore, India.
6. Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., & Ohlman, B. (2012). A survey of information-centric networking. *IEEE Communications Magazine*, 50(7), 26–36.
7. Xylomenos, G., Ververidis, C. N., Siris, V. A., Fotiou, N., Tsilopoulos, C., Vasilakos, X., et al. (2014). A survey of information-centric networking research. *IEEE Communications Surveys & Tutorials*, 16(2), 1024–1049.
8. Wu, L., Zhang, T., Xu, X., Zeng, Z., & Liu, Y. (2015). Grey relational analysis based cross-layer caching for content centric networking. In *IEEE/CIC International Conference on Communications in China (ICCC)*, November 2–4 (pp. 1–6). Shenzhen, China.

9. Quan, W., Xu, C., Guan, J., Zhang, H., & Grieco, L. A. (2014). Social cooperation for information-centric multimedia streaming in highway VANETs. In *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, June 19–19 (pp. 1–6). Sydney, Australia.
10. Tang, Y., Guo, K., Ma, J., Shen, Y., & Chi, T. (2019). A smart caching mechanism for mobile multimedia in information centric networking with edge computing. *Future Generation Computer Systems*, *91*(2), 590–600.

Sizing of Wireless Networks with Sensors for Smart Houses with Coverage, Capacity, and Interference Restrictions



Jhonatan Fabricio Meza Cartagena, Deepa Jose, and J. S. Prasath

Abstract This research work proposes an effective solution for sizing in wireless networks supported by IEEE 802.15.4g, thus presenting the possibility of communicating networks with remote sensors in a completely transparent way for end devices, considering for this analysis restriction parameters such as the capacity, coverage, and interference knowing that the field of application of Wireless Networks is recently emerging, considerably gaining much popularity, which is increased as their features increase and new applications are discovered for them. This article has also included the application of a practical scenario, which demonstrates the communication of the devices remotely.

Keywords Wireless sensor networks · ZigBee · IEEE 802.15.4g · HAN domestic area networks · Sizing

1 Introduction

In this document, we have tried to provide an optimal solution to the sizing of wireless sensor networks for smart homes (Smart home), Fig. 1, based on the IEEE 802.15.4g, taking into account the characteristics and its restrictions allowing us to analyze the different parameters that may exist [1].

Considering the scenario based on the internal location of wireless sensors, we say that these are small devices that consume very little power, low data speed, and low mobility and are capable of measuring environmental parameters and communicating

J. F. M. Cartagena
Instituto Superior Tecnológico 17 de Julio, Ibarra, Ecuador
e-mail: jmeza@ist17dejulio.edu.ec

D. Jose (✉)
Department of ECE, KCG College of Technology, Chennai, India

J. S. Prasath
Department of EIE, KCG College of Technology, Chennai, India
e-mail: prasath.ei@kcgcollege.com



Fig. 1 Architecture of a WSN home network

with each [2]. Therefore, the environment of these sensors can be associated with their positions allowing the creation of many interesting applications.

On the other hand, within a group of simple protocols emerges the IEEE 802.15.4, where one of the most relevant characteristics of this standard are the flexibility of the network, low costs, and low energy consumption. It should be recognized that this standard has many home automation and industrial applications [3], which require a low data transmission rate, the key to its use being the reduction of expenses and information exchange with minimal effort [4, 5]. There is also talk of the introduction of 4G in communication technologies in the fourth generation.

On the other hand, if we mention the IEEE 802.15.4, we also have to mention that it is ZigBee [6]. It is a network technology that emerges as a low-cost wireless communication standard [7] that is capable of satisfying needs, such as the advantages of low cost, low power, and a broader coverage, making it very suitable for the development and use in a house.

In other words, smart grids today make up a very sophisticated energy system connecting a number of networks and devices that imply a tendency to be driven by their great ability to integrate wireless components, resembling the success of mobile wireless communication systems, considering intelligent systems capable of analyzing a situation, making calculated decisions, and responding to anomalies based on implemented models or algorithms.

It is necessary to know the methods that exist to extend the useful life of the devices of a WSN wireless sensor network [8], the processing of their data, routing, topologies, and the placement of the control devices in the determination of coverage.

Therefore, the problem presented below aims to provide a solution to the detection of a minimum energy consumption in order to improve the excess energy in a connected wireless sensor network (WSN), and designed for improving performance and coverage for indoor users (where most mobile users spend much of their time), as well as overall system capacity by downloading data [9].

On the other hand, considering our study, we have to talk about the wireless access points (Wireless Access Point) known or called AP for its acronym in English, since these will serve as intermediaries serving as transmitters and receivers of the signals sent in the radio of the network.

We also have to consider the interference that the devices of our wireless network system may receive, considering that this parameter can be caused by other services that also operate within our band, and it is imperative to know that a primary requirement of IEEE 802.15 applications 0.4 is to try to control and reduce interference caused by communications within a common channel. Therefore, it is very important in the design and efficient control within the topologies of a wireless network of network sensors. The rest of the work is organized as follows. In Sect. 2, the work related to the sizing of wireless networks is presented. Section 3 describes the formulation of the optimization problem to deploy a network infrastructure with wireless sensors. In Sect. 4, we propose the optimization of the problems posed in a sensor network based on the analysis of results. Finally, we conclude this document in Sect. 5.

2 Network Dimensioning Wireless

In this section, we briefly examine smart home networks, and based on their main contributions, we discuss the main task approaches based on capabilities described in Sect. 3. Furthermore, in this document, a series of methods have been proposed to find coverage, capacity, and interference connected to a wireless sensor network (WSN). The scenario may be optimal under certain criteria, therefore, the main objective is to cover a total detection area by placing a minimum of sensors, which when deployed tend to be efficiently connected so that each sensor can find a connection path and can arrive at a station considered as a base. In fact, in a WSN, where the devices are densely deployed, a subset of the devices can cope with coverage and connectivity issues. The rest of the devices can be switched to a sleep state to conserve power. With the use of device placement and routing techniques, it is difficult to find coverage, when using heterogeneous WSNs. Due to this a connectivity problem occurs. Therefore, we consider all the traffic generated by the applications that run smart grids with a heterogeneous nature in terms of minimum performance, reliability, etc. [10].

3 Problem Formulation

Wireless sensor networks constitute the fundamental basis of our study, thus determining the necessary parameters to be able to achieve a modeling focused on the optimal dimensioning of the WSN, considering restrictions such as coverage, capacity, and interference.

For the proposed problem, a set N of users located randomly in the region is considered. Assuming likewise M possible locations (candidate sites) of access points in the region. A possible location is a place where an AP could be placed, but not where one will necessarily be located. A coverage radius is highlighted depending on the configured power of the AP equipment, depending on the objective of covering a certain area.

a. Coverage Problem

In this way, for this case, we propose a metaheuristic approach considering an algorithm that allows us to cover a 2-dimensional area (length and width), and where it can be.

Identify a potential position for the placement of a sensor point, so that our target points are k -coverage and all sensors are our n -connectors.

The algorithm that we propose [11] considers technical and modeling criteria, necessary to carry out a sizing process where a design problem is formulated in a concrete way. Therefore, the type of objective function and restrictions must meet some characteristics or certain initial parameters such as:

- For each candidate site i , the quantity is defined $Z_i \in \{0,1\}$. If the value is 1, it indicates that the candidate site i is an active site.
- For each sensor u_j , the quantity is defined $Y_j \in \{0,1\}$. If the value is 1, it indicates that the user is covered by at least one candidate site.
- The relationship between the list of active sites and user coverage is as follows:

$$Y_j \leq \sum_{i=1}^M \alpha_{i,j} \sum Z_i \quad (1)$$

The above equation is defined as a constraint. If there is no active candidate site covering the sensor, Y_j is necessarily equal to 0, that is, the sensor is not covered. To indicate that it covers at least 90% of the sensors, another restriction must be added, which counts the percentage of sensors covered.

$$N \sum_{j=1}^{1N} Y_j \geq 0.9 \quad (2)$$

Finally, the goal of the problem is to minimize the number of candidate sites that are active. To do this, we write another equation that is now called the objective function.

Therefore, we write what we call the optimization problem that we completely express as:

$$\min \sum_{i=1}^M Z_i \tag{3}$$

$$\text{s.t. } Y_j \leq \sum_{i=1}^M \alpha_{i,j} Z_i; \forall j \in U \tag{4}$$

$$N \sum_{j=1}^{1N} Y_j \geq 0.9 \tag{5}$$

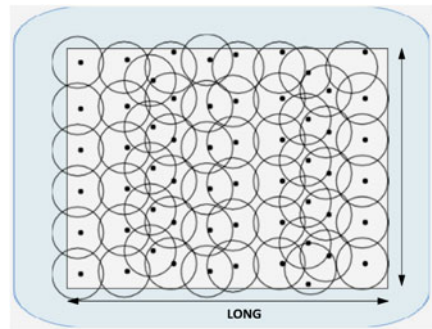
where, Eq. (3) seeks to minimize the number of candidate sites that are active. Equation (4) defines the coverage constraint. Equation (5) tells us that 90% of the sensors can be covered (Fig. 2).

b. Capacity Coverage Problem

The next step that we can consider with the same algorithm raised above is to include a capacity restriction in the existing network wireless local (WLAN) claiming benefit of AP devices placed on stage. In this parameter, our scenario raises the problem considering the capacity of the WLAN network, arbitrarily placing APs, determining a radius and a limited capacity to the sensors. We describe the optimization problem below.

$$\min \sum_{i=1}^M Z_i \tag{6}$$

Fig. 2 Illustration of a full coverage



$$\text{s.t. } Y_j \leq \sum_{i=1}^M X_{i,j}; \forall j \in U \quad (7)$$

$$\sum_{j=1}^N X_{i,j} \leq C Z_i; \forall i \in S \quad (8)$$

$$\sum_{j=1}^N Y_j \geq N \cdot Porc \quad (9)$$

$$X_{i,j} \leq \alpha_{i,j} Z_i; \forall i \in S; \forall j \in C \quad (10)$$

Equation (6) corresponds to the objective function. Equation (7) is the one that is subject to the restrictions.

iii. Coverage, Capacity, and Interference Problem

This section further investigates the impact of interference on the performance of the desired connection. We base the interference reduction algorithm on the ZigBee home network architecture, where one or multiple network-related functionalities are implemented. In the architecture, we can assume that at least one WLAN AP is associated with the proposed scenario.

In this scenario, a network is proposed where nearby APs interfere with each other, this being the base model for others who were looking for assign channels.

In fact, an AP is considered to allow connectivity with sensors that are within a certain coverage radius [12, 13]. The capacity of an AP is limited and is given by the number of simultaneous sensors that can be attended, considering that the APs operate on the same channel, so there cannot be two APs within an interference distance, since they would produce interference with each other. The sensors can be located in any position within the region and seek to connect to an AP that is available and capable of serving it.

It is therefore necessary to consider the parameters previously raised in the problem of coverage with capacity, and we assign the following consideration:

Fig. 3 Scenario of the problem to perform the optimal sizing of networks with wireless sensors



$$Z_i + Z_k \leq 2 - \beta_i, k; \forall i \in S; \forall k \in S \quad (11)$$

Universal data aggregation point optimization algorithm considering coverage, capacity, and interference

Step 1: R = Coverage Radius.

Step 2: P = We define the coverage probability between 0 and 10, we define a capacity for the second restriction, and we define a percentage of interference.

Step 3: Determine the number of candidate sites $S = (S_1, S_2, \dots, S_n)$ in the i -nth position given by (X_{Si}, Y_{Si}) .

Step 4: We define a set $U = (U_1, U_2, \dots, U_n)$ of sensors. The j -nth sensor position given by (X_{Uj}, Y_{Uj}) .

Step 5: We establish the relationship between AP condition and sensor coverage.

Step 6: The problem performs AP minimization by means of the function.

Objective (Fig. 3).

4 Analysis of Results

Figure 4 shows a scenario with n meters by n meters, where we place WLAN APs to provide connectivity to the different sensors considering a coverage radius of R meters in any position within our scenario. In addition, the scenario includes 21 sensors and 45 candidate sites for the deployment of the wireless sensor network within a map made under the AutoCAD platform.

In fact, the main limitation of our proposed scenario is its scalability. The problems that due to their simplicity can generate millions of decisions, and according to the models, both real and approximate solutions can be found. To simulate the coverage optimization problem, the MATLAB programming language was used, and the parameters of the problem are initialized with previously determined values, while the degree of coverage in the destination field and dimension of the access point were entered by the sensor, starting from the previous problem with 21 sensors and 45 candidate sites in the scenario.

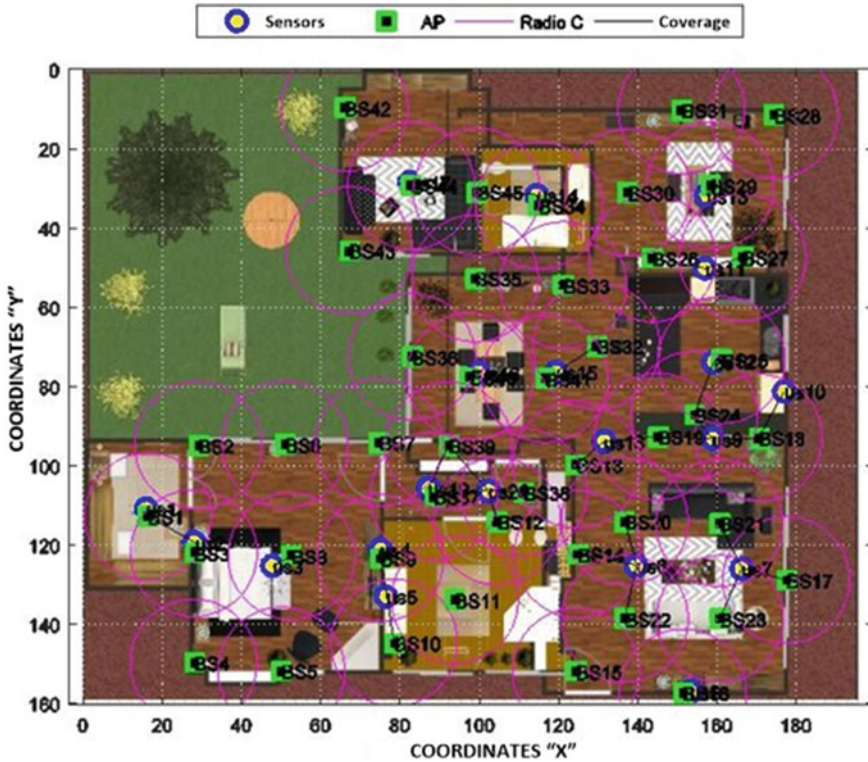


Fig. 4 Location of multiple APs to cover WSN

Figure 5 finds a minimum of AP active sites that cover at least a percentage of sensors at starting from the previous problem with 21 sensors and 45 candidate sites in the scenario.

Figure 6 locates an AP arbitrarily with each access point considering a certain radius and a capacity limited to this with the number of sensors.

Therefore, when analyzing the results obtained in the scenario in Fig. 6, we consider that to correct this problem and maintain the minimum speeds required, there are some tips and recommendations that they make, such as finding the best location for the access point, placing either your antenna or even change the latter for a better performance one. It is evident that if we choose the perfect place to place it, we will have more opportunities to improve (or optimize) the wireless network (Fig. 7).

Figure 8 shows us the last scenario where we place the same conditions as the previous ones with the addition of the interference restriction, which will allow us to understand what happens in this case.

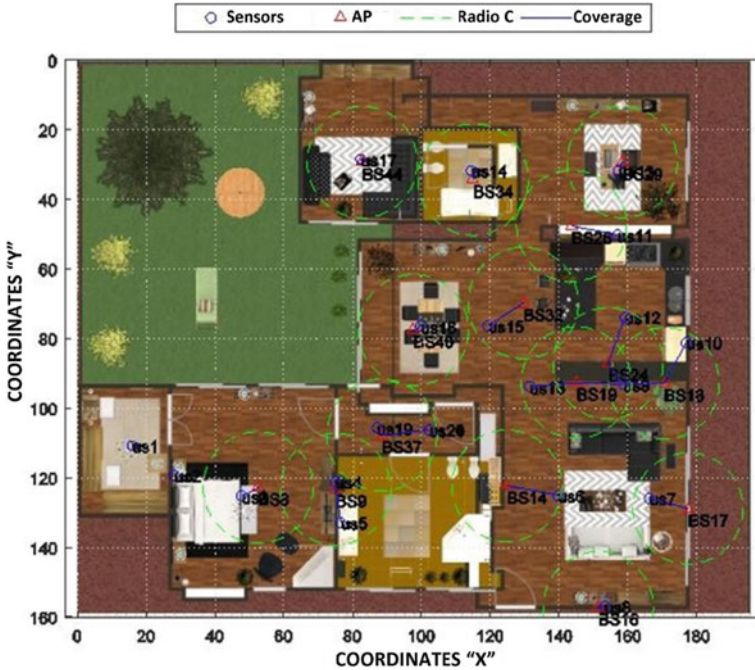


Fig. 5 Coverage of a minimum percentage of area

Figure 9 solves the interference problem that arises in our scenario, but clearly shows the coverage problem. This means that the APs that are within a distance cannot interfere with each other but limits the coverage.

As we can see that under the restriction of coverage, capacity, and interference, the solution is not as simple as putting another channel. The channels overlap each other, to avoid 100% interference, it should be more than 5 channels apart.

That is, if we take channels 6 and 11 as an example, they do not interfere, but 6 and 7 do. Some APs have an automatic channel selection function that tries to automatically select the least saturated channel, which in other words is meant to say that wireless connections are made over radio waves and there are several channels available. If your wireless network connection shows interference or disconnects regularly, then an attempt is made to change channels.

5 Conclusions

In this document, we propose an algorithm with heuristic techniques that seeks a WSN deployment with optimal coverage through an optimal initial distribution of the wireless network, which through computer simulations found a better solution to

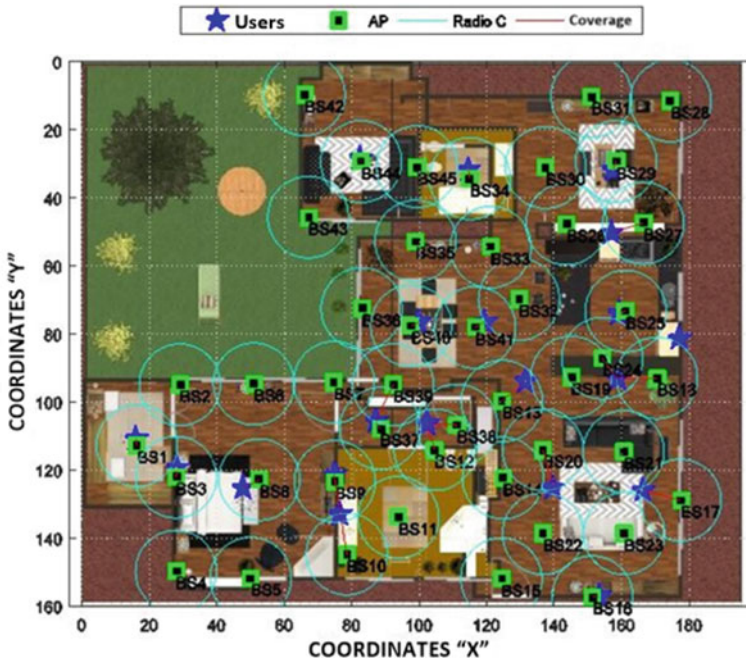


Fig. 6 Scenario graph not optimized

move our AP devices with maximum coverage and a minimum of energy consumption for communication. On the other hand, the analysis in the proposed scenario leads us to say that the scope of communication and the range of interference in IEEE 802.15.4g is based on channel models. The realistic intention is to include analysis with interference minimization techniques that are capable of relaxing the range of interference. However, the WSN application at various smart grid sites presented in this document consider the architecture of wireless sensors for integration with candidate sites to sustainably satisfy the smart grid. Analyzing the themes and scenarios carried out in relation to Wireless Sensor Networks and taking into consideration the restrictions previously raised, it is important to highlight that it is necessary to take into account that without the implementation of said technology in these environments, many of the homes, they may not have the benefits that these technologies have provided us.

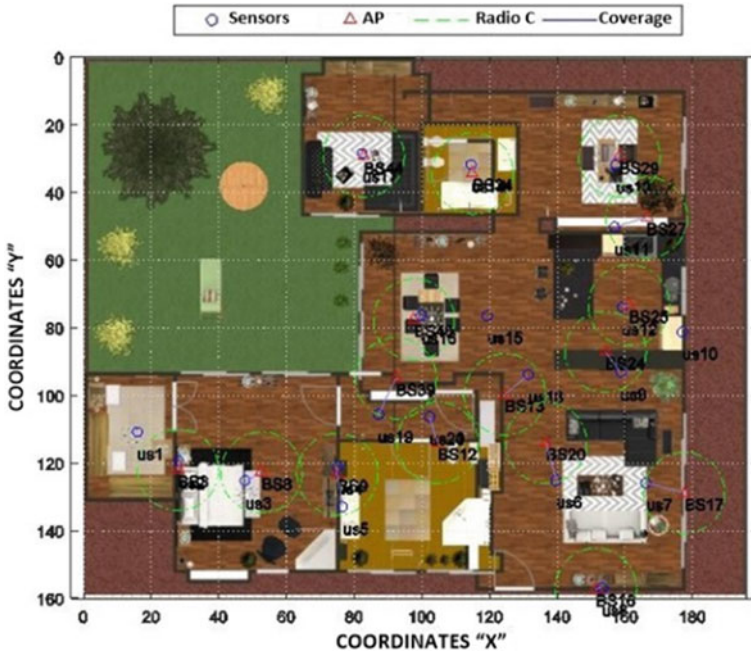


Fig. 7 Network capacity simulation with APs that connect each sensor

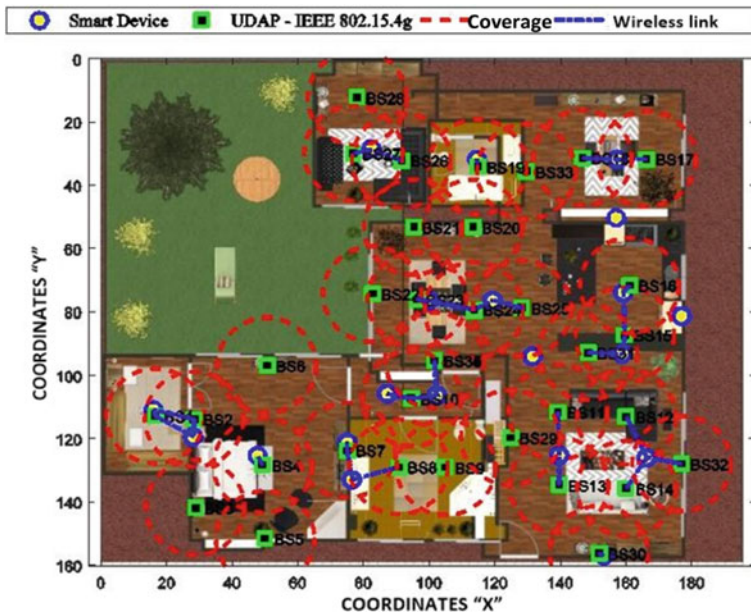


Fig. 8 Scenario graph not optimized



Fig. 9 Interference simulation with coverage limitation

References

1. Cuomo, F., Cipollone, E. & Abbagnale, A. (2009). Performance analysis of IEEE 802.15.4 wireless sensor networks: An insight into the topology formation process. *Computer Networks*, 53(18), 3057–3075. <https://doi.org/10.1016/j.comnet.2009.07.016>
2. Cheng, L., Wu, C. D., & Zhang, Y. Z. (2011). Indoor robot localization based on wireless sensor networks. *IEEE Transactions on Consumer Electronics*, 57(3), 1099–1104. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6018861
3. Erol-kantarci, M., & Mouftah, H. T. (2011). Wireless sensor networks for cost-efficient residential energy management in the smart grid. 2(2), 314–325
4. Asad, O., Erol-kantarci, M., & Mouftah, H. (2011). Sensor network web services for demand-side energy management applications in the smart grid. 1176–1180.
5. Galinina, O., et al. (2014). Optimizing energy efficiency of a multi-radio mobile device in heterogeneous beyond-4G networks. *Performance Evaluation*, 78, 18–41. <http://linkinghub.elsevier.com/retrieve/pii/S016653161400056X>. Garroppo, R. G., et al. (2015). The impact of the access point power model on the energy-efficient management of infrastructured wireless LANs. *Computer Networks*, 1–13. <http://linkinghub.elsevier.com/retrieve/pii/S138912861500451X>
6. Cuomo, F., Abbagnale, A., & Cipollone, E. (2013). Ad hoc networks cross-layer network formation for energy-efficient IEEE 802.15.4. ZigBee wireless sensor networks. *Ad Hoc Networks*, 11(2), 672–686. <https://doi.org/10.1016/j.adhoc.2011.11.006>
7. Baronti, P., Pillai, P., Chook, V. W. C., Chessa, S., Gotta, A., & FunHu, Y. (2006). Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards.
8. Akilandeswari, M., & Srikanth, U. (2012). Metaheuristic approach for maximizing the lifetime of heterogeneous wireless sensor networks.

9. Dudnikova, A., Panno, D., & Mastro Simone, A. (2015). Measurement-based coverage function for green femtocell networks. *Computer Networks*, 83, 45–58. <http://www.sciencedirect.com/science/article/pii/S1389128615000730>
10. Amin, R., & Martin, J. (2012). Smart grid communication using next generation heterogeneous wireless networks, 229–234.
11. Cheng, H., et al. (2016). Energy-efficient node scheduling algorithms for wireless sensor networks using Markov random field model. *Information Sciences*, 329, 461–477. <http://linkinghub.elsevier.com/retrieve/pii/S0020025515006945>
12. Murugan, R., & Mohan, M. R. (2012). Artificial bee colony optimization for the combined heat and power economic dispatch problem. *ARPN Journal of Engineering and Applied Sciences*, 5(7), 1353–1366.
13. Jose, D., Kumar, P. N., & Ramkumar, S. (2014). Reliability aware self-healing FFT system employing partial reconfiguration for reduced power consumption. In *Proceedings of the 2014 IEEE Students' Technology Symposium* (pp. 31–36).

Cloud-Based Parkinson Disease Prediction System Using Expanded Cat Swarm Optimization



Ramaprabha Jayaram and T. Senthil Kumar

Abstract Parkinson disease is identified as the second most severe neurodegenerative disorder that affects the nerve system of people. This disease could mainly affect the walking, speech, and vision of patients followed by body nervousness, handwriting, harsh voice quality, depression, and sleeping problems. The proposed research study focuses on early detection and diagnosis of disease from the accelerometer sensor-based data by evaluating the deviations present in patient's motor symptoms. A cloud-based Parkinson disease prediction system is developed for a clinical decision-making process that helps the doctor to diagnose the Parkinson-influenced patient from a remote place. Gait parameters of the patient were extracted along to provide input vectors to the classifier model for onboard Parkinson disease prediction and diagnosis. An effective expanded cat swarm optimization (ECSO)-based feature selection technique is explored to overcome the problem of dataset dimensionality. It could select the most appropriate features from the patient dataset according to a logically inspired evolutionary algorithm. Using this feature selection technique in the k-Nearest Neighbor (k-NN), classifier model could significantly improve the disease prediction accuracy, and also minimizes the disease prediction time against the existing classifier models.

Keywords Parkinson disease · Feature selection · Cloud-based prediction · Cat swarm optimization · k-nearest neighbor classifier

1 Introduction

Parkinson disease is treated as the most severe neurological disease caused due to increase in motor dysfunction. This disease can be recognized by exploiting handwriting dynamics, speech, motion, and gait signals. To support early identification

R. Jayaram (✉) · T. Senthil Kumar
SRM Institute of Science & Technology, Kattankulathur, Chengalpattu, India
e-mail: ramapraj@srmist.edu.in

T. Senthil Kumar
e-mail: senthilt2@srmist.edu.in

and diagnosis of Parkinson disease, there is a need for continuous monitoring and assessment of changes occurring in the patient's gait observation. Motor symptoms of this disease can be easily characterized by the occurrence of resting tremor, body stiffness, and bradykinesia (slowness of movement) [1]. Some of the non-motor symptoms can be characterized by the depression, fatigue, and sleep disorder. In addition, the Parkinson patient may also have some common problems like postural instability, coordination, and balance. Also, the patient having disease used to walk with leisurely gait speed, decreased step length, and height. In order to improve the healthcare system, accelerometer and gyroscope sensors were used to monitor the physical activity and fall prediction of remote patients by measuring the spatial and temporal parameters of gait prototype [2]. The key parameters such as gait velocity, cadence, time, distance, stride length, and step length were considered to be the major variables for analyzing the gait cycle.

Most researchers put emphasis on using the accelerometer sensor to measure the balancing factor and posture transitions. In some cases, gait evaluation was done through live video recording of patients to track the gait movements for human fall prediction and real-time activity monitoring. Recognizing such human activity through gait analysis could help the therapists in general context, but does not provide any kind of clinical assistance for them. There are different types of feature sets like time domain and frequency domain to encode the gait signals in the regression-based model [3]. To make better assistance to the clinicians during prediction and diagnosis of Parkinson disease, various machine learning classifiers such as Bayesian, decision tree, support vector machine, and artificial neural networks were used to identify the hidden patterns associated with the dataset. In order to avoid the over fitting problem and improve the accuracy and prediction time classification, various feature selection techniques such as information gain, principal component analysis, linear discriminant analysis, and correlation feature subset selection were deployed over the dataset [4]. Afterwards, various stages of Parkinson disease are identified by the classifier model for further diagnosis and rehabilitation monitoring purposes.

The key contributions of this research study includes: (1) a cloud-based Parkinson disease prediction system for real-time prediction and diagnosis of Parkinson disease; (2) a novel ECSO-based feature selection method for handling dimensionality reduction and over fitting problems; and (3) an efficient k-NN-based classifier model is proposed for predicting various stages of Parkinson disease severity. The remainder of this research study is organized as follows. The next section presents the related works pertaining to the Parkinson disease prediction system. Section 3 briefs the architecture of cloud-based Parkinson disease prediction system along with ECSO-based feature selection and k-NN based classifier model proposed for this research study. In Sect. 4, results and discussions of the proposed healthcare system is assessed in the context of prediction time and accuracy. The final section gives the conclusion and future enhancement of this research study.

2 Related Works

Nowadays, cloud-based remote patient monitoring and diagnosis have become more popular with various diseases. Especially, people suffering from neurological disorder need to be identified for regular monitoring and diagnosis according to severity stages [5]. In order to predict the intended target variable, a feature selection method is used to minimize the number of input variables by removing non-sensitive information from the data samples. The Principal Component Analysis (PCA) feature selection is used to find the ways of maximum variation. It performs well in case of having very less data samples for the available classes. But, sometimes it ignores the class labels also due to unsupervised methods [6]. In some cases, Linear Discriminant Analysis (LDA)-based feature selection method is preferred for the sake of dimensionality reduction. This linear transformation technique takes the supervised methods to find feature subspace by maximizing the class separation based on the normally distributed classes assumption [7].

The Naive Bayes' classifier has independent features among the classes, and expresses the event probability according to its prior knowledge. At the time of analysis, this classifier could make the classification based on the combined source of information such as likelihood and prior probability values. Next, a multiplicative function is used to combine these values to produce the posterior probability as a product [8]. The Support Vector Machine (SVM) classifier can be used to make classification on both linear and nonlinear data samples. It needs to identify the hyper plane that could differentiate the classes through maximum origin [9]. Further, the SVM is upgraded with particle swarm optimization and genetic algorithms to select the optimal features for searching among the appropriate continuous variables [10]. Therefore, the proposed research focused on developing optimization selection techniques along with k-NN classifier models to enhance the Parkinson disease prediction systems.

3 Cloud-Based Parkinson Disease Prediction System

An architecture of cloud-based Parkinson disease prediction system is proposed as presented in Fig. 1. It clearly shows that the remote patient can interact with the doctor using smart phones or video-enabled devices over the cloud platform. All the monitoring and diagnosis activities related to the healthcare system is controlled by the cloud manager through appropriate resource allocation policy. Usually, the medical data samples obtained from the remote patient using smart phones will be controlled by the multimedia content server to control the live streaming of data over the internet. Then, the Parkinson disease manager service could handle the relevant operations by running the feature extraction and classifier models deployed in the virtual machines. In this research study, an ECSO-based feature selection method is explored to reduce the dimensionality of medical data before going for disease

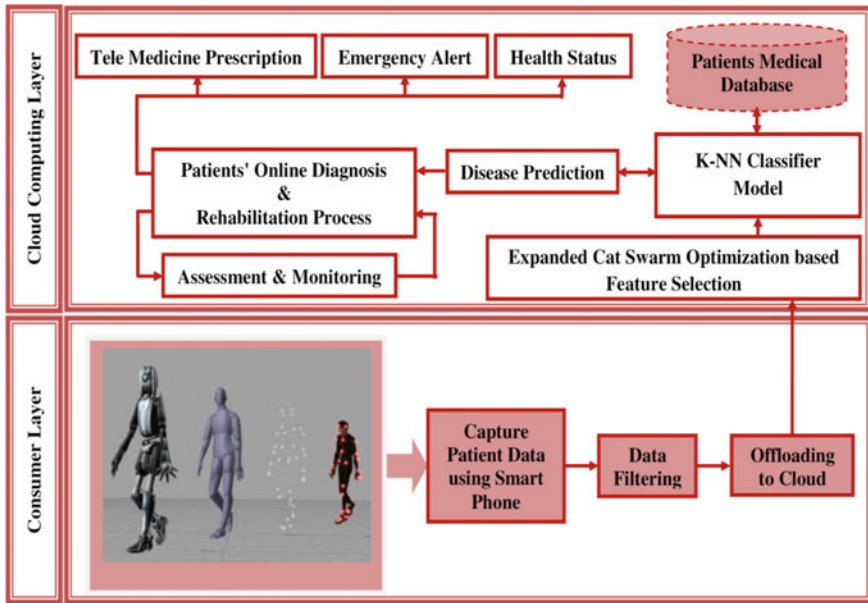


Fig. 1 Cloud-based Parkinson disease prediction system

prediction. Then, the proposed k-NN-based classifier model is used for predicting the severity of Parkinson disease. Accordingly, it will suggest the therapist to make appropriate diagnosis pertaining to severity levels.

As shown in the architectural process, the remote patient has to upload the medical data sample using a smartphone to know whether he/she is affected by the Parkinson disease or not. After receiving the data sample at cloud layer, the foremost thing is to remove the noise present in the signal, and then go for further feature selection and classification process. After prediction, discriminated medical data is stored in the electronic health record over the cloud platform as patient history. So that the healthcare professionals and therapist could periodically monitor and diagnose the patient from remote places. When the user tries to access the healthcare system using the internet, then the cloud manager will authenticate the user and initiate the privacy preserving access controls to establish the collaboration between the user and healthcare system. Afterwards, virtual machines will be launched with ECSO-based feature extraction and k-NN-based classifier model deployments to handle the medical session in the form of web services. After disease prediction, the multimedia content server will notify the severity and assessment values during diagnosis.

3.1 ECSO-Based Feature Selection

An N-dimensional integer vector represents the important features of lively captured gait data of patients. It can belong to integer space as given in Eq. (1).

$$S^N = \{(x_1, \dots, x_n, \dots, x_N) \mid x_n \in I, x_n^{\min} \leq x_n \leq x_n^{\max}, 1 \leq n \leq N\} \quad (1)$$

Let I denote the set of integers, x_n^{\min} and x_n^{\max} represent the lower and upper limit ranges of integer variables. Here, the multi-objective optimization of this research problem can be mathematically expressed as shown in Eq. (2).

$$f(X) = \begin{cases} \text{Min}_X f_1(X) \\ \text{Min}_X f_2(X) \\ \dots \\ \text{Min}_X f_n(X) \end{cases} \quad (2)$$

where X be the vector is present over the integer space S^N . Here, the m th objective function can be represented as $f_m(X)$ for all $m \in \{1, 2, \dots, n\}$ with N-dimensional search vectors X . The solution to the minimization problem is said to be the dominate vector if and only if satisfies Eqs. (3) and (4).

$$f_p(X_u) \leq f_p(X_v), \forall p \in \{1, 2, \dots, n\} \quad (3)$$

$$f_p(X_u) < f_p(X_v), \exists p \in \{1, 2, \dots, n\} \quad (4)$$

Let X_u and X_v be the two candidate solution with its dominant vector $f(X_u)$ (represented by $f(X_u) < f(X_v)$). A solution X_u is said to Pareto optimal if it is not dominated by any other possible solution. It means there are no other possible solutions available in the space with respect to multiple objectives. Therefore, an ECSO is proposed with a multi-objective integer optimization context.

The proposed ECSO feature selection uses two sets of cats with seeking and tracing mode to find the best solution. A mixing ratio MR parameter is defined as the ratio of cats in tracing mode to the cats in seeking mode. Other parameters like MI denote the mixing iteration, φ denote the probability of asynchronous updating, N be the decision variables, and SS be the swarm size denoting the number of cats available in the pool. Let us consider the position and velocity of i th cats available in the N-dimensional space $N = \{1, 2, \dots, n\}$ can be represented as $P_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$ and $V_i = \{v_{i1}, v_{i2}, \dots, v_{in}\}$. In tracing mode, the position of copied j^{th} cat is defined by $P_j = \{p_{j1}, p_{j2}, \dots, p_{jn}\}$, and the best global position of cat swarm is defined by $P_g = \{p_{g1}, p_{g2}, \dots, p_{gn}\}$. During the seeking mode of the cats, the parameters like seeking memory pool (SMP) denotes the copies of current cat, self-position consideration (SPC) denotes whether the cat is selected again or not, count dimension change (CDC) indicates varying dimension of cat position, and selected dimension

range (SDR) represents the mutation ratio of desired dimension. Then, create $i = SMP - SPC$ copies of j th cat such as $p_{ln} : 1 \leq l \leq i, 1 \leq n \leq N$, where p_{ln} denotes the n th dimension of l th copy with mutation operation. All copies will be checked for bounds as defined in Eq. (5).

$$p_{ln} = \begin{cases} \text{round}(p_{ln}) & \text{if } \text{round}(p_{ln}) \in R_n \\ p_{ln} & \text{otherwise} \end{cases} \quad (5)$$

where R_n denotes the set of integer value range $R[p_n^{\min}, p_n^{\max}]$ with condition $p_n^{\min} \leq p_n \leq p_n^{\max}$. Finally, choose any candidate randomly from the i th copies and place it in the appropriate position of j th cat.

At the tracing mode, the i th cat will start chasing the target according to the updating of velocity as defined in Eq. (6).

$$v'_{in} = v_{in} + c_1 * r_1 * (p_{gn} - p_{in}) \quad (6)$$

Let p_{in} and v_{in} be the position and velocity of the i th cat with n th dimension, c_1 be the constant, and r_1 be the uniformly distributed range of random numbers [0, 1]. Then, the position of i th cat will get updated as stated in Eq. (7).

$$p_{in} = \begin{cases} \text{round}(p_{in} + v'_{in}) & \text{if } \text{round}(p_{in} + v'_{in}) \in R_n \\ p_{in} & \text{otherwise} \end{cases} \quad (7)$$

According to the evaluation of Pareto front from the newly built i th cat fitness, further updating is done until the current solution is non-dominating.

3.2 Decision Prediction Classifier Model

The K-NN classifier model is proposed for classification of Parkinson disease. The patient data containing incomplete patients with p assessment. Assume the patient data d_i consisting of class information represented as $\{\omega_1, \omega_2, \dots, \omega_p\}$. To estimate the d_i^g and ω_g class related to the training set T^g in which each patient fits in to ω_g . So, the incomplete patient of k-NN can be expressed as given in Eq. (8).

$$T_i^g = \{t_1^g, t_2^g, \dots, t_k^g\} \subset T^g \quad (8)$$

Let the parameter t_k^g denote the k th neighbor of the pattern d_i belonging to training subset T^g . As a result, the total p assessment of the pattern d_i can be defined by d_i^g . To obtain the class estimation ω_g with a bigger distance will lead to a discounting factor with weight function as defined in Eq. (9).

$$\alpha_k^g = \frac{e^{-\|d_i - t_k^g\|}}{\sum_{k=1}^K e^{-\|d_i - t_k^g\|}} \tag{9}$$

where $\|\cdot\|$ denotes the Euclidean distance. After obtaining the weight α_k^g and class ω_g of each neighbor by the k-NN classifier, the missing pattern of ω_g can be estimated as given in Eq. (10).

$$d_{io}^g = \sum_{k=1}^K \alpha_k^g \cdot t_{ko}^g \tag{10}$$

where d_{io}^g denotes the o th missing attribute of d_i pertaining to complete attribute in t_k^g . Based on the complete estimation vector, the sub-classification result of the classifier model over p pieces could result in d_i as given in Eq. (11).

$$P_i^g = \lceil(d_i^g|T) \tag{11}$$

Let \lceil be the standard classifier model exploited to classify the d_i^g , P_i^g be the output classification obtained in the form of probability value of d_i^g .

To measure the performance of k-NN classifier models, evaluation metrics like sensitivity, specificity, and accuracy are measured as defined in Eqs. (12), (13), and (14), respectively.

$$C_{\text{Sensitivity}} = \frac{T^P}{T^P + F^N} \tag{12}$$

$$C_{\text{Specificity}} = \frac{T^N}{T^N + F^P} \tag{13}$$

$$C_{\text{Accuracy}} = \frac{T^P + T^N}{T^P + F^P + T^N + F^N} \tag{14}$$

where T^P , T^N , F^P , and F^N represents the number of true positive, true negative, false positive, and false negative obtained during classification. In future, research study can be enhanced with behavioral learning negotiation framework for choosing the best and cost-effective healthcare service by negotiating with multiple healthcare service providers [11, 12].

4 Experimental Evaluation

Experimental implementation of ECSO feature selection and k-NN classifier model is implemented using python coding. In order to verify the effectiveness of the proposed healthcare system, benchmark dataset containing extracted gait features are taken from the public repository [13]. The dataset consists of a 10 s accelerometer measurement of Parkinson disease patients taken by keeping their hands in the position of postural tremor. Also, their arms are outstretched with palms facing in upward direction. Afterwards, this dataset is supplied to k-NN classifier model by extracting the necessary features using ECSO approach for further training purpose. After training, the new patient's sample is given at run time that acts as a testing dataset. Then, the model will predict whether the new patients' samples are affected by Parkinson or not. The observed results of various classifier models with respect to prediction time and prediction accuracy are given in Table 1.

The obtained result shows that the proposed k-NN classifier model with ECSO feature selection method takes very less prediction time over the existing classifiers. Similarly, the proposed k-NN classifier model outperforms the existing classifier models with respect to prediction accuracy. This type of disease severity level prediction could bring some impact on diagnosing the Parkinson disease. Moreover, the proposed cloud-based Parkinson disease prediction system presented a prediction model with the aid of machine learning to make improvement over the human accuracy and scope of early detection. It can be more useful in developing rapid prediction models that can help a doctor to minimize the long process of diagnosis and also eradicate the human error. In future, a hybrid wrapper feature selection method can be used to improve the performance of proposed classifier model to implement in the real-time tele-monitoring healthcare applications. Therefore, this research study provides disease identification from remote places, and also diagnoses the disease with low cost investment by saving doctors risk and increasing patient appointments.

Table 1 Performance of classifier models

Classifier models	Feature selection	Prediction time (Seconds)	Prediction accuracy (%)
Proposed k-NN	PCA	69	84.56
	LDA	64	91.65
	ECSO	62	93.82
NB	PCA	85	74.34
	LDA	82	79.20
	ECSO	78	81.57
SVM	PCA	89	81.77
	LDA	82	83.20
	ECSO	76	90.98

As per observation, the performance of various classifiers were measured with respect to the disease prediction time as shown in Fig. 2. It is very much clear from the graph, the proposed k-NN classifier with ECSO feature selection method takes very less prediction time compared to existing NB and SVM classifier models with PCA and LDA feature selection methods. Since the proposed ECSO feature selection method arrives to optimally possible solutions available in the space while comparing to any other methods in terms of multiple objectives. Similarly, the performance measurements of various classifiers were made with respect to the disease prediction accuracy as shown in Fig. 3. More evidently, the performance graph shows that the proposed k-NN classifier with ECSO feature selection method obtains high prediction accuracy against the existing classifier models with PCA and LDA feature selection

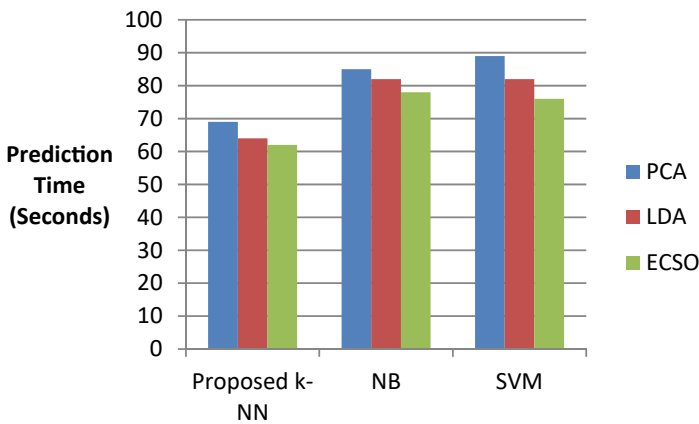


Fig. 2 Cloud-based Parkinson disease prediction system

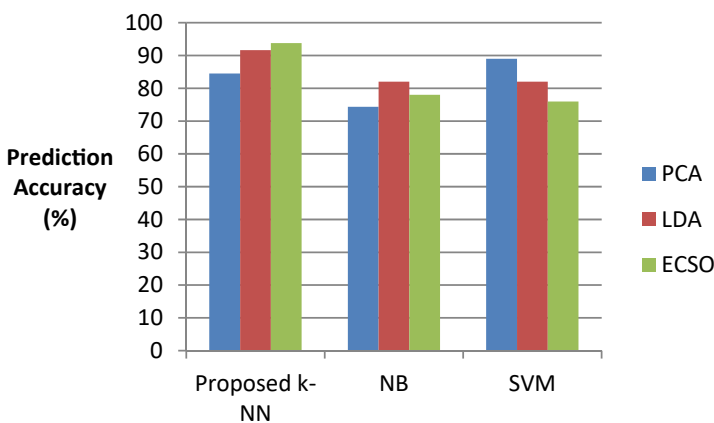


Fig. 3 Cloud-based Parkinson disease prediction system

methods. This dramatic improvement is achieved by the proposed classifier model due to appropriate management of medical data containing incomplete patents with an assessment.

5 Conclusion and Future Works

The proposed research study presents the cloud-based Parkinson disease prediction system to identify and diagnose the various stages of Parkinson disease. In order to minimize the misclassification of disease and to assist the neurologist, a k-NN classifier model is explored. This research problem needs to handle the multiclass classification and quantify the Parkinson disease severity stages over the higher dimensional data. Therefore, the proposed ECSO-based feature selection technique from the medical data set could identify the optimized features for directly applying in k-NN classifier model. Moreover, the performance of the proposed classifier model with feature selection technique is compared against the existing classifier models. According to experimental evaluation, the proposed k-NN classifier outperforms the existing models in terms of prediction time and accuracy. In future, deep feature selection and Bayesian-based approaches can be explored to rank the features of original data and diagnose the Parkinson disease problem.

References

1. Gondim, I. T. G. O., Souza, C. C. B., Rodrigues, M. A. B., Azevedo, I. M., Coriolano, M. G. W. S., & Linsa, O. G. (2020). Portable accelerometers for the evaluation of spatio-temporal gait parameters in people with Parkinson's disease: An integrative review. *Archives of Gerontology and Geriatrics*, *90*, 104097.
2. Balaji, E., Brindha, D., & Balakrishnan, R. (2020). Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease. *Applied Soft Computing*, *94*, 106494.
3. Asuroglu, T., Acıci, K., Erdaş, C. B., Toprak, M. K., Erdem, H., & Ogul, H. (2018). Parkinson's disease monitoring from gait analysis via foot-worn sensors. *Biocybernetics and Biomedical Engineering*, *38*(3), 760–772.
4. Tejeswinee, K., Jacob, S. G., & Athilakshmi, R. (2017). Feature selection techniques for prediction of neuro-degenerative disorders: A case-study with Alzheimer's and Parkinson's disease. *Procedia Computer Science*, *115*, 188–194.
5. AlMamun, K. A., Alhusein, M., Sailunaz, K., & Islam, M. S. (2017). Cloud based framework for Parkinson's disease diagnosis and monitoring system for remote healthcare applications. *Future Generation Computer Systems*, *66*, 36–47.
6. Li, Y., Dai, W., & Zhang, W. (2020). Bearing fault feature selection method based on weighted multidimensional feature fusion. *IEEE Access*, *8*, 19008–19025.
7. Tumar, I., Hassouneh, Y., Turabieh, H., & Thaher, T. (2020). Enhanced binary moth flame optimization as a feature selection algorithm to predict software fault prediction. *IEEE Access*, *8*, 8041–8055.

8. Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(281), 1–16.
9. Ozcift, A. (2012). SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. *Journal of Medical Systems*, 36, 2141–2147.
10. Ye, F. (2018). Evolving the SVM model based on a hybrid method using swarm optimization techniques in combination with a genetic algorithm for medical diagnosis. *Multimedia Tools and Applications*, 77, 3889–3918.
11. Rajavel, R., & Mala T. (2021). Agent-based automated dynamic SLA negotiation framework in the Cloud using the stochastic optimization approach. *Applied Soft Computing*, 101, 107040.
12. Rajavel, R., & Mala, T. (2016). Adaptive probabilistic behavioural learning system for the effective behavioural decision in cloud trading negotiation market. *Future Generation Computer Systems*, 58, 29–41.
13. Devpost. (2017). Parkinson's telemonitoring—monitor severity of patients' Parkinson's tremors through their mobile phones. *MedHacks*. <https://devpost.com/software/parkinson-s-telemonitoring>

Electric Vehicle Monitoring System Based on Internet of Things (IoT) Technologies



Yogesh Mahadik, Mohan Thakre, and Sachin Kamble

Abstract Electric vehicles (EVs) are emerging as a preferred way to reduce environmental concern's needs. Concern and energy insufficiency, and in the foreseeable term, this pattern is expected to grow. However, this inadequate charging infrastructure is now a major barrier to the adoption of EVs. Deployment of this infrastructure is expected to maximize the adoption of EVs to promote community access. Connectivity between charging substations (CS) is therefore mandatory. Entertainment real-time status of CSs can provide useful information, such as availability of charging provisions, reserves, and time to meet the SC. The purpose of this paper is to have a better solution related to EV charging mechanism leveraging the benefits of the Internet of Things (IoT) technologies. The IoT is a model that gives the present facilities a real-time worldwide communication view of the physical world employing the sensors and the transmitting networks. This article suggests a real-time server-based prediction of EV infrastructure.

Keywords Charging substations · Forecasting of system · EV charging · Internet of things

1 Introduction

Countries around the world have introduced in recent years regulatory criteria for solving transport problems, such as the increasing non-renewable fuel demand and metropolitan emissions. So, a lot of attention was paid to eco-friendly EVs due to availability issues of fossil fuel, climate, and pollution control. Many countries have laid down essential criteria for facilitation of EV extension and approval [1, 2]. For instance, several North American cities are in the evolving time to easily substitute hybrid plug-in conventional cars and EVs [3], then. EVs have great cost advantages and convenience. EVs have ease, in comparison to internal combustion engine (ICE)

Y. Mahadik (✉) · S. Kamble
Government Polytechnic Malvan, Malvan, MH, India

M. Thakre
K. K. Wagh Institute of Engineering Education and Research, Nashik, India

automobiles, and maintenance cost reduces for an EV [4]. The range of EVs also expands faster than in the past. The lightweight Mahindra Reva car with 16 kWh power will have 80 miles of driving range. In Model Tesla X (SUV) vehicles, the mileage with a power of 94 kWh is 400–450 km. Figure 1 indicates a block diagram of the system. Though the driving ranges of the EVs depend on the capability of the battery, the complexity of the powertrain plays a significant part. New recent study concentrates on improving driving battery control and management systems [5]. EVs spectrum needed a detailed review on the Li-ion batteries used. In [6], discussed with EVs are the fundamental principles of service, assembly, and the efficiency of several Li batteries. In [7], fast CSs can load battery from 20% to about 80% of the state of charge level (SOC) 30 min. Naomi et al. studied the best loader mode in [8]; this ensures quality and minimal time for charging electric Cars and Hybrid Vehicles (HEVs). SOC is one of the major players in delivering the present battery condition for safe charging and discharge that increases battery life in turn [9, 10]. Different methods of SOC estimation are discussed in [11]. In the Lobby Automotive business, intelligent communication networks (1) vehicle-to-infrastructure (V2I), (2) vehicle-to-driver (V2D), and (3) Vehicle to vehicle (V2V) are the types of networks. Due to major government expenditures, connectivity systems such as V2V and V2I were well studied. The key purpose of using these communication mechanisms is to minimize congestion of roads, preservation of road transport [12], and avoid collisions with cars [13, 14]. However, few works confirm the reliability of vehicle-to-device systems in an EV, and additional work is required in this field of study. The data correspondence deliberately serves the EV customer and the CS effective power control scheduling. The SOC calculation information to assign the slot is transmitted to the closest CS based on grid power demand. In [15, 16], the approximate SOC information is sent to three separate operators, to exchange the EV information, the supplier of charging facilities (CSP), delivery system retailer (RET), and Operator (DSO). EV user info is transmitted to an automatic charge controller infrastructure of connectivity such as the Internet or cellular networks [17]. After all, information has been obtained; the charge controller has an extra degree of flexibility to tax the EV. In [18], the article aggregator has a precondition to alert the aggregator besides the planned departure with the available SOC period when a car is attached in or out. The lightweight charging of EVs was carried out in [19]. Optimum online preparation was introduced in [20] using a cost-responsive fee method. Thereby, the EV plug-in pattern can be explored by the aggregator. Along with the advice of the CS, the algorithm suggested can alert the EV driver to the safe SOC limit, and a detailed calculation of the EV charging time can be given. For EV, the geographical positioning and path views of various available CSs are required. The disadvantages of the current EV Charging methods are – EV location, user phone number, information of the vehicle, and sufficient charging power can trigger different. EV load estimation is a crucial problem for efficient operation and energy business executives. EV charging has been advised based on an accurate range predictor system in real-time. Not in the static, the device cannot provide the driver with any details of CS, i.e., CS works or does not work, how many cars are waiting for charging, and what charging points are available. From the above, it was observed that an application

is in real-time Battery status EV needed to collect transportation details and the CS status for the estimation of future CS.

This work suggests a server-based forecast in real-time Application: (i) the administration of scheduling to avoid waiting time; and (ii) have a CS recommendation in real-time.

Economic risks, lowered charges for night time EVs, these benefits relative to other approaches—this is what we are talking about. The PHP programming generates a synergetic application language of the operating system Linux UBUNTU 16.04 LTS and cloud processes and handles all related details of Google cloud platform structured query language (CSQL). The utility of this application is also verified by a low-cost LTC 4150, Wi-Fi ESP 8266 trial device, and Adriano.

2 System Description

The emphasis of this paper is on the IoT aspect of deciding the value of the SoC and data transfer to IoT networks. The data can be accessed inside the App. The user can also find the Locations of the nearest charging station using the smartphone. The user once knowing the state of his car battery, can quickly assess whether to continue with the supply of power to the grid or based on the tariff rates, sinks electricity from the grid. There will be various prices for selling electricity to the grid and power is taken from the grid. Bidirectional, the grid would have converters for power transfer [5, 6]. Some grids are available, which uses solar energy as a source as well [7]. Figure 1 indicates a block diagram of the system.

The significant and visible connectivity, sensors, and small devices are all IoT features. New enabling networking technologies, especially IoT networking is a kind of network that is not connected to major network suppliers [8]. Small and large networks are created by IoT. IoT uses sensors to collect the data from its machine modules over a wide area. It is the primary working machine, sensor or controller is used. More controllers, such as Arm Mbed and Arduino, achieve more in the area of IoT, such as phones or tablets. The output or outcomes are used to view the output that minimizes the effort to get the details.

2.1 EV Load Forecasting and SOC Measurement

Load prediction is an important process that can improve distribution network performance. The precision of load forecasting has an immense effect on the functioning of an organization. In general, the cost depends on the peak request. During a precise time, a particular CS may have additional power demand. It must be noted that the unit cost is expressed in kilowatt-hour rupees (Rs/kWh), where the rupee is India's official currency.

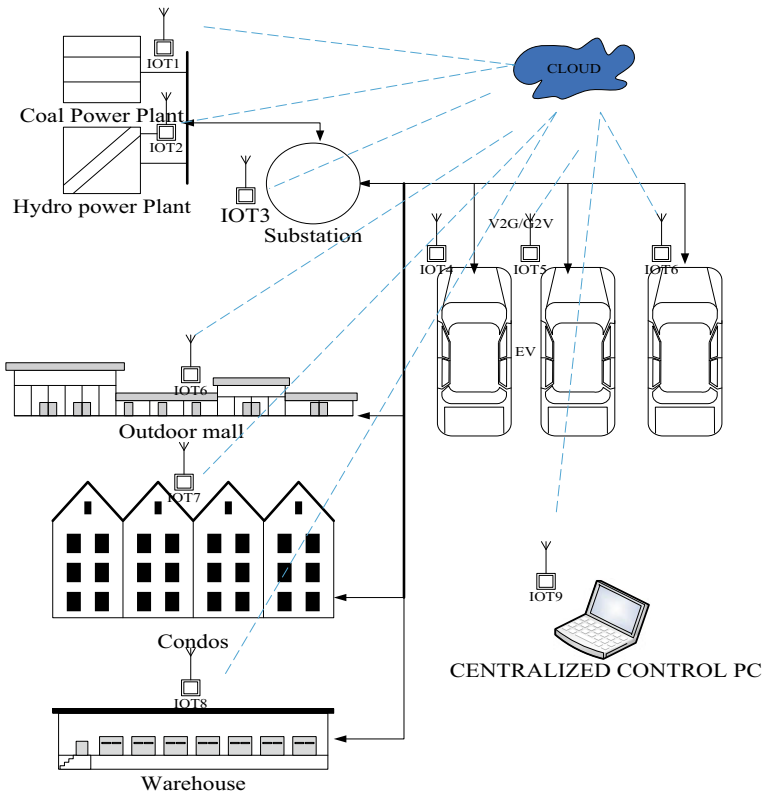


Fig. 1 Block diagram of IoT-based V2G and G2V system

The communication network proposed would work to minimize the variations. It updates the user about the accessibility of cost and slots for each CS. A consumer can therefore prefer a station where there is local supply. Of course, if intimated, the accessibility of slots will prevent congestion over a precise period impressively during peak hours at the CSs. Therefore, inadequate number of CSs can lead to load forecasting complications in the delivery of energy to EV consumers. The energy delivery crisis can arise from the debate above and affect the recipient of the EV. A stronger battery has been able to solve this problem. Therefore, the battery SOC is to be tracked. Evaluation of the SOC is a big concern for the use of the battery. The available SOC of a battery provides main control method parameters [21–23]. The output of the battery is calculated by the SOC so that the exact SOC calculation can improve life and also save energy by activating the device intelligent management strategy [24]. But the important thing is the battery problem that is driven by chemical energy, which cannot be directly measured. This contributes to the difficulty in the SOC calculation.

There are several methods for determining a battery's SOC Setting [10, 11]. These techniques can be categorized further as: (I) electrochemical, (II) adaptive, (III) Electrical. The electrochemical techniques are extremely precise but need access to the chemical composition of the battery. So, the technology is difficult to introduce. Techniques of adaptation need a circuit model equivalent and solution procedure battery. The precision of the adaptive techniques is determined by the equivalent model efficiency. Electrical techniques only need measurable parameters like this present load/to unload terminal voltage and internal resistance. Because of its simple application and low complexity, Coulomb counting (electrical) is one of the most common SOC estimation techniques. This approach is known as well as the new integration technique of ampere-hour (Ah). An algorithm is complex and easy to deploy. The current that is charged or is measured in a cell is known as the current integration technique. If the first SOC is known, correct SOC can then be obtained [25, 26]. The current readings are incorporated mathematically, SOC values can be used to calculate as normal the usable SOC is found by measuring the moving charge. When determining the battery cell, say, it has 60 Ah, 216,000 A seconds can be estimated. Coulomb count 0% SOC thus means battery counts 0 C and 100% SOC is a count of the 216,000, which is used for calculating the charge rate. The method of counting coulomb is simple and of low complexity. In theory, the procedure for coulomb counting is incorporation and charging quantification by detecting the battery's current input and output. It operates by integrating the current active over time to measure the cumulative amount of battery power input or output. This leads to a capacity measured usually in ampere-hours. The method is sound, as long as it is consistent with current measurements, it can be used on all the EV device batteries [27, 28]. The instrument tracks the current through an external resistor between the positive end of the battery and the load. On the Volt-Frequency converter, it transforms the present sensory stress through several interrupt pin output pulses. These pulses fit at a fixed charge quantity to or from the battery. The device has an Arduino connection. The exit of the digital format is available for Arduino codes. Surveillance of the SOC is important because the remaining is used to measure energy and the maximum possible driving range to achieve the best output (speed or distance).

3 Proposed IoT-Based EV Management

By the year 2030, the GSM Association has estimated that every EV will be connected, and this will earn living, earnings of over 19B euros. Several criteria and standards have been set in recent years. Now technologies have emerged for V2V and V2I communications. Various forms of communications already available for exchange are V2G contact to achieve a successful EV charging based on standards of the Society of Automotive Engineers (SAE), the International Association for Standards (ISO), and Proposals from the International Electrotechnical Commission (IEC) [29, 30]. Several standards have been defining different communication

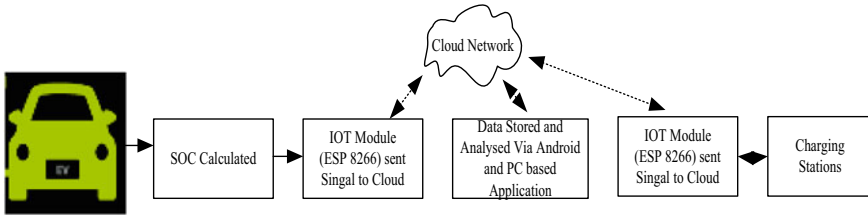


Fig. 2 EV management unit

processes. The basics are different norms, such as charging contact costs, protocols, connector charging, protection, and charging topology [31, 32]. The developed application has three divisions (web, android), (i) SOC estimation, (ii) CS selection, and (iii) database selection and the processing of data. The total planned flowchart of the scheme is portrayed in Fig. 2.

The Coulomb counter of the LTC is related to the Adriano and this will measure the battery’s SOC. The information read by the laptop or PC will be stored in the server database using the IoT Wi-Fi module ESP8266. It will run via an Internet protocol at the same time. Depending on the request of the user, the web page will display the CSs positions with the aid of GPS. Based on availability, the user will have the choice of slots, charging costs, and distance, choosing a specific CS. Once the request is received by the CS from the EV, the charging slot will be assigned based on the requirements for the infrastructure for EV charging. This system is built to ensure the ideal EV charging, and it would also boost the grid by avoiding peak charges.

3.1 Test Method for Calculating SOC

The experiment is conducted with a battery of 9 V, 300 mAh Li-ion, Arduino Uno, ESP 8266, and a 9 V DC motor with a coulomb counter. The Li-ion battery is currently rated at 300 mAh, which means, for 60 min, it can supply 300 mA, and this is said to be the battery cell’s initial charge. The load on the motor is 70 mAh, and with the charge available, it can run for 4.28 h. Now the price for the battery charge is given by:

$$C_c = R_p * E_c(t) \tag{1}$$

where, $E_c(t)$ is the cost of energy at a given hour.

If the motor is running for 1 h, so 70 mAh will be consumed.

The remaining SOC value, therefore, is 76.66% of the observed capacity, it is 0.63 kWh.

3.2 Algorithm of CS Selection

Relevant information will be given by the CS selection algorithm to the end user's EV. When engaging the charging points in the CSs with the EV, the points convey data to the server. Large networks known as network service providers (NSPs) help in sharing of knowledge between CSs and the server. The NSP is linked to exchanges in the metropolitan area (MAEs) via the access points of the network (NAP's). The initial NAP's were the internet links the dots. The references to both NAPs and MAEs are as points for internet sharing (IXs). Following the data, the domain name system will be passed through (DNS). The DNS records the names of the computers and their corresponding IPs on the internet, emails. Finally, to guarantee the transferred data stays private, it uses a stable socket layer (SSL). In the meantime, the EV position information is tracked via GPS. A significant parameter is the battery drain rate. To calculate the SOC that is the available form of Coulomb counting research. It is used to measure the rate of battery drain and the SOC available at routine intervals. With the open SOC, it is possible to calculate the remaining distance which the EV will cover. The results are stored in the localhost and the same will be communicated with the application that was created. The position of EVs and CSs are carried to the application using the Maps and the API. An API is a code that makes two distinct APIs computer applications for communicating with one another using a hypertext transfer protocol to access the API (HTTP). The website uses protected information to ensure the security of the information Protocol of hypertext transfer (HTTPS). Therefore, everything in the application has reviews from the CSs and the EVs. An Index of the Availability of the slot is given, and this will be compared with the remaining power of the battery to find the optimized CSs.

Below is the step-by-step method.

Step 1: Dynamically collecting the input from the CSs and EVs for a one-minute cycle.

Step 2: The percentage of the SOC is the key input from which the battery will calculate the drain rate (BDR).

Step 3: The approximate distance to empty, based on the BDR (DTE). As given in Eq, it will be determined.

Capacity of the battery (B_c) = X (kWh)

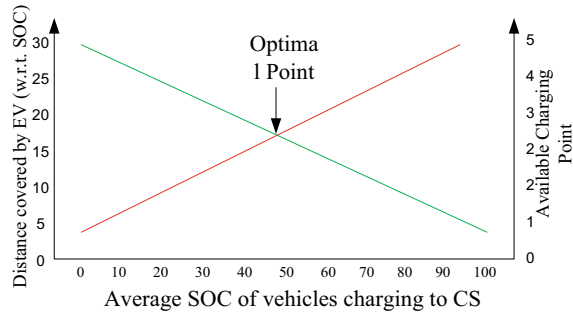
$$\text{Approximate Range } (E_r) = Y \text{ (km/kWh)} \quad (2)$$

$$(\text{DTE}) = \text{km } (X * Y * \text{SOC}) \quad (3)$$

Step 4: The overall distance measured (DTE) will be that with the available SOC and any EV up to CS, the EV will drive inside the limit will be drawn.

Step 5: Then in the CS, an index of availability (AI) will be retrieved for each station.

Step 6: Both the DTE and AI details are taken into consideration and optimized, as seen in the sample as shown in Fig. 3.

Fig. 3 Optimal point finding

Step 7: The EV will fly up to a maximum distance of 27 km SOC Remaining. The battery would be empty at the average speed, 40 min, and the EV would not be able to move. In the meantime, the availability of charging slots also varies with the individual CS.

Step 8: The booking option and all related details of the booking option CS are shown for each CS, this process is repeated.

Step 9: The station will be provided with the lowest optimal point as the best choice, and in ascending order, the results will be sorted. Now as needed, the EV user can pick the CS. (Note: The calculation of the total time is given in the appendix for two cases in which a user chooses a CS based on the shortest distance required).

It should be noted that if an EV user hits the booked user, CS will use reserve points approved by the CS for EV earlier, no-booking, or early-arrival users. Such points of reserve support the CS with quick chargers in the event of an emergency and anomalies. Non-flexible booking drivers can pay, however, using the program rather than drivers. Time-based arrival in such instances uses the priority (ATP) algorithm. ATP enables the Once vehicles to arrive, they enter service, and therefore reduce the wait. It takes time and prevents the charger from being idle. Lineups for ATP algorithms continued until all the charging points are connected by all the vehicles. Engaged based on the vehicles arrival time. Once a vehicle is present, the algorithm, assigned to a charging point, updates the waiting points. For other cars, algorithm will repeat with the same procedure before it charges all the cars. Figure 4 reveals the Flowchart of the algorithm of ATP, the EV charging plan.

3.3 Applications

(i) Provide schedule management to reduce waiting time; and (ii) provide real-time management; CS advised for less distance and shortened charging time. Moreover, the proposed scheme prevents interference by third parties and preserves the anonymity and complexity of EV users. Exchange of information between the customer and the CS. End-users can use the CS conveniently based on their requirements for this. This synergistic application is built in the programming language

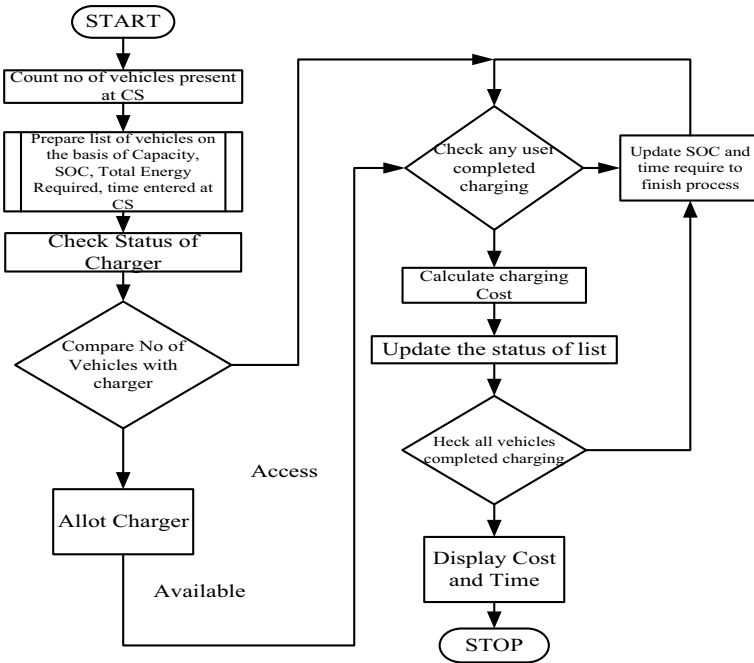


Fig. 4 ATP-based algorithm

of PHP, Linux UBUNTU 16.04 LTS operating system, and all related information shall be stored and handled with Cloud Structured Query Language (CSQL) from the Google cloud platform.

4 Conclusion

Increasingly, the IoT paradigm is integrated with real-world applications, particularly in the real world, in the field of electricity. In this article, a real-time solution is suggested that, the charge scheduling of EVs should be improved based on modern state-of-the-art with reliable communication between vehicles and CSs-Technologies-the-art. More protection is offered by the proposed scheme, it is possible to stop unnecessary rush in charge stations. The user will receive the data of all nearby CSs based on various variables, such as consumer comfort, electricity variations, the user’s ability to pay, and prices This robust framework not only is useful for EV users, but also CSs to predict future loading, congestion avoidance, and energy consumption with reducing and/or balancing loads. To ensure the stability of the grid, V2V, and V2V in future work, the developed application can be integrated into V2G. The introduction of smarter grids has been slowed down. The mentality is that we need to provide some form of contact from each unit back to the operator of

the grid. Furthermore, the efficient processing of data from both EVs and CSs takes time. To achieve a solution that is optimized, the appliances, therefore, by linking to a web page anywhere, it can be made intelligent for that client, and then a computer is programmed to respond to that client, also suitable for the grid operator to better balance the load and grid.

The complexity and computation pressure of this centralized approach would exponentially increase due to the rising number of EVs. Since the cellular network is inexpensive, it can cause over-congestion that will degrade the consistency of the relationship.

References

1. Thomas, C. E. S. (2009). Transportation options in a carbon-constrained world: Hybrids, plug-in hybrids, biofuels, fuel cell electric vehicles, and battery electric vehicles. *International Journal of Hydrogen Energy*, 34(23), 9279–9296.
2. Dickerman, L., & Harrison, J. (2010). A new car, a new grid. *IEEE Power Energy Magazine*, 8(2), 55–61.
3. Vrazic, M., Vuljaj, D., Pavasovic, A., & Paukovic, H. (2014). Study of a vehicle conversion from an internal combustion engine to electric drive. In *IEEE International Energy Conference* (pp. 1544–1548).
4. Dost, P. K-H., Spichartz, P., & Sourkounis, C. (2018). Charging behavior of users utilizing battery electric vehicles and extended-range electric vehicles within the scope of a field test. *IEEE Transactions on Industry Applications*, 54(1).
5. Hannan, M. A., Hoque, Md. M., Hussain, A., Yusof, Y., & Ker, P. J. (2018). State-of-the-art and energy management system of lithium-ion batteries in electric vehicle applications: Issues and recommendations. *IEEE Access*, 2018, 19362–19378.
6. Navid, R., Kashanizadeh, B., Vakilian, M., & Barband, S. A. (2013). Optimal placement of fast-charging station in a typical microgrid in Iran. In *International Conference on the European Energy Market* (pp. 1–7).
7. Mohamed, N., Aymen, F., Mouna, B. H., & Alassaad, S. (2017). Review on the autonomous charger for EV and HEV. In *International Conference on Green Energy Conversion Systems*, 2017, 1–6.
8. Ramadan, H. S., Becherif, M., & Claude, F. (2017). Energy management improvement of hybrid electric vehicles via combined GPS/rule-based methodology. *IEEE Transactions on Automation Science and Engineering*, 14(2), 586–597.
9. Gholizadeh, M., & Salmasi, F. R. (2014). Estimation of state of charge unknown nonlinearities and state of health of a lithium-ion battery based on a comprehensive unobservable model. *IEEE Transactions on Industrial Electronics*, 61(3), 1335–1344.
10. Tang, X., Liu, B., & Gao, F. (2017). State of charge estimation of LiFePO₄ battery based on a gain-classifier observer. In *Energy Proceedings-105* (pp. 2071–2076). <https://doi.org/10.1016/j.egypro.2017.03.585>.
11. Di Lecce, V., & Amato, A. (2011). Route planning and user interface for an advanced intelligent transport system. *IET*, 5(3), 149–158.
12. Taleb, T., Benslimane, A., & Letaief, K. B. (2010). Toward an effective risk-conscious and collaborative vehicular collision avoidance system. *IEEE Transactions on Vehicular Technology*, 59(3), 1474–1486.
13. Barrios, C., Motai, Y., & Huston, D. (2015). Trajectory estimations using smartphones. *IEEE Transactions on Industrial Electronics*, 62(12), 7901–7910.

14. The International Electrotechnical Commission (IEC) (2001) Electric vehicle conductive charging system—Part 1: General requirement. In *IEC 61851-1 the International Electrotechnical Commission*.
15. Clairand, J., Rodríguez-García, J., & Álvarez, B. C. (2018). Smart charging for electric vehicle aggregators considering users' preferences. *IEEE Access*, 6, 54624–54635.
16. Roterling, N., & Ilic, M. (2011). Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets. *IEEE Transactions on Power System*, 26(3), 1021–1029.
17. Han, S., Han, S., & Sezaki, K. (2010). Development of an optimal vehicle-to-grid aggregator for frequency regulation. *IEEE Transactions on Smart Grid*, 1(1), 65–72.
18. Sundstrom, O., & Binding, C. (2012). Flexible charging optimization for electric vehicles considering distribution grid constraints. *IEEE Transactions on Smart Grid*, 3(1), 26–37.
19. Shaobing, Y., & Yang, S. (2019). Price-responsive early charging control based on data mining for the electric vehicle online scheduling. *Electric Power Systems Research*, 167, 113–121.
20. Xydas, E. S., Marmaras, C. E., Cipcigan, L. M., Hassan, A. S., & Jenkins, N. (2013). Forecasting electric vehicle charging demand using support vector machines. In *Proceedings of the 48th International Universities' Power Engineering Conference*.
21. Quiros-Tortos, J., Ochoa, L., & Lees, B. (2015). A statistical analysis of EV charging behavior in the UK. In *Innovative Smart Grid Technologies Latin America (ISGT LATAM) 2015 IEEE PES* (pp. 445–449).
22. Cai, Z. H., Liu, G. F., & Luo, J. (2010). Research state of charge estimation tactics of nickel-hydrogen battery. In *Proceedings of International Symposium Intelligence Information Processing and Trusted Computing* (pp. 184–187).
23. He, H. W., Xiong, R., & Guo, H. Q. (2012). Online estimation of model parameters and state-of-charge of LiFePO₄ batteries in electric vehicles. *Applied Energy*, 89(1), 413–420.
24. Hu, X., Zou, C., Zhang, C., & Li, Y. (2017). Technological developments in batteries: A survey of principal roles types and management needs. *IEEE Power Energy Magazine*, 15(5), 20–31.
25. Maurya, D. S., Jadhav, P. D., Joshi, R. S., & Bendkhale, R. R. (2020). A detailed comparative analysis of different multipulse and multilevel topologies for STATCOM. In *International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1112–1117). Coimbatore, India. <https://doi.org/10.1109/ICESC48915.2020.9155708>.
26. Hadke, V. V., & Thakre, M. P. (2019). Integrated multilevel converter topology for speed control of SRM drive in plug-in-hybrid electric vehicle. In *Proceedings of the 3rd IEEE International Conference on Trends in Electronics and Informatics (ICOEI 2019)* (pp. 1013–1018). SCAD Clg. of Engg. & Tech., Tirunelveli, Tamil Nadu, India CFP19J32-ART; ISBN: 978-1-5386-9439-8.
27. Mahadik, Y. V., & Vadirajacharya, K. (2017). Battery life enhancement using hybridization of battery and UC. In *International Conference on Circuit, Power and Computing Technologies (ICCPCT)*. <https://doi.org/10.1109/iccpct.2017.8074393>.
28. Mahadik, Y., & Vadirajacharya, K. (2019). Battery life enhancement in a hybrid electrical energy storage system using a multi-source inverter. *World Electric Vehicle Journal*, 10(2), 17. <https://doi.org/10.3390/wevj10020017>
29. Vaibhav, S. G., & Mohan, P. T. (2019). A novel 4-level converter for switched reluctance motor drive in plug-in HEVs. In *IEEE International Conference on Intelligent Computing and Control Systems (ICICCS 2019)* (pp. 626–631). Vaigai College of Engineering, Madurai, Tamil Nadu, India. ISBN: 922-1-7281-2214-6.
30. Mahadik Y. V., Yeole D. S., & Chowdhary P. K. (2020). Fast charging systems for the rapid growth of advanced Electric Vehicles (EVs). In *International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India* (pp. 1–6). <https://doi.org/10.1109/ICPECTS49113.2020.9336979>.
31. Chowdhary, P. K., & Thakre, M. P. (2020). MMC-based SRM drives for hybrid-EV with decentralized BESS in battery driving mode. In *International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)* (pp. 1–6). Chennai, India. <https://doi.org/10.1109/ICPECTS49113.2020.9337029>.

32. Mane, J., & Hadke, V. (2020). Performance analysis of SRM based on asymmetrical bridge converter for plug-in hybrid electric vehicle. In *International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)* (pp. 1–6). Chennai, India. <https://doi.org/10.1109/ICPECTS49113.2020.9337059>.

Dynamic Analysis and Projective Synchronization of a New 4D System



M. Lellis Thivagar, Ahmed S. Al-Obeidi, B. Tamarasaran,
and Abdulsattar Abdullah Hamad

Abstract A new 4D dissipative hyperchaotic system with an unstable equilibrium point is introduced. The proposed system consists of ten terms including three quadratic nonlinearities which constructed through using a nonlinear state control algorithm in the known Lorenz system, exhibits self-excited attractors and chaotic system attractors, with two positive derivations of Lyapunov. The dynamical properties of this system are analyzed using theoretical and numerical simulations based on equilibrium points, stability, dissipative, Lyapunov exponents, and phase portrait. Besides, various coexisting attractors or multistability with different initial conditions under the same parameters are investigated. Furthermore, Projective Synchronization (PS) of an identical proposed system is realized by nonlinear control strategy and Lyapunov stability theory.

Keywords 4D hyperchaotic system · Self-excited attractors · Multistability · Projective synchronization

1 Introduction

The attractors of nonlinear dynamic systems have attracted a lot of attention from researchers due to their possible control and optimization applications [1, 2], electronic circuits [3–6], conforming dynamics [7], and encryption [8–11]. One of the important applications in secure communication is messages with simple chaotic systems are not always safe. To overcome this problem, a higher n -dimensional system ($n \geq 4$) with more than two positive Lyapunov exponents are required to increase randomness and unpredictability [12–17].

Since 1963, Lorenz system is the first physical of a 3D chaotic attractors, and other 3D chaotic systems were discovered such as the systems of Chen (1999) Lu

M. L. Thivagar · B. Tamarasaran · A. A. Hamad (✉)
Madurai Kamaraj University, Madurai, India

A. S. Al-Obeidi
Specialty of Mathematics, Gifted School of Nineveh, Nineveh, Iraq
e-mail: ahmedsedeeq@uomosul.edu.iq

(2002), Liu (2004), Pan (2009), all of these mentioned systems contain only one positive Lyapunov exponent (LE) [18–20]. Rössler system is considered the first 4D hyperchaotic system with two positive LE which was introduced in 1979. The 4D hyperchaotic systems are more complex and interesting compared with other chaotic systems because it contains two positive LE to take advantage of it in the secure communication. Several systems have been introduced with various types of the dimension: 4D [12–15], 5D [21–23], and 6D [24–27].

The exponent of these systems is calculated from Wolf Algorithm [27], and it plays an essential and effective role in classification of the attractors of these systems as periodic, quasi-periodic, chaotic, hyperchaotic, and chaotic with 2-torus. Every 4D system is possible to be chaotic or hyperchaotic attractors based on the sign of LEs. So, the sign of LEs (+, +, 0, –) refers to hyperchaotic, whereas the sign (+, 0, –, –) indicates to chaotic attractor. A hyperchaotic system with n -dimensional ($n \geq 4$) and $n-2$ positive Lyapunov exponent is considered the best system.

The above discussion motivated us to introduce a new 4D system. The following points summarize the results of the paper:

- A new 4D dissipative system with $n - 2$ of positive Lyapunov exponent is constructed;
- The proposed system has self-excited attractors;
- Multistability of the new system is studied with various types of initial conditions; and
- Projective Synchronization (PS) of an identical proposed system is realized utilizing nonlinear control and Lyapunov stability theory.

2 Description of the System

Lorenz chaotic system is the first-ordinary differential equation which has three-dimensional depicted as:

$$\begin{cases} \dot{x}_1 = -ax_1 + ax_2 \\ \dot{x}_2 = cx_1 - x_1x_3 - x_2 \\ \dot{x}_3 = -bx_3 + x_1x_2 \end{cases} \quad (1)$$

In which $x_i, i = 1, 2, 3$ are variables, and a, b, c positive parameters, device (1) is chaotic at the parameters $a = 10, b = \frac{8}{3}, c = 28$. By adding nonlinear feedback controller strategy to Lorenz system, a new 4D system with (+, +, 0, –) sign of LEs is obtained and consists of ten terms including three nonlinearities which is depicted as:

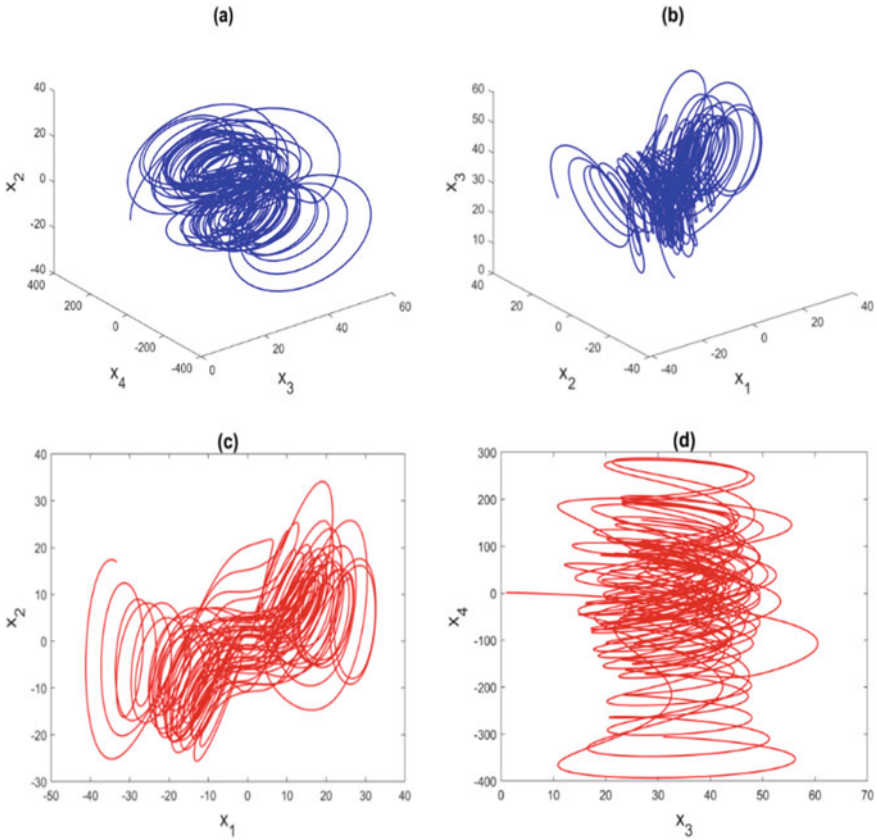


Fig. 1 Phase portrait of new system with $a = 10, b = \frac{8}{3}, c = 35, h = 2.39$

$$\begin{cases} \dot{x}_1 = -ax_1 + ax_2 + x_4 \\ \dot{x}_2 = cx_1 - x_1x_3 - x_2 \\ \dot{x}_3 = -bx_3 + x_1x_2 \\ \dot{x}_4 = hx_4 - x_1x_3 \end{cases} \quad (2)$$

In which $x_i, i = 1, 2, 3, 4$ are variables, and a, b, c are positive parameters, and h is the control parameter. Based on phase portrait, the attractors of system (2) are shown in Fig. 1.

2.1 Equilibrium Points, the Jacobian Matrix, and Divergence

The system (2) possesses three equilibrium points, only one is real as

$$\begin{cases} P_1(0, 0, 0, 0) \\ P_2(-13.9721i, +6.7725i, 35.4847, -207.4866i) \\ P_3(+13.9721, -6.7725i, 35.4847, +207.4866i) \end{cases} \quad (3)$$

The Jacobian matrix, characteristic equation, and roots at P_1 , and typical parameters are Eqs. (4), (5), and (6), respectively, as

$$J = \begin{bmatrix} -a & a & 0 & 1 \\ c-x_3 & -1 & -x_1 & 0 \\ x_2 & x_1 & -b & 0 \\ -x_3 & 0 & -x_1 & h \end{bmatrix} \quad (4)$$

$$(-b - \lambda)(\lambda^3 + (a + 1 + h)\lambda^2 + (a - h(a + 1) - ca)\lambda + ah(c - 1)) = 0 \quad (5)$$

$$\begin{cases} \lambda_1 = -8/3 \\ \lambda_2 = 2.39 \\ \lambda_3 = -24.7419 \\ \lambda_4 = 13.7419 \end{cases} \quad (6)$$

Since some roots are positive, the equilibrium point P_1 is unstable, that means it is classified as self-excited attractors. The trace or divergence of system (2) is written as Eq. (7):

$$\nabla v = \sum_{i=1}^4 \frac{\partial \dot{x}_i}{\partial x_i} = -a - 1 - b + h \quad (7)$$

So, system (2) is dissipative under the condition $-(a + 1 + b) > h$. The value of control parameter h plays an important role in divergence and $\nabla v = -11.2767$.

2.2 Lyapunov Exponents and Dimensions

The numerical simulation was implemented via Wolf Algorithm with MATLAB R2020a, under $a = 10, b = \frac{8}{3}, c = 35$ and $h = 2.39$, the initial value at $(0.1, 0.1, 0.1, 0.1)$ with step = 0.5 and tend = 200. The proposed system has (+, +, 0, -) sign of Lyapunov exponents and of these exponents are Eqs. (8) and (9), respectively, as

$$\begin{cases} LE_1 = 0.8246 \\ LE_2 = 0.2776 \\ LE_3 = 0.0000 \\ LE_4 = -12.376 \end{cases} \quad (8)$$

And

$$\sum_{i=1}^4 LE_i = -11.2737 \tag{9}$$

It is clear that the sum of Lyapunov exponents is almost equal to divergence, i.e., Eq. (7), which indicates the validity of our results and the sign of the Lyapunov exponents as (+, +, 0, -), refer to the new system as hyperchaotic. The plot of the LEs is shown in Fig. 2. Table 1 gives more details about compute LEs with various values of control parameters. Dimensions of Lyapunov are found as:

$$D_{LE} = j + \frac{1}{|LE_{j+1}|} \sum_{i=1}^j LE_i = 3 + \frac{LE_1 + LE_2 + LE_3}{|LE_4|} = 3.0891 \tag{10}$$

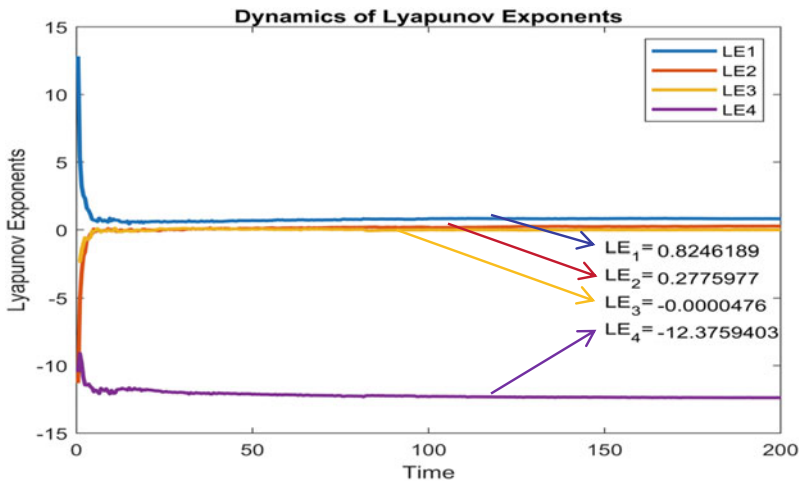


Fig. 2 Lyapunov exponents of 4D hyperchaotic system

Table 1 LE_S of system (2) various initial $(0.1, 0.1, 0.1, 0.1)^T$

No.	h	LE_1	LE_2	LE_3	LE_4	behavior	Sign of the LE_S
1	0.093	0.2378	0.0000	-0.9989	-12.8085	chaotic	(+, 0, -, -)
2	0.107	0.3056	0.0000	-1.0800	-12.7813	chaotic	(+, 0, -, -)
3	1.269	0.3550	0.0007	-0.0913	-12.6580	chaotic	(+, 0, -, -)
4	1.287	0.2477	0.1377	0.0002	-12.7618	hyperchaotic	(+, +, 0, -)
5	3.034	1.1918,	0.0880	-0.0003	-11.9103	hyperchaotic	(+, +, 0, -)
6	3.516	1.2018	0.0008	-0.3634	-10.9886	chaotic	(+, 0, -, -)
7	3.55	1.1030	-0.0002	-0.3919	-10.8268	chaotic	(+, 0, -, -)

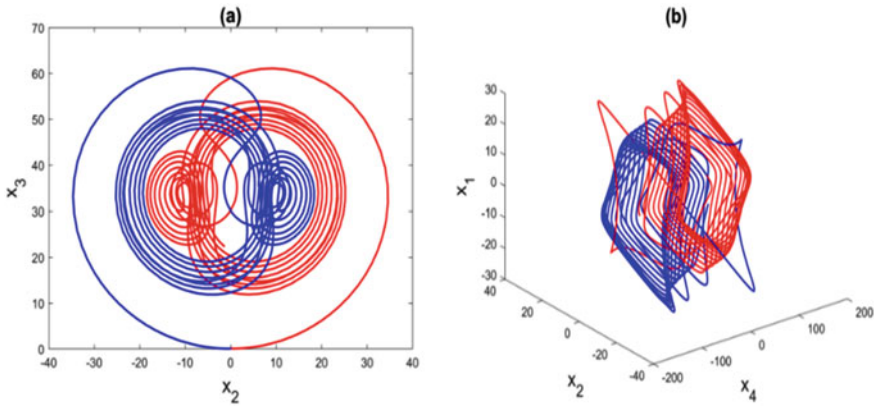


Fig. 3 Phase portraits of different coexisting for conditions of initial $(\pm 0.1, 0.1, 0.1, 0.1)$ (red, blue) in the: (a) $x_2 - x_3$ plane, (b) $x_1 - x_2 - x_4$ space

2.3 Multistability

A change in the initial conditions under the same set of parameters is known as coexistence attractors or multistability. System (2) exhibits coexistence attractors with the conditions of initial $(\pm 0.1, 0.1, 0.1, 0.1)$ (red, blue), and the parameters as $a = 10, b = \frac{8}{3}, c = 35,$ and $h = 1$ as shown in Fig. 3.

It is clear that from Fig. 3, the red attractors refer to one symmetric initial conditions $(0.1, 0.1, 0.1, 0.1)$, whereas the blue attractors correspond to one symmetric conditions of initial $(-0.1, 0.1, 0.1, 0.1)$, under the same parameters $a = 10, b = \frac{8}{3}, c = 35,$ and $h = 1$.

2.4 Projective Synchronization

The word synchronization means that one system follows another system, therefore, two systems are required. The first and second systems are known as the drive and response systems, respectively. Consider system (2) as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \underbrace{\begin{bmatrix} -a & a & 0 & 1 \\ c & -1 & 0 & 0 \\ 0 & 0 & -b & 0 \\ 0 & 0 & 0 & h \end{bmatrix}}_A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_B \underbrace{\begin{bmatrix} -x_1x_3 \\ x_1x_2 \\ -x_1x_3 \end{bmatrix}}_C \quad (11)$$

A is parameters matrix and B.C is the nonlinear part, whereas the response system is:

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} = A \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \left(B \begin{bmatrix} -y_1y_3 \\ y_1y_2 \\ -y_1y_3 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \right) \tag{12}$$

and let $U = [u_1, u_2, u_3, u_4]^T$ is the controller that should be constructed. The error dynamics for synchronization between the 4D systems (11) and (12) are

$$e_i = y_i - Sx_i, \quad i = 1, 2, 3, 4,$$

S is scaling matrix $S1 = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$ and satisfied that, $\lim_{t \rightarrow \infty} e_i = 0$.

Then we obtain the error dynamics as:

$$\begin{cases} \dot{e}_1 = ae_2 + ax_2 - ae_1 + e_4 + 3x_4 + u_1 \\ \dot{e}_2 = ce_1 - cx_1 - e_2 - y_1y_3 + 3x_1x_3 + u_2 \\ \dot{e}_3 = -be_3 + y_1y_2 - 4x_1x_2 + u_3 \\ \dot{e}_4 = he_4 - y_1y_3 + 5x_1x_3 + u_4 \end{cases} \tag{13}$$

Theorem 1 System (12) follow to system (11) (Projective Synchronization) with the following nonlinear control:

$$\begin{cases} u_1 = -ax_2 - 3x_4 - ce_2 \\ u_2 = -ae_1 + cx_1 + y_1y_3 - 3x_1x_3 \\ u_3 = -y_1y_2 + 4x_1x_2 \\ u_4 = -e_1 - 2he_4 + y_1y_3 - 5x_1x_3 \end{cases} \tag{14}$$

Proof Rewrite the error dynamics (13) with control (14) as:

$$\begin{cases} \dot{e}_1 = -ae_1 + (a - c)e_2 + e_4 \\ \dot{e}_2 = (c - a)e_1 - e_2 \\ \dot{e}_3 = -be_3 \\ \dot{e}_4 = -e_1 - he_4 \end{cases} \tag{15}$$

Construct a Lyapunov function as:

$$V(e) = e^T P e = \frac{1}{2}e_1^2 + \frac{1}{2}e_2^2 + \frac{1}{2}e_3^2 + \frac{1}{2}e_4^2 \tag{16}$$

$$\dot{V} = e_1\dot{e}_1 + e_2\dot{e}_2 + e_3\dot{e}_3 + e_4\dot{e}_4$$

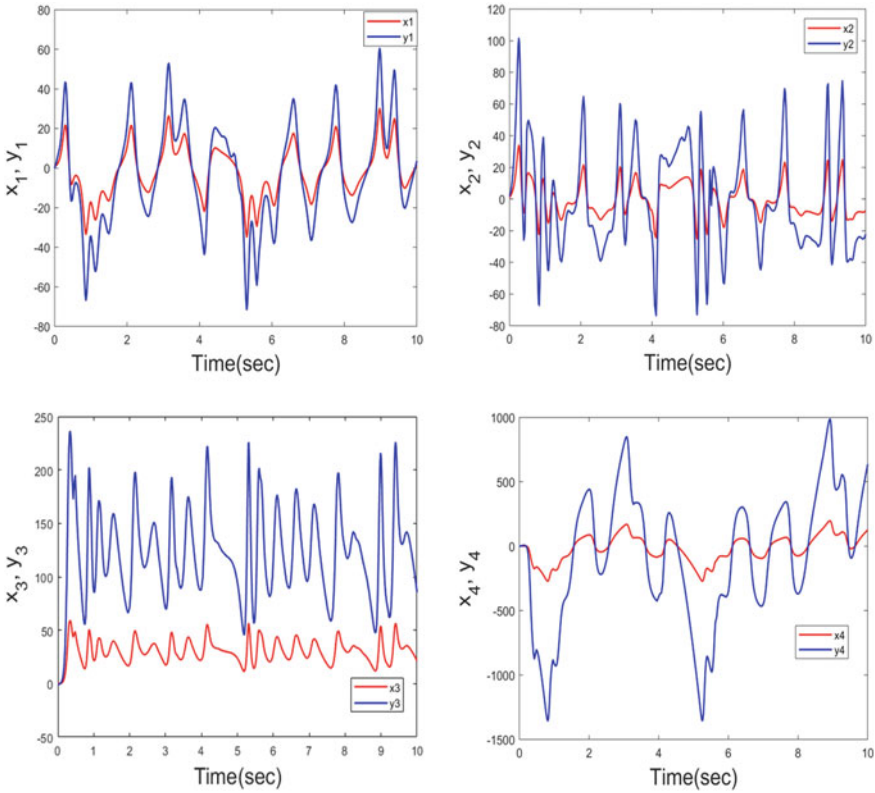


Fig. 4 Projective synchronization between systems (11) and (12) with control (14)

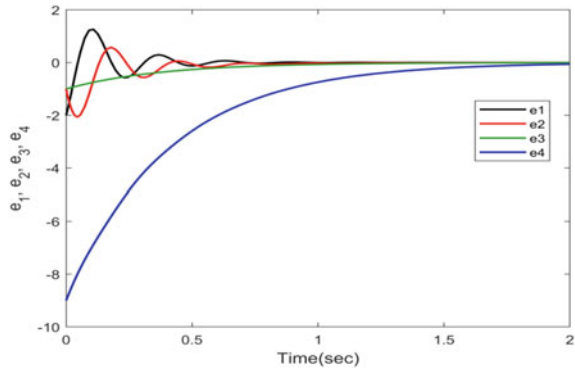
$$\begin{aligned} \dot{V} &= e_1(-ae_1 + (a-c)e_2 + e_4) + e_2((c-a)e_1 - e_2) + \\ &\quad e_3(-be_3) + e_4(-e_1 - he_4) \\ \dot{V} &= -ae_1^2 - e_2^2 - be_3^2 - he_4^2 = -e^T Q e \end{aligned} \tag{17}$$

where $Q = \text{diag}(a, 1, b, h)$, so $Q > 0$. Therefore, $\dot{V}(e_i) > 0$. The controller (14) is acceptable and the PS is realized as shown in Figs. 4 and 5.

3 Conclusions

A new 4D dissipative hyperchaotic system with only one unstable equilibrium point is constructed from the famous 3D Lorenz system by using a state feedback control strategy. The proposed system belongs to self-excited attractors and consists of ten terms including three quadratic nonlinearities. The dynamics behavior of the system

Fig. 5 The convergence of system (13) with controllers (14)



are analyzed theoretically and numerically, i.e., sensitivity to initial conditions, equilibrium points, stability analysis, Lyapunov exponents, Kaplan-Yorke dimension, multistability, and phase portrait. Moreover, a projective synchronization on an identical new system is carried out.

References

1. Abed, K. A., & Ahmad, A. A. (2020). The best parameters selection using Pso algorithm to solving for Ito system by new iterative technique. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(3), 1638–1645.
2. Ahmad, A. A. (2020). By using a new iterative method to the generalized system Zakharov-Kuznetsov and estimate the best parameters via applied the Pso algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2), 1055–1061.
3. Zhang, S., Zeng, Y., Li, Z., Wang, M., Zhang, X., & Chang, D. (2018). A novel simple no-equilibrium chaotic system with complex hidden dynamics. *International Journal of Dynamics and Control*, 6(4), 1465–1476.
4. Vaidyanathan, S., Dolvis, L. G., Jacques, K., Lien, C. H., & Sambas, A. (2019). A new five-dimensional four-wing hyperchaotic system with hidden attractor, its electronic circuit realization and synchronisation via integral sliding mode control. *International Journal of Modelling, Identification and Control*, 32(1), 30–45.
5. Zhang, S., Wang, X., & Zeng, Z. (2020). A simple no-equilibrium chaotic system with only one signum function for generating multidirectional variable hidden attractors and its hardware implementation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(5), 053129.
6. Thivagar, M. L., Ahmed, M. A., Ramesh, V., & Hamad, A. A. (2020). Impact of nonlinear electronic circuits and switch of chaotic dynamics". *Periodicals of Engineering and Natural Sciences*, 7(4), 2070–2091.
7. Al-Kateeb, Z. N., & Jader, M. F. (2020). Encryption and hiding text using DNA coding and hyperchaotic system. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2), 766–774.
8. Al-Kateeb, Z. N., & Mohammed, S. J. (2020). Encrypting an audio file based on integer wavelet transform and hand geometry. *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, 18(4), 2012–2017.
9. Al-Kateeb, Z. N., & Mohammed, S. J. (2020). A novel approach for audio file encryption using hand geometry. *Multimedia Tools and Applications*, 79(27–28), 19615–19628.

10. Al-Kateeb Z. N., AL-Shamdeen M. J., & Al-Mukhtar F. S. (2020). Encryption and steganography a secret data using circle shapes in colored images. In *Proc. FISCAS, Karbala, Iraqi Journal of Physics: Conference Series* (vol. 1591, p. 012019).
11. Singh, J. P., & Roy, B. K. (2017). Coexistence of asymmetric hidden chaotic attractors in a new simple 4-D chaotic system with curve of equilibria. *Optik, 145*, 209–217.
12. Hamad, A. A., Al-Obeidi, A. S., Al-Taiy, E. H., Khalaf, O. I., & Le, D. N. (2020). Synchronization phenomena investigation of a new nonlinear dynamical system 4D by Gardano's and Lyapunov's methods. *66*, 3311–3327.
13. Al-Talib, Z. S., & AL-Azzawi, S. F. (2020). Projective synchronization for 4D hyperchaotic system based on adaptive nonlinear control strategy. *Indonesian Journal of Electrical Engineering and Computer Science, 19*(2), 715–722.
14. Al-Azzawi, S. F., Mohamed, M. A., Rubiani, H., Mamat, M., Titley, J., & Langi, Y. A. R. (2020). Chaotic Lorenz system and it's suppressed. *Journal of Advanced Research in Dynamical and Control Systems, 12*(2), 548–555.
15. Wang, Z., Cang, S., Ochola, E. O., & Sun, Y. (2012). A hyperchaotic system without equilibrium. *Nonlinear Dynamics, 69*(1–2), 531–537.
16. Zhang, S., Zeng, Y., Li, Z., Wang, M., & Xiong, L. (2018). Generating one to four-wing hidden attractors in a novel 4D no-equilibrium chaotic system with extreme multistability. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 28*(1), 013113.
17. Wei, Z. (2011). Dynamical behaviors of a chaotic system with no equilibria. *Physics Letters A, 376*(2), 102–108.
18. Thivagar, M. L., & Abdullah Hamad, A. (2020). A theoretical implementation for a proposed hyper-complex chaotic system. *Journal of Intelligent & Fuzzy Systems, 38*(3), 2585–2595.
19. Abed, K. A. (2020). Controlling of jerk chaotic system via linear feedback control strategies. *Indonesian Journal of Electrical Engineering and Computer Science, 20*(1), 370–378.
20. Singh, J. P., Rajagopal, K., & Roy, B. K. (2018). A new 5D hyperchaotic system with stable equilibrium point, transient chaotic behaviour and its fractional-order form. *Pramana, 91*(3), 33.
21. Ojoniyi, O. S., & Njah, A. N. (2016). A 5D hyperchaotic Sprott B system with coexisting hidden attractors. *Chaos, Solitons & Fractals, 87*, 172–181.
22. Hu, G. (2009). Generating hyperchaotic attractors with three positive Lyapunov exponents via state feedback control. *International Journal of Bifurcation and Chaos, 19*(02), 651–660.
23. Yang, W. M., Osman, & Chen, C. (2015). A new 6D hyperchaotic system with four positive Lyapunov exponents coined. *International Journal of Bifurcation and Chaos, 25*(4), 1550060.
24. Al-Azzawi, S. F., Thivagar, M. L., Al-Obeidi, A. S., & Hamad A. A. 2020. Hybrid synchronization for a novel class of 6D system with unstable equilibrium points. *Materials Today: Proceedings*.
25. Al-Obeidi, A. S., & Al-Azzawi, S. F. (2020). Hybrid synchronization of high-dimensional chaos with self-excited attractors. *Journal of Interdisciplinary Mathematics, 23*(8), 1569–1584.
26. Al-Azzawi, S. F., & Al-Obeidi, A. S. (2020). Chaos synchronization in a new 6D hyperchaotic system with self-excited attractors and seventeen terms. *Asian-European Journal of Mathematics, 2150085*.
27. Wolf, A., Swift, J. B., Swinney, H. L., & Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena, 16*(3), 285–317.

Improving the Protection of Wireless Sensor Network Using a Black Hole Optimization Algorithm (BHOA) on Best Feasible Node Capture Attack



Ankur Khare, Rajendra Gupta, and Piyush Kumar Shukla

Abstract Wireless Sensor Network (WSN) is an area of research that connects mutually huge subareas of communication, routing, security, and attacks. WSN is conceivably the most susceptible network to node capture attack due to its dynamic nature in huge area. A node capture attack is introduced by seizing few nodes through an intruder to capture entire WSN by extracting the useful information like keys, routing mechanism, and data from WSN. To improve the protection of WSN, we proposed a Black Hole Optimization Algorithm (BHOA) on best feasible node capture attack to discover the optimal nodes having superior possibility of attack. The BHOA is applied on a function Vertex Participation. The experiment is performed on MATLAB 2019a environment, and the results show the better quality, efficiency of BHOA against MA, OGA, MREA, GA, and FFOA based on traffic compromised ratio, power consumption cost, and attacking time.

Keywords Black hole · Compromised · Feasible · Seizing · Wireless Sensor Network (WSN)

1 Introduction

The hasty growth of WSN (Wireless Sensor Network) [1, 2] has been achieved in the last few years for enormous application fields like disaster monitoring, remedial use, and manufacturing sections. The vast deployment of WSN has been provided the greater chance to attacker for attacking on network vulnerabilities to devastate the total communication of WSN [3, 4]. The MA (Matrix Algorithm) is developed to increase the efficiency of node capture attack [5, 6] by generating node-path group with minimum execution time. Node-path group is evaluated on the basis of direct and indirect combination of vertices and paths, and after that the cost of confiscating vertices is obtained. Therefore, MA is used to obtain minimum cost of confiscating

A. Khare (✉) · R. Gupta
Rabindranath Tagore University, Raisen, India

P. K. Shukla
University Institute of Technology, RGPV, Bhopal, India

vertices for malicious use of WSN. The attacking effectiveness is not obtained by MA [7]. So, another Opti-Graph Attack (OGA) approach is applied to generate advanced efficiency of node capture attack with least cost. Energy expenditure is not considered in OGA [8], further MREA (Minimum Resource Expenditure node capture Attack) [9] is implemented with shortest path calculation through traveling sales person and Hamilton path approach. The performance of MREA is further enhanced in terms of least node capturing by utilizing an optimization approach GA (Genetic Algorithm) [10]. GA is applied on functions having more than one objective like vertex and key contribution and resource expenditure to obtain malicious vertices [11]. Moreover, the efficiency of attacking method of GA is increased by using another optimization approach FFOA (Fruit Fly Optimization Algorithm) in terms of compromised traffic and attacking rounds [12].

Therefore, we developed a Black Hole Optimization Algorithm (BHOA) to get better quality outcomes on node capture attack in WSN. BHOA utilizes vertex participation for evaluating the optimal value of each vertex for selection of malicious vertices over WSN.

2 BHOA for Node Capture Attack

Various vertices (sensor nodes) V_{sn} and links L_{sn} are merged to generate a wireless sensor network $W_{sn} = (V_{sn}, L_{sn})$ with single and multipath ways for message passing. A message is transmitted in the form of packets in which a packet is passed through a link $L_{e,f}$ between vertices V_e and V_f within an area of transmission T^a after encryption via common key $K_{e,f} = K_e \cap K_f$. Different keys $K_e \subset K_{sn}$ are used to perform encryption of different packets before passing via vertex $V_e \in V_{sn}$. The intruder model is a multifarious structure to initiate on WSN for attacking the vertices to take away the essential information like cryptographic keys, and encryption methods. A single key $K_{e,f} \in K^{cm}$ of a link $L_{e,f}$ is compromised to capture that link $L_{e,f} \in L_{sn}$. A single link $L_{e,f} \in L^{cm}$ of a path P_e is compromised to capture that path.

$P_e \in P_{sn}$. A single path $P_e \in P^{cm}$ of a route $R_{source,dest}$ is compromised to capture that route $R_{source,dest} \in R_{sn}$.

The key objective of intruder is to attack small number of vertices for controlling the whole WSN. Therefore, we implemented a BHOA to pick the optimal vertices based on maximum vertex participation for compromising the full WSN.

2.1 Vertex Participation

A vertex $e \in V_{sn}$ has a direct connection (link) (e, vp_{sn}) (symbolized as $(e \rightarrow vp_{sn})$) having a vertex of path $vp_{sn} \in P_{sn}$. A vertex e is seized to capture the vp_{sn} , which is called as direct connection from e to vp_{sn} , and represented by a function $P^{cm}(e, vp_{sn}) : V_{sn} \times P_{sn} \rightarrow R_{sn}^{\{0,1\}}$ through Eq. 1.

$$P^{cm}(e, vp_{sn}) = \begin{cases} 1, & e \rightarrow vp_{sn} \\ 0, & \text{Otherwise} \end{cases} \quad (1)$$

A vertex $e \in V_{sn}$ has an indirect connection having a vertex of path $vp_{sn} \in P_{sn}$. A vertex e is seized to capture the vp , which is called as partial compromise (indirect connection) from e to vp gratifying some situation:

- vp is not captured.
- At least one key is provided to vertex e which is mutually used through a connection in the path vp . A connection (f, g) of vp . is shown as $1 \leq |K_{f,g} \cap K'^{cm}| < |K_{f,g}|$, in which K'^{cm} is compromised key set with vertex e by attacker. It is called as partial compromise (indirect connection) from e to vp

The direct and indirect vertex participation are calculated. Direct connections are utilized to obtain number of captured paths for generating Direct Vertex Participation ($DP_v(e)$), and indirect connections are utilized to obtain Indirect Vertex Contribution ($IP_v(e)$). $DP_v(e)$. is calculated through Eq. 2 by utilizing Eq. 1.

$$DP_v(e) = \sum_{e \in V_{sn}, vp_{sn} \in P_{sn}} P^{cm}(e, vp_{sn}) \quad (2)$$

$IP_v(e)$ is calculated for the vertex e and entire path connections through Eq. 3.

$$IP_v(e) = \sum_{e \in V_{sn}, vp_{sn} \in P_{sn}} IP^{cm}(e, vp_{sn}) \quad (3)$$

Here, $IP^{cm}(e, vp_{sn}) : V_{sn} \times P_{sn} \rightarrow R_{sn}^{\{0,+\infty\}}$ is calculated through Eq. 4.

$$IP^{cm}(e, vp_{sn}) = \frac{|K'_{vp_{sn}}|}{|K_p(vp_{sn})|} \sum_{(e,f) \subset p_e} \frac{1}{|L(vp_{sn})|} \left(1 - \frac{|K'_L(e,f)|}{|K_{e,f}|} \right) \quad (4)$$

Here,

$|L(vp_{sn})|$ = Number of connections in vp_{sn} .

$|K'_{vp_{sn}}|$ = Number of ks of vertex of path vp_{sn} not acquired by the intruder.

$|K'_L(e,f)|$ = Number of keys not obtained by intruder.

$|K_{e,f}|$ = Number of shared Keys by vertex V_e and V_f .

$|K_p(vp_{sn})|$ = Number of shared keys of whole connections.

The $|K'_L(e, f)|$ (Eq. 5), $|K'_{vp_{sn}}|$ (Eq. 6), and $|K_p(vp_{sn})|$ (Eq. 7) values are calculated as below:

$$K'_L(e, f) = K_{e, f} - K^{cm} \quad (5)$$

$$K'_{vp_{sn}} = \bigcup_{(e, f) \subset P_{sn}} K'_L(e, f) \quad (6)$$

$$K_p(vp_{sn}) = \bigcup_{(e, f) \subset P_{sn}} K_{e, f} \quad (7)$$

Finally, we combined the $DP_v(e)$ and $IP_v(e)$ to obtain Vertex Participation (VP) through Eq. 8.

$$VP(e) = \delta \times DP_v(e) + (1 - \delta) \times IP_v(e) \quad (8)$$

Here,

$VP(e)$ = Vertex Participation value of vertex e .

δ = constant in (0, 1) denoted the consequence of direct and indirect vertex participation. The maximum value of δ is obtained greatest destructiveness of WSN. The VP value is also superior for greatest destructiveness of WSN.

2.2 Black Hole Optimization Algorithm (BHOA)

After obtaining the Eq. 8 for each and every vertex in WSN, BHOA is utilized to evaluate the maximum VP values for providing best vertices V^{cm} from available vertices. An area of space–time (x, y, t) with sturdy and influential gravitational field is known as black hole in which not anything can get away from it. An event horizon is a mathematically explained region about a black hole that results in the point of no return. If something goes near to the horizon, then it would be fascinated into the black hole and eternally vanish. The mass, charge, and momentum are three general characteristics of black hole. The momentum and charge are not present in basic black hole but black hole has larger mass with maximum gravitational force.

2.2.1 Basic Preliminaries

Black Hole. The optimal aspirant among the entire aspirants at every repetition is nominated as a black hole in black hole optimization.

Stars. All the aspirants are known as stars excluding the black hole. The black hole is not formed casually and it is a genuine population aspirant.

Movement. All aspirants are stimulated near to the black hole on the basis of their present position and an arbitrary number.

1. BHOA initiates with primary aspirant solutions population to an optimization predicament and an evaluated objective function for optimization problem.
2. The optimal aspirant is nominated as black hole and other aspirants are normal stars in every repetition. The black hole initiates to pull the stars in the region of it after the initial stage.
3. Black hole absorbs the highly nearer stars and new aspirant solution (star) is arbitrary obtained for starting the new search after locating the new star to the search space.

2.2.2 Fitness Value Evaluation

- Initially, the aspirant solutions (stars) population $P(X) = \{X_1^t, X_2^t, X_3^t, \dots, X_N^t\}$ is arbitrarily obtained and located in search space for optimization.
- Evaluate the entire population fitness by using Eqs. 9 and 10.

$$F_a = \sum_{a=1}^{PS} E(P(t)) \quad (9)$$

$$F_{BH} = \sum_{a=1}^{PS} E(P(t)) \quad (10)$$

Here,

F_a = ath star fitness value.

F_{BH} = Black hole fitness value.

$E(P(t))$ = Evaluation functions for population at time t .

PS = Population Size.

The population is evaluated and the optimal aspirant with optimal fitness value F_a is nominated as black hole and other are normal stars. Black hole starts enthraling the highly nearer stars and other stars begin movement near to the black hole after the initial phase.

Black Hole Fascination Rate of Stars. The black hole begins fascinating the stars in the region of it and other stars begin movement near to the black hole. The black hole fascination rate is evaluated by using Eq. 11.

$$X_a(t + 1) = X_a(t) + \text{rand} \times (X_{BH} - X_a(t)) \quad (11)$$

Here,

$a = 1, 2, 3, \dots N$.

$X_a(t)$ = a th star position at repetition t .

$X_a(t + 1)$ = a th star position at repetition $(t + 1)$.

X_{BH} = Black hole position in search space.

rand = random number within $(0, 1)$.

N = Number of aspirant solutions (Stars).

A star may attain a minimum cost position than black hole though moving near to the black hole. Therefore, black hole shifts to that new star position and vice versa. The process will be continued with new black hole, and other stars begin movement near to that new black hole position.

Possibility of the Event Horizon Passage through moving Stars. The possibility of the event horizon passage moving stars is utilized to obtain more optimal value for the optimization problem. Each aspirant solution (star) that passages the black hole's event horizon will be fellated by black hole and an aspirant dies by blowing through black hole each slice of time. Therefore, new aspirant is populated and dispersed arbitrarily in search space to provide new search in population with constant aspirant solutions. The next repetition performs later than the movement of all stars. The horizon radius (R) is evaluated by using Eq. 12.

$$R = \frac{F_{BH}}{\sum_{a=1}^N F_a} \quad (12)$$

Here, F_a , F_{BH} , and N have been previously explained. If distance between an aspirant solution (star) and optimal aspirant (black hole) is found smaller than R , then aspirant is shrunken and a new aspirant is formed and dispersed arbitrarily in the search space.

Black Hole Optimization Algorithm (BHOA)

Input: The Stars Population with arbitrary positions located in search space for optimization.

Output: the optimal aspirant (black hole)with fitness value

```

START
Initialize the stars population with arbitrary positions
 $P(X) = \{X_1^t, X_2^t, X_3^t, \dots, \dots, X_N^t\}$ 

WHILE (until the end situation) DO
    FOR a = 1 to N DO
        Calculate the population fitness (eq. 9 and 10)
        Nominate the black hole (optimal aspirant) with maximum fitness value.
        Modify every star position by utilizing eq. 11.
        IF (A star may attain a minimum cost position than black hole)
            THEN
                Black hole shifts to that new star position and vice versa.
            IF (an aspirant (star) passages the event horizon)
                Calculate the horizon radius (R) by utilizing eq. 12.
                IF (Distance between an aspirant solution (star) and optimal aspirant (black hole) is found smaller than R)
                    THEN
                        Aspirant is shrunken and a new aspirant is formed and dispersed arbitrarily in the search space. Substitute it with a new aspirant in an arbitrary position in search space.
                    ELSE
                        BREAK;
                END IF;
            END IF;
        END FOR;
    END WHILE;
STOP
    
```

3 Results and Analysis

The efficiency of BHOA is analyzed and compared with MA, OGA, MREA, GA, and FFOA in terms of several circumstances shown in Table 1.

Table 1 Circumstances of simulation

Circumstances	Values
Number of vertices (V_{sn})	250
Network region	100 m \times 100 m
Source vertices (S^v)	20
Area of transmission (T^a)	25
Destination vertices (D^v)	7
Complete keys (K_{sn})	250
Population size (PS)	250
Iterations (N)	250

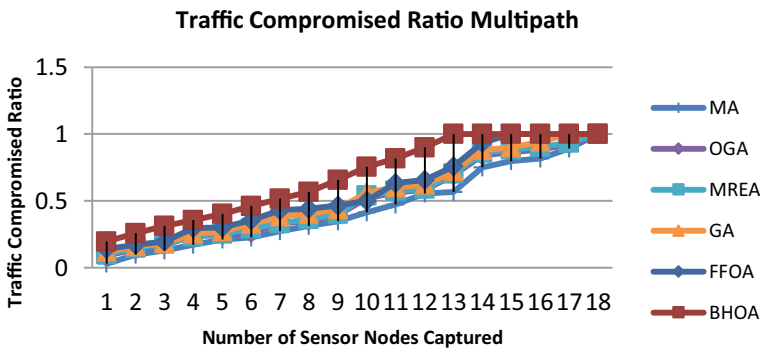


Fig. 1 Traffic compromised ratio (multipath routing)

3.1 Traffic Compromised Ratio

Traffic Compromised Ratio is defined as the proportion of captured paths to the all paths in WSN for single and multipath routing. Figure 1 elaborates the better quality of attacking ability of BHOA in terms of traffic compromised ratio with compromising least sensor nodes to provide 9.09 and 13.33% superior performance against the FFOA, 23.08 and 23.53% superior performance against GA, 28.57 and 27.78% superior performance against MREA, OGA, and MA for single and multipath routing, respectively, to capture complete WSN.

3.2 Power Consumption Cost

The Power Consumption Cost is the cost of power consumption for electing the sensor nodes to capture for co-operating the complete WSN. Figure 2 elaborates the greater efficiency of BHOA in terms of power consumption cost by means of number of keys to capture least sensor nodes to provide 20.06 and 19.47% superior

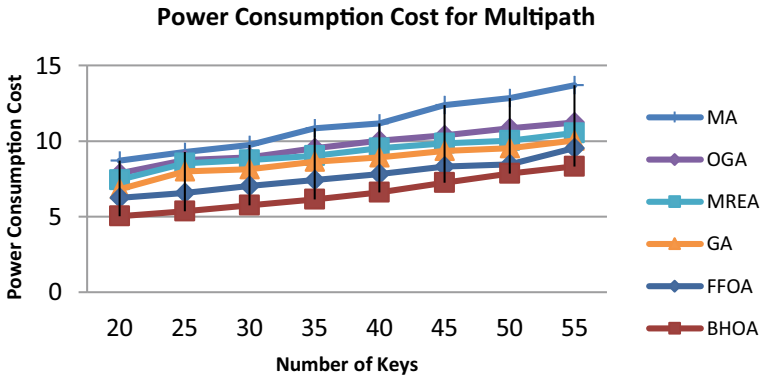


Fig. 2 Power consumption cost (multipath routing)

performance against the FFOA, 25.16 and 26.20% superior performance against GA, 33.18 and 32.42% superior performance against MREA, 40.65 and 36.09% superior performance against OGA, and 53.75 and 42.26% superior performance against MA for single and multipath routing, respectively, to capture complete WSN.

3.3 Attacking Time

The Attacking Time is described as the number of attacking rounds to co-operate the complete WSN by capturing the minimum sensor nodes. It is totally based on traffic compromised ratio. Hence, Fig. 3 elaborates the greater performance of BHOA in terms of attacking time by means of number of keys to capture least sensor nodes to provide 9.09 and 13.33% improved performance against the FFOA, 23.08 and

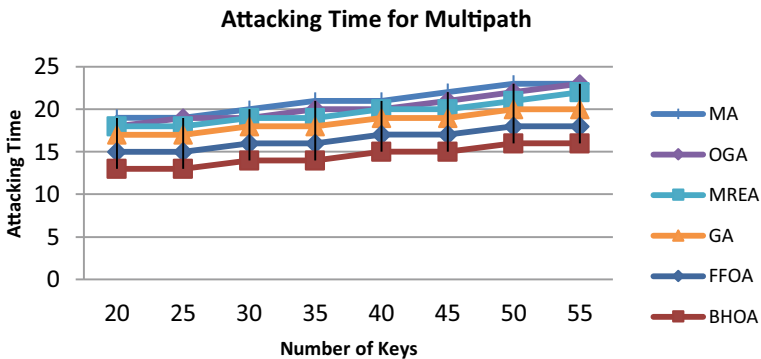


Fig. 3 Attacking time (multipath routing)

23.53% improved performance against GA, 28.57 and 27.78% improved performance against MREA, OGA, and MA for single and multipath routing, respectively, to capture complete WSN.

4 Conclusion

Wireless Sensor Network (WSN) is a combination of numerous research fields like communication, routing, protection, and attacks. The dynamic behavior of WSN makes it highly vulnerable to node capture attack in which an attacker captures some nodes to compromise whole WSN for stealing the important data, routing method, and keys. To enhance the protection of WSN, a Black Hole Optimization Algorithm (BHOA) is introduced in most feasible node capture attack to determine the optimal nodes having advanced opportunity of attack. The BHOA is applied on a function Vertex Participation, and results represent the superior efficiency of BHOA against MA, OGA, MREA, GA, and FFOA on the basis of traffic compromised ratio, power consumption cost, attacking time on MATLAB 2019a environment.

References

1. Xiong, C., Li, S., Liu, L., Li, R., & Jin, Y. (2019). A hybrid key pre-distribution scheme for wireless sensor networks. In *IOP Conf. Series, Journal of Physics* (pp. 1–18). <https://doi.org/10.1088/1742-6596/1229/1/012066>.
2. Shaukat, H. R., Hashim, F., Shaukat, M. A., & Alezabi, K. A. (2020). Hybrid multi-level detection and mitigation of clone attacks in mobile wireless sensor network (MWSN). *Sensors, MDPI*, 20, 1–23.
3. Chowdary, K., & Satyanarayana, K. V. V. (2017). Malicious node detection and reconstruction of network in sensor actor network. *Journal of Theoretical and Applied Information Technology*, 95(3), 582–591.
4. Ehdai, M., Alexiou, N., Ahmadian, M., Aref, M. R., & Papadimitratos, P. (2017). Mitigating node capture attack in random key distribution schemes through key deletion. *Journal of Communication Engineering*, 6(2), 1–10.
5. Alshammari, M. R., & Elleithy, K. M. (2018). Efficient and secure key distribution protocol for wireless sensor networks. *Sensors, MDPI*, 18, 1–25. <https://doi.org/10.3390/s18103569>.
6. Ahlawat, P., & Dave, M. (2018). An attack resistant key predistribution scheme for wireless sensor Networks. *Journal of King Saud University—Computer and Information Sciences, Elsevier*, 1–13.
7. Lin, C., & Wu, G. (2013). Enhancing the attacking efficiency of the node capture attack in WSN: A matrix approach. *The Journal of Supercomputing, Springer Science & Business Media*, 1–19. <https://doi.org/10.1007/s11227-013-0965-0>.
8. Lin, C., Wu, G., Yu, C. W., & Yao, L. (2015). Maximizing destructiveness of node capture attack in wireless sensor networks. *The Journal of Supercomputing, Springer Science & Business Media*, 71, 3181–3212. <https://doi.org/10.1007/s11227-015-1435-7>.
9. Lin, C., Qiu, T., Obaidat, M. S., Yu, C. W., Yao, L., & Wu, G. (2016). MREA: A minimum resource expenditure node capture attack in wireless sensor networks. *Security and Communication Networks, Wiley Online Library*, 9, 5502–5517. <https://doi.org/10.1002/sec.1713>.

10. Bhatt, R., Maheshwary, P., Shukla, P., Shukla, P., Shrivastava, M., & Changlani, S. (2020). Implementation of fruit fly optimization algorithm (FFOA) to escalate the attacking efficiency of node capture attack in wireless sensor networks (WSN). *Computer Communications, Elsevier, 149*, 134–145.
11. Shukla, P. K., Goyal, S., Wadhvani, R., Rizvi, M. A., Sharma, P., & Tantubay, N. (2015). Finding robust assailant using optimization functions (FiRAO-PG) in wireless sensor network. *Hindawi Publishing Corporation, Mathematical Problems in Engineering*, 1–8. <https://doi.org/10.1155/2015/594345>.
12. Molina, D., Poyatos, J., Ser, J. D., Garcia, S., Hussain A., & Herrera, F. (2020). Comprehensive taxonomies of nature- and bio-inspired optimization: inspiration versus algorithmic behavior. *Critical Analysis and Recommendations. cs.AI*, pp. 1–76. [arXiv:2002.08136v2](https://arxiv.org/abs/2002.08136v2).

A Systematic Analysis of the Human Activity Recognition Systems for Video Surveillance



Sonika Jindal, Monika Sachdeva, and Alok Kumar Singh Kushwaha

Abstract In recent years, human activity recognition has become a prominent research area in numerous fields such as healthcare, smart home activity analysis, suspicious activity recognition, robotics, surveillance, and security. The focus of the current research work is the analysis of human activity recognition systems for video surveillance. The human activity recognition system involves the detection of normal as well as abnormal activities. The recognition of human activities is still considered a challenging issue despite the contributions of numerous researchers. The erratic human behavior and complexities of the video datasets create numerous challenges to precisely observe the human activities with significant performance. The analysis of the human activity detection systems for video surveillance is conducted on the basis of state-of-art contributions by different researchers in the field. The paper also describes the taxonomy of human activity detection. It ends with a discussion of the challenging issues in the field along with the concluding remarks.

Keywords Human activity recognition · Video surveillance · Security surveillance · Human computer interaction · Visual surveillance

1 Introduction

Human activity is the movement of the human body parts and the recognition of these activities belong to the development of the autonomous system as the application of the human–computer interaction. Human activities can be recognized as vision-based devices or sensor-based devices. The sensor-based devices can be contact or

S. Jindal (✉) · M. Sachdeva
Department of Computer Science & Engineering, IKG Punjab Technical University, Jalandhar,
India
e-mail: sonikajindal@sbsstc.ac.in

M. Sachdeva
e-mail: monika@ptu.ac.in

A. K. S. Kushwaha
Department of Computer Science & Engineering, Guru Ghasidas Vishwavidyalaya, Bilaspur, India

contactless gadgets with any kind of special sensor or the inbuilt sensors available in mobile phones or other advance electronic gadgets. The vision-based devices are the contactless devices in which a camera is considered as the major source of information collector. The captured data can be processed with the help of computer vision techniques to recognize human activities.

The focused section of the current research work is the recognition of human activities with vision-based video data. The video data is useful for the recognition of normal as well as abnormal human activities. The taxonomy of human activity recognition is illustrated in Fig. 1. The normal activities involve the daily routine benign activities like sleeping, walking, playing, standing, sitting, etc. that can be considered as the smart home activities. The activities other than the normal benign activities are referred to the abnormal activities such as health issues (like sudden falling, chest pain, vomit, faint, etc.) of elder people, abnormal crowd activities (like running crowd, punching, slap, etc.), or any kind of suspicious activities (like the appearance of weapon at public places, leftover baggage at public places, theft, etc.).

The surveillance with video data is the major source of visual data for human activity recognition. The video data analysis can be considered as the process with the modules of object recognition, human action recognition, and the classification of the actions. These surveillance systems serve as the protective element for the need of the future. For instance, the aging population of the world is estimated to increase from 0.461 billion to 2 billion by the year 2050 [1]. But the extraction of video data from all the private and public places can lead to the processing of huge data. The manual loop up for the daily activities is a tedious task for a human and the requirement of huge manpower with the expertise to analyze the live data. Moreover, the storage of such huge data can also increase the financial burden. This makes to develop the autonomous video surveillance system that can lesser down the storage

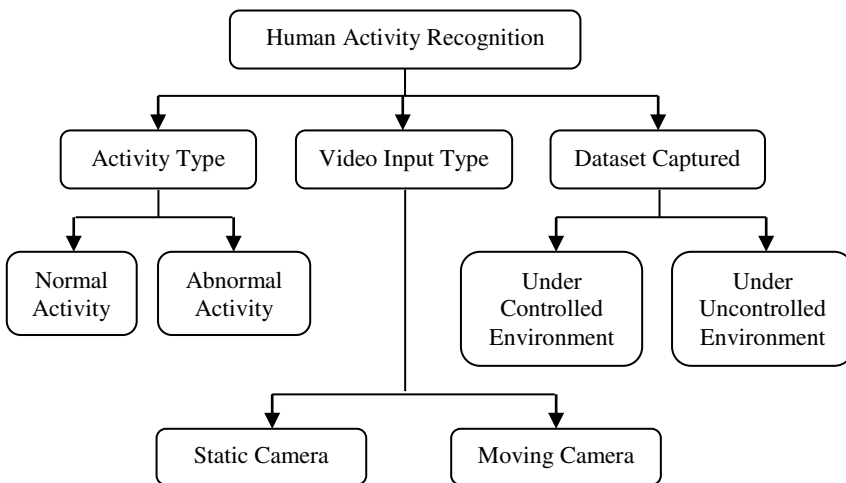


Fig. 1 Taxonomy of human activity recognition

cost, human errors, and huge manpower expertise. The autonomous system should be able to perform the task with ease with the ability to analyze the live abnormal activities so that the immediate problem solution can be determined. It is also vital to perform the object recognition [2] to determine the suspicious activities.

The present work discusses the different state-of-art techniques of human activity recognition systems for video surveillance. These research contributions are described in Sect. 2. The analysis of the human activity recognition systems is conducted on the basis of discussed research contributions presented in Sect. 3. The analysis is described with the discussion of the significant human activity recognition techniques for video surveillance and challenging issues for the same. The paper ends with the concluding remarks of the work.

2 Human Activity Recognition Systems for Video Surveillance

The section describes the different human activity recognition systems for video surveillance. The growing usability of autonomous human activity recognition in different fields has drawn the attention of researchers. The quality contributions of the researchers from the year 2011 to 2020 are discussed with the focused region of video surveillance. Moreover, the findings from the different research contributions are also highlighted in Table 1.

Zhang et al. [3] have employed the improved variant of the Gaussian mixed model that adds the attributes of k-means clustering along with the envelope shape presentation to handle the varying viewpoints in different video backgrounds. The average recognition of 93.23% was evaluated by authors. Gao et al. [4] have presented the network transmission based (NTB) algorithm with a focus on the recognition of abnormal activities. The experimentation was conducted on the recording data obtained from two manually settled cameras and attained the efficient results. Shieh and Huang [5] have focused to care the aging people with video surveillance. The authors have determined the falling activities using the pattern recognition based autonomous falling detection algorithm. The system throughput was improved with the software pipelining system. Chaturamali and Rodrigo [6] have addressed the problems of huge feature vector size and longer training time by adapting the support vector machine classifier. The classifier was implemented for the dataset of UIUC1 and Weizmann and achieved the comparing results. The training time of SVM was a bit higher than the other techniques for the case of fewer data samples. Gupta et al. [7] have conducted human activity recognition using the Gait pattern recognition model. The method utilizes the real-life scenario based on the Hu-moments [8] to evaluate human activities. The average recognition rate of 91.36% and 95.01% was evaluated for the Weizmann and KTH datasets, respectively.

Further, Bengalur [9] have worked for the recognition of 3D joint structures model for the recognition of human activities. The model utilizes the depth feature along

Table 1 Findings from the human activity recognition systems for video surveillance

Author and year	Technique	Dataset	Activity type	Results (%)	Remarks
Zhang et al. [3]	Gaussian mixed model	CASIA Activity analysis	Normal activities	Accuracy 93.23%	No state-of-art comparison
Gao et al. [4]	Network transmission algorithm	Self prepared dataset	Normal and abnormal	Equal Error rate 11.4%	Need to reduce error rate
Shieh and Huang [5]	Autonomous falling detection system	Self prepared dataset	Falling and non-falling activities	Accuracy 93% non-falling & 90% Falling	Possibility of extension for more activities and performance improvement
Chaturamali and Rodrigo [6]	Support vector machine	UIUC1 and Weizmann datasets	Normal activities	average recognition rate 99.32%	SVM classifier takes more time for training the sample
Gupta et al. [7]	Gait pattern recognition	KTH and Weizmann actions	Normal activities	Average recognition rate 93.19%	No comparison with state-of-art
Bengalur [9]	Support vector machine	Self prepared dataset of 650 videos	13 Normal activities	Accuracy 89%	No evaluation on standard benchmark datasets
Kaur and Singh [10]	Kalman filter	Weizmann dataset	10 Normal activities	Not available	No performance evaluation
Htike et al. [11]	KNN, SOM, MLP, FCM, and KMA	Human posture dataset	Normal human postures	MLP accuracy 97% for 5 & 98% SOM for 2 posture	No Consideration of standard benchmark dataset
Taha et al. [12]	Hidden Markov models	Cornell CAD-60 and CAD-120	Normal activities	CAD-60 82.3% & CAD120 94.4%	The work needs improvement for special activities like eating, drinking, etc
Patwardhan [13]	Dynamic probabilistic networks	Self prepared dataset	Normal activities	Average accuracy 75.5%	DPN lacks the HMM for the walking activity

(continued)

Table 1 (continued)

Author and year	Technique	Dataset	Activity type	Results (%)	Remarks
Kamal et al. [14]	Modified Hidden Markov model (MHMM)	IM-daily depth activity and MSR daily 3D activity	Normal daily activities	IM-Daily accuracy 68.3%, MSR 91.3%	Accuracy of the MHMM model was only 68.3% that can be improved in real life evaluation
Khemchandani and Sharma [15]	RLSTW-SVM approach	IXMAS, UMD, UIUC1, & Weizmann	Normal Activities	Average Accuracy 93.76%	The work can be extended for the interaction of multiple people
Ijjina and Chalavadi [16]	Convolutional neural network	Weizmann, SBU Kinect interaction, NATOPS gesture, & MIVIA action	Normal activities	Average accuracy 92.73%	Work can be extended for group activities & categories of human behaviour recognition
Zhao et al. [17]	Region-based mixture model	Weizmann and UT-tower datasets	Normal activities	Average accuracy 98.99	It can be made autonomous for the adjustment of regions in the RMM model
Kamthe and Patil [18]	Semantic based approach	CAVIAR (PETS 2004) and PETS 2006	Suspicious activities	Average accuracy 94.5%	Results evaluated only for two activities
Bouachir et al. [19]	NB, LSVM, LDA, and RBF-SVM	Self prepared dataset	Suicide attempt activity	100% Accuracy with NB & RBF-SVM	The occlusion problem affects the reliability for the detection of the 3D skeleton
Singh et al. [21]	Hidden Markov model	WVU multi-view, MSR view-point, i3D Post, and KTH action	Normal activities	Average accuracy 99.66%	It can be extended for the recognition of group activities recognition

(continued)

Table 1 (continued)

Author and year	Technique	Dataset	Activity type	Results (%)	Remarks
Kong et al. [22]	Three stream convolutional neural network	UR Fall detection and multiple camera Fall	Falling activities	Average recall 99.4%	Method fails for bad quality videos. The hardware requirement is also expensive
Jaouedi et al. [23]	Gated recurrent neural network	KTH, UCF101, and UCF Sports Action	Normal and sports activities	Average accuracy 91.54%	Accuracy needs to improve for the UCF101 and UCF sports datasets
Silva and Marana [24]	Bag of poses approach	KTH and Weizmann datasets	Normal activities	Average accuracy 97.51%	It can be extended for the recognition of special activities and group based interactions

with the support vector machine for skeleton attribute analysis. Kaur and Singh [10] used the Kalman filter to track human activities. The overall process involved the steps of background analysis by calculating the mean of ‘N’ frames, human detection by subtracting the background, and final human activity recognition. The visuals of the work were significant, but the numerical performance analysis was not conducted by authors. Htike et al. [11] have explored different classifiers for the recognition of human activity recorded with a single camera. The considered classifiers were k-nearest neighbor (KNN), self-organizing maps (SOM), multi-layer perceptron feed-forward neural network (MLP), fuzzy c-means (FCM), and k-means algorithm (KMA). Among the mentioned classifiers, multiplayer perceptron was evaluated as a significant classifier due to higher recognition accuracy. Taha et al. [12] have employed the Hidden Markov Models (HMM) for the recognition of human activity. The experimentation was conducted for the 3D position with depth parameter on the basis of skeleton structure. Patwardhan [13] have investigated the dynamic probabilistic networks (DPN) for normal activity recognition. The method was evaluated for the basic activities of walking, sitting, standing, and walking whose performance comparison was conducted with HMM. The author indicated more accurate results in the indoor light controlled environment than the outdoor environment. Kamal et al. [14] have presented the modified HMM for the recognition of the 3D human body and their respective activities. The feature set of the system was enhanced by fusing the temporal joint features and spatial depth shape attributes. The HMM model was modified to acquire the moving body features along with the static human body features.

Furthermore, Khemchandani and Sharma [15] have proposed the robust least square twin support vector machine (RLSTW-SVM) for human activity recognition.

The method was proposed to handle the heteroscedastic noise and incorporation of the outlier effect. The optic flow and silhouette were also utilized as the feature descriptors. Ijjina and Chalavadi [16] have used the deep learning based convolutional neural network along with the investigation of features through depth stream videos and RGB motion streams. The approach was multi-model with the ability to tolerate the noise in a robust manner. Zhao et al. [17] have proposed the Region based Mixture Model (RMM) for human activity recognition from low quality video clips. The method of Layered Elastic Motion Tracking was adapted for the extraction of motion features and body shape with the estimation of raw edges. Kamthe and Patil [18] have focused on the detection of suspicious human activities using the semantic approach. The suspicious activities of loitering at ATM and abandoned bag detection were evaluated with the accuracy of 93% and 96%, respectively. Bouachir et al. [19] have described an intelligent system for the recognition of suicide attempts. The motion and pose features especially the 3D joint features are extracted which were utilized for the classification with different classifiers of Naïve Bayes (NB), Linear Support Vector Machine (LSVM), Linear Discriminant Analysis (LDA), and SVM with radial basis kernel (SVM-RBF). The human face features are also vital to detect the face movement [20]. Among the mentioned classifiers, the SVM-RBF has attained the superior results.

Recently, Singh et al. [21] have presented a multi-view based human activity recognition system with the usage of the hidden Markov model (HMM) as the activity detector. The overall process detects the people by removing the background, extract combinational features (linear binary pattern, optical flow motion, and contour based distance signal features), and determine the activity with the HMM approach. The accuracy of the discussed system was better than the other comparison techniques. Kong et al. [22] have focused to determine the falling activities from the multi-view cameras using the three stream convolutional neural network. The approach considers the first two streams adapt the Silhouettes and motion history images as the input and the third stream considers the dynamic images. In the overall scenario, the method was effective in terms of accuracy but lacks due to inefficient results for the bad representation of video clips. Jaouedi et al. [23] have explored the Gated Recurrent Neural Network for the recognition of human activities. The major focus of the paper was the extraction of significant features using Kalman Filter and Gaussian Mixture Model related to moving body. The significant results were noted for the recognition of normal as well as sports activities. Silva and Marana [24] have employed the Bag of Poses approach along with the Spatio-temporal features for the recognition of 2D human poses. The method adapts the indirect features by converting the line segments into the parameter space.

3 Systematic Analysis and Discussion

The analysis of the human activity recognition systems for video surveillance is conducted on the basis of discussed research contributions. The section describes the overall major focused activities, significant research technique, and challenging issues related to the discussed work.

The usability of different activities in the discussed studies is illustrated in Fig. 2. The activities of falling, suicide attempts, suspicious activities, and other alarming activities are considered in the list of abnormal activities. The daily routine activities and sports activities are taken under the category of normal activities. On the basis of Fig. 2, it can be noticed that there are major research contributions for the recognition of normal activities. There is lesser work available for the detection of abnormal activities.

On the basis of different activities, the authors have selected their respected dataset and used the different techniques for the classification and detection of activities. The significance of different research techniques depends on different scenarios such as the video quality, individual activity or group interaction, availability of feature components, and type of activity. In a comprehensive manner, the deep learning models have attained outperformed recognition accuracy.

Further extending the discussion, there is always some limitation or the possible extension that can be considered to improve the work. The possible directions and limitations of the individual studies are mentioned in the remarks column of Table 1. In a comprehensive manner, the following challenging issues are observed:

- The detection of abnormal activities that demand alarming attention needs to be installed with GPU based higher configured systems that can make it expensive for real time implementation.
- The installation of cameras in real time raise issues related to privacy, usability expertise, acceptance in society, and large scale applicability.
- In the discussed work, most of the contributions can be noticed related to work on the specific datasets for specific activities. The real time implementation of these

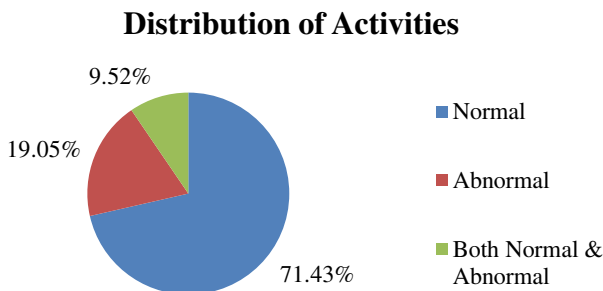


Fig. 2 Distribution of the different human activities in discussed work

systems demands to have higher recognition accuracy regardless of the color, gender, age, size, shape, and other body gadgets.

- The systems should also be resolved for the view-point issues as the human cannot stay focused towards the camera for the activity recognition. This raises the need to install multiple cameras at the same location with the ability of the system to observe the activity from side poses as well.

4 Conclusion

Human activity recognition is one of the significant applications in the digital age to adapt for different applications such as video surveillance, health monitoring, automated security, and human computer interaction. This research paper has focused on video surveillance to monitor different normal and abnormal activities. The contributions of the different researchers related to human activity recognition for video surveillance are thoroughly analyzed and significant outcomes are discussed. In the discussed techniques, the deep learning models are considered as efficient to recognize both the normal and abnormal activities. The different challenging issues determined with the existing techniques are described which can be addressed to implement the human activity recognition for video surveillance in real life.

References

1. Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, 8, 210816–210836. <https://doi.org/10.1109/ACCESS.2020.3037715>.
2. Gupta, S., Kumar, M., & Garg, A. (2019). Improved object recognition results using SIFT and ORB feature detector. *Multimedia Tools and Applications*, 78(23), 34157–34171.
3. Zhang, F., Wang, Y., & Zhang, Z. (2011). View-invariant action recognition in surveillance videos. In *Proceeding of the First Asian Conference on Pattern Recognition* (pp. 580–583), Beijing, China.
4. Gao, H., Lin, W., Yang, X., Li, H., Xu, N., Xie, J., & Li, Y. (2011). A new network-based algorithm for multi-camera abnormal activity detection. In *Proceedings of the 2011 IEEE International Symposium of Circuits and Systems (ISCAS)* (pp. 361–364), Rio de Janeiro, Brazil.
5. Shieh, W. Y., & Huang, J. C. (2012). Falling-incident detection and throughput enhancement in a multi-camera video-surveillance system. *Medical engineering & physics*, 34(7), 954–963.
6. Chathuramali, K. M., & Rodrigo, R. (2012). Faster human activity recognition with SVM. In *Proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer2012)* (pp. 197–203), Colombo, Sri Lanka.
7. Gupta, J. P., Singh, N., Dixit, P., Semwal, V. B., & Dubey, S. R. (2013). Human activity recognition using gait pattern. *International Journal of Computer Vision and Image Processing (IJCVIP)*, 3(3), 31–53.
8. Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2), 179–187.

9. Bengalur, M. D. (2013). Human activity recognition using body pose features and support vector machine. In *Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1970–1975), Mysore, India.
10. Kaur, R., & Singh, S. (2014). Background modelling, detection and tracking of human in video surveillance system. In *Proceedings of the 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity* (pp. 54–58), Ghaziabad, India.
11. Htike, K. K., Khalifa, O. O., Ramli, H. A. M., & Abushariah, M. A. (2014). Human activity recognition for video surveillance using sequences of postures. In *Proceedings of the Third International Conference on e-Technologies and Networks for Development* (pp. 79–82), Beirut, Lebanon.
12. Taha, A., Zayed, H. H., Khalifa, M. E., & El-Horbaty, E. S. M. (2015). Skeleton-based human activity recognition for video surveillance. *International Journal of Scientific & Engineering Research*, 6(1), 993–1004.
13. Patwardhan, A. S. (2015). Walking, lifting, standing activity recognition using probabilistic networks. *International Research Journal of Engineering and Technology (IRJET)*, 3(9), 667–670.
14. Kamal, S., Jalal, A., & Kim, D. (2016). Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM. *Journal of Electrical Engineering & Technology*, 11(6), 1857–1862.
15. Khemchandani, R., & Sharma, S. (2016). Robust least squares twin support vector machine for human activity recognition. *Applied Soft Computing*, 47, 33–46.
16. Ijjina, E. P., & Chalavadi, K. M. (2017). Human action recognition in RGB-D videos using motion sequence information and deep learning. *Pattern Recognition*, 72, 504–516.
17. Zhao, Y., Di, H., Zhang, J., Lu, Y., Lv, F., & Li, Y. (2017). Region-based mixture models for human action recognition in low-resolution videos. *Neurocomputing*, 247, 1–15.
18. Kamthe, U. M., & Patil, C. G. (2018). Suspicious activity recognition in video surveillance system. In *Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (pp. 1–6), Pune, India.
19. Bouachir, W., Gouiaa, R., Li, B., & Noumeir, R. (2018). Intelligent video surveillance for real-time detection of suicide attempts. *Pattern Recognition Letters*, 110, 1–7.
20. Gupta, S., Thakur, K., & Kumar, M. (2020). 2D-human face recognition using SIFT and SURF descriptors of face's feature regions. In *The Visual Computer* (pp. 1–10). <https://doi.org/10.1007/s00371-020-01814-8>.
21. Singh, R., Kushwaha, A. K. S., & Srivastava, R. (2019). Multi-view recognition system for human activity based on multiple features for video surveillance system. *Multimedia Tools and Applications*, 78(12), 17165–17196.
22. Kong, Y., Huang, J., Huang, S., Wei, Z., & Wang, S. (2019). Learning spatiotemporal representations for human fall detection in surveillance video. *Journal of Visual Communication and Image Representation*, 59, 215–230.
23. Jaouedi, N., Boujnah, N., & Bouhleb, M. S. (2020). A new hybrid deep learning model for human action recognition. *Journal of King Saud University-Computer and Information Sciences*, 32(4), 447–453.
24. Silva, M. V. da, & Marana, A. N. (2020). Human action recognition in videos based on spatiotemporal features and bag-of-poses. *Applied Soft Computing*, 95. <https://doi.org/10.1016/j.asoc.2020.106513>.

Integrating IoT with Blockchain: A Systematic Review



Malvinder Singh Bali, Kamali Gupta, and Swati Malik

Abstract With the advent of IoT, Internet dominance has been extended far and wide, resulting in management of billions of smart devices online. However, all the management frameworks in IoT developed so far are based on centralized models, which have their own set of issues like single point of failure and security constraints. To encounter major issues like this, Blockchain provides an effective alternative to encounter the issues of security and privacy. Blockchain is being a distributed and decentralized ledger framework, when used with IoT helps to encounter a lot of security related constraints. In this paper, a detailed study related to Blockchain has been conducted and how well it works in collaboration with IoT to overcome all the major security and privacy issues in an IoT ecosystem. Many other applications of Blockchain in an IoT environment were also studied, with the intention of exploring all the potential benefits it can provide while working with IoT.

Keywords Blockchain · Decentralization · Smart contract · IoT

1 Introduction

With the advancement of Digital Technology, IoT has fostered various technologies like Cloud Computing and Machine Learning to Data Analytics and Information modelling and thus becoming an integral part of IoT world. All these technologies are dependent on data which is provided by IoT with the help of sensor-based devices. With the passage of time, very smaller to smaller devices are brought under the ambit

M. S. Bali (✉) · K. Gupta · S. Malik
Chitkara University Institute of Engineering and Technology, Chitkara University Rajpura,
Patiala, India
e-mail: malvinder.singh@chitkara.edu.in

K. Gupta
e-mail: Kamali.singla@chitkara.edu.in

S. Malik
e-mail: Swati.malik@chitkara.edu.in

of IoT domain including household devices, health care, industry, precision agriculture and autonomous vehicles. Data collected from various IoT devices contains Critical and Private data, and current IoT architectures are not able to make data reliable and secure [1]. Most of the IoT based architectures are heavily centralized and liable to face failures due to single point which brings Scalability issue and brings lot of issues like Privacy and Security. The introduction of Blockchain brings new opportunity to overcome the challenges faced by IoT where decentralized approach is followed and seen as a long-term solution for the rise of IoT and helps in avoiding centralized failures. Presently Blockchain is seen as the suitable option to provide Secure and Distributed network to IoT [2].

The term Blockchain appeared in article “a cryptographic secured chain of blocks” [3]. Later it was universally introduced by Nakamoto in [4] and brought into implementation in 2008 and 2009 as the main part of Cryptocurrency Bitcoin. With the slow and steady adoption of Blockchain features to decentralized IoT communication and eradicate the need for depending on central trusted servers. Various researches are undergoing through provide reliable, trustworthy, scalable and secure IoT based network using Blockchain. Several benefits of decentralizing IoT with the help of Blockchain are as follows:

- IoT move from centralized server system to Decentralized Blockchain increases fault tolerance and eradicates central point of failure.
- With Blockchain network, users can authenticate data integrity and also the identity of the sending participant.
- IoT data stored in Blockchain are Immutable which ensures accountability.

2 IoT

2.1 Introduction of IoT

It extends internet power beyond computers and smart phones. The concept of IoT in simple terms is to connect all the electronic devices with the internet world. When these things are connected to the internet world, they can send and receive information. This ability makes these things smart. In IoT, objects strongly interconnected with the internet can be categorized as object which gather data and send it, object which receive data and act upon and objects that can do both. The IoT is a system of interconnected digital device, machines, animals or people provided with unique identifiers and the ability to transmit and share data over the network without human to computer or human to human interaction. IoT technology stack is divided into four layers involved in making IoT work. (1). Device hardware: They act as an interface between the real and digital world. From small objects like microphone to big size objects, everything can be practically turned into a connected device by adding sensors or actuators to calculate and gather the required data. (2). Device software: responsible for implementing device communication with cloud, collect

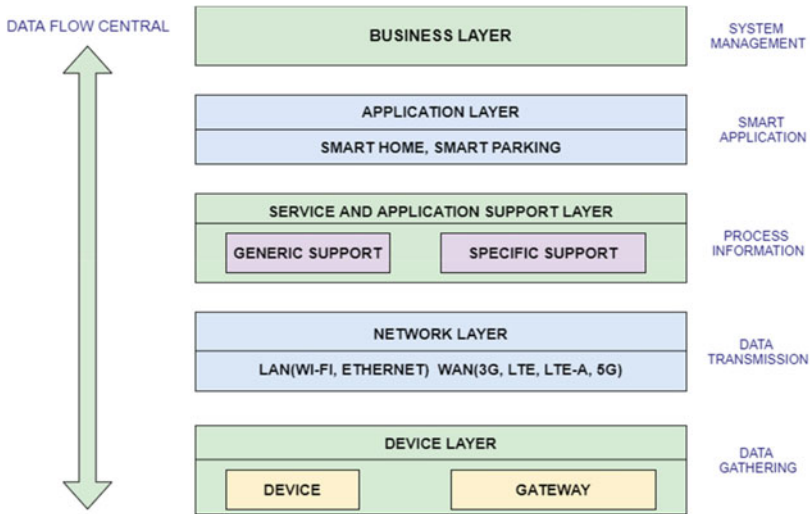


Fig. 1 IoT layered architecture [1]

data and integrate device for performing real time data analysis with IoT network. (3). Communication: which comprises of Cellular and satellite communication and specific protocol used in IoT environment (Zigbee, Zwave, etc.). Choosing relevant solution is one of the major challenges of every IoT stack. (4). Platform: space in which data is gathered, managed, processed, analyzed, and presented in a user friendly way. Several architectures for IoT were present but it never got standardized. Famous IoT architectures which made their presence in the IoT market were Microsoft Azure, RAMI4.0, WS02 and Internet of everything model. The paper highlights an architecture comprised of five layers as shown in Fig. 1. Network layer will provide network connectivity using Wired LANs, WANs, and Radio Satellites. It also facilitates with the feature of AAA (Authentication, Authorization and Accounting). Service and Application Support Layer (SASL) supports two capabilities such as Generic support capabilities and Specific support capabilities provides support for different IoT applications. Business Layer can build various business-related models, graphs, flowcharts and reports by gathering information from application layers [5] to make better decisions.

2.2 Challenges in IoT

2.2.1 Data Privacy

Due to diversified integration of IoT based services, lot of data is being generated sending vast majority of information of users to third parties who observe

the pattern/behavior of the users making users data vulnerable [6]. To encounter this problem, various policies have been framed for providing privacy but challenge remains to provide solutions to ensure privacy by design.

2.2.2 Management of data

Tremendous amount of data generated by IoT devices is huge and becomes unmanageable. So, to store bulk data, we need scalable infrastructure to efficiently handle growing data [7].

2.2.3 Centralized IoT Network

Continuous expansion of IoT nodes with reliability on central network for data storage will lead to single point of failure issue. If centralized server is down, whole network nodes will be down.

2.2.4 Data Originality Issue

For IoT, it is very tedious to authenticate the originality of the data. It can be altered by anyone without user having the proof of its tampering.

2.2.5 Network Complexity

In IoT, there are difficult network protocols used as per the application. Protocols such as NB-IoT, Sigfox, Lora, NFC, and Bluetooth offer different coverage ranges creating an interoperability issues [7, 8].

2.2.6 Third Party Data Breach

In IoT network, data is uploaded on the cloud server which is mostly handled by the 3rd party. They can also violate the privacy of the user by observing continuous pattern of the user's data [9].

2.2.7 IoT Device Resource Constraints

Most of the devices lack storage, computation and battery power which makes IoT devices vulnerable to different kinds of attacks.

2.2.8 Heterogeneity

Heterogeneity of protocols and IoT data types brings different issues related to incompatibility.

3 About Blockchain Technology

Blockchain is a distributed and decentralized public ledger (Fig. 2).

The answer is very simple. As the name suggests it is literally a chain of blocks as shown in Fig. 3. It consists of many blocks connected to each other in such a way that each block is connected to the next and previous block.

Each block mainly contains three things:

- **Data** The information like sender’s address, receiver’s address and the amount of transaction is stored in this bit of block.
- **Hash of own block** A unique long identifier consisting on numbers and alphabets which acts as the fingerprint of the block is stored in this bit. For example, “7d86c5626f1516e01182b5ed12e6924c0d2e908cba8ce3cd062f062b8d989 18b”
- **Hash of previous block** Each block has the previous block’s hash stored in it. The reason for this is to increase security. If someone tries to tamper with a block, its hash changes and then the hash will not match the hash stored in the previous block and we would get to know which block has been tampered with.

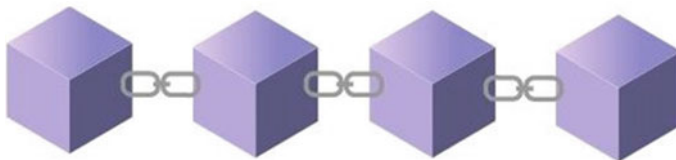


Fig. 2 Simplified blockchain. Source Selfmade

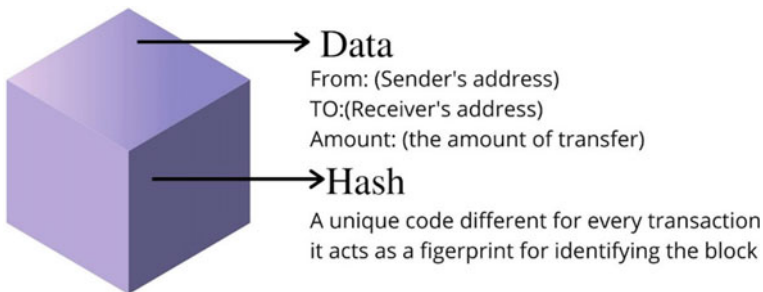


Fig. 3 Structure of a block Source Selfmade

3.1 Security Features of Blockchain

3.1.1 SHA256 Hash

The hashes are calculated using the sha256 algorithm which has something called as the avalanche effect. In the avalanche effect if we change even one alphabet in the data the hash changes completely and we get a 100% new hash. The sha256 hash is only one way. We can enter the data and get a hash but not the other way around.

3.1.2 Proof of Work

Nowadays we have a lot of computing power so just using a secure hash is not enough as computers can quickly tamper with the data and recalculate all the hashes of next blocks so fast that it would be impossible to detect that someone has tampered with the data. So, to overcome this flaw proof of work was introduced, this slows down the creation of new blocks and if you tamper with a block you would have to recalculate the proof of work for all other blocks as well which is time consuming.

3.1.3 Decentralization

This is the most important concept in blockchain technology. A blockchain has no owner rather it is distributed in a network of many computers/users, and each computer/user is called a 'node'. There can be millions of nodes in a network and each node has a copy of the blockchain stored in it. So, if we take a hypothetical situation where a hacker logs onto a node's pc and tampers with the blockchain successfully changing all its hashes and proof of work too. As soon as the blockchain is refreshed it will notice that that 1 node's blockchain does not match with the other millions of nodes and revert it to normal.

So, in conclusion, to tamper with the blockchain a hacker has to hack into millions of computers at the same time and change all the hashes and proof of work within seconds which is next to impossible. That is why blockchain is a very secure network.

3.2 Key Characteristics of Blockchain

3.2.1 Decentralization

All the data is stored at and verified by all initiated parties.

3.2.2 Immutability

Data once committed (stored) to the Blockchain cannot be changed or deleted. Once a transaction is entered in the database and the accounts are updated, the records cannot be attached because they are linked to every transaction record that came before them.

3.2.3 Transparency

In Blockchain, users involved in the Transaction process can see the transaction value and their authenticity (node or user) is provided by three plus alphanumeric address. Also, users of the system can reveal their identity in the system or remain silent.

3.2.4 Peer-to-Peer Transmission

Communication takes place between one system to other system without the support of a centralized system. Each node stores and forwards information to all other nodes.

3.2.5 Smart Contract

A smart contract is just a computer program that controls the digital transfer of currency or assets between two or more parties under some given conditions just as in a normal contract, but more securely and 100% fraud-free (Fig. 4). The smart contract states the conditions of the contract and enforces them on its own. They work a lot like food machines; you put the required amount of crypto currency in the machine and you get what you paid for.

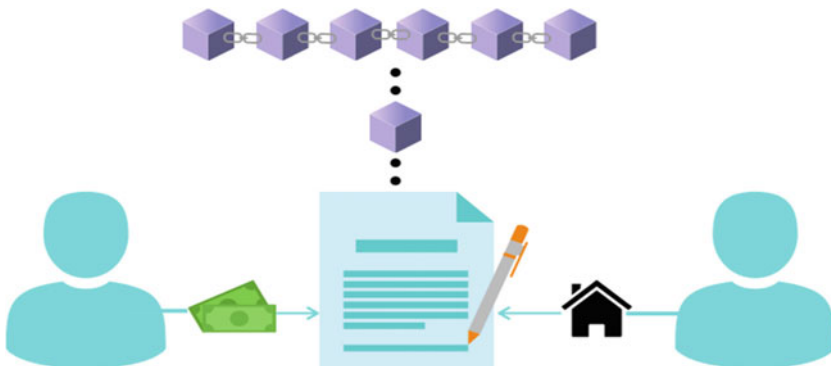


Fig. 4 Smart contract *Source* Selfmade

For example, if someone want to buy a piece of land from a person the standard procedure is that the two parties agree on the price, the land owner gets paid and then they go to a govt. office to get the land ownership changed in the govt. ledger which is just a huge register stored in the govt. archives. And for changing this ownership details, the govt. charges 5–7% of the land value. But on the other hand, if we use a smart contract, we can save this 5% fee. As soon as the buyer pays the smart contract the money gets transferred into the owner's crypto wallet and the land ownership details are now stored in a block, and this block becomes a part of a very large chain and copies itself onto millions of nodes hence securing the data. This is a lot safer than the govt. ledger as it cannot be destroyed due to any natural calamity. See Fig. 2 for illustration of this example.

- **Security of a smart contact:**

The conditions of a smart contract are 100% transparent, and there are no chances of fraud. You always get what you pay for, and the contract is totally unbiased. It is just a computer code written in a programming language e.g., python, so the contract will always execute and there is no chance of failure.

- **Speed of a smart contract:**

In the above property example, if we go with the traditional way of transferring ownership, it takes a huge amount of paperwork and travelling to many desks in govt. offices which is very time consuming and difficult. But if we use a smart contract the transaction takes place in a few seconds and the information gets stored in the blockchain immediately and hassle-free. Also, the smart contracts are 100% accurate and have no chances of human error.

4 Literature Review of Integration Between IoT and Blockchain

In our literature survey, we have undergone through various research work done in the amalgamation of IoT and Blockchain. Major research proposals have appeared that highlight the Integration of IoT with Blockchain. With the help of Blockchain Technology features, IoT can utilize the benefits of cost reduction, enhanced security, accelerated data exchange and trust building [7]. Conoscenti et al. [10] did a systematic literature review to highlight the features of Blockchain adopted by IoT. Atzori et al. [11] highlighted the drawback of applying BC in the IoT environment and also presented solutions that have emerged in the recent years. In our review of the previous work done, we have classified our literature research into subsections highlighting the research work done by various researchers to address the challenges in integrating IoT and Blockchain and also various proposals from them to combat the issue.

4.1 Privacy of IoT via Blockchains

4.1.1 Privacy Issues in Centralized IoT

With the introduction of IoT, data, people, virtual environments and different physical objects like wearable sensors, medical devices and smart objects interconnect with each other, resulting in the development of smart environments like, smart cities, smart transport systems, smart agriculture and so on.

However, there is a downside to this technological development also. Numerous challenges like privacy of individuals, security of the processes and the developed system arise. Let us take an example of a smart healthcare application which involves the use of different bio-sensors attached to the patient's body, regularly sending the important data related to the patient like blood pressure, heart rate to the cloud for analysis and evaluation and RFID based tag sensors for user authentication. Data after analysis will be sent to the consulted doctor's smart device and the patient for improvement and feedback.

If we analyze this application intricately, we observe that all the sensitive information regarding location of the patient, medical history and the present medical state of the patient is sent to the cloud using sensors, so there is great need to protect the privacy of the patient, particularly in case of high-profile patient.

The data collected from various IoT sensors can be used for diverse purposes. Certain organizations can predict the user's behavioral pattern and habits as well as the metadata. Customers using IoT based devices put trust on the companies providing smart services and are totally unaware of their data being shared or sold to the third-party.

4.1.2 Decentralization via Blockchain for IoT Privacy

Various ongoing researches in making data decentralized to resolve privacy violation are done and lot of proposed frameworks, architecture and algorithms have been designed. In [11], the author provided a framework called PISCES aiming to provide a strict separation between data provider and data controller. Here providers manage their data privacy and the data controllers are accountable to the privacy and data protocol for IoT. In [12], paper describes PLATIBART, a platform for exchanging IoT Blockchain based application with continuous testing. Using Plaribart, researcher shows how IoT Blockchain based application is developed, tested and then analyzed how effective IoT based application will be in terms of privacy.

5 Convergence of BC and IoT

5.1 Issues Faced in the IoT

5.1.1 Data Privacy

Due to diversified integration of IoT based services, lot of data is being generated sending vast majority of information of users to third parties who observe the pattern/behavior of the users making users data vulnerable [6]. To encounter this problem, various policies have been framed for providing privacy but challenge remains to provide solutions to ensure privacy by design.

5.1.2 Management of Data

Tremendous amount of data generated by IoT devices is huge and becomes unmanageable. So, to store bulk data, we need scalable infrastructure to efficiently handle growing data [13].

5.1.3 Centralized IoT Network

Continuous expansion of IoT nodes with reliability on central network for data storage will lead to single point of failure issue. If centralized server is down, whole network nodes will be down.

5.1.4 Data Originality Issue

In IoT network, it is very difficult to authenticate the originality of the data. It can be altered by anyone without user having the proof of its tampering.

5.1.5 Network Complexity

In IoT, there are difficult network protocols used as per the application. Protocols such as NB-IoT, Sigfox, Lora, NFC, and Bluetooth offer different coverage ranges creating an interoperability issues [10, 14].

5.1.6 Third Party Data Breach

In IoT network, data is uploaded on the cloud server which is mostly handled by the 3rd party. They can also violate the privacy of the user by observing continuous pattern of the user's data [9].

5.1.7 IoT Device Resource Constraints

Most of the devices lack storage, computation and battery power [11] which makes IoT devices vulnerable to different kinds of attacks.

5.1.8 Heterogeneity

Heterogeneity of protocols and IoT data types brings different constraints like privacy and interoperability issues.

5.2 Benefits of Integrating Blockchain with IoT

Following the features provided by Blockchain and its potential impact in the current phase, lots of research work is being done to amalgamate the Blockchain technology with IoT to counter the challenges faced in centralized IoT. Below are the benefits offered by BC based IoT framework.

5.2.1 Enhanced Security

Data of IoT can be secured via blockchain as data is stored in the form of Blockchain transactions which are encrypted. Smart contracts in Blockchain technology also help to improve security of IoT. Any data breach will update the users to prompt necessary action thereby improving security [15].

5.2.2 Minimizes Third Party Risks

Decentralized distribution of data removes third party risks as data is stored only at authenticate nodes and cannot be altered without approval of user [16].

5.2.3 Fault Tolerance

Since network of Blockchain based IoT devices is decentralized with point-to-point network where every device attached has the same data. So, failure/fault in one node will not hamper entire network.

6 IoT-Blockchain Applications

6.1 Chain of Things [17]

Chain of Things (CoT) is a Hong Kong-based blockchain + IoT venture production studio involved in a number of ventures in different industry verticals. An integrated blockchain and IoT hardware solution to solve IoT's issues with identity, security, and interoperability.

6.2 My Bit [18]

Provides the first IoT ecosystem in the world. It works on Smart contract system using Ethereum is built by industry veterans in Switzerland.

6.3 Slock.It [19]

Slock is an application combining the technology of Internet of things and Block Chain. This means that you can put advertisement on a blockchain, so the ad is connected to you. The reply to add is also on blockchain, and with internet of things, that transaction of the physical good can be confirmed, without a middleman. Slock is a platform that support all needed to share or sell. Suppose you want to share a room, a bicycle, a car. You find the ad on the blockchain, and you communicate with the owner without needing to give personal data. Only the counterparty, will be involved in all steps to complete a transaction.

6.4 Riddle and Code [20]

It helps various Stakeholders like companies from different spheres, to individuals and institutions to reap the benefits of digital technology of Blockchain and IoT which includes cybersecurity, data trust, avoiding online frauds, etc. The Riddle and

Code amalgamates smart card security features with Bitcoin Technology and IoT. Also, to tackle the projects related to Cryptography, they have hardware and software built in Stacks.

6.5 *Modem [21]*

Provides real time solutions in Intelligence of Supply chain and Automation making strong ecosystem where features like IoT, Blockchain based Network and Artificial Intelligence can be provided to various Industries.

7 Integration Challenges in Blockchain and IoT

7.1 *Scalability Issue*

Every time, there is a constant rise in the transaction of Blockchain ledger which might lead to centralization of huge data. Thus, it requires proper management of records to avoid future issues in Blockchain.

7.2 *Processing Power and Time*

IoT ecosystem has a lot of diversification as compared to other computing networks. IoT network encompasses various devices with varied computing capabilities, some of them lagging the capability of running an identical encryption algorithm at the desired speed.

7.3 *Storage*

Storage will be tedious. Although in Blockchain there is no need for a centralized server to store device ID's and transactions, the ledger needs to be stored on the nodes themselves and with an increase in time duration, the size of the ledger is also going to increase. Smart devices such as sensors lack this capability due to low storage capacity available.

7.4 Interoperability

The private and public nature of Blockchain will create a problem as their interoperability is also a challenge.

7.5 Legal and Compliance Issues

The absence of any legal and compliance code, due to blockchain being a new territory, poses a difficult threat to be dealt by manufacturers and the service providers in this field. Due to this serious problem, many business domains refrain from using Blockchain technology.

8 Conclusion and Future Work

IoT has revolutionized the digital world by connecting every device with Internet and generating huge amount of data which can be used in expert decision making by observing user's device patterns. Practically, lot of issues like interoperability, scalability, Security and Privacy, Trust and Heterogeneity issues have jeopardized IoT Technology. Also, different Protocols used for machine to machine interaction have made the task of vendors very difficult as interoperability issues will lead to issues like security breach. To overcome all the challenges, Blockchain does provide a solution to overcome these challenges but adoption of regulations by regulatory bodies is key to the Inclusion of Blockchain and IoT as part of Government Infrastructure. The adoption will help citizens, companies and government to adopt twin technology hazzle free. Lastly certain issues have also been highlighted like scalability and storage capacity which will affect Blockchain and IoT. So, research efforts need to be made in this direction to make this twin technology become the next big thing in I.T. world. In the next phase, we will incorporate the features of both the twin technologies and create a secure and Reliable IoT Application using NB-IoT. Also, the tradeoff between Public and Private Blockchain is an area of research where more emphasis will be to select the best Blockchain system for Real time Applications.

References

1. Alaba, F. A., Othman, M., Hashem, I. A. T., & Alotaibi, F. (2017). Internet of Things security: A survey. *Journal of Network and Computer Applications*, 88, 10–28.
2. Huckle, S., Bhattacharya, R., White, M., & Beloff, N. (2016). Internet of Things, blockchain and shared economy applications. *Procedia Computer Science*, 98, 461–466.

3. Haber, S., & Stornetta, W. S. (1990). How to time-stamp a digital document. In *Conference on the Theory and Application of Cryptography* (pp. 437–455). Springer.
4. Nakamoto, S. (2019). *Bitcoin: A peer-to-peer electronic cash system*. Manubot.
5. Kim, T., Kim, D. M., Pratas, N., Popovski, P., & Sung, D. K. (2017). An enhanced access reservation protocol with a partial preamble transmission mechanism in NB-IoT systems. *IEEE Communications Letters*, 21(10), 2270–2273.
6. Zhou, J., Cao, Z., Dong, X., & Vasilakos, A. V. (2017). Security and privacy for cloud-based IoT: Challenges. *IEEE Communications Magazine*, 55(1), 26–33.
7. Gopal, S. (2016). *Blockchain for the internet of Things. Tata Consultancy Services White Paper, TCS Design Services*.
8. Conoscenti, M., Vetro, A., & De Martin, J. C. (2016). Blockchain for the Internet of Things: A systematic literature review. In *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1–6). IEEE.
9. Akpakwu, G. A., Silva, B. J., Hancke, G. P., & Abu-Mahfouz, A. M. (2018). A survey on 5G networks for the Internet of Things: Communication technologies and challenges. *IEEE Access*, 6, 3619–3647.
10. Stankovic, J. A. (2014). Research directions for the internet of things. *IEEE Internet of Things Journal*, 1(1), 3–9.
11. Foukia, N., Billard, D., & Solana, E. (2016). PISCES: A framework for privacy by design in IoT. In *2016 14th Annual Conference on Privacy, Security and Trust (PST)* (pp. 706–713). IEEE.
12. Walker, M. A., Dubey, A., Laszka, A., & Schmidt, D. C. (2017). Platibart: A platform for transactive IoT blockchain applications with repeatable testing. In *Proceedings of the 4th Workshop on Middleware and Applications for the Internet of Things* (pp. 17–22).
13. Mohammadi, M., & Al-Fuqaha, A. (2018). Enabling cognitive smart cities using big data and machine learning: Approaches and challenges. *IEEE Communications Magazine*, 56(2), 94–101.
14. Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks*, 76, 146–164.
15. Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the internet of things. *IEEE Access*, 4, 2292–2303.
16. <http://www.techracers.com/blockchain-key-features>
17. Chain of Things. (2017). Retrieved February 1, 2018, from <https://www.chainofthings.com/>.
18. My Bit. (2017). Retrieved February 1, 2018, from <https://mybit.io/>.
19. Bahga, A., & Madiseti, V. K. (2016). Blockchain platform for industrial internet of things. *Journal of Software Engineering and Applications*, 9(10), 533–546.
20. Riddle and Code. (2017). Retrieved February 1, 2018, from <https://www.riddleandcode.com>.
21. Modum. (2017). Retrieved February 1, 2018, from <https://modum.io/>.
22. Lu, X., Niyato, D., Jiang, H., Kim, D. I., Xiao, Y., & Han, Z. (2018). Ambient backscatter assisted wireless powered communications. *IEEE Wireless Communications*, 25(2), 170–177.

Quality Assisted Spectrum Allocation in Cognitive NOMA Networks



D. Prasanth Varma and K. Annapurna

Abstract Future communications are developed with new communication standards for multi access technologies, where spectrum sharing and Non-orthogonal multiple access (NOMA) are the two latest approaches. To transfer the data from one end to another end through this wireless communication medium interference take place, which affects the total system performance. To minimize the interference in these communications the efficient spectrum sharing technique was developed. Cooperative relaying in NOMA system with spectrum sharing using threshold modeling were proposed in past. To improve the resource allocation in NOMA with cooperative relaying, instantaneous signal to noise ratio were used in obtaining higher outage throughput in a spectrum sharing cognitive radio-NOMA (CR-NOMA) system. In the previous methods constant thresholding is considered in the energy detection model and the quality of signaling is not observed, which results in lower throughput. In this work, a Quality Assisted Spectrum Allocation (QASA) based on loss probability is proposed and a dynamic threshold modeling is suggested for spectrum sensing under dynamic channel condition. Here the Rayleigh and Rician fading models are considered and used the Lagrange mathematical concept to calculate the bit error and to minimize the distortion. The throughput of the proposed work performs well when compared with the conventional method of Simultaneous wireless information and power transfer (SWIPT).

Keywords Cognitive radio · NOMA · Throughput · Dynamic thresholding · Spectrum allocation

1 Introduction

An approach of spectrum sharing for nodes with poorer channel condition is presented in [1]. This approach defines a node as primary with node having poorer channel condition. Here a two slot communication model is presented, which performs the spectrum sharing in the first slot part and perform a data exchange in the second slot.

D. P. Varma (✉) · K. Annapurna
Vignan's Foundation for Science, Technology and Research, Vadlamudi, India

The present approach is developed with the objective of maximization of network throughput by the optimization of time slot in maximization of network throughput. A NOMA communication based on overlay cooperative cognitive radio with the channel estimation and successive interference cancellation (SCI) is presented in [2]. The approach derives an analytical expression for outage probability computation for the primary and secondary receiver as a function of channel error and SCI. The modeling analysis of downlink NOMA network with an independent and non-identical Rayleigh channel with signal to noise ratio (SNR), and data rate is presented in [3]. In [4] the approach of total power over sum-SINR minimization is presented. The analysis of this presented approach defines an accurate approximation of achievable rate and derives the closed form of outage probability. For a near optimal solution to spectrum sensing and sharing, opportunistic interference regulated spectrum utilization is presented in [5]. This approach presented a non-binary coding in spectrum sensing for receiver centric modeling. The model is developed in reference to the interference temperature metric in constraint to spectrum usage. This approach offers an additional usage of spectrum using interference temperature metric in spectrum sharing. The effect of multiple channel interference on the signal propagation in CR-NOMA is presented in [6]. The interference of a PU and SU, Tx-Rx based data rate allocation is presented. [7] outlined a power-dominant multiplexing approach in a CR-NOMA based on time slotting over a real network. A cooperative relaying for the spectrum utilization enhancement in a non-orthogonal multiple access (NOMA) is proposed. A suboptimal power allocation in NOMA network with average rate of coding is presented. The presented approach is developed on independent Rayleigh fading channel interference. [8] outlined an analytical framework for computing the achievable rate by Gauss-Chebyshev integration developed for Rician fading channel. In [9] a full duplex approach with NOMA network is presented to improve network capacity for next generation communication. A secondary user access based on spectrum sharing is outlined in [10]. This approach optimizes the quality-of-service of a PU based on the aggregated rate allocation in a wireless network model. In [11] QoS based coding in consideration to individual user is presented. This approach proposes a CR based NOMA system for energy efficient coding.

In [1] an approach for throughput improvement with optimal power transfer approach is presented. This method outlined the optimal computation of power and information communication in a two slot manner. The presented approach controls the data rate allocation by monitoring the channel gain and power allocation. Wherein throughput is improved, the quality of the received data is unaddressed. In addition to throughput, quality of data exchanged is needed. The impact of QoS on the sharing spectrum has a greater impact on the offered quality of service in the network. In order to achieve the objective of low error resource allocation in a CR-NOMA network, here in this paper, a new approach of resource allocation in CR-NOMA based on observed QoS metric using distortion monitoring parameter is presented. The rest of this paper is outlined in five sections. Section 2 presents the existing approach of joint spectrum resource allocation; Sect. 3 outlines the proposed approach of

quality assisted spectrum allocation. Obtained experimental results for the proposed approach is discussed in Sect. 4. The conclusion of this work is presented in Sect. 5.

2 Joint Spectrum Resource Allocation

The block diagram of CR-NOMA operation is shown in Fig. 1 where each frame consists of two subsections. The first subsection is engaged in the spectrum sensing and sharing operation, where a “simultaneous wireless information and power transfer (SWIPT)” is performed. In the second subsection the data exchange takes place.

In spectrum sensing process, each secondary user (SU) measures the energy of the received signal is given by

$$E_n = \frac{1}{N} \sum_{k=1}^N |R(k)|^2, \tag{1}$$

where E_n is the signal energy and it is measured from all the signals which are coming from N multiple paths. The computed signal energy is compared with the given threshold in deriving the engagement of channel of PU. The detector performs the comparison as

$$\begin{aligned} E_n < Th; & H_0, \\ E_n \geq Th; & H_1. \end{aligned} \tag{2}$$

The spectrum which is sensed S_f to be free is allocated to SU for communication. The threshold limit I_{max} is defined such that interference should be minimum, and it should be tolerable caused due to SUs. The transmitting power P_t is used for

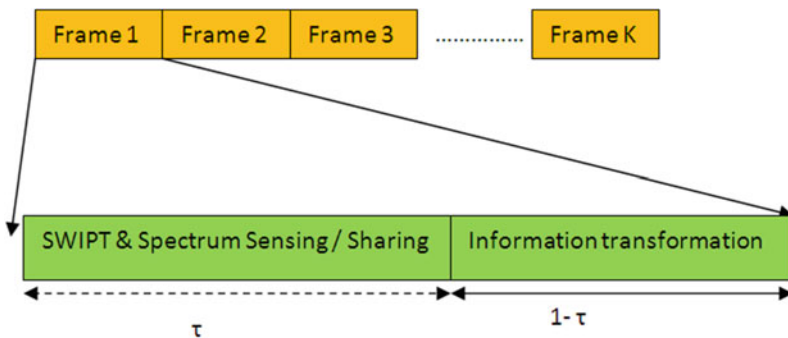


Fig. 1 Structure of CR-NOMA operation

information uplink and a channel gain of P_g is observed for the PU signal. The optimization equation for the rate is given by

$$\max(DR(t)) = P(H_0)(1-t) \left(1 - Q\left(\alpha + \sqrt{t f_s \vartheta}\right)\right) K_0 \quad (3)$$

The optimization function is represented with respect to power transmitted, minimum rate and interference. It is represented by

$$\max DR(t) = (1-t) \sum_{i=1}^N \log_2 \left(1 + \frac{P_i \vartheta_i}{P_s g_s + \sum_{j=i+1}^N P_j \vartheta_j} \right). \quad (4)$$

Subjected to

$$\sum_{i=1}^N P_i g_s \leq I_{\max}, \quad (5)$$

$$\sum_{i=1}^N P_i \leq P_t \quad (6)$$

$$\log_2 \left(1 + \frac{P_i \vartheta_i}{P_s g_s + \sum_{j=i+1}^N P_j \vartheta_j} \right) \geq R_i, \quad (7)$$

where the allocation of maximum rate is governed by allocated transmission power P_i and SNR ϑ_i over the given transmission power P_s with a channel gain of g_s . The optimization operation gives an optimal data rate allocation with respect to the interference observed in the channel and power allocated; however, the conventional method has no consideration on the offered QoS with outage probability. The losses of information result in lower QoS in such system. In developing an optimization solution in reference to allocated data based on loss probability, a new approach of Quality Assisted Spectrum Allocation is proposed.

3 Quality Assisted Spectrum Allocation

In the process of cognitive radio operation, spectrum sensing is performed using energy detector. Energy detector is a threshold based correlation of measured signal energy from PU with a given threshold. The decision is made as spectrum free when

signal energy is below the threshold limit as given in Eq. (8).

$$E_n = \begin{cases} \frac{1}{N} \sum_{i=1}^N \left(\frac{R_i}{\sigma}\right)^2 > \text{Thresh}, & (H_1), \\ \frac{1}{N} \sum_{i=1}^N \left(\frac{R_i}{\sigma}\right)^2 < \text{Thresh}, & (H_0). \end{cases} \quad (8)$$

This threshold has a major impact in the detection of signals and accuracy in spectrum sensing. In the scenario where channel has a fixed interference effect, a scalar threshold is applicable. However, the probability of estimation under varying channel condition has variation in signal energy and hence a constant threshold results in false detection. To improve the estimation accuracy, in this paper, dynamic threshold is derived by using the likelihood ratios and considering the signal scaling, shaping and probability of detection.

Probability of estimate $P(R | H_0)$ defines the detection probability of the received signal, and estimation of free spectrum is computed by a generalized likelihood ratio derived over a Gaussian distribution. The detection probability of the estimate is developed by the false alarm probability, defined as the estimate probability.

Probability of false alarm P_{F1} is defined by

$$P_{F1} = \text{Prob}\{E > \text{Thresh} | H_0\} \quad (9)$$

The threshold is derived for estimate based on Neyman-Pearson rule is defined by

$$\text{Thresh} = F_{\text{Est} | H_0}^{-1}(1 - P_{F1}, \theta_0, k_0). \quad (10)$$

Here, θ_0 and k_0 are the shaping and scaling parameter, respectively. During the communication when the signal is propagated in medium, it will undergo some fading effects. The Rayleigh fading is defined by

$$f(r) = \frac{r}{\sigma^2} \exp\left\{-\frac{r^2}{2\sigma^2}\right\}, \quad r \geq 0 \quad (11)$$

and the Rician fading is given by

$$k(\text{dB}) = 10 \log_{10}\left(\frac{k_d^2}{2\sigma^2}\right) \quad (12)$$

Due to the fading effect the interference takes place and it leads to a packet loss. This loss is measured by using the distortion parameter which is defined as bit error rate. By using the Lagrange method the objective function which is defined as distortion is minimized and updated. The cumulative similarity index is used to

calculate the similarity between the signals. The optimization function in reference to similarity index is defined by

$$\min\{J\} = D_{\text{istrSI}} + \lambda_{\text{SI}} \cdot R_a, \tag{13}$$

where J is the Lagrange optimization cost function.

Optimal rate is allocated by considering the distortion, rate requirements and power allocated. The final equation for maximum data rate allocation in addition to the existing constraint is subjected by

$$\begin{aligned} \max(DR) &= \min(\text{Distr}, R_a, P_i) \\ &= (1 - t) \sum_{i=1}^N \log_2 \left(1 + \frac{P_i \vartheta_i}{P_s g_s + \sum_{j=i+1}^N P_j \vartheta_j} \right) \geq R_{av}. \end{aligned} \tag{14}$$

4 Experimental Results

The developed approach is compared with the conventional method in terms of throughput. Simulation is carried out by using Matlab software, and the nodes are distributed randomly in the network. Each node has its bandwidth and the power which is required for communication. The network topology is shown in Fig. 2. The

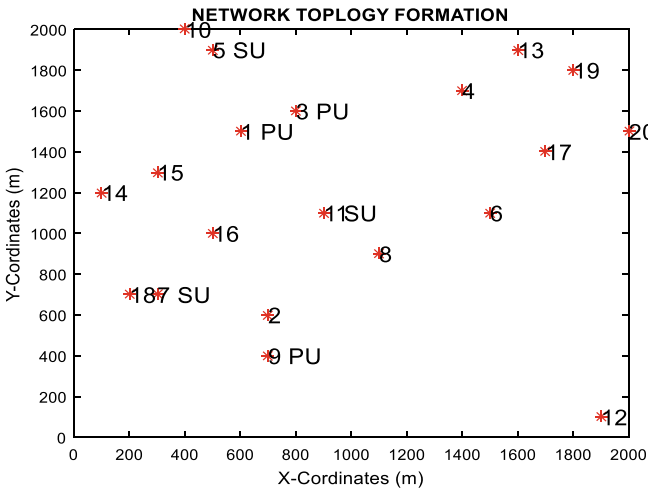


Fig. 2 Network distribution for simulation

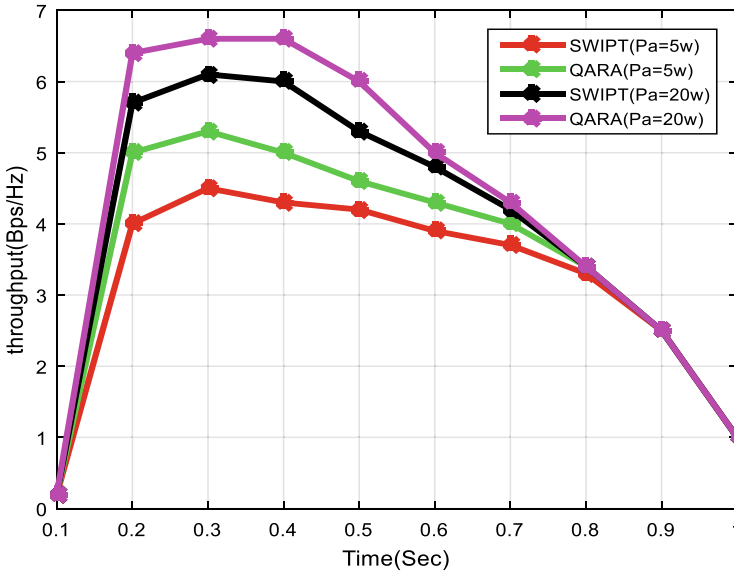


Fig. 3 Throughput with time for varying P_a

SWIPT and QASA are developed for optimal rate allocation. Compilation is carried out for one second. For communication the PU transmission power P_s is ranged from [5, 10, 30] and gain is 4. The P_a is set to [5, 20] W and [10, 30] W and the fading factor is 100. The SNR is [-10:8] dB.

The system throughput for an observation time of 1 s with varying P_a is presented in Fig. 3. The result for the developed approach illustrates an increase in throughput for a initial time period, however with course of time the channel get congested and interference builds. The interference leads to lower data rate allocation hence throughput decreases.

The network throughput with the varying channel gain is presented in Fig. 4. The channel gain increases the power level of the received signal and raises the interference level. The higher interference level builds an overhead in transmission and hence the throughput is decreased.

The throughput of PU with respect to simulation time and by varying P_a is shown in Fig. 5. With an increase in the downlink power, the throughput of PU is decreased. In case of SU with an increase in downlink power the throughput will increase.

The impact of channel gain on the network throughput is presented in Fig. 6 below. Initially the channel gain curve starts at peak point and when the channel gain is increased the throughput started falling down. With an increase in the channel gain, there will be more interference regarding PU which results in lower throughput.

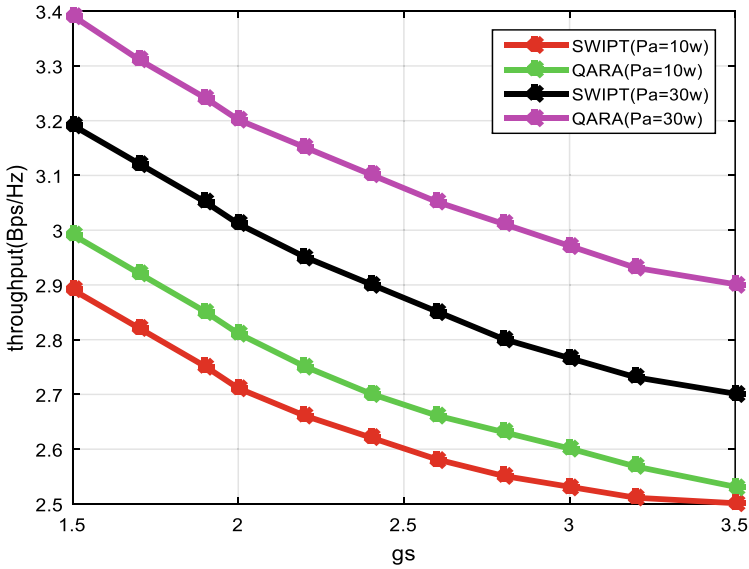


Fig. 4 Network throughput with variation of channel gain g_s

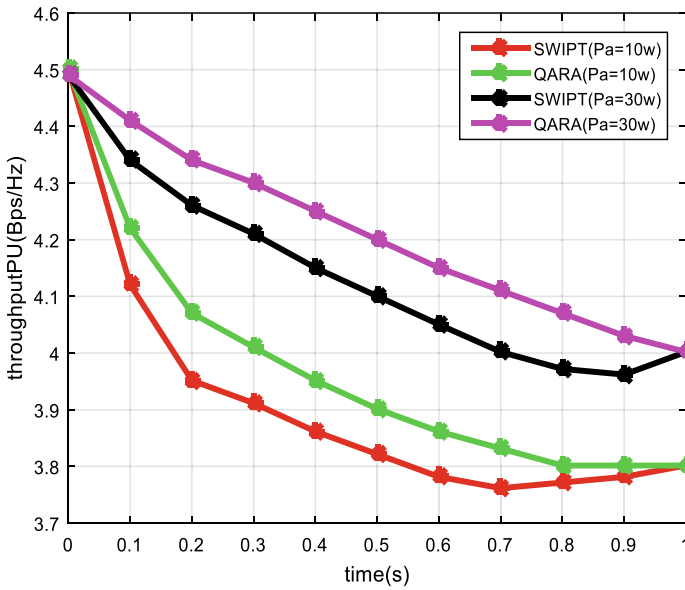


Fig. 5 Throughput PU for different P_a w.r.t. time variation

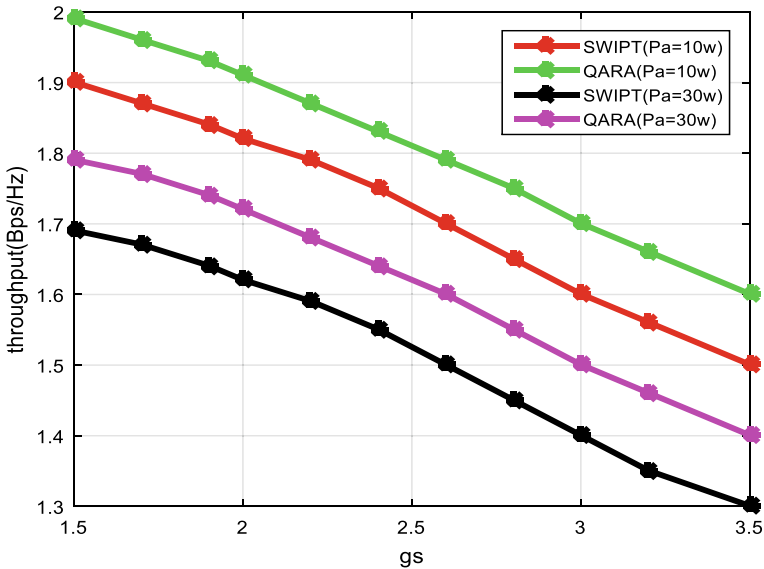


Fig. 6 Throughput over different channel gain with P_a varying

5 Conclusion

The optimum rate is allocated by minimizing the distortion using the developed approach Quality Assisted Spectrum allocation. The distortion is minimized by using the Lagrange optimization cost function and calculated the bit error rate using Lagrange multipliers. Similarity index is used to find the similarity between the transmitted signals and received signals and then using this index the cost function is updated. However in the existing method constant threshold is considered, but as the channels are very dynamic, the dynamic threshold is computed by using the Neyman–Pearson rule. The presence and absence of signal is detected clearly by using this rule with signal scaling and shaping parameters. Simulation is done by varying P_a , P_s , and time, and SNR is considered as $[-10:8]$ dB. Here the interference is reduced in this work, and the proposed work outperforms the conventional method.

References

1. Song, Z., Wang, X., Liu, Y., & Zhang, Z. (2019). Joint spectrum resource allocation in NOMA-based cognitive radio network with SWIPT. *IEEE Access*, 7, 89594–89603.
2. Kim, N.S. (2019). Cooperative overlay cognitive radio NOMA network with channel errors and imperfect SIC. *International Journal of Intelligent Engineering and Systems*, 12(5).

3. Zhang, Y., Dong, Y., Wang, L., Liu, J., Peng, Y., & Feng, J. (2019). Outage capacity analysis for cognitive non-orthogonal multiple access downlink transmissions systems in the presence of channel estimation error. *CMC*, 60(1), 379–393.
4. Baidas, M. W., Alsusa, E., & Hamdi, K. A. (2019). Performance analysis and SINR-based power allocation strategies for downlink NOMA networks. *IET Research Journals*, 1–13.
5. Bator, J., Tan, H.-P., Brown, K. N., & Doyle, L. (2007). Modelling interference temperature constraints for spectrum access in cognitive radio networks. In *2007 IEEE Int. Conf. Commun. (ICC)* (pp. 6493–6498).
6. Ghasemi, A., & Sousa, E. S. (2007). Fundamental limits of spectrum-sharing in fading environments. *IEEE Transactions on Wireless Communications*, 6(2), 649–658.
7. Kim, J. B., & Lee, I. H. (2015). Capacity analysis of cooperative relaying systems using non-orthogonal multiple access. *IEEE Communications Letters*, 19(11), 1949–1952.
8. Jiao, R., Dai, L., Zhang, J., MacKenzie, R., & Hao, M. (2017). On the performance of NOMA-based cooperative relaying systems over Rician fading channels. *IEEE Transactions on Vehicular Technology*, 66(12), 11409–11413.
9. Mohammadi, M., Shi, X., Chalise, B. K., Ding, Z., Suraweera, H. A., Zhong, C., & Thompson, J. S. (2019). Full-duplex non-orthogonal multiple access for next generation wireless systems. *IEEE Communications Magazine*, 57(5), 110–116.
10. Chen, B., Chen, Y., Chen, Y., Cao, Y., Zhao, N., & Ding, Z. (2018). A novel spectrum sharing scheme assisted by secondary NOMA relay. *IEEE Wireless Communications Letters*, 7(5), 732–735.
11. Zhang, Y., Yang, Q., Zheng, T., Wang, H., Ju, Y., & Meng, Y. (2016). Energy efficiency optimization in cognitive radio inspired non-orthogonal multiple access. In *IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (pp. 1–6).

Proficient Dual Secure Multi Keyword Search by Top-K Ranking Based on Synonym Index and DNN in Untrusted Cloud



Rosy Swami and Prodipto Das

Abstract Recent developments in cloud services have increased number of data owners to store their encrypted data in the cloud whereas equal or more data users participate to retrieve data. Secure retrieval of relevant data has become a challenging issue. In this paper, a secure ranking based multi-keyword search using semantic index is being developed. Initially, owner builds an index file by semantic representation of keyword. Security key is provided by Trusted Authority (TA) for decrypting the obtained results at the user side. TA manages dual security processes such as managing secret keys and issuing security devices to the data users. User query reaches proxy server, and it checks whether any frequent keyword matches with given query using Boolean Search. If it does not match, query enters into the main server which stores all document and index files to obtain relevant result using Deep Learning Neural Network. In deep learning neural network, the query is processed with vector space model to retrieve the relevant documents. Finally, user decrypts the relevant results obtained from deep neural network. The experimental result shows that our proposed model provides better performance in terms of recall, ranking privacy, precision, and searching time.

Keywords Cloud service provider · Semantic search · Secure · Multi-keyword search · Trusted authority

1 Introduction

Cloud computing is a new wave in information technology that probes enormous users as well as researchers towards it. In the cloud environment, the users store their data to utilize the flexibility and economic savings provided as well as to handle storage problems. The outsourced data may be in the form of documents, text, voice, multimedia, and so on and at the same time may also contain sensitive data. In general, the sensitive data includes personal identification such as PAN number, passport number, health information, financial information, bank account

R. Swami (✉) · P. Das
Department of Computer Science, Assam University, Silchar, Assam, India

details, and so on. These sensitive data are vulnerable to several attacks and require confidentiality. Thus in cloud storage, outsourcing of sensitive information to cloud server becomes a risk. To ensure security and privacy of the outsourced sensitive data, it is necessary to encrypt the data before outsourcing. Encryption techniques such as Advanced Encryption Standard (AES) method, Rivest-Shamir-Adleman (RSA) algorithm, Attribute-Based Encryption (ABE) method, and Blowfish algorithms are utilized to ensure security in cloud computing [1]. Comparisons made among these algorithms are based on key size, block size, scalability, encryption and decryption speed, and quality of security. In many cases, researchers have concluded that AES algorithm performs better than other algorithms in which small key size is required for high-level security. Most of the ABE schemes incur high cost of computation [2]. In multi-keyword search, security and privacy plays vital role which can be provided with the consideration of multiple attributes [3].

In cloud storage where multi owners are involved, key management is a complex process [4]. In mass information retrieval from cloud, the term is weighted based on location of the particular term in the document [5]. This method is often classed as N-level vector model (NVM). Keyword frequency is also considered as a main feature in multi-keyword searching process.

The major contributions of this paper are summarized as follows:

- Proxy server is introduced in Cloud Service Provider (CSP) to reduce the searching time, and searching efficiency is increased using Boolean search in the proxy server.
- Main server supports multiple users at a time with the help of Deep learning based Neural Network (DNN), which provides an accurate result.
- Trusted Authority (TA) is employed to provide secure document retrieval for authorized users. TA manages dual security processes as key management and Security Device Issuing (SDI).
- Secure top k ranking is achieved using Euclidean distance calculation, and accuracy of document retrieval is improved.

The rest of the paper is organized as follows: Section 2 surveys previous work regarding multi-keyword search methods and synonym index-based ranking methods. Section 3 presents proposed work and discusses its processes and procedure. Section 4 describes the result and output of proposed system. Section 5 concludes the contributions of the proposed system.

2 Related Work

Personalized search is presented in [6], with the support of Semantic Web Ontology WordNet. User's history is analysed for building user interest model for each user and also scoring mechanism adopted for users to express their interest. Automatic evaluation of keyword priority is realized through the interest model. Personalized Rank Search Encryption (PRSE) is employed with the semantic extension of user's

query based on interest model and privacy-preserving. However, building the interest model became complex due to user's dynamic activities.

The authors in [7, 8] have discussed the index structure. Generalized Inverted Index (Ginix) for keyword search, which is efficient structure, is presented in [7]. In Ginix to save storage space all consecutive IDs are in inverted list into intervals. To improve Ginix it is necessary to reorder the documents in datasets using scalable algorithms. Tree-based index structure is used in [8] to facilitate search and updating process for the concern of efficiency. The authors to enhance the security employed dominant function-hidden-inner product encryption by precluding search patterns. But in this scheme revocation process is challengeable. Vector Space Model is used in [9] to construct searchable index to achieve accurate search results, and tree-based index structure is employed to improve efficiency in multi-keyword search. Here index file is secured but has inefficient key management scheme which provides encryption and decryption key for both owner and user.

Authentication is necessary at user side to ensure security of system. In [10], a new Identification Based Authentication protocol based on hierarchical model of identity is presented. It is lightweight and efficient protocol on user side. In this scheme hierarchical model is comprised of root Public Key Generation (PKG) (level-0), data center with sub-PKG (level-1), and bottom level (level-2) with users. Users are provided with unique names for identification. It is a time-consuming process to assign unique name to users since in cloud number of users increase rapidly. In [11], a security device and PKG for user authentication reduces time and also maintain security. The user must have these two factors for authentication. If the security device is lost or stolen, the user is provided with new security device, and lost/stolen device is revoked to prevent encryption and decryption through old device. The server performs some algorithms to revoke the device.

3 Proposed Work

3.1 System Overview

Our work concentrates on searching efficiency and secures ranking process in untrusted cloud server. Our overall architecture is comprised of Cloud Service Provider (CSP), Data Owners (DOs), Data Users (DUs), and Trusted Authority (TA). Figure 1 illustrates our architecture with multiple DOs and multiple DUs. Outsourcing encrypted documents and index file to cloud server is the initial process of DOs. Here multiple DOs are allowed to upload any numbers of documents and single index file. Initially DO separate the documents into keywords and build an index with semantic representation of keywords using WordNet to increase search efficiency. DO calculate Term Frequency (TF), Inverse Document Frequency (IDF), and $TF \times IDF$ values for each keyword and store its weight in index table. Then DO construct all documents into one index file with particular keyword location. After

constructing index file DO outsources encrypted document files and index files to cloud server with its secret key and provides the encrypted keys to TA.

Encryption at DO side is performed by using the DO's secret key. Encryption at DO side consists of *index encryption* and *document encryption*. The entire encryption process is shown in Fig. 2. Index encryption is performed by Diffie-Hellman Algorithm (DHA) which encrypts the keywords in index file with their corresponding document. In Document encryption, each document in document set is encrypted by AES algorithm using DO's key. DO provides the encryption keys to TA for key management. AES takes block of data of 128 bits and a key and produces Ciphertext as output.

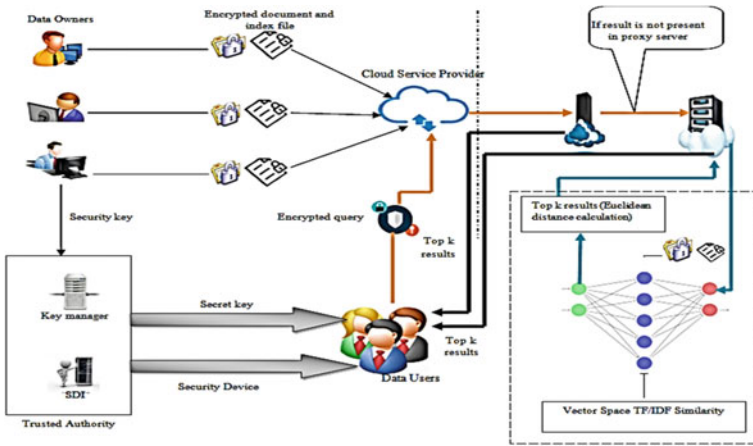


Fig. 1 Overall architecture of our proposed work

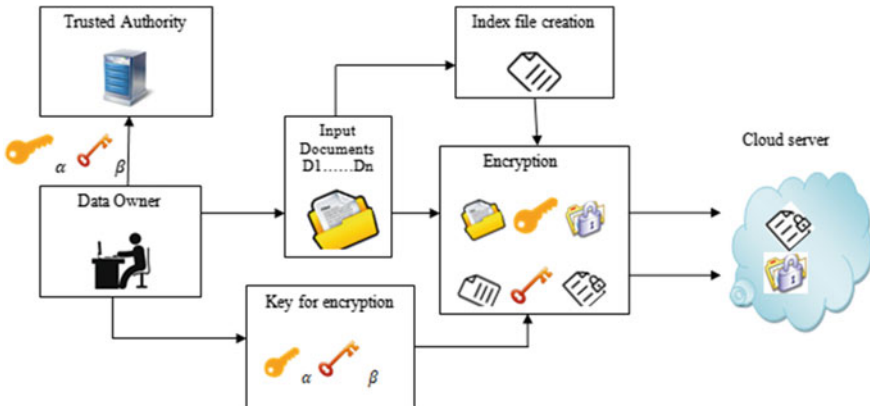


Fig. 2 Encryption process

3.2 Multi-keyword Search

The user query (encrypted multi-keyword) is searched in cloud server. In proposed system, CSP is comprised of main server and proxy server. Hereby we have introduced proxy server for minimizing searching time of user queries in cloud. The relevant result search is first performed in proxy server, if the result is not found for query in proxy, then the query is moved into main server. Proxy server stores the recently searched keywords and top k relevant results for those keywords. User query is arrived at proxy server and searched if there any keyword matches with the query. Consider a query Q with keywords $w_1, w_3,$ and w_4 . Let $C = (c_1, c_2, c_3)$ is the recently searched multi-keyword set (candidate set) in proxy server. For Q present in proxy then the candidate set in proxy should satisfy the following condition:

$$c_i = \begin{cases} 1 & \text{for } w_1 \wedge w_3 \wedge w_4 \\ 0 & \text{for otherwise} \end{cases} \tag{1}$$

Let $c_1 = (w_1, w_3, w_5); c_2 = (w_1, w_2, w_3); c_3 = (w_1, w_3, w_4)$

Boolean search is performed between Q and each multi-keyword set in C . The top k result of multi-keyword is retrieved to DU when all keywords in query are present in recently searched multi keyword. For above example, the proxy server retrieves the top k result of c_3 . The query is moved into main server when it is not able to find any matched keyword in proxy server. In main server, DNN is employed for similarity measurement and to minimize search time.

Main server effectively stores the encrypted documents and index files in the allocated space of each DO. Similarity between query and keyword is processed by DNN. In Fig. 3 queries from multiple users are fed into input layer and in hidden layer similarity between each query and index files are measured by Vector Space Method (VSM).

The DNN allows multiple users at the same time and processes user’s queries very effectively and quickly. Input layers process multiple users’ queries, then hidden layer performs similarity measurement of each user’s query and output layer provides the similarity results. After similarity measurement, the top k relevant results are obtained by calculating distance between keywords based on term location which is provided in index file. In hidden layer similarity between user’s query and keywords is calculated by Vector Space $TF \times IDF$ model. Consider x_i is input and w_i is weight of input at each layer, then the

Output form hidden layer is

$$z = \sum_{i=0}^n w_i x_i \tag{2}$$

Output from output layer is

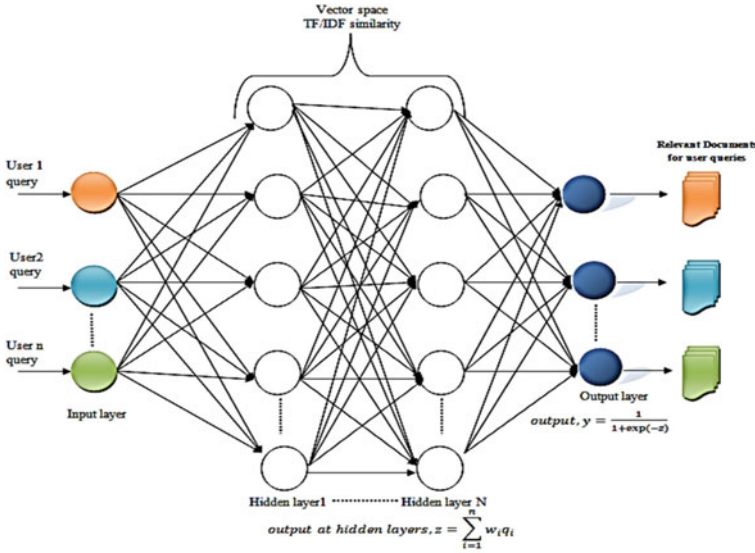


Fig. 3 DNN based keyword search

$$y = \sigma(z) = \frac{1}{1 + \exp(-z)} \tag{3}$$

Our proposed work concentrates on semantic multi-keyword search. Hence in hidden layer, the semantic representation of keywords in query is performed. This increases the relevancy of the results and precision factor. Then similarity measurement takes place in another hidden layer to obtain relevant results for query. At output layer, the relevant documents are obtained from DNN.

Vector Space Method VSM is the most popular way of similarity measurement in keyword search. In VSM, M-dimensional vectors are used to represent documents and queries using each dimension for representing a different keyword. The document vector is constructed using weight of keywords present in that document. The weight of each keyword is given in index file by DO based on $TF \times IDF$ value. The document vector for document D is

$$\vec{D} = (t_{D,w1}, t_{D,w2}, \dots, t_{D,wn}) \tag{4}$$

where $t_{D,w1}$ represents the weight of keyword1 in document D . The query vector is denoted as

$$\vec{Q} = (q_1, q_2, \dots, q_n) \tag{5}$$

The next step after constructing document and query vector is to measure similarity between all the documents and the query. The cosine measure is employed for similarity measurement. In cosine measurement similarity between document $D1$ and query Q is given by

$$Sim(Q, D1) = \cos(\theta) = \frac{\vec{Q} \cdot \vec{D1}}{|\vec{Q}| \times |\vec{D1}|} \tag{6}$$

When keyword in query $Q1$ lies in document, then Q_i equals t'_{wi} ; otherwise it equals zero.

$$\cos(\vec{D1}, \vec{Q1}) = \frac{\sum_{i=1}^n t_{D,wi} t'_{wi}}{\sqrt{\sum_{i=1}^n (t_{D,wi})^2} \sqrt{\sum_{i=1}^n (t'_{wi})^2}} \tag{7}$$

Top k result retrieval Euclidean distance is calculated between keywords in query and in document to improve retrieval accuracy. The top k relevant result is retrieved by calculating Euclidean distance between keywords in document vector and query vector. Euclidean distance between two keywords is given by

$$d_{EQ} = \sqrt[3]{(q_1 - q_2)^2} \tag{8}$$

$$d_{ED} = \sqrt[2]{(t_{w1} - t_{w2})^2} \tag{9}$$

where

d_{EQ} Distance between keywords in Query.

d_{ED} Distance between keywords in Document.

3.3 Document Retrieval

The initial stage of document retrieval is to generate queries on user side. The DUs encrypt the query using AES algorithm and send the query to CSP with their security device provided Security Device Issuer which is maintained by TA. The security device is a secure random number provided for user identification and the user without the security device will not be allowed to search in cloud server. SDI revokes the security device when DUs lost their security device. When query enters into CSP the multi-keyword search algorithm is performed and top k relevant result is obtained. DU decrypts the retrieved top k documents using secret provided by key manager which is maintained by TA.

4 Experimental Evaluation

4.1 Simulation Setup

Java is used to implement the proposed scheme. The software tool NetBeans 6.9.1, to develop java programs, is installed for our simulation. Wamp server (MySQL 5.1.36) is also installed as server in our work. CloudSim, which offers essential classes for defining data centers, computational resources, and virtual machines, is used for deploying CSP and proxy server. WordNet is a tool for computational linguistics and natural language processing used in our scheme. For simulation, 25–30 numbers of owners, 100–150 numbers of users, 8 numbers of keywords, and 200 numbers of documents are considered.

4.2 Comparative Analysis

In this section, we compared our proposed multi-keyword search system with existing works MRPS method [9], N-level vector model [5], Enhanced-TFIDF method [12], and Space Efficient Semantic search [13].

Effectiveness of Precision Precision is one of the standard measures in keyword search. Precision depends on the accuracy of retrieved documents.

Figure 4 shows that the performance of Enhanced-TFIDF method and MRPS method is significantly changed with respect to number of documents retrieved. These methods are not able to achieve better results with the increasing number of documents. Our proposed scheme using DNN achieves precision of 89% which is significantly higher than previous work.

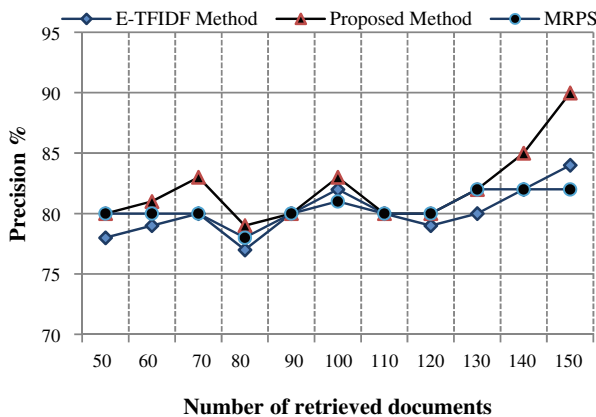
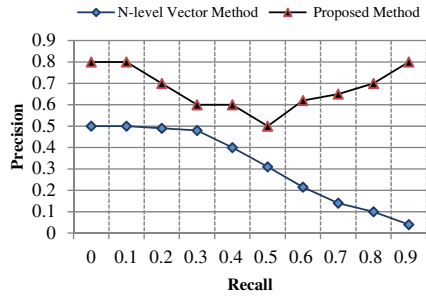


Fig. 4 Comparative results for precision

Fig. 5 Comparative analysis for precision-recall



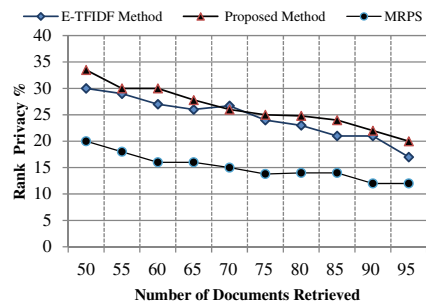
Effectiveness of Recall In information retrieval, recall(sensitivity) is one of the performance metrics which increases the accuracy. It depends on the number of related documents retrieved for query. But the precision and recall are contradictory (i.e.) upon definite threshold the increase of recall causes reduce in precision and increase of precision causes decrease in recall.

Figure 5 shows that our proposed scheme achieves high precision for increased recall compared with previous works. In our work, we have achieved high precision for increased recall compared to N-level vector method. In our work a precision of 0.8 is achieved for increased recall a factor of 0.2.

Effectiveness of Rank Privacy Rank privacy is one of the significant performance metrics of privacy. It depends on rank of each document in the retrieved top k result. Generally, ranking is processed with non-encrypted documents for accurate ranking. But in our proposed work, we have processed ranking over encrypted document to ensure security. Figure 6 shows that E-TFIDF and MRPS methods achieved significantly low-rank privacy. In our work, we provide accurate ranking over encrypted data. The relevant result from DNN is further processed with the calculation of Euclidean distance which provides accurate rank for retrieved documents. Our proposed work achieves rank privacy comparatively higher than previous work. We have achieved rank privacy of 33.5%.

Effectiveness of Search time Search time includes the time taken for index searching time and query processing time. Figure 7a–c show that our proposed scheme reduces

Fig. 6 Comparative results for rank privacy



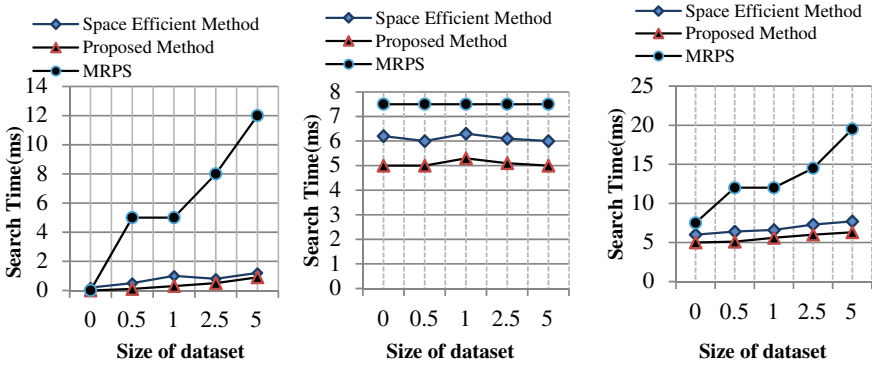


Fig. 7 a Comparative results for Hash index search time; b Comparative results for query processing time; c Comparative results for total search time

the searching time in terms of index searching time, query processing time, total search time due to effective indexing with synonym set. Finally, our proposed scheme achieves high precision, recall, and rank privacy with reduced search time.

5 Conclusion

In this paper, we presented a secure top k ranking based multi-keyword search in untrusted cloud. The authorized user can be allowed to search over encrypted data by using security device provided by TA. Proxy server reduces the searching time by storing recently searched keywords with their top k results. DNN in main server effectively supports multiple users at a time to search and also gives accurate relevant results using vector space model for similarity calculation in hidden layers. Retrieved relevant documents are again processed by calculating Euclidean distance between keywords and query to rank the retrieved documents. Encryption is performed using AES algorithm. Finally, performance evaluation indicates that our proposed scheme is effective in multi-keyword search and top k result retrieval is also effective in precision, recall, ranking privacy, and search time.

References

1. Sonam Darda, M., & Kulkarni, M. K. (2015). Study of Multi-keyword ranked searching and encryption technique over cloud. *International Journal of Computer Science and Information Technologies*, 6(6), 5417–5420.
2. Wang, S., Jia, S., & Zhang, Y. (2019). Verifiable and multi-keyword searchable attribute-based encryption scheme for cloud storage. *IEEE Access*, 7, 50136–50147.

3. Zhang, L., Zhang, Y., & Ma, H. (2018). Privacy-preserving and dynamic multi-attribute conjunctive keyword search over encrypted cloud data. *IEEE Access*, 6, 34214–34225.
4. Zhang, W., Lin, Y., Xiao, S., Wu, J., & Zhou, S. (2016). Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing. *IEEE Transactions on Computers*, 65(5), 1566–1577.
5. Peng, J., Tang, S., Zhang, L., & Liu, R. (2017). Information retrieval of mass encrypted data over multimedia networking with N-level vector model-based relevancy ranking. *Multimedia Tools and Applications*, 76(2), 2569–2589.
6. Fu, Z., Ren, K., Shu, J., Sun, X., & Huang, F. (2016). Enabling personalized search over encrypted outsourced data with efficiency improvement. *IEEE Transactions on Parallel and Distributed Systems*, 27(9), 2546–2559.
7. Wu, H., Li, G., & Zhou, L. (2013). Ginix: Generalized inverted index for keyword search. *Tsinghua Science and Technology*, 18(1), 77–87.
8. Yan, J., Zhang, Y., & Liu, X. (2016). Secure multi-keyword search supporting dynamic update and ranked retrieval. *China Communications*, 13(10), 209–221.
9. Fu, Z., Sun, X., Liu, Q., Zhou, L., & Shu, J. (2015). Achieving efficient cloud search services: Multi-keyword ranked search over encrypted cloud data supporting parallel computing. *IEICE Transactions on Communications*, 98(1), 190–200.
10. Li, H., Dai, Y., Tian, L., & Yang, H. (2009). Identity-based authentication for cloud computing. In M. G. Jaatun, G. Zhao & C. Rong (Eds.), *Cloud Computing. CloudCom*, LNCS 5931 (pp. 157–166.). Beijing, China.
11. Liu, J. K., Liang, K., Susilo, W., Liu, J., & Xiang, Y. (2016). Two-factor data security protection mechanism for cloud storage system. *IEEE Transactions on Computers*, 65(6), 1992–2004.
12. Fu, Z., Sun, X., Linge, N., & Zhou, L. (2014). Achieving effective cloud search services: Multi-keyword ranked search over encrypted cloud data supporting synonym query. *IEEE Transactions on Consumer Electronics*, 60(1), 164–172.
13. Woodworth, J., Salehi, M. A., and Raghavan, V., (2016). S3C: An architecture for space-efficient semantic search over encrypted data in the cloud. In *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data)* (pp. 3722–3731). Washington D.C., USA.

Transfer Learning-Based Detection of Covid-19 Using Chest CT Scan Images



Aryaman Chand, Khushi Chandani, and Monika Arora

Abstract The global pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus has had historical impact on the world. The virus causes severe respiratory problems and with an R_0 of 5.7, spreads at a rapid rate. At the time of writing, there were over 85 million cases and 1.8 million deaths caused by COVID-19. In the proposed methodology, Deep Convolutional Neural Networks (DCNNs) have been trained, with the help of transfer learning, to learn to identify whether a suspected patient is suffering from this disease using their chest CT scan image. Transfer learning technique enables the transfer of knowledge from pre-trained models which have been previously trained on extremely large datasets. Various DCNN models have been applied such as AlexNet, ResNet-18, ResNet-34, ResNet-50, VGG-16, and VGG-19. The DCNNs were evaluated on a set of 2,481 chest CT scan images. Various performance metrics (Accuracy, MCC, Kappa, F1 score, etc.) were calculated for all DCNN models to enable their comparative evaluation. After extensive testing, ResNet50 was found to give the best results in this binary classification task. The highest accuracy achieved was 97.37% and highest kappa was 0.947. Identification of presence of COVID-19 using this method would provide great benefit to society and mankind.

Keywords Convolutional neural network · SARS-CoV-2 · Coronavirus · COVID-19 · Transfer learning

1 Introduction

Amid the ongoing global pandemic caused by the SARS-CoV-2 virus, the world has seen loss of life and property to an extreme degree. The disease caused by this virus, colloquially termed as COVID-19, causes several respiratory problems in humans, including but not limited to dry coughs, chest pain, chest pressure, shortness of breath, and sore throat. In serious cases, this disease can also cause loss of speech and movement [1]. Moreover, this disease spreads at a rapid pace. A recent study

A. Chand (✉) · K. Chandani · M. Arora
Amity University, Noida, Uttar Pradesh, India

shows that COVID-19 has a basic reproductive number (R_0) of 5.7, which means that one infected carrier of this disease can potentially transmit to an average of 5.7 people [2]. The onset of this pandemic has prompted governments all over the world to impose lockdowns and travel bans. However, the spread has been widespread and so extensive that small, remote island nations in Oceania have reported cases as well. At the time of writing, there are over 85 million cases and 1.8 million deaths due to this disease [3]. Hence, healthcare systems worldwide are overburdened and there is a dearth of doctors, nurses, and medical support staff. Moreover, conventional testing methods of COVID-19 diagnosis have issues of their own. The molecular tests (PCR tests/nucleic acid tests/viral RNA tests) are unreliable with a false negative rate of as high as 37% [4]. A professional lab and a few days are required to generate a result from this type of test. Moreover, since the outbreak of COVID-19, medical laboratories are operating at capacity, thus increasing the time taken to generate a result from a normal two days up to ten days in several places [5, 6]. They are also a challenge to mass produce since the testing process requires additional reagents. Antigen tests provide faster results but are even more unreliable, with a reported false negative rate as high as 50% [7].

Therefore, considering these factors, there raises a case for a COVID-19 testing mechanism that offers quick and reliable results. In this research, we explore the possibility of a testing mechanism that is based on the chest CT scan images of the suspected patient. Studies show that an infected patients' chest shows changes in a CT scan even before the onset of symptoms [8]. The infrastructure to get a CT scan done is also widespread and prevalent all over the world, and costs are moderate when compared to conventional testing methods.

Although there has been some research in this space, certain design and execution challenges were identified. Related work in medical applications has been done to identify COVID-19. Transfer learning-based detection of COVID-19 was presented by [9]. They achieved a maximum accuracy of 87%. Deep learning based segmentation system to quantify lung infection areas was presented by [10] and obtained a maximum accuracy of 91%. An automatic DL-based COVID-19 diagnostic system was proposed by [11] and achieved a maximum accuracy of 85%. A DL-based technique to quantify the severity of COVID-19 using segmentation was presented by [12], achieving a maximum accuracy of 94.8%.

In this paper, a model based on transfer learning of state-of-the-art Deep Convolutional Neural Network (DCNN) models is proposed. Transfer learning enables the transfer of knowledge from pre-trained networks which have been trained on massive image datasets, such as ImageNet. The idea behind transfer learning is to transfer the weights of the layers of the DCNN and apply these layers to a new classification task. The detection of COVID-19 from chest CT scan images was casted as a binary classification problem, with the dataset used containing chest CT scans of healthy patients and those suffering from COVID-19. The main contribution of this research was to utilize several off-the-shelf and pre-trained DCNN models via transfer learning approach. Another major contribution of this research was to find

the right image transformation techniques and tune the model hyperparameters—variables such as batch size, learning rate, and number of epochs—to obtain better classification results.

The chapter scheme is as follows. Section 2 contains details about the proposed methodology. Section 3 discusses the results. Section 4 concludes the research.

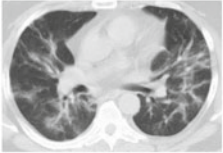
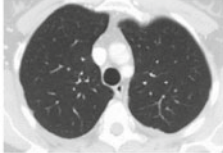
2 Methodology

In the proposed methodology, the goal is to leverage transfer learning technique and apply the pre-trained models in the binary classification task of identifying the presence of COVID-19 in suspected patients. By using these pre-trained models, their demonstrated capabilities in image classification can be transferred to the task at hand [13]. This can be done by training the models without updating the weights of the weighted layers. The training of all of the models was done using the fastai library in Python. The fastai library automatically removes the classification head of the pre-trained DCNN model and replaces it with a new head suitable for the database. It also sets automatic defaults for the learning rate, optimizer, etc. [14]. The execution of the code was done in Google Colaboratory environment due to the need of high-powered GPUs for training DCNNs. The hyperparameters of each of the DCNNs were tuned extensively to improve the results obtained. Hyperparameters include variables such as number of epochs, batch size and learning rate. These variables help control the learning process of the DCNNs. The dataset used to train and evaluate the models consists of 2,481 images of chest CT scans. The dataset has been described in detail in Sect. 2.1. To increase the robustness and integrity of the DCNN models, data augmentation technique was used. Data augmentation virtually increases the size of the dataset by applying certain transformations to each image of the dataset in order to create several copies of it. In theory, this technique should improve the classification accuracy of the DCNN models since they have more data to train on. Various permutation and combinations of different data augmentation techniques were explored in order to obtain the best possible results. Data augmentation has been further elucidated in Sect. 2.2. The several DCNN models used in this research have been explained in detail in Sect. 2.3. After training all the DCNN models, several performance metrics were calculated in order to enable the comparative analysis of the several models. Figure 1 shows the block diagram of the methodology proposed.

2.1 Data Description

The images were collected by researchers in Brazil to encourage the research and development of deep learning based methods of detection of COVID-19. CT scan images of real patients were collected from hospitals in Sao Paulo, Brazil [15]. There

Table 1 Details of the dataset with sample image of both classes

Class name	Number of images	Sample image
COVID-19	1252	
Non-COVID-19	1229	

are a total of 2,481 images in the dataset, split into two classes—COVID-19 and non-COVID-19. There are 1252 chest CT scans of patients suffering from COVID-19 and 1229 chest CT scan images of healthy patients. All of the CT scan images are in the transverse anatomical plane; hence, there was little need for preprocessing in that regard. Sample images of infected and healthy patients have been shown in Table 1. Since the DCNNs were trained and evaluated on fastai, there was no need to specify a separate test dataset. The validation split was set to twenty percent for all DCNN models.

2.2 Data Augmentation

Data augmentation is the process of manipulating the images of the dataset with certain image transformations such that several copies of each image are created and the size of the dataset is artificially increased. The copies of each image of the dataset contain the same contextual information from the human perspective. However, the copies are not identical down to the pixel levels; hence, they act as more training data for the DCNN model to train on—since the models are fed with several variations of the same image. There are several transformations available in the fastai library. However, it is also critical to apply these transformations with caution so as to retain the important data within the images of the dataset. The following transformations were found to work best with the dataset—rotation, horizontal flip, lighting, zoom and warp. These transformations were applied to a random image from the dataset, and are shown in Fig. 2.

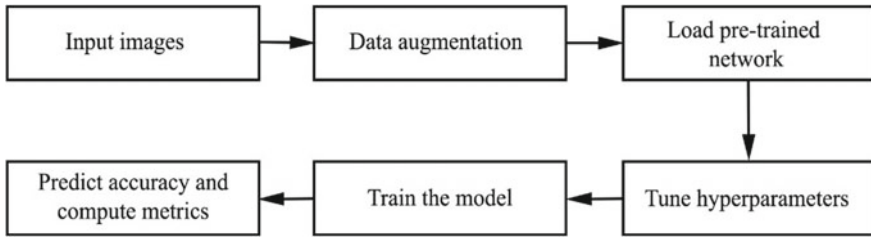


Fig. 1 Block diagram of proposed methodology

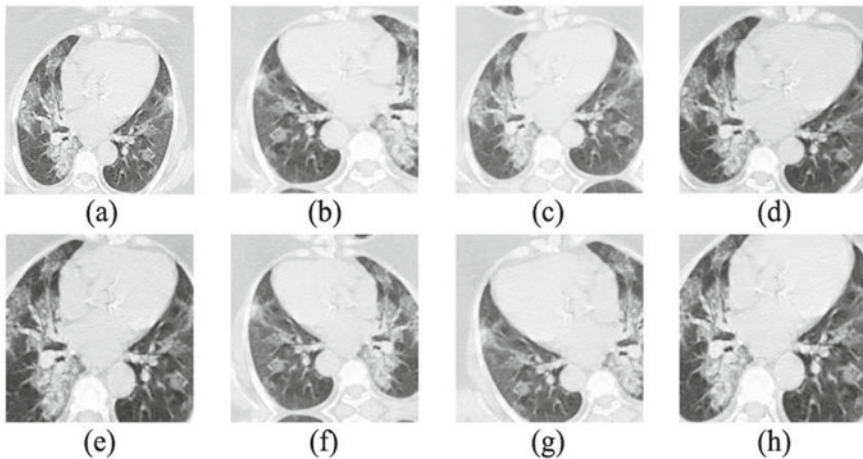


Fig. 2 a Original image b horizontal flip c warp and lighting d zoom, rotation and warp e zoom and lighting f horizontal flip, rotation and warp g horizontal flip and warp h zoom and lighting

2.3 Model Architectures

The DCNN models used in this research were VGG16, VGG19, AlexNet, Resnet18, ResNet34, and ResNet50

AlexNet

The architecture of AlexNet DCNN comprises of eight layers, including five convolutional layers and three fully connected layers. AlexNet was the first DCNN to make use of the ReLU activation function. There are 60 million parameters in this architecture. It utilizes data augmentation and dropout to counteract overfitting [16].

VGG

VGGnets created by the Visual Geometry Group at Oxford University were derived from the AlexNet architecture. They utilize smaller 3×3 receptive fields and a stride of one to preserve spatial resolution. They use the ReLU activation function, akin to

AlexNet. Various versions of VGGnets exist, but the two best performing models, VGG16 and VGG19, with sixteen and nineteen layers, respectively, were used in this research [17].

ResNet

ResNets are fundamentally different to linear DCNN architectures such as AlexNet and VGG. ResNets utilize the residual block and rely on micro architecture modules. The identity mapping, or the skip connection, feeds the output of the previous layer to the layers ahead, ensuring that the performance of subsequent layers does not fall. This concept enabled much deeper models to be trained [18].

3 Results

The binary classification task of identifying COVID-19 from chest CT scans was completed with very promising results. ResNet50 performed the best overall, as discussed further in this chapter.

Using the learning rate finder graph of the fastai library, the optimal starting learning rate was found out to improve the results and optimize the training time. From Fig. 3, a learning rate slice of ($1e^{-3}$, $7e^{-2}$) was chosen for the training of ResNet50. In this range, the loss has a steep negative slope, ideal for the training process. Extensive testing with various hyperparameters and data augmentation transformations resulted in some notable and predictable results. The optimal batch size was found to be sixteen, and tests with higher batch sizes of thirty-two and sixty-four yielding worse results. It was observed that around seven to eight epochs were needed to converge to local minima, and tests with more training epochs did not yield any better results. Therefore, for the final results obtained for each model, the number of training epochs was set to eight. Further, minimal tweaking of the learning rate was required since

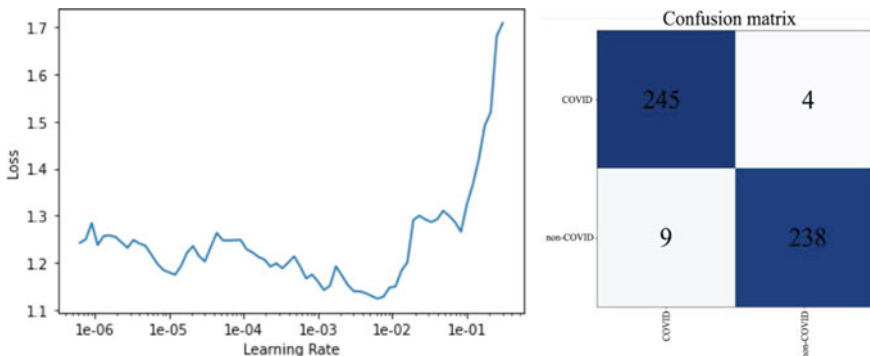


Fig. 3 Learning rate plot and confusion matrix for ResNet152

the in-built learning rate finder of the fastai library provided a good head start and the learner object automatically updated it during training.

Twenty percent of the dataset was reserved for validation, while eighty percent was used for training. The DCNN's predictions on the validation images were represented in the form of a confusion matrix, as shown in Fig. 3. The confusion matrix is a 2×2 matrix that represents the true positives (infected patients that were classified as having the disease), true negatives (healthy patients that were classified as not having the disease), false positives (healthy patients that were classified as having the disease), and false negatives (infected patients that were classified as not having the disease). Minimizing false negatives is one of the most critical targets of this research since conventional testing methods of COVID-19 have a high rate of false negatives.

Performance metrics were calculated using the confusion matrix obtained for each DCNN model. Firstly, basic metrics such as accuracy, specificity, and recall were calculated (Eqs. 1–3). Recall highlights the probability of detection of the disease and specificity highlights the selectivity of rejection.

$$\text{Accuracy} = \frac{TN + TP}{TN + FN + TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

Further evaluation needs to be done in order to compare the DCNN models thoroughly since the metrics calculated above are only true indicators for a balanced dataset. Since the dataset used in this research is unbalanced, true comparison between the DCNN models requires more metrics. These metrics are precision, F1 score, false negative rate (FMR), error, MCC and Cohen's Kappa index (Eqs. 4–9). Precision represents the positive predictive value. The F1 score, or the Dice coefficient, is the harmonic mean of precision and recall. FNR is perhaps one of the most critical performance metrics for such a classification task since a false negative on the test has serious public health risks associated with it. Moreover, conventional testing methods of COVID-19 have high tendencies to return false negative results, hence minimizing this metric would prove the efficacy of this model over molecular and antigen tests. MCC and Kappa index are two metrics that are completely agnostic to the size of each class and hence work best for unbalanced datasets such as the one used in this research. The value of MCC lies between -1 and $+1$, with a higher score being more desirable. Cohen's Kappa index is a measure of the performance delta between the trained classifier and a random classifier that makes random predictions according to the frequency of occurrence of each class.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Table 2 Accuracy results of all models

Model	Accuracy (%)	Specificity	Recall	Precision	F1 score	FNR	Error	MCC	Kappa
AlexNet	87.50	0.928	0.821	0.918	0.866	0.178	0.125	0.753	0.750
VGG16	92.94	0.949	0.907	0.943	0.925	0.092	0.07	0.859	0.851
VGG19	93.75	0.933	0.941	0.930	0.935	0.058	0.062	0.874	0.875
ResNet18	92.13	0.928	0.914	0.925	0.919	0.085	0.078	0.842	0.842
ResNet34	96.12	0.973	0.951	0.976	0.963	0.048	0.038	0.923	0.923
ResNet50	97.37	0.983	0.983	0.964	0.964	0.016	0.026	0.947	0.947

$$F1 \text{ score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

$$\text{False Positive Rate (FNR)} = \frac{FN}{TP + FN} \quad (6)$$

$$\text{Error} = \frac{FP + FN}{TN + FN + TP + FP} \quad (7)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TN + FP)(TN + FN)(TP + FP)(TP + FN)}} \quad (8)$$

$$Kappa = \frac{\text{Accuracy} - \text{random Accuracy}}{1 - \text{random Accuracy}} \quad (9)$$

The highest classification accuracy was given by ResNet50. ResNet50 also provided the best MCC and Kappa index scores, useful for evaluating DCNN models on an unbalanced dataset. Further, a false negative rate of just 1.6% was achieved, which is a massive improvement over conventional tests such as molecular and antigen tests, which have a reported false negative rate of as high as 50%. The final results with all performance metrics have been depicted in Table 2.

4 Conclusion

Limited work has been done on the detection of COVID-19 through the CT scan image of the patient. In this research, the prospect of using DCNN models via transfer learning approach to detect COVID-19 was explored. After extensive testing and computing performance metrics, very promising results were obtained in this binary classification task. The proposed model was able to successfully identify the presence or absence of the disease with a maximum accuracy of 97.37%. Further, false negatives were at an exceptionally low 1.6%. These results could be further improved with a larger dataset for the DCNN models to train on. However, the proposed model

contributes to a possibility of a quick, cost effective and most importantly, accurate diagnosis of COVID-19 using chest CT scan images. Practical use of this model would provide great benefit to society and mankind.

References

1. World Health Organization. (2020). Coronavirus Disease (COVID-19), <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses#:~:text=symptoms>.
2. Sanche, S., Lin, Y., Xu, C., Romero-Severson, E., Hengartner, N., & Ke, R. (2020). High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerging Infectious Diseases*, 26(7), 1470–1477. <https://doi.org/10.3201/eid2607.200282>.
3. Wikipedia. (2020). COVID-19 pandemic data, https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data.
4. Interpreting a covid-19 test result, 2020. *BMJ*, 369, m1808. <https://doi.org/10.1136/bmj.m1808>.
5. NBC News. (2020). Frustration over long coronavirus test result wait times grows in Florida, <https://www.nbcnews.com/health/latino/frustration-over-long-coronavirus-test-result-wait-times-grows-florida-n1234493>.
6. Times of India. (2020). Delhi: Up to 10-day waiting time for Covid-19 test at private labs, <https://timesofindia.indiatimes.com/city/delhi/delhi-up-to-10-day-waiting-time-for-covid-19-test-at-private-labs/articleshow/76521078.cms#:~:text=June%2023%2C%202020-Delhi%3A%20Up%20to%2010%2Dday%20waiting%20time%20for%20Covid%2D,19%20test%20at%20private%20labs>.
7. Science. (2020). Coronavirus antigen tests: quick and cheap, but too often wrong? <https://www.sciencemag.org/news/2020/05/coronavirus-antigen-tests-quick-and-cheap-too-often-wrong>.
8. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395(10223), 514–523.
9. Sarker, L., Islam, M., Hannan, T., & Ahmed, Z. (2020). COVID-DenseNet: A deep learning architecture to detect COVID-19 from chest radiology images. preprint. <https://doi.org/10.20944/preprints202005.0151.v1>.
10. Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., Xue, Z., Shen, D., & Shi, Y. (2020). Lung infection quantification of COVID-19 in CT images with deep learning. arXiv Preprint [arXiv:2003.04655](https://arxiv.org/abs/2003.04655).
11. Wang, S., Zha, Y., Li, W., Wu, Q., Li, X., Niu, M., Wang, M., Qiu, X., Li, H., Yu, H., & Gong, W. (2020). A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *European Respiratory Journal*. <https://doi.org/10.1183/13993003.00775-2020>.
12. Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., Bernheim, A., & Siegel, E. (2020). Rapid AI development cycle for the Coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv Preprint [arXiv:2003.05037](https://arxiv.org/abs/2003.05037).
13. Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.
14. Howard, J., & Gugger, S. (2020). Fastai: A layered api for deep learning. *Information*, 11, 2.
15. Kaggle. (2020). SARS-COV-2 Ct-Scan Dataset, <https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset>.
16. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>. ISSN 0001-0782.

17. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In *Proceedings of International Conference on Learning Representations*, <http://arxiv.org/abs/1409.1556>.
18. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition* (pp. 770–778), <https://doi.org/10.1109/CVPR.2016.90>.

A Hybridized Machine Learning Model for Optimal Feature Selection and Attack Detection in Cloud SaaS Framework



Reddy Saisindhutheja and Gopal K. Shyam

Abstract Cloud computing offers several profitable services as and when required to the customer, and so it is growing up as a forthcoming drift in the IT sector. Software-as-a-service (SaaS) is one of the outstanding and fastest-growing fields in cloud computing history. It is a license to acquire cloud applications via the Internet. Out of all this curiosity, security is found as one of the key issues that delay the growth of SaaS. The motivation of this research work is to provide security for SaaS by handling massive amounts of data. At first, feature selection is accomplished by Oppositional Crow Search Algorithm (OCSA). These nominated features are sent for detecting the attacks via Deep Belief Network (DBN). The main objective is to introduce an innovative, secure framework for SaaS by performing attack detection when there is enormous traffic. The proposed and conventional models are evaluated using a benchmark dataset. Results prove that the proposed OCSA + DBN outperforms the other existing methods with respect to precision, sensitivity which are positive measures and False Positive Rate (FPR) and False Detection Rate (FDR); the negative measures. Moreover, the proposed work performs better with the existing works with a performance indicator of 3% for all the metrics.

Keywords Cloud computing · SaaS · Attack detection · Crow search algorithm · Deep belief network

1 Introduction

Over the years, digital appliances play a major role in human life [1]. The sudden increase in the world-wide internet usage needs a new strategy to manage the size, range, and availability of data [2, 3]. Cloud computing is an upcoming expertise that is substantial in delivering on-demand, scalable and consistent services to its customers

R. Saisindhutheja (✉)
Sreenidhi Institute of Science & Technology, Hyderabad 501301, Telangana, India

R. Saisindhutheja · G. K. Shyam
REVA University, Bengaluru 560064, Karnataka, India
e-mail: gopalkrishnashyam@reva.edu.in

at low cost [4]. Besides, it offers a simple strategy to use servers, databases, software, and analytics over the Internet and accomplishes all the application services [5–7].

The service delivery models of cloud are Infrastructure-as-a-service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). SaaS is the top most layer of cloud which is present on PaaS and IaaS. As the characteristics obtained, it has a chance of acquiring the threats as well. Hence, SaaS is highly vulnerable to attacks. Attackers may harm the systems with the most common attacks like data theft, DDoS attack, illegal access, etc. [8–11].

In contempt of this, machine learning algorithms are effectively used for attack detection. Out of these models, deep-learning resembles the most useful method in finding the attack detection. The key contributions of this work are as follows: (i) Selecting the best features using OCSA algorithm, (ii) Performing the attack detection via DBN, and (iii) Comparing the results of proposed work with existing methods. The SaaS providers using this framework will design and develop the detection phase, which fits best in finding various threats.

The remaining part of the paper is organized as follows: Sect. 2 illustrates the literature review. The proposed attack detection system is demonstrated in Sect. 3. The results acquired with the proposed model are discussed in Sect. 4. At the end, conclusions and future work are given in Sect. 5.

2 Literature Review

In [12], the authors have developed a new architecture for detecting DDoS attacks in the cloud. This architecture is accomplished by coupling the “sparse AE for feature extraction and DNN for classification”. Further, by using a relevant tuning model, the parameters are fine-tuned for AE and DNN. As a result, the overfitting problem is resolved with the least error.

A new security system is proposed in [13, 14] to handle the activated DDoS attack by unauthorized IoT servers. By considering SDN with cloud, the DDoS attack is mitigated on the IoT servers. Results attained with high accuracy in detecting DDoS attacks.

The DDoS attack detection and prevention system was introduced in [15] using ICRPU and feature selection. To examine the traffic, Hellinger distance function was used. Further, the packets were distinguished as valid groups and extracted features regarding DDoS.

A new method formulated on the study of network flow in the cloud-based services is introduced in [16] for FRC attack detection and prevention. Results of the proposed work are very effective with low overhead and high accuracy.

CRESOM-SDNMS was developed in [17] to counterattack the DDoS attacks in the cloud scenario. After initialization, the improvised topology preservation is used in the SOM-based categorization process to remove the vector quantization problems. The results show that the proposed work had minimized the FPR at DDoS prevention.

The challenges that identified from the literature review are poor attack detection rate, low classification accuracy, redundant, huge and data with high dimensionality which are not managed properly. The proposed work overcomes all these problems.

3 Proposed Attack Detection Framework

The proposed work aims to design and develop a secure framework for detecting the attacks which is given in Fig. 1. Following are the steps used to develop this framework.

Step 1: The input data $Input^{Data}$ is subjected to RSA [18] for carrying out encryption, and the encrypted data after evaluation $Encrypt^{Data}$ is decrypted by mapping $Encrypt^{Data}$ to 9 RSU's. The applications and RSU count will not differ. After decryption the resultant is $Decrypt^{Data}$.

Step 2: As $Decrypt^{Data}$ consists of more features, the training phase might be very complex. Hence, the features F_n are selected from $Decrypt^{Data}$. OCSA is used to accomplish feature selection.

Step 3: As soon as the optimal features F_n are chosen, they are given for DBN to detect the attacks, in which it predicts whether the feature selected is attack free or not.

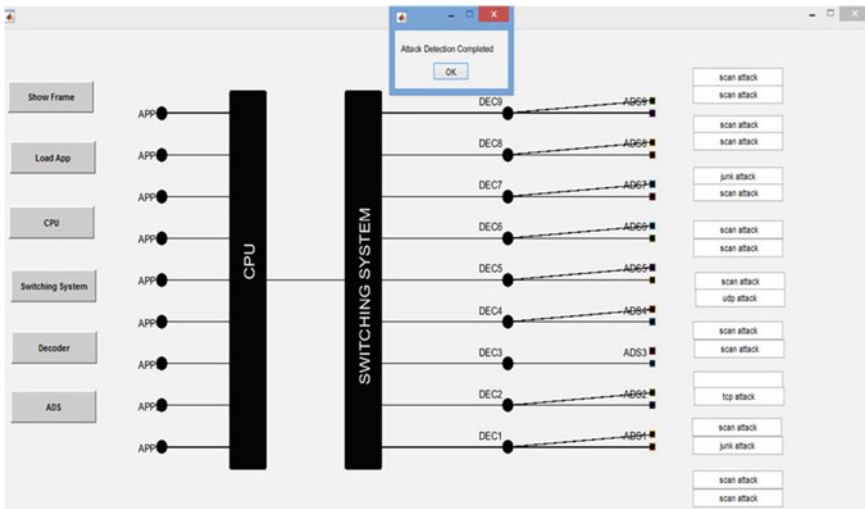


Fig. 1 The architecture of the proposed work

3.1 Feature Selection Using OCSA

The selection of the appropriate features of $Decrypt^{Data}$ improves the capability of the system potential to detect the attacks. In the presented work, the relevant attributes are selected by discarding the unnecessary and duplicate attributes. To complete this target OCSA [30] is used which is combination of Crow Search Algorithm (CSA) and Opposition Based Learning (OBL).

CSA is an advanced meta-heuristic method. It depends on the social aspects and intellectual approach of crows [19–21]. The advantage of CSA is motivated starting with the practice of crow redeeming good enough food in a mysterious location and restoring it at particular time. To safeguard its stuff, it conceals and misguides others away from its actual spot. The major asset of CSA it has only two parameters: flight length and awareness probability, to fix optimization problems.

OBL is a competent method to break the calculation in certain problems of optimization [22, 23]. This method is encouraged by the opposite association with the real-world objects. Without using the notion of counteraction, it would be tough to describe couple of entities like north–south, day–night, east–west, etc., in the existing world. Followings are the steps adapted in OCSA:

Step 1: Pope represents the total quantity of population of the crows (search agents), Flight length $Flight^{length}$, maximum quantity iteration (Ki) and $aware^{porb}$ as awareness probability are set. Moreover, crows (i.e., columns of the dataset) are initiated via Eq. (1).

$$Crow_i = \begin{bmatrix} Attribute\ 11 & Attribute\ 12 & \dots & Attribute\ 1D \\ Attribute\ 21 & Attribute\ 22 & \dots & Attribute\ 2D \\ \dots & \dots & \dots & \dots \\ Attribute\ n1 & Attribute\ n2 & \dots & Attribute\ nD \end{bmatrix} \tag{1}$$

Step 2: The location with respect to crows are initialized in random way. In this work, the total attributes used are 115 so $Population = 115$. These are assigned as 1 or 0 randomly. If the crow’s i th location is 0, then its respective field is not opted to perform classification, otherwise considered.

Step 3 (Generating opposite solution): The opposite position is evaluated using Eq. (2).

$$Attribute'_i = a_i + b_i - Attribute_i \tag{2}$$

For every solution $Crow_i$, there exists a unique opposite solution $OCrow_i$. The evaluation of the opposite location is $OCrow_i$ ($Attribute'$) is denoted using Eq. (3). Here, $i \in 1, 2, \dots, m$.

$$OCrow_i = \begin{bmatrix} Attribute\ 11' & Attribute\ 12' & \dots & Attribute\ 1D' \\ Attribute\ 21' & Attribute\ 22' & \dots & Attribute\ 2D' \\ \dots & \dots & \dots & \dots \\ Attribute\ n1' & Attribute\ n2' & \dots & Attribute\ nD' \end{bmatrix} \quad (3)$$

Step 4: The memory of crows is initialized and noted. Further, the count is given as $K_i = 1$.

Step 5: Do for $i = 1$ to n , a crow is chosen arbitrarily to track n th crow.

Step 6: R_m , the random number is produced and position is updated using Eq. (4) if $R_m > = aware^{prob(t)}$. The position update of thief crow m that follows the owner crow n .

$$S_m^{t+1} = S_m^t + R_m \times F_m^{length(t)} \times (B_n^t - S_m^t) \quad (4)$$

Here, B_n^t represents the food source of owner crow m and R_m is a random number in the interval $[0, 1]$. At the iteration t , the flight length x is denoted as $F_m^{length(t)}$. In case, if $R_n < aware^{prob(t)}$, then revise the random position.

Step 7: By assessing the updated location of crows their memories are updated. Further results are seen and best position is opted. Out of 115 attributes, OCSA selects 36 features which are further given to DBN for detecting the attacks.

3.2 DBN Based Attack Detection

DBN is designed and developed by smolensky in 1986 with multiple layers, in which each layer constitutes visible and hidden neurons which are completely interrelated. The neuron's result is based on probability in Boltzmann machine which is denoted as *outcome*. The *outcome* in Eq. (5) is given with a probability of $prob(\tau)$ in Eq. (6). τ denotes pseudo-temperature of the parameter and used for governing the noise levels.

$$out = \begin{cases} 0 & \text{with probability of } 1 - prob(\tau) \\ 1 & \text{with probability of } prob(\tau) \end{cases} \quad (5)$$

$$prob(\tau) = \frac{1}{1 + e^{-\frac{\tau}{T}}} \quad (6)$$

The energy state which is configured is denoted as *State*, E^{Fun} which is the energy function of the Boltzmann machine is given in Eq. (7). $W_{A,B}$ is the weight function between the neurons A and B , and φ_A is their biases.

$$E^{Fun}(state) = - \sum_{A < B} State_A State_B W_{A,B} - \sum_A \varphi_A State_A \quad (7)$$

The energy differential function is evaluated with ΔE^{Fun} . In the Restricted Boltzmann Machine (RBM), for visible and hidden neurons VIS , HID , the updated energy function attained using Eqs. (8)–(10).

$$E^{Fun}(\vec{VIS}, \vec{HID}) = \sum_{A,B} wt_{A,B} VIS_A HID_B - \sum_A VIS_A P_A - \sum_B HID_B Q_B \quad (8)$$

$$E^{Fun}(VIS_A, \vec{HID}) = \sum_B wt_{A,B} HID_B + P_A \quad (9)$$

$$E^{Fun}(\vec{VIS}, HID_B) = \sum_A wt_{A,B} VIS_A + Q_B \quad (10)$$

Unsupervised training is performed in RBM. Training set is represented as $Train$, where the rows of matrix denote \vec{VIS} . The highest probability is obtained with assigning the weights with $Weight_m$ which is given in Eq. (11).

$$Weight_m = \max_{Weight} \prod_{\vec{VIS} \in Train} prob(\vec{VIS}) \quad (11)$$

Using Eq. (12), probability is given to every (VIS, HID) pair. Z is partition function which is mathematically given in Eq. (13).

$$prob(\vec{VIS}, \vec{HID}) = \frac{1}{z} e^{-E^{Fun}(\vec{VIS}, \vec{HID})} \quad (12)$$

$$Z = \sum_{\vec{VIS}, \vec{HID}} e^{-E^{Fun}(\vec{VIS}, \vec{HID})} \quad (13)$$

The error is calculated using Root Mean Square Error (RMSE) as per Eq. (14), and it is the variation of predicted and actual values. Objective function regarding fitness is evaluated in Eq. (15).

$$RMSE = Predicted - Actual \quad (14)$$

$$Obj = Min(RMES) \quad (15)$$

The Contrastive Divergence (CD) is used to initialize visible states systematically. The updated weights are given in Eq. (16).

$$wt_{VIS,HID} = \Delta wt_{VIS,HID} + wt_{VIS,HID}. \quad (16)$$

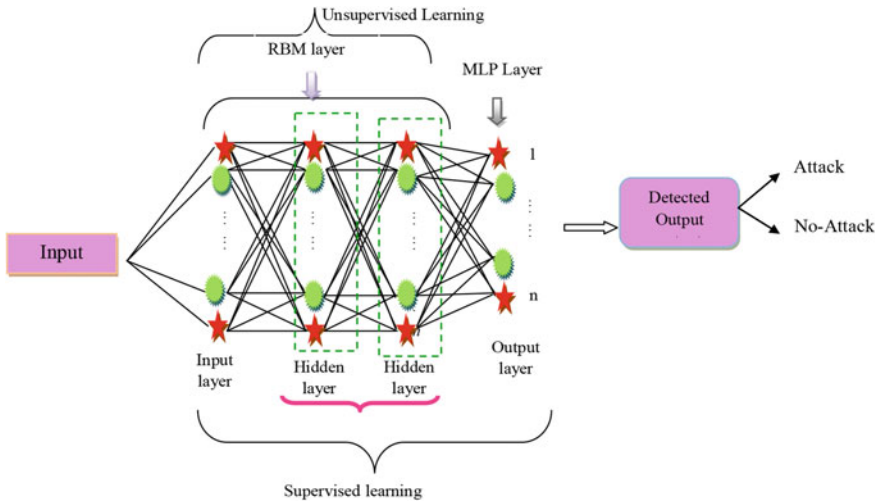


Fig. 2 Architecture of DBN based attack detection

The existence of the attack in the network is determined by DBN. The architecture of attack detection via DBN is given in Fig. 2.

4 Experimental Results

4.1 Execution Setup

The proposed framework was implemented in the MATLAB 2019. The dataset is downloaded from [24]. The performance analysis regarding the optimal feature selection is evaluated for the proposed and other existing works like PSO [25], GA [26], WOA [27], FF [28] and ROA [29]. Additionally, the performance of the classifiers is evaluated using NN [30], SVM [31], and CNN [32]. Evaluation is performed with respect to specificity, precision; the positive measures and FDR, FPR; the negative measures.

4.2 Performance Metrics

Following are the performance metrics used for evaluating the proposed work.

Precision: The proportion of standard data and the total quantity of standard and attack data. Calculated as $TP / (TP + FP)$.

Specificity: The ratio of original positive values that are identified as positive. Calculated as $TN/(TN + FP)$.

False Positive Rate (FPR): Possibility of incorrectly ignoring the null hypothesis. Calculated as $FP/(FP + TN)$.

False Detection Rate (FDR): Actual ratio of type I errors. Calculated as $FP/(TP + FP)$.

4.3 Performance Analysis with Respect to Feature Selection

Analysis using positive measures: The obtained results using positive measures are given in Fig. 3a, b. The accuracy is maximum for OCSA at all the learning percentages. It is observed that the precision and specificity metric recorded an outstanding performance with 98% when compared with other traditional models.

Analysis using Negative measures: The outcomes of FPR and FDR with OCSA and other conventional works are given in Fig. 4c, d.

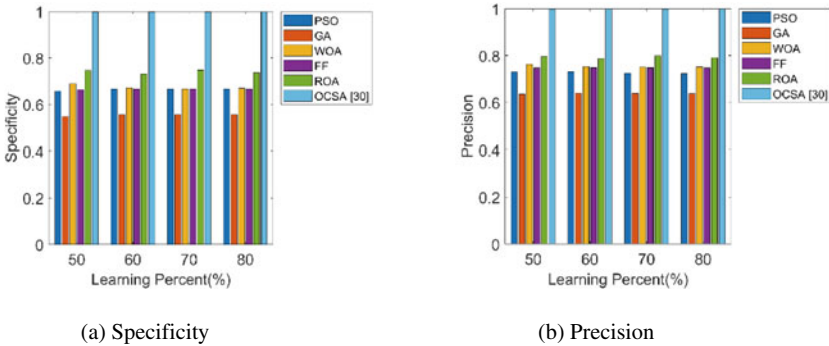


Fig. 3 Evaluation of the proposed and conventional algorithms with positive metrics

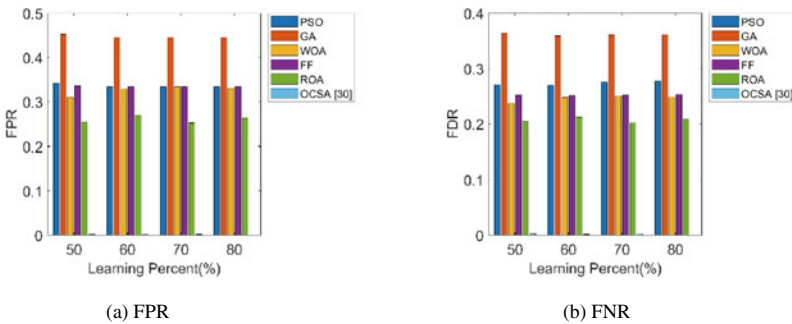


Fig. 4 Evaluation of the proposed and conventional algorithms with negative metrics

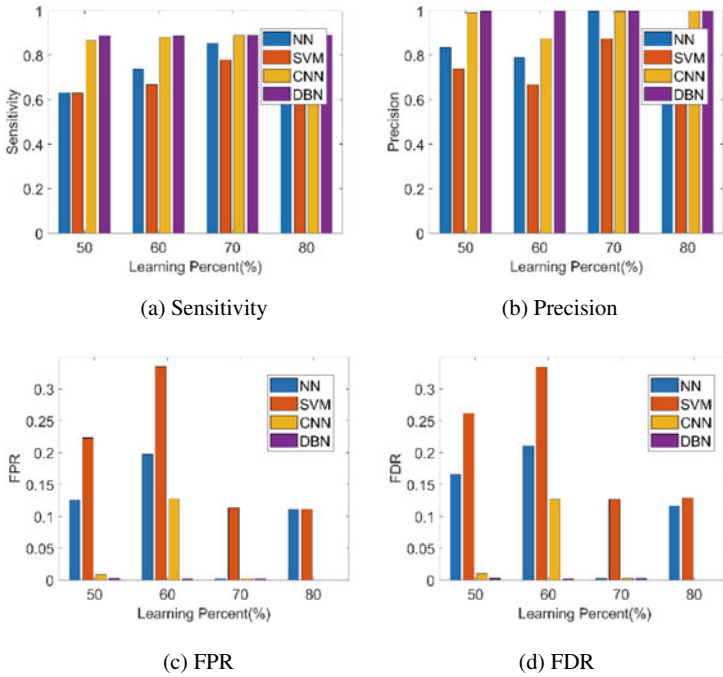


Fig. 5 Evaluation of the proposed and Traditional classifiers

4.4 Performance Analysis of the Classifier

The performance of the proposed work with DBN is computed with NN, SVM and CNN. The results obtained using the positive measures is given in Fig. 5a, b. On the other hand, evaluations of FPR and FDR are given in Fig. 5c, d. Results prove that the OCSA scores the low values for erroneous measures.

5 Conclusions and Future Work

This work intends to develop a secure framework for SaaS. As a benchmark, the N_BaIoT dataset is utilized. This dataset contains nine different applications, and each application comprises 115 features. It is subjected to OCSA to perform optimal feature selection. OCSA selects 36 best features from the dataset. After feature selection, the selected features were sent to DBN for detecting the attacks. The proposed OCSA and DBN based attack detection system outperformed with precision, sensitivity, FPR, and FDR over other conventional models and classifiers.

We aim to focus our future work on executing the entire work in the cloud environment. Also, perform cross validation, which states how accurate the proposed model

is. Further, aiming at real-time identification of threats and explore the various attacks in the multi-cloud environment.

References

1. Ashok Kumar, C., & Vimala, R. (2020). Load balancing in cloud environment exploiting hybridization of chicken swarm and enhanced raven roosting optimization algorithm. *Multimedia Research*, 3(1), 45–55.
2. Alkadi, O., Moustafa, N., & Turnbull, B. (2020). A review of intrusion detection and block chain applications in the cloud: Approaches, challenges and solutions. *IEEE Access*, 8, 104893–104917.
3. Dong, S., Abbas, K., & Jain, R. (2019). A survey on Distributed Denial of Service (DDoS) attacks in SDN and cloud computing environments. *IEEE Access*, 7, 80813–80828.
4. Vhatkar Kapil Netaji, B. G. P. (2020). Optimal container resource allocation using hybrid SA-MFO algorithm in cloud architecture. *Multimedia Research*, 3(1), 11–20.
5. Wang, W., Du, X., Shan, D., Qin, R., & Wang, N. (2020). Cloud intrusion detection method based on stacked contractive auto-encoder and support vector machine. *IEEE Transactions on Cloud Computing*. <https://doi.org/10.1109/TCC.2020.3001017>
6. Ravi, N., & Shalinie, S. M. (2020). Learning-driven detection and mitigation of DDoS attack in IoT via SDN cloud architecture. *Internet of Things Journal*, 7(4), 3559–3570.
7. Tian, Z., Luo, C., Qiu, J., Du, X., & Guizani, M. (2020). A distributed deep learning system for web attack detection on edge devices. *IEEE Transactions on Industrial Informatics*, 16(3), 1963–1971.
8. Mishra, P., Varadharajan, V., Pilli, E. S., & Tupakula, U. (2020). VMGuard: A vmi-based security architecture for intrusion detection in cloud environment. *IEEE Transactions on Cloud Computing*, 8(3), 957–971.
9. Virupakshar, K. B., Manjunath Asundi, D. G., & Narayan. (2020). Distributed Denial of Service (DDoS) attacks detection system for open stack-based private cloud. *Proceedings of the Procedia Computer Science, Elsevier*, 167, 2297–2307.
10. Garre, J. T. M., Perez, M. G., & Ruiz-Martinez, A. (2020). A novel Machine Learning-based approach for the detection of SSH botnet infection. *Future Generation Computer Systems*, 115, 387–396.
11. Ninu Preetha, N. S., Brammya, G., Ramya, R., Praveena, S., Binu, D., & Rajakumar, B. R. (2018). Grey wolf optimisation based feature selection and classification for facial emotion recognition. *IET Biometrics*, 7(5), 490–499.
12. Bhardwaj, A., Mangat, V., & Vig, R. (2020). Hyperband tuned deep neural network with well posed stacked sparse autoencoder for detection of DDoS attacks in cloud. *IEEE Access*, 8, 181916–181929.
13. Ravi, N., & Shalinie, S. M. (2020). Learning-driven detection and mitigation of DDoS attack in iot via SDN cloud architecture. *IEEE Internet of Things Journal*, 7(4), 3559–3570.
14. Pillutla, H., & Arjunan, A. (2019). Fuzzy self organizing maps-based DDoS mitigation mechanism for software defined networking in cloud computing. *Journal of Ambient Intelligence and Humanized Computing*, 10, 1547–1559.
15. Bharot, N., Verma, P., Sharma, S., & Suraparaju, V. (2018). Distributed denial-of-service attack detection and mitigation using feature selection and intensive care request processing unit. *Arabian Journal for Science and Engineering*, 43, 959–967.
16. Bhushan, K., & Gupta, B. B. (2019). Network flow analysis for detection and mitigation of Fraudulent Resource Consumption (FRC) attacks in multimedia cloud computing. *Multimedia Tools and Applications*, 78, 4267–4298.
17. Harikrishna, P., & Amuthan, A. (2020) SDN-based DDoS attack mitigation scheme using convolution recursively enhanced self organizing maps 45(104). <https://doi.org/10.1007/s12046-020-01353-x>.

18. Zhang, H., et al. (2020). Efficient and secure outsourcing scheme for RSA decryption in internet of things. *IEEE Internet of Things Journal*, 7(8), 6868–6881.
19. Zamani, H., Nadimi-Shahraki, M. H., & Gandomi, A. H. (2019). *CCSA: Conscious neighborhood based crow search algorithm for solving global optimization problems*. *Applied Soft Computing*. Elsevier.
20. Rizk-Allah, R. M., Hassanien, A. E., & Bhattacharyya, S. (2018). Chaotic crow search algorithm for fractional optimization problems. *Applied Soft Computing*, 71, 1161–1175. Elsevier.
21. Sayed, G. I., Hassanien, A. E., & Azar, A. T. (2019). Feature selection via a novel chaotic crow search algorithm. *Neural Computing and Applications*, 31, 171–188. Springer.
22. Mahdavia, S., Rahnamayana, S., & Debb, K. (2018). Opposition based learning: A literature review. *Swarm and Evolutionary Computation*, 39, 1–23. Elsevier.
23. Ventresca, M., Rahnamayan, S., & Tizhoosh, H. R. (2010). A note on opposition versus randomness in soft computing techniques. *Applied Soft Computing, Elsevier*, 10(3), 956–957.
24. UCI Machine learning Repository, https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT.
25. Tanweer, M. R., Suresh, S., & Sundararajan, N. (2015). Self regulating particle swarm optimization algorithm. *Information Sciences*, 294, 182–202.
26. McCall, J. (2005). Genetic algorithms for modelling and optimization. *Journal of Computational and Applied Mathematics*, 184(1), 205–222.
27. Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, 95, 51–67.
28. Iztok Fister, I., Xin-SheYang & JanezBrest. (2013). A comprehensive review of firefly algorithms. *Swarm and Evolutionary Computation*, 13, 34–46.
29. Sai Sindhu Theja, R., & Shyam, G. K. (2020). An efficient metaheuristic algorithm based feature selection and recurrent neural network for DoS attack detection in cloud computing environment. *Applied Soft Computing Journal*, 100, 1–11. Elsevier.
30. Binu & Kariyappa. (2018). RideNN: A new rider optimization algorithm-based neural network for fault diagnosis in analog circuits. *IEEE Transactions on Instrumentation and Measurement*, 1–25. IEEE.
31. Gangappa, M., Kiran Mai, C., & Sammulal, P. (2019). Enhanced crow search optimization algorithm and hybrid NN-CNN classifiers for classification of land cover images. *Multimedia Research*, 2(3), 12–22.
32. Thomas, R., & Rangachar, M. J. S. (2019). Fractional rider and multi-kernel-based spherical SVM for low resolution face recognition. *Multimedia Research*, 2(2), 35–43.

English Master AMMU: Advanced Spoken English Chatbot



A. N. Gayathri and V. Viji Rajendran

Abstract Everyone needs to express their ideas, thoughts, and emotions. For a professional or a student, they have to express or communicate their thoughts effectively. To make communication effective, it is important to understand the English language. As English is an international language, all had to have fluent English-speaking skills. Making improvements in spoken English can be done in various ways. Talking to friends or family members in English, practicing phrases and sentences, etc. are some of them. But most people feel some awkwardness talking in front of a crowd or even with friends. Here comes the helper Assistant, Miss AMMU Teacher. Miss AMMU Teacher is the advanced version of AMMU spoken English chatbot. In this work, a chatbot called AMMU, Automatic Mega-agent Managed User guide is built for helping this purpose. AMMU will communicate with the user and will help to improve the spoken English capability of users.

Keywords Chatbot · Dialogflow · Integromat · Google assistant · Bot

1 Introduction

Communication is a very basic need of every individual. A healthy and effective communication makes a relationship or a professional perform better. English may not be the most widely used language in the world, but it is the official language of the world's 53 countries, and nearly 400 million people speak it. If you wish to speak to someone from another country, there is a good chance that you both speak English. As a second language, so many more people are now spending time to learn English. Several nations incorporate English in their school syllabus and children acquire English knowledge at their younger age.

The pandemic outbreak of COVID-19 (coronavirus) [1] demands us to stay home for safety reasons. Children are the most affected by these scenarios. All schools have online classes, and that is the new normal way of teaching. Most of the students face many issues while attending these kinds of classes. And for the professionals also,

A. N. Gayathri (✉) · V. Viji Rajendran
NSS College of Engineering, Palakkad, India

they have to stay home and do their work and acquire their knowledge through online platforms. As the situation changes, technologies have to change. Here professionals and children both are referred to as students as they both need knowledge. Practicing English with friends and family is a good idea. But most people feel awkward while speaking like this and are afraid of making mistakes. Spoken English courses and classes are very costly and are not affordable for ordinary people.

Here comes the importance of Chatbots. A chatbot is an assistant that can communicate with the human through voice or text, interpret the human queries and make a feeling to the user as they are interacting with a real human. Chatbots, in the near future, will witness a high boom in the field of education [2] because the current COVID situation made all students study through multimedia and telecommunication media. Chatbots for the education system can be developed in various ways. Before going on to the development of English Master AMMU, the following section discusses the most famous and developer-friendly chatbot development frameworks from renowned companies that help to make efficient chatbots. A variety of chatbot development platforms are available for developers. After experimenting with many of them, Google Dialogflow seems to be developer-friendly and is selected for building English Master AMMU. The experimented chatbot development platforms are discussed in the next section.

2 Chatbot Development Framework

There are a variety of chatbot building platforms that help the developer to easily start building chatbots from scratch. Some of them helped the developer to create a chatbot completely without any coding. Also, there are so many chatbot building platforms that help in building chatbots with some code and with some inbuilt features. The chatbot building frameworks from renowned companies are IBM Watson Assistant from IBM, Azure Cognitive services from Microsoft Azure, and Google Dialogflow from Google. The following subsections discuss each framework separately.

2.1 IBM Watson Assistant

IBM Watson Assistant is offered by IBM as a part of the IBM Watson suite. Watson assistant is mainly an AI-powered engine that is capable of providing several NLP services. Watson Assistant incorporates with several cognitive services provided by the IBM Watson platform to enhance the conversation through spell checks, a tone analyzer, and other services. The chatbot that was built using Watson assistant has the ability to know when to search for an answer from a knowledge base when to ask for clarity, etc. [3].

2.2 *Azure Cognitive Services*

Cognitive services are actually a group of REST APIs (Application Programming Interfaces) [4]. Azure cognitive services, a part of Azure and Microsoft public cloud, guarantees the most secure, highly available, and smooth platform for chatbot building. Using Azure Cognitive Services will help the developer force machine learning capabilities in the developer's application, with Microsoft expertise in clearing and training the data, with no load of developing complicated machine learning algorithms. Developers can use the Azure cognitive service QnA maker for making simple Frequently Asked Question (FAQ) bots or question-answer bot. Also, Microsoft Azure provides the web App Bot service, which is a union of web app and bot service [5].

QnA maker is actually an important service provided by Microsoft Azure. The purpose of the QnA maker is to help to enhance the user service with question and answer capability that can be extracted or that can be gathered from semi-structured or structured contents like FAQ, product manuals, etc., that all are available in Uniform Resource Locator (URLs), documents and so on [6].

In Microsoft Azure, it is possible to create a bot into C# or Node.js. The Bot framework provides the core component for creating bots [7]. The main components of it are the Bot connector, Bot framework SDK, and the Bot directory. The main functions of a bot connector are routing the messages, managing the states of the bot, bot registration management, directory tracking, etc. Bot Framework Software Development Kit (SDK) V4 is an open-source SDK that enables the developers to model and builds sophisticated conversations using their favorite programming language [8]. The bot directory is where the user will be able to find bots.

2.3 *Google Dialogflow*

Google Dialogflow is a tool that helps to build smart conversational agents offered by Google. Dialogflow helps the developer to design conversational interfaces and integrate them with other applications. Dialogflow incorporates Google's different AI experiences like learning expertise, search capabilities, speech recognition, and of course the NLU [9]. The developer can create free agents (bots) and make it efficient by creating several intents (use case verb or action). If the bot wants to give dynamic responses, fulfillment features of Dialogflow are used [10].

3 System Design

3.1 System Architecture

The proposed system aims to build a chatbot that helps the user to improve their spoken English skills through simple activities, learning lessons, practice sections and also correcting the user’s grammar. Figure 1 shows the diagram for the proposed system.

The proposed chatbot is implemented to be deployed in Google assistant [11]. That is, the user interaction medium for this chatbot system is Google assistant. A user-friendly Google assistant platform is chosen for this work due to the fact that it can be easily accessible for the user, without making any account or linking any Google accounts. The user interacts with the chatbot through Google Assistant, either by means of text messages or by means of voice messages. Then, the Google assistant application sends this message to the Dialogflow, which is the main part of the Chatbot. It is possible to make multiple agents (bots) in for the same chatbot. The agent will categorize the user utterances or input and make a match to the relevant intent. After that, corresponding responses will be sent back to the user. This agent works well if the user only needs static responses from the agent. If the building agent needs more dynamic responses, the Fulfillment feature of Dialogflow can be used. That is, in some cases, it is needed to extract information or data from external databases or services, and the fulfillment facility of Dialogflow is used. In this work, a tool called Integromat [12] is used for the realization of fulfillment in Dialogflow.

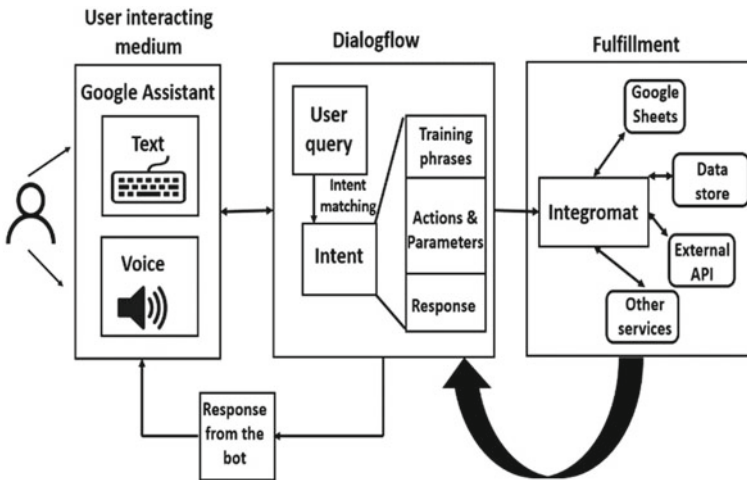


Fig. 1 Proposed system architecture

3.2 Implementation

Initial Version AMMU: English Master AMMU is the advanced version of the Spoken English chatbot AMMU (Automatic Mega-agent Managed User guide). The initial version is a Google Assistant-based chatbot that makes some conversation with the user like taking permission from the user to ask their details, making them comfortable, and will make a friendly relation with the user. After that, it will give small practice activity to the user which helps them to gain confidence with speaking English.

English Master AMMU design: English Master AMMU is designed to be a Master, that is a friendly master, who makes friendly chats with the user and also gives some activities, practice sections, and study materials. AMMU will collect user data for understanding user improvements and for further analysis. The tool Integromat is used for the external API connection, data store integration, and integration of Google sheets for collecting user data.

English Master AMMU has two agents and one mega agent. Using the feature Mega agent, multiple Dialogflow agents, called sub-agents, can be combined into a single agent. The first agent contains friendly conversations with the user. It includes the agent asking the user for their details like name, place, hobbies, etc. These details are collected to the Google sheets for further analysis. Apart from these, this agent also includes some practice activities and learning lessons. The first agent has more than 20 intents for giving the above-mentioned features. The second agent’s name is Jokes. It is intended for saying jokes when the user asks. It makes relationships with the user easier.

Development: As mentioned above, the majority of the development of English Master AMMU is done using Google Dialogflow. The main steps for the development of AMMU are as follows:

- Create an account in Dialogflow and create the first agent, here named it as EnglishBot.
- 20 intents are built for this agent.

Fig. 2 Different intents in EnglishBot agent

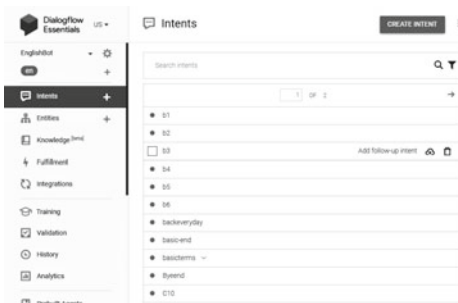




Fig. 3 Google sheet integration in Integromat

- For each intent, give appropriate training phrases and responses. Coding for voice, image, audio, and video are all done in JavaScript or JSON language. Figure 2 gives the intents that are used in EnglishBot agents.
- During the creation of each intent, necessary entities are created. The System entities like @sys.given-name and @sys.location are used for extracting user name and place. Further, necessary entities are created.
- The second agent Jokes is created.
- A mega agent named MegaEnglishBot is created, and the first and second agents are added to it.
- Integromat account is created for fulfillment implementation.
- A scenario is created in Integromat.
- Google sheet is integrated into the Dialogflow chatbot through the webhook in Integromat. Figure 3 shows integration.
- While setting up the Integromat, Integromat will produce a webhook URL. Copy and paste that URL to the Dialogflow webhook fulfillment for all the agents and mega agents.
- Also, the fulfillment should be enabled for the appropriate intents whose data are to be stored. Here, the name, place, and activity intent's fulfillment are enabled.

Testing of English Master AMMU: The last but frequently needed step for building a chatbot is testing. Chatbot AMMU is frequently tested using the “try it now” panel in the Dialogflow interface. After initial testing is done, the created agents are combined with the mega agent. Then the mega agent is integrated into the Google assistant. This will lead to the Action on the Google console page. There is a testing tab, where the fully developed chatbot can be tested. AMMU is perfectly working without errors. At last, the current version is deployed on to the Google assistant. Version 6, now available for users, has been approved by Google.

4 Experimental Results

This section discusses the experimental results that have been obtained after the testing and integration of English Master AMMU with the Google Assistant. The comparison of the current application and the English master AMMU is also discussed here.

English master AMMU takes five versions to get improved and get approved by Google. Version 6 of AMMU is a teaching assistant that initially makes a friendly conversation with the user and after that gives users some teaching or learning lessons and practice activities. Figure 4 shows the welcome UI of English Master AMMU. The current version is perfect and works with dynamic and static user utterances. Figure 5 shows the conversation of AMMU with the user.

After the initial friendly conversation and the first simple activity, the current phase 1 version of AMMU gives the user a provision for studying the basics of English grammar. Then, AMMU asks the user to continue with studying the basics of grammar and practice the activity or not. If the user selects “yes”, they will be provided four options for study purposes. If the user selects “No” for the previous question, AMMU will end the conversation. In future versions, AMMU will give some more activities like vocabulary practice and quizzes when the user is not interested in further studying. The study options that AMMU provided are as follows:

- Basic terms in English
- Everyday sentences
- English grammar tenses
- Some helpful phrases.

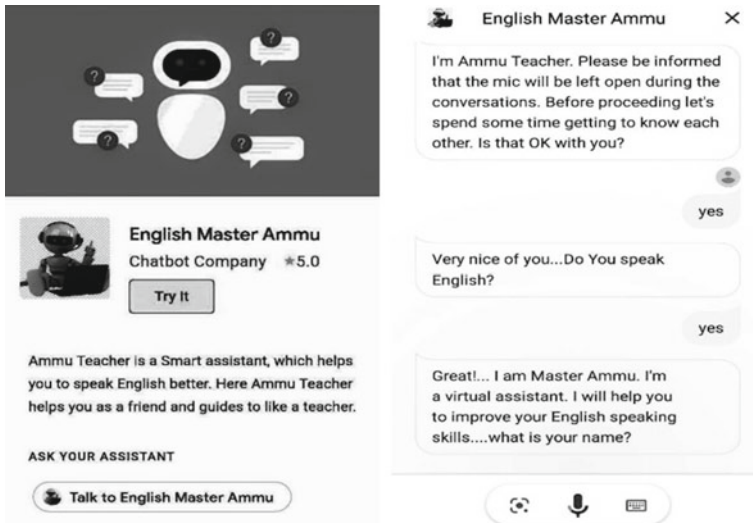


Fig. 4 Welcome page of English master AMMU

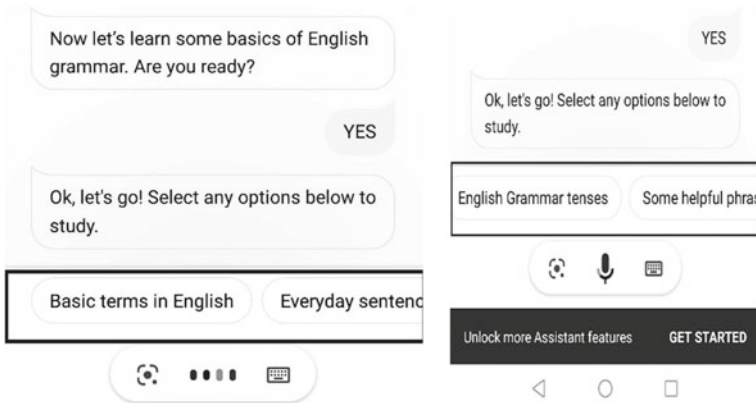


Fig. 5 Working demonstration of latest chatbot release

Figure 5 shows the above-mentioned options that are provided to users. Upon selecting each option, the user is taken to the next page, where each page contains the detailed material about the given main options. Each option has sub-options like the basic terms that given here are nouns, pronouns, verbs, adverbs, adjectives, and prepositions. Additional sub-options are there for deep knowledge.

There are some similar chatbots available in Google Assistant, named English Teacher and Learn English club, both being used to improve English. The comparison between them is shown in Table 1. Learn English Club [13] and English Teacher [14] are the two chatbots that have the same functionality as AMMU. Basically, AMMU, Learn English Club, and English Teacher are integrated with Google Assistant. AMMU will support both text-based and voice-based communication. As Google Assistant chatbot, English Teacher and the Learn English Club also support text and voice-based communication. These three are also available as an open-source facility.

AMMU initially makes a friendly conversation with the user. This creates a pleasant atmosphere for the user. Learn English Club is just giving some knowledge about the basic terms in English. No more features are there. In the case of the English Teacher, there is no friendly conversation between the user and agent. The agent gives some learning activities from the start itself. The learning activities are just one or two practice activities, but on the other hand, AMMU gives the user a very pleasant relationship and then makes them more comfortable to study. There are both study materials and practice activities in English Master AMMU.

Table 1 Comparison on different chatbots

AMMU	Learn english club	English teacher
Google assistant chatbot	Google Assistant chatbot	Google assistant chatbot
Both text and voice-based	Both text and Voice-based	Both text and Voice-based
Open-source	Open-source	Open-source
Makes a friendly conversation to make the user comfortable before start learning	Give basic terms of English grammar. Nothing more	No friendly conversation with the user. Gives some learning activities from the start itself
Focused on correcting user’s grammar and vocabularies	No more features to expose	Not trying to correct grammar

5 Conclusion

Chatbots are an intelligent way of communicating with the user. They can make an efficient relationship with the user by interacting like a human. In this paper, an advanced version of Spoken English Chatbot AMMU is discussed. This is capable of making users friendlier with the usage of the language English and makes them more confident in speaking English. Apart from this, practice sections are very helpful for the user for improving their speaking ability, and also learning sections make their grammar parts very strong. Currently, the latest version is available for users in Google assistant in the name “English Master AMMU”.

In the next version of AMMU, a data store will be added for data analyzing and Integromat will be utilized for giving the users their improvements on their English learning. The external API connection is also made possible through the help of Integromat and make use of Dialogflow NLU (Natural Language Understanding) for adding grammar correcting feature to the AMMU.

References

1. Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics*, 87(4), 281–286.
2. Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, 151, 103862.
3. Sabharwal, N., Barua, S., Anand, N., & Aggarwal, P. (2020). Building your first bot using Watson assistant. In *Developing cognitive bots using the IBM Watson engine* (pp. 47–102). Berkeley, CA: Apress.
4. Ed-Douibi, H., Daniel, G., & Cabot, J. (2020, June). OpenAPI Bot: A chatbot to help you understand REST APIs. In *International Conference on Web Engineering* (pp. 538–542). Cham: Springer.
5. Shaikh, K. (2019). Eagle-eye view of azure cognitive services. In *Developing bots with QnA maker service* (pp. 1–29). Berkeley, CA: Apress.
6. Shaikh, K. (2019). The what, when, why, and how of the QnA maker service. In *Developing bots with qna maker service* (pp. 31–59). Berkeley, CA: Apress.

7. Etaati, L. (2019). Bot framework. In *Machine learning with microsoft technologies* (pp. 335–353). Berkeley, CA: Apress.
8. Sabharwal, N., Barua, S., Anand, N., & Aggarwal, P. (2020). Bot frameworks. In *Developing cognitive bots using the IBM Watson engine* (pp. 39–46). Berkeley, CA: Apress.
9. Sabharwal, N., & Agrawal, A. (2020). Introduction to Google dialogflow. In *Cognitive virtual assistants using google dialogflow* (pp. 13–54). Berkeley, CA: Apress.
10. Sabharwal, N., & Agrawal, A. (2020). Advanced concepts in google dialogflow. In *Cognitive virtual assistants using google dialogflow* (pp. 55–117). Berkeley, CA: Apress.
11. Tulshan, A. S., & Dhage, S. N. (2018). Survey on virtual assistant: Google assistant, siri, cortana, alexa. In *International Symposium on Signal Processing and Intelligent Recognition Systems* (pp. 190–201). Singapore: Springer.
12. Hughes, K., Lecky-Thompson, J., Ammon, M., & Murphy, H. (2017). Track the impact of your publications.
13. English Learn Club, https://assistant.google.com/services/a/uid/0000002a507e2159?hl=en_in&source=web.
14. English Teacher, https://assistant.google.com/services/a/uid/00000082469a1382?hl=en_in&source=web.

Investigation of CNN-Based Acoustic Modeling for Continuous Hindi Speech Recognition



Tripti Choudhary, Atul Bansal, and Vishal Goyal

Abstract Recently, Convolutional Neural Network (CNN) gains more popularity over hybrid Deep Neural Network (DNN) and Hidden Markov Model (HMM) based acoustic models. CNN has the ability to deal with speech signals and it makes appropriate choice for the Automatic Speech Recognition (ASR) system. The sparse connectivity, weight sharing, and pooling allow CNN to handle a slight position shift in the frequency domain. This property helps to manage speaker and environment variations. CNN works well for speech recognition, but it was not appropriately examined for the Hindi speech recognition system. The activation functions and optimization techniques play a vital role in CNN to achieve high accuracy. In this work, we investigate the impact of various activation functions and optimization techniques in the Hindi-ASR system. All the experiments were performed on the Hindi speech dataset developed by TIFR, with the help of the Kaldi and Pytorch-Kaldi toolkit. The experiment results show that the ELU activation function with Rmsprop optimization techniques gives the best Word Error Rate (WER) 14.56%.

Keywords Automatic speech recognition · GMM · FBANK · CNN

1 Introduction

ASR is the process of converting human speech into text sequences. The speech recognition field has dramatically benefited from DNN as it estimates the millions of parameters without overfitting [1]. The high-performance gain with less need for engineered features (i.e., MFCC, PLP) makes it the foremost favorite choice for Large Vocabulary Continuous Speech Recognition (LVCSR) system. DNN based acoustic models replace the statistical models (HMM, GMM) in the field of ASR [2]. DNNs are good enough to perform class-based discrimination [3]; on the other hand, it suffers from some serious drawbacks. First, due to the feed-forward fully connected nature, it cannot capture the structural locality of feature space [4], Second, DNN's cannot persist the past information. Finally, DNN has suffered from a vanishing

T. Choudhary (✉) · A. Bansal · V. Goyal
GLA University, Mathura, India

gradient problem [5]. Except above, the fully connected nature of DNN introduces the millions of learnable parameters that increase the computational load of the ASR system. Generally, neural networks need a huge amount of data to train a large number of parameters. Recently, more emphasis has been given to those models that require fewer parameters. These models have become very popular in low resource languages and show high performance. In this work, we have only a few hours of data, and CNN requires fewer parameters compared to other models. In this work, we choose CNN based acoustic model to test the Hindi ASR performance.

CNN [6] is a modification of neural network (NN). CNN can efficiently seize the structural locality from the feature space [7]. CNN has the capability to capture the position invariant features. CNN is not good enough to handle variable-length inputs. Basically, CNN consist of three layers, these are pooling layer, convolutional layer and Fully Connected (FC) layers. In convolutional layers the filter is glide across the width and height of the input and dot product of input and filter are count at every spatial position. The pooling layer is responsible for reducing the dimensionality of the feature map without losing any useful information. This layer minimizes the computational complexity of the system and control the matter of overfitting. The last layers of CNN are FC layer in this input in the form of a vector. In FC the output of previous layer is compress and form a single vector.

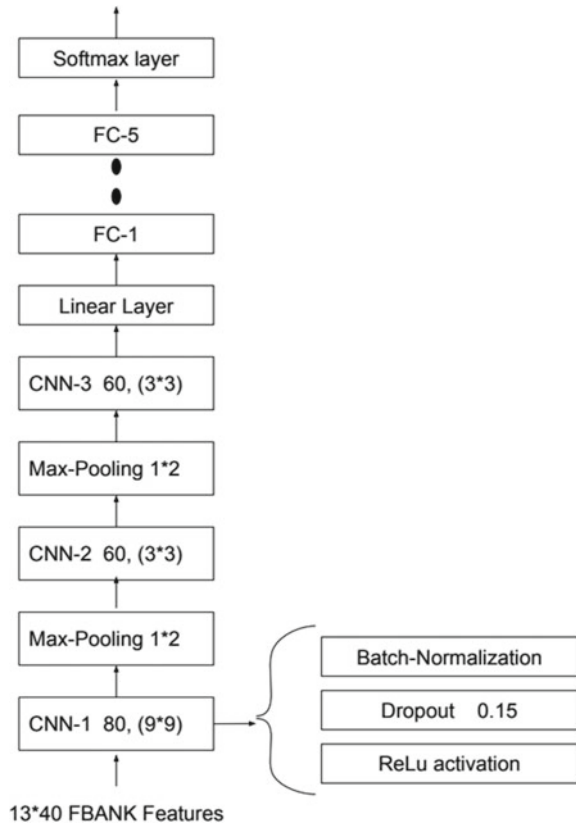
The overall performance of any CNN based acoustic model highly depends on the type of activation function and optimization techniques. The optimization problem becomes more challenging as the number of parameters increases in limited resource conditions. We have tested three different optimization techniques to investigate their role in low or limited resource conditions in this work. This work aims to study the impact of activation function and optimization techniques on limited resource Hindi speech data-set. In this work, we investigate the Word Error Rate (WER) of the Hindi ASR based on five activation functions (Sigmoid, tanh, relu, leaky-relu, and elu) and three optimization techniques (Adam, SGD, and Rmsprop). We have been used the Hindi speech data-set [8] for this work. In comparison to the other European languages, a small amount of work has been done for the Hindi language [4, 9–16]. The major difficulty for Hindi ASR is the lack of speech data available for research. All the hyper-parameters and CNN architecture remained the same for all experiments.

The remaining paper is sorted out as follows: In Sect. 2, CNN-based acoustic modeling for the Hindi language was discussed. In Sect. 3, speech corpus details were explained. Sect. 4 covers the results and experimental setup details. In Sect. 5, the conclusion is discussed.

2 CNN-Based Acoustic Modeling for Hindi Language

The architecture of CNN acoustic modeling is illustrated in Figure 1. The convolutional layers are the main building blocks of any CNN architecture, in which a small size of filters was applied to the input to generate feature maps. 40-FBANK features

Fig. 1 CNN based acoustic modeling



were used as an input to the CNN architecture throughout this work. We have applied the filter of size 9×9 to the first convolutional layer to get the output feature map. Eighty feature maps were used in the first convolutional layer. The dimensions of the feature maps are reduced by applying the pooling layer known as max pooling layer with a filter size of 1×2 . The pooling was only applied in the frequency domain. The input context of 13 frames was used in this work.

Max pooling layer contains the most prominent features of the previous feature map. The final dimension has been reduced to 1×5 . Finally, a fully connected layer was used. In this work, we used the FC layer of size five, in which the last layer used a softmax activation function to generate the posterior probabilities. The other details are mentioned in Table 1.

Table 1 CNN architecture configurations

Layers	#Layers	Feature map size	Filter size	#Neurons
Convolution	3	80, 60, 60	9, 3, 3	–
Pooling	3	80, 60	2, 2	–
Fully connected	3	–	–	1024

Table 2 Hindi corpus details

SN	Dataset	#Sentences	Duration
1	Training	800	2.1 h
2	Development	100	10 Min
3	Test	100	10 Min

3 Dataset Details

In this work, we used the Hindi speech dataset created by TIFR, Mumbai, in 2000 [8]. This dataset contains 100 speaker utterances. 10 sentences were recorded by each speaker in an isolated room with two microphones on 16 kHz sampling frequency. This dataset is small in approximately 2.5 hours. We split this dataset into three parts, training set, development set, and test set. We used 80 speakers for training and 10 speakers for development, remaining speakers for the testing. Further details of this dataset can be found in Table 2.

4 Experimental Setup and Results

In this work, 40 log-Mel Filterbank features with a 13-frame context window were used as an input. To extract the features, we used the Kaldi [17] toolkit. The CNN based acoustic modeling was done by the Pytorch-Kaldi [18] toolkit. The speech features and alignments generated via the Kaldi toolkit were used as an input for the Pytorch-Kaldi toolkit. We used the mini batch of size eight. The various optimization techniques were used. The training has been done for 24 epochs. The system performance was measured, after each epoch on the dev set. We have chosen the threshold value as 0.001 for updating the learning rate. We used the initial learning rate of 0.08. Except these, we use the dropout of value 0.2 in all layers for all experiments. The pronunciation dictionary of the Hindi language contains 2803 words made up of 68 Hindi phones. In this work, only closed vocabulary experiments have been done for investigation. For language modeling, the different text was collected from various sources to train the language model. The SRILM [19] toolbox was utilized to prepare the language model. For language modeling, only the labeled data of tarin set was used.

4.1 Experiments with Baseline System

In this section, we prepared the baseline system using the Kaldi toolkit. The alignments were generated by a tri-phone based GMM system was used to train the CNN based acoustic model. The Pytorch-kaldi toolkit was utilized in training of CNN based acoustic modeling. The Sigmoid activation function used to train all the acoustic models is presented in this section. The Stochastic Gradient Descent (SGD) algorithm was used in optimization. Table 3 reports the ASR system performance using different acoustic modeling. First, GMM-HMM-based acoustic modeling took place using 39 MFCC features. We that found the tri-phone-based acoustic model improves the ASR system performance. For DNN based acoustic modeling, 40-dimensional FBANK features were used. We used five-layer architecture with 1024 hidden neurons for this purpose. We used the sigmoid activation function in hidden layers and the softmax activation function in the output layer. DNN works better compare to the conventional GMM-HMM model. Finally, CNN based acoustic model takes place. For CNN, we used 40 FBANK features and 13 context window sizes as an input to the CNN model. All other configurations remained the same as describes in the previous section. Table 3 reports the best WER Of 23.04% was found with CNN based acoustic model. Figure 2 demonstrates the evaluation of loss and accuracy of CNN based acoustic modeling for above mention setup.

Table 3 Baseline system performance evaluation

Acoustic models	Features	Language model	WER(%)
Monophone-GMM	MFCC	tri-gram	30.20
Triphone-GMM	MFCC	tri-gram	28.85
DNN	FBANK	tri-gram	25.50
CNN	FBANK	tri-gram	23.06

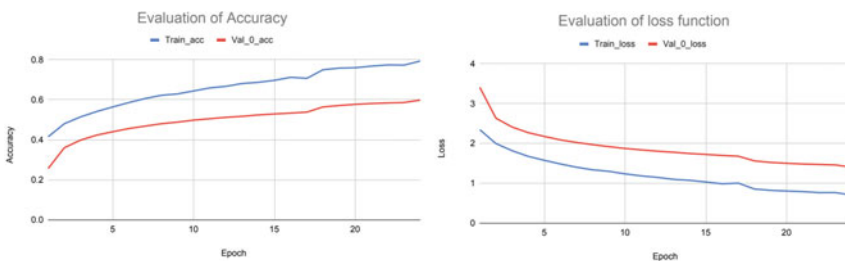


Fig. 2 Baseline CNN based acoustic modeling, (left) accuracy evaluation, (right) loss function evaluation

Table 4 Experiments with different activation function and optimization techniques

Optimization	Sigmoid	Tanh	ReLu	Leaky-ReLu	'Elu
SGD	23.06	16.48	15.36	15.88	14.56
Adam	25.08	17.65	16.88	17.44	15.66
Rmsprop	24.66	16.96	16.04	16.88	15.04

4.2 Experiments with Different Activation Function and Optimization Techniques

In this study, we explore the role of three optimization techniques (SGD, Rmsprop, Adam). The impact of five activation function on accuracy was measured. We found that Sigmoid activation function performs very poorly compared to the other activation functions. Tanh activation function works slightly better than the Sigmoid activation function. Except for Sigmoid and Tanh, we tested ReLu, Leaky-ReLu, and Elu activation function. In the same series, we tested our model with different optimization techniques. We found SGD optimization works better compared to the other two optimization techniques. We found elu activation function with SGD optimization technique gives best WER of 14.56%. All other configurations were remained same throughout all the experiments (Table 4).

5 Conclusion

Training of neural network acoustic models on limited resource data mostly suffers from generalization issues. A careful selection of the acoustic model and fine-tuning will address the issue. In this paper, the role of activation functions and optimization techniques were investigated on limited resource Hindi dataset. Overall, 9% relative reduction on WER was measured using elu activation function for Hindi speech recognition. In optimization techniques, SGD technique performs well in comparison to Rmsprop and adam optimization. This work can be investigated to explore the different hyperparameters of the neural networks. The accomplished WERs were still moderately high however inferred the right way for subsequent work.

References

1. Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2011). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing*, 20(1), 30–42.
2. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82–97.
3. Mohamed, A. R., Dahl, G., Hinton, G. (2009) Deep belief networks for phone recognition. In Proceedings of Nips workshop on deep learning for speech recognition and related applications. vol. 1(9) (p. 39)
4. Passricha, V., & Aggarwal, R. K. (2019). A hybrid of deep CNN and bidirectional LSTM for automatic speech recognition. *Journal of Intelligent Systems*, 29(1), 1261–1274.
5. Glorot, X., Bengio, Y. (2010) Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).
6. Rumelhart, D. E. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing*, 1, 318–362.
7. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
8. Samudravijaya, K., Rao, P. V., Agrawal, S. S. (2000) *Hindi speech database*. In *Sixth International Conference on Spoken Language Processing*
9. Sahu, P., Dua, M., Kumar, A. (2018) Challenges and issues in adopting speech recognition. In *Speech and Language Processing for Human-Machine Communications* (pp. 209–215). Singapore: Springer.
10. Kuamr, A., Dua, M., Choudhary, A. (2014) Implementation and performance evaluation of continuous Hindi speech recognition. In *2014 International Conference on Electronics and Communication Systems (ICECS)* (pp. 1–5). India: IEEE.
11. Kumar, A., Aggarwal, R. K. (2020) A time delay neural network acoustic modeling for Hindi speech recognition. In *Advances in Data and Information Sciences* (pp. 425–432). Singapore: Springer.
12. Dua, M., Aggarwal, R. K., & Biswas, M. (2019). Discriminatively trained continuous Hindi speech recognition system using interpolated recurrent neural network language modeling. *Neural Computing and Applications*, 31(10), 6747–6755.
13. Dua, M., Aggarwal, R. K., & Biswas, M. (2019). GFCC based discriminatively trained noise robust continuous ASR system for Hindi language. *Journal of Ambient Intelligence and Humanized Computing*, 10(6), 2301–2314.
14. Kumar, A., Aggarwal, R. K. (2020) A hybrid CNN-LiGRU acoustic modeling using raw waveform sincnet for Hindi ASR. *Computer Science*, 21(4).
15. Kumar, A., Aggarwal, R. K. (2020) Hindi speech recognition using time delay neural network acoustic modeling with i-vector adaptation. *International Journal of Speech Technology*, 1–12.
16. Kumar, A., & Aggarwal, R. K. (2020). Discriminatively trained continuous Hindi speech recognition using integrated acoustic features and recurrent neural network language modeling. *Journal of Intelligent Systems*, 30(1), 165–179.
17. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J. (2011) The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding 2011 (No. CONF)*. *IEEE Signal Processing Society*
18. Ravanelli, M., Parcollet, T., Bengio, Y. (2019) The pytorch-kaldi speech recognition toolkit. In *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6465–6469). IEEE.
19. Stolcke, A. (2002) SRILM—an extensible language modeling toolkit. In *Seventh international conference on spoken language processing*.

Energy Conserving Techniques of Data Mining for Wireless Sensor Networks—A Review



Pragati Patil Bedekar, Atul Raut, and Abhimanyu Dutonde

Abstract Now a day the application of Data mining in many areas has been tremendously increased. Data management and its processing is becoming the active research area for research community. In this paper the major concern is to study and analyse the impact and performance of various data mining techniques over sensor networks (wireless), as the characteristics of sensor nodes and its wireless nature are the primary concerns. The volume and rate for data generation is huge, which is variable in nature; therefore, it is very much essential to design and implement data mining methods for Wireless Sensor Network. This paper mainly focused on to the comprehensive survey of existing data mining techniques, wherein the various limitations and its probable solutions are highlighted in detailed. A transition of traditional mining techniques to newly introduced research is also analysed in this paper. Detailed working and description of compression algorithm has been stated in this paper.

Keywords Data mining · Energy conservation · System efficiency · Wireless sensor network

1 Introduction

To extract useful or meaningful information from the large databases; the concept of data mining is used. Therefore, it relates to the meaningful or relevant information from the warehouse.

P. P. Bedekar (✉)

Department of Computer Science and Engineering, Abha Gaikwad-Patil College of Engineering, Nagpur, India

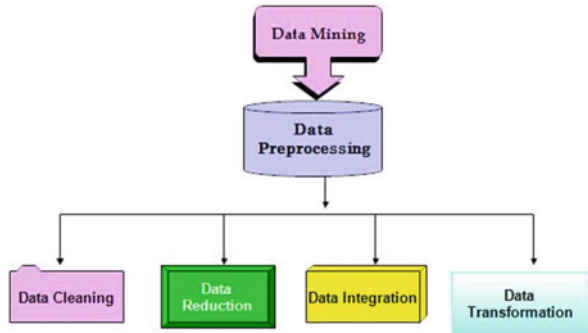
A. Raut

Department of Information Technology, Jawaharlal Darda Institute of Engineering and Technology, Yavatmal, India

A. Dutonde

Department of Information Technology, Tulsiramji Gaikwad-Patil College of Engineering and Technology, Nagpur, India

Fig. 1 Data processing techniques of data mining



One of the broader areas of Data Mining is Data Pre-processing which includes techniques like

- **Data Cleaning** This technique is used to remove noise and correct inconsistencies in data.
- **Data Integration** This technique is used to combine data from various sources into single warehouse.
- **Data Reduction** Data reduction can be observed by Data aggregation, clustering methods.
- **Data Transformation** Changing format, values, data (transforming data from one form to another) (Fig. 1).

Reliable and real time monitoring plays a vital role in many applications like field monitoring, health and habitant monitoring, tracking of objects, analysis of data etc. Practical monitoring of above mentioned applications is very impotent because maximum time it deals with the fast rate of dynamic and huge amount of data streams which can be geographically distributed all over network. This huge amount of variable data is then analysed, processed and transformed using mining techniques. In this paper existing and revised data mining techniques has been studied.

2 Overview of Methods Used

Abdelmoghith and Mouftah [1] The proposed work is focused on the compression of data which is being sent between sink and sensor node. The mathematical results proved that the data reduction is up to 70% and energy saving is up to 37%. The data compression algorithm, S-LZW, LEC, Huffman, ND-Encode, RQT etc., has been studied and analyzed, but the proposed MBC has been proved to be the energy efficient scheme. Mamurjon and Ahn [2] The authors have discussed about Data Collection methods in WSN. Generally gathering sensed data from each node is a tedious job. In general, base node starts from node nearby base station, wherein it has travel all over the network for collecting data from each node, return back

and upload the data to base node. This entire process is energy and time consuming. Hence the author has introduced the concept of mobile sink, where in the mobile sink used to visit every cluster in the network, return back and send sensed information to sink node. The introduced method has decreased energy consumption of nodes. Mudgule et al. [3] have done review on different data reduction techniques like adaptive filter, tree based method, cluster technique, data prediction method, etc. It has been observed that to reduce energy consumption the data reduction technique of Data Mining is important. Data reduction is used to remove unnecessary data while transmission. In this technique, it reduces the repeated data, its processing and analyzed it before transmission and hence as the amount of data reduce it leads to energy preservation in WSN.

3 Classification of Compression Techniques

Two algorithms are discussed in this paper that is compression and reconstruction algorithm [4]. Compression algorithm takes X as input and generates a compressed X_c whereas reconstruction algorithm executes on compressed representation X_c and gives output as the reconstruction Y .

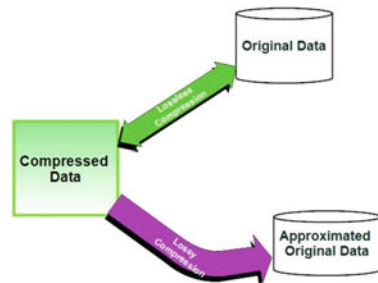
Based on requirement of reconstruction, a data compression technique has been classified into two broader classes as shown in Fig. 2.

1. Lossless compression scheme, reconstruction of original data from compressed data, in which Y is identical to X and
2. Lossy compression scheme, provide higher compression than aforesaid compression but allow Y to be different from X .

The ultimate aim is to reduce energy consumption of the Network, which can be achieved by using compression techniques. It deals with conversion of data strings of characters into another representation which may have same data in small length as much as possible. The main objective is efficient coding and minimize bandwidth required for transmission.

The main objective of this review is to show performance of different lossless and lossy compression techniques and their comparative study. This technique helps to

Fig. 2 Working of compression technique



save battery power and data storage capacity or transference capacity. It has been studied that more energy is utilized in transmission rather than the processing of information therefore the sending compressed data can work out at some extent. Following is the review on different algorithms/schemes used for data compression.

3.1 String Based Compression Techniques

LZW is a general technique used for compression of data. This technique is a universal lossless data compression algorithm. It is also called as dictionary based algorithm [5–10]. In the year 1977, A. Lempel and J. Ziv published this algorithm.

Welch [11], Singh and Singh [12], Saidani et al. [13] The LZW algorithm: it never send character, and it tries to send codes for already known strings. Each time a new code is sent, a new code is attached to the dictionary. Working of algorithm is as follow:

- a. Reading of incoming text (no analysis will be performed here)
- b. Check and verify whether the string has been read before or not
- c. If no in step b, then read string and add it to the dictionary
- d. Compression of text will be done if the new code is written instead of string.

The start dictionary contains standard character set, that is, ASCII Codes. Two control codes namely 256 and 257 are attached to this dictionary. The 256 is used as an EOF (end of file) character, and the 257 is used as an EOD (end of dictionary) character. Example,

Let us consider the string c a b b a c a b b a.

Index	Dictionary	Step	X	y	x+y	Code	
1	A	1		c	C		
2	B	2	C	a	Ca	3	
3	C	3	A	b	Ab	1	
4	Ca	4	B	b	Bb	2	
5	Ab	5	B	a	Ba	2	
6	Bb	6	A	c	Ac	1	
7	Ba	7	C	a	Ca		String Matched
8	Ac	8	Ca	b	Cab	4	
9	Cab	9	B	b	Bb		
10	Bba	10	Bb	a	Bba	6	String Matched
11	Bba	11	A		A	1	

At the start, the dictionary contains all possible individual characters and x is empty; y contains the next character in the char stream. Now, let us find out whether the string $x + y$ is present in the dictionary or not. If it is not found in the dictionary then add $x + y$ string to the dictionary.

Hence after applying concept of LZW algorithm over the above text the code 3 1 2 2 1 4 6 1 is produced as output.

S-LZW here **S** stands for Sensor. S-LZW is a dictionary/string based lossless compression algorithm which is used to support data compression in WSN. It is stated that to transmit plain text data, more energy is required as compared to compressed text. Compressed form data gives more reliability comparatively. It has been observed that in transmission process of data, packet loss occur in WSN at more rate; hence, S-LZW algorithm follows the strategy to divide the incoming data stream into small and independent blocks as shown in the figure. While doing this the algorithm can assure that if packet loss happened in transmission process then it only affect the packet loss following packets in its own block and therefore rest of blocks will be secure enough. With reference to [8] as resources are limited in WSN, therefore a dictionary of size 512 entries is adopted to fit in small memory of a sensor node. At the same time, the algorithm said to compressed data into the block of 528 Bytes as shown in the figure which helps to improve performance. Now the next step is to divide the compressed data into many blocks of independent packets (Figs. 3 and 4).

Fig. 3 Flowchart for LZW algorithm

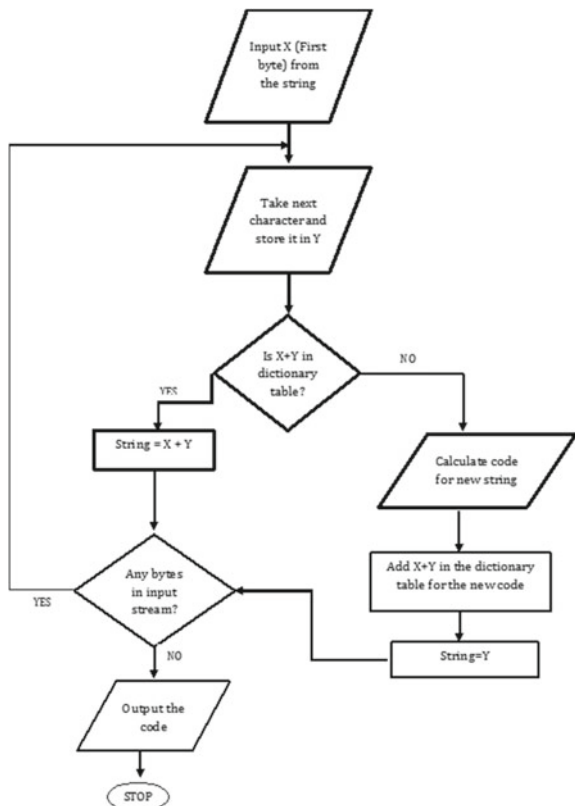
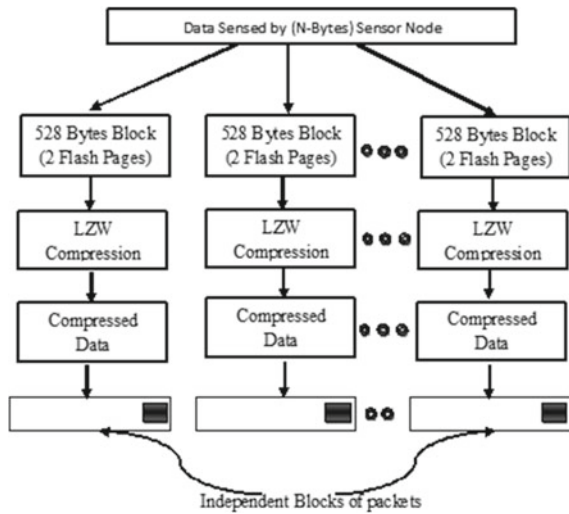


Fig. 4 Working of S-LZW algorithm



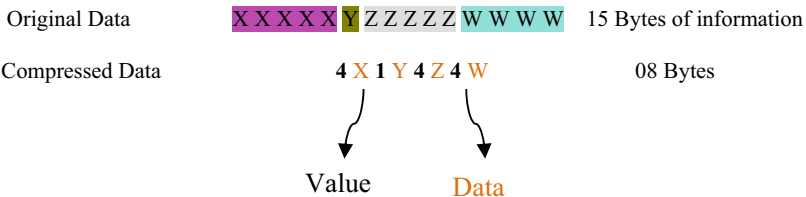
3.2 RLE and K-RLE

Run Length Encoding (RLE). As mentioned in S-LZW algorithm, transmission and reception of data consume more power. In order to reduce overhead of energy consumption while data transmission which is also called as in-network processing technique, enhanced compression techniques are used—RLE is one of them. First we will see the working of RLE in detail [9].

RLE (lossless) is a very easy technique which is used to compressed digital data. The data is stored as single data in the form of Value and Count. It is more benefited when data carries the sequences of same data value occurrences in the original runs or data stream; hence, instead of storing same original run, the data will be stored in terms of data value and count. This method is needed to reduce the amount of data needed for storage and transmission.

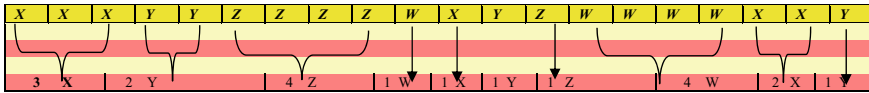
General definition is *d* is a data item occurs at *n* consecutive times in the input stream, replace *n* occurrences with single pair *nd*.

Example: Consider the following original run length.



RLE is used where we expect the consecutive sequence of same runs. In the given example, similar occurrences of X, Y, Z and W are counted and compressed as Count followed by data or vice versa.

- With this example we can see the clear difference of using LZW and RLE.



K-RLE means RLE with K-Precisions. In this technique if data items d , $d + K$, or $d - K$ occurs n consecutive times in the input stream then replace the n occurrences with the single pair nd [8].

3.3 Coding by Ordering

Coding by ordering is a method used to minimize amount of data need to send for further communication. The main objective is to increase network lifetime by using two schemes mainly a. Data Aggregation b. Data Compression [7].

Step 1. The controller breaks down the space into different regions also known as cuboids and transmit the interested data packets to each region.

Step 2. After step 1, maximum nodes transmit sensed data to the controller after the specified intervals.

Step 3. As almost all nodes send data at the specified time, then it will become easier to combine all the readings into a single packet.

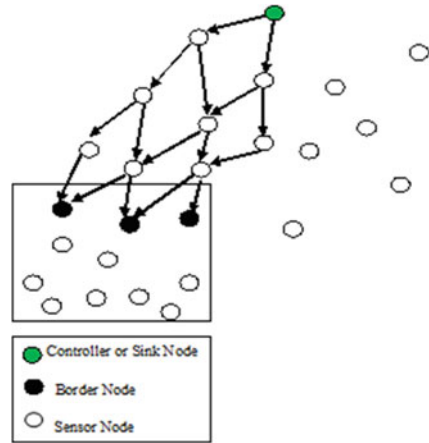
Step 4. One data packet with only one adder will travel from the region to the controller (Fig. 5).

Data Funneling (routing algorithm) has been introduced [7] to reduce the communication overhead while sending the set of sensed data by the number of sensor nodes to the single destination called as sink node. The proposed routing algorithm not only benefited in energy conservation but also used to reduce the data packet collision in the network.

For the compression of data ‘Coding by Ordering’ scheme has been used which losslessly compressed data encode learning information in the sequence or order.

The nodes (nodes) disbursed in large region will send data to the border node. Border node is point where all data from the region (send by n nodes) will be collected. In this case sensor nodes will compute the communication cost for border node only instead of sink (controller) node. As border nodes needs to communicate with controller so it will compute its communication cost.

Fig. 5 Coding by order



Border node is the responsible one to place each node’s packet into a big super-packet containing sensed information from environment then send only one big-packet to control node. The packet (node) contains the NODE ID and node ID contains its position and payload. The Border node contains the nodes ID convey belonging of payload assign to which particular node. The authority to choose ordering of packets to super packets is managed by border node, and it will not affect application as the all data reached at controller at the same time.

Example: Assume that there are four nodes in the region with ID’s 1, 2, 3, and 4. Each of these sensor ranging from 1 to 4 sensed some number values let us say {0, ..., 5}. It is stated earlier that the border node has freedom, in the below example the data packets of node 4 has been suppressed by border node, and now it will select possible permutations combination for the remaining 3 nodes that is $3! = 6$ which is shown in below table [7] (Table 1).

Table 1 Mapping from permutations to integers

Permutations	Integer
123	0
132	1
213	2
231	3
312	4
321	5

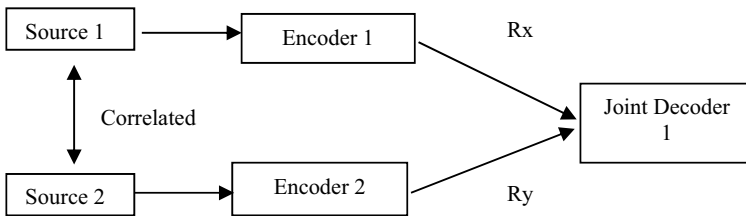
3.4 Pipelined In-Network Compression

The authors' motivation behind this compression technique is to save energy required for the communication in wireless network. Thereby reducing the repetitive data stream the said objective has been achieved by Pipelined In-Network compression method.

1. Sensed data by node (mote) is buffered at own node itself
2. Delay for some appropriate time satisfying End-to-End latency-bound
3. Remove same or repeated data from the buffered data.

3.5 Distributed Compression

Distributed Compression as the name suggests it is used to do compression in a distributed manner. Here without the communication between n motes or sensor nodes the data redundancies can be removed in a distributed manner.



As it is clear from the above figure, information can jointly compress correlated observations, even with no communication between nodes.

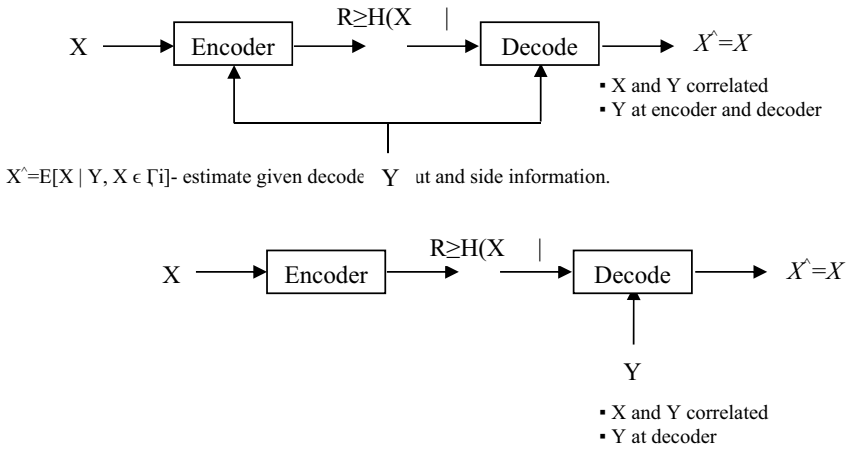
- Source coding with side information-discrete alphabets

X and Y are two data sets likely equal with length of three binary data. Hamming Distance between X and Y is 1 at most.

$$X = [0\ 1\ 0], Y = [0\ 1\ 0], [0\ 1\ 1], [0\ 0\ 0], [1\ 1\ 0].$$

X can be compressed to two bits. If both decoder and encoder know Y, what will happen to the compression rate of X? Then if Y is known to the decoder.

According to this method, since the Decoder know Y and X is only one Hamming Distance apart from Y, it is not sufficient to differentiate $X = 111$ from $X = 0000$ [6].



For same reason, $X = 001$ and 110 , $X = 010$ and 101 , and $X = 011$ and 110 do not need to distinguished from each other. These set of two X values are grouped as four coset and assigned four different binary index numbers:

coset 1 = $(000, 111)$: 00, coset 2 = $(001, 110)$: 01, coset 3 = $(010, 101)$: 10, coset 4 = $(011, 100)$: 11.

If input is $X = 010$ and $Y = 110$, Decoder received $Y = 110$ as a side information from Y and $X = 10$ as a partial information from X . Then at the Decoder, $X = 010$ is selected from coset 2 since 110 has a hamming distance of 2 from 110 . Whether X know Y or not, X can still reduce or compress three bits information into two bits.

This method is particularly used in discrete sources and continuous sources. The same method useful for lossless and lossy compression schemes.

4 Conclusion

The major concern to study compression techniques is the constrained of battery of sensor nodes. However, many researchers has proposed the use of compression techniques of data mining to extend the life of sensor networks. The studies stated that comparatively lossless data compression methods are easy to implement and understand, also gives better compression efficiency. This paper talks about the comprehensive survey of various compression algorithms and its detailed working. The compression algorithm of both categories lossless and lossy has been studied and analysed here. The main objective to study different types of compression algorithm is to understand how energy can be saved with the help of it. Data gathering, its processing, aggregation and compression have become the interested field of research.

With the use of basic data mining techniques, in future the paper will focus on various types of aggregation methods, fusion techniques, funnelling techniques,

storage, analysis, and knowledge extraction, and compression of huge volume of data is also the key concern in future.

References

1. Abdelmoghith, E. M., & Mouftah, H. T. (2013). A data mining approach to energy efficiency in wireless sensor networks. In *International Symposium on Personal, Indoor and Mobile Radio Communication: Mobile and Wireless Networks* (pp. 2612–2626).
2. Mamurjon, D., & Ahn, B. (2013). A novel data gathering method for large wireless sensor networks. In *ELSEVIER-International Conference on Electronic Engineering and Computer Science. IERI Procedia*, 4, 288–294.
3. Mudgule, C. B., Nagaraj, U., & Ganjewar, P. D. (2014). Data compression in wireless sensor network: A survey. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(11), 60–63.
4. Preethi, A., & Ramakrishnan, A. (2015). A survey on data compression techniques in wireless sensor network. *International Journal of Electrical and Computer Engineering*, 2(3).
5. Vidhyapriya, R., & Vanathi, P. (2009). Energy efficient data compression in wireless sensor networks. *The International Arab Journal of Information Technology*, 6(3), 297–303.
6. Sadler, C. M., & Martonosi, M. (2006). Data compression algorithms for energy—constrained devices in delay tolerant networks (pp. 265–278). ACM
7. Kimura, N., & Latifi, S. (2005). A survey on data compression in wireless sensor networks. In *Proceedings of the International Conference on Information Technology: Coding and Computing* (pp. 8–13).
8. Alwadi, M., & Chetty, G. (2015). Sensor selection scheme in temperature wireless sensor network. *International Journal of Wireless & Mobile Networks (IJWMN)*, 7(3), 47–53.
9. Kusuma, J., Doherty, L., & Ramchandran, K. (2001). Distributed compression for sensor networks. In *Proceedings of the IEEE Conference on Image Processing, Thessalonki, Greece, October 2001* (pp. 82–85).
10. Capo-Chichi, E. P., & Friedt, J.-M. (2009). K-RLE: A new data compression algorithm for wireless sensor network. In *Third International Conference on Sensor Technologies and Applications* (pp. 502–507).
11. Welch, T. A. (1984). A technique for high-performance data compression. *IEEE*, 8–18.
12. Singh, N., & Singh, H. (2019). Energy efficient NK-RLE data compression scheme. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(6), 4014–4021.
13. Saidani, A., Jianwen, X., & Mansouri, D. (2020). A lossless compression approach based on delta encoding and T-RLE in WSNs. *Hindawi Wireless Communications and Mobile Computing*, 2–10.

IoT Based Healthcare System for Patient Monitoring



S. Saravanan, M. Kalaiyarasi, K. Karunanithi, S. Karthi, S. Pragaspathy, and Kalyan Sagar Kadali

Abstract The Internet of Things gadgets can acquire and transmit the data straightforwardly with different gadgets through the cloud, giving a gigantic measure of data to be assembled, stored and investigated for data-analytics processes. The aim of this paper is to enhance the patient's quality of life by accessing real time visibility of the patient's condition, through measuring the physiological parameters like systolic, diastolic, pulse rate and body temperature values. The key idea is to administer care to the patients by constantly monitoring the medical parameters include blood pressure, pulse rate and body temperature without the need for the patient to move from facility to facility for constant supervision of their health. Data gathered from the blood pressure sensor and temperature sensor is analyzed and stored in the cloud, which can be monitored by the caregivers of the patient from any location and respond appropriately, based on the alert received.

Keywords Internet of things · Healthcare system · Blood pressure · Temperature · Arduino UNO

1 Introduction

The Internet of things (IoT) is the inter-connecting of different gadgets in order to communicate and share data with other networks and computers. It is utilized in

S. Saravanan

B V Raju Institute of Technology, Narsapur, India

e-mail: Saravanan.s@bvrit.ac.in

M. Kalaiyarasi (✉) · S. Karthi

V.S.B. Engineering College, Karur, India

K. Karunanithi

Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Tamil Nadu, India

S. Pragaspathy

Vishnu Institute of Technology, Bhimavaram, India

K. S. Kadali

Shri Vishnu Engineering College for Women, Bhimavaram, India

different areas such as e-health [1], Cultural Heritage [2–6] not failing to remember legitimate area [7–9], Public organization domain [10–13], and Humanitarian Help and Disaster Relief [14–16], yet in addition home computerization, self-ruling and associated vehicles [17], and wearable innovation.

The IoT devices are developed with different types of sensors, actuators, processors and transceivers [18]. With the increasing utilization of sensors by clinical gadgets, remote and constant observing of a patient's health is getting conceivable. This organization of sensors, actuators and other portable specialized gadgets, alluded to as Internet of Things for Medical Devices (IoT-MD), is ready to revolutionize the functioning of the medical care industry. Associated health technologies have acquired conspicuousness because of the broad event of ongoing illnesses and the significant requirement to control medical care expenses of a maturing populace. An associated medical care climate advances the speedy progression of data and empowers easy access to it. Improved home care facilities and customary health updates to clinicians diminish the odds of repetitive or improper consideration, improve tolerant consideration and security, and decrease generally expenses of care. It can likewise be utilized to follow way of life infections, for example, hypertension, diabetes and asthma which need ceaseless observing.

Technological giant Cisco predicts that 25 billion gadgets will be connected with IoT by 2015 and 50 billion by 2020. The IoT is another innovative, and there is still no steady definition for it [19]. IoT can offer incredible guarantee in the field of medical services. It may be utilized broadly to improve admittance to nature of mind and in addition give mindfulness, backing and help to patients in distant areas. IoT can likewise give help to patient's ongoing sicknesses. IoT can offer incredible guarantee in the field of medical services. It may be utilized broadly to improve admittance to nature of mind and in addition give mindfulness, backing and help to patients in distant areas. IoT can likewise give help to patient's ongoing sicknesses. Auto-ID Center represents the joint effort among industry and private area to deal with new innovation for following merchandise all around the world. Presently Auto-ID Labs are driving scholarly examination network on IoT.

A remote versatile multi-boundary gadget was proposed to procure physiological signals and send them to a nearby worker through Bluetooth remote innovation. Four sorts of screen units were intended to impart by means of the WiFi remote innovation, including a neighborhood screen unit, a control place, cell phones and a web page. The utilization of different screen units is intending to meet distinctive clinical prerequisites for various clinical work force. This framework was exhibited to advance the portability and adaptability for both the patients and the clinical faculty, which further improves the nature of medical services [20].

ECG instruments calculate the electrical function of the heart to assess heart problems. ECG signal consistency affects when a patient has any cardiac problems. Be that as it may, the accessibility of ECG gadgets is causing ECG conclusion moderate and troublesome. The issues found in many medical clinics are that nonstop observing of indispensable parameters are accomplished for ICU patients, yet the screens are nearby to the room in which the patient is conceded. Doctor needs to consult the patient regularly and assess his/her situation by dissecting the deliberate

criteria, e.g., temperature, circulatory strain, beat oximeter, E.C.G. and more [21]. Latest and progressed installed gadgets have demonstrated their capacity to analyze health boundaries which are significant and to be persistently checked in clinical fields. Advancement sheets in clinical and non-clinical apparatuses have arisen with quicker and less expensive organization measure.

The IoT-MD gives an environment where a patient's significant information gets communicated by clinical gadgets through a passage onto secure cloud based stages where it is stored and dissected. It helps to store information for a large number of patients and perform investigation continuously, eventually advancing a proof based medication framework. In this paper, IoT based healthcare system for patient monitoring has been proposed.

This paper is further arranged in the following manner. Sect. 2 gives the objective of the proposed healthcare system. Sect. 3 presents the proposed methodology. Implementation steps and results are discussed in Sect. 4, and the conclusion of this paper has been given in Sect. 5.

2 Objectives

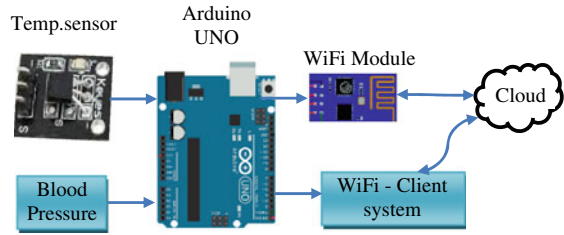
- To create IoT-driven framework for making it conceivable to drastically lessen costs and improve health by expanding the accessibility and quality of care.
- To create an embedded innovation for use all through IoT-driven medical healthcare system, including:
 - Sensor that gather quiet information.
 - Microcontrollers that process analyze and remotely communicate the information.
- To create Healthcare-explicit passages through which sensor information is further dissected and transmitted to the cloud.

3 Proposed System

Wellbeing and health management is one of the most encouraging application zone of IoT innovation. Remotely observing of patients prompts more viable and ideal treatment, prompting better administration of health. Likewise, patients are engaged by getting more perceivability into their genuine ailments, empowering them to assume a functioning part in controlling and impacting their treatment. The proposed method presents the idea of solving medical problems utilizing most recent innovation, Internet of Things.

The block diagram of smart real time health care monitoring and alert system using IoT has been given in Fig. 1. It presents the smart medical care framework utilizing IoT which is intended to give Quality Health Care to everyone. Through this proposed system, patient's significant needy parameters such as blood pressure

Fig.1 Monitoring and alert system



(BP) and temperature can be measured progressively. The workflow of the proposed system has been given in Fig. 2. It has been explained as follows:

Step:1 Initially, connect the blood pressure sensor and temperature sensor to the Arduino UNO.

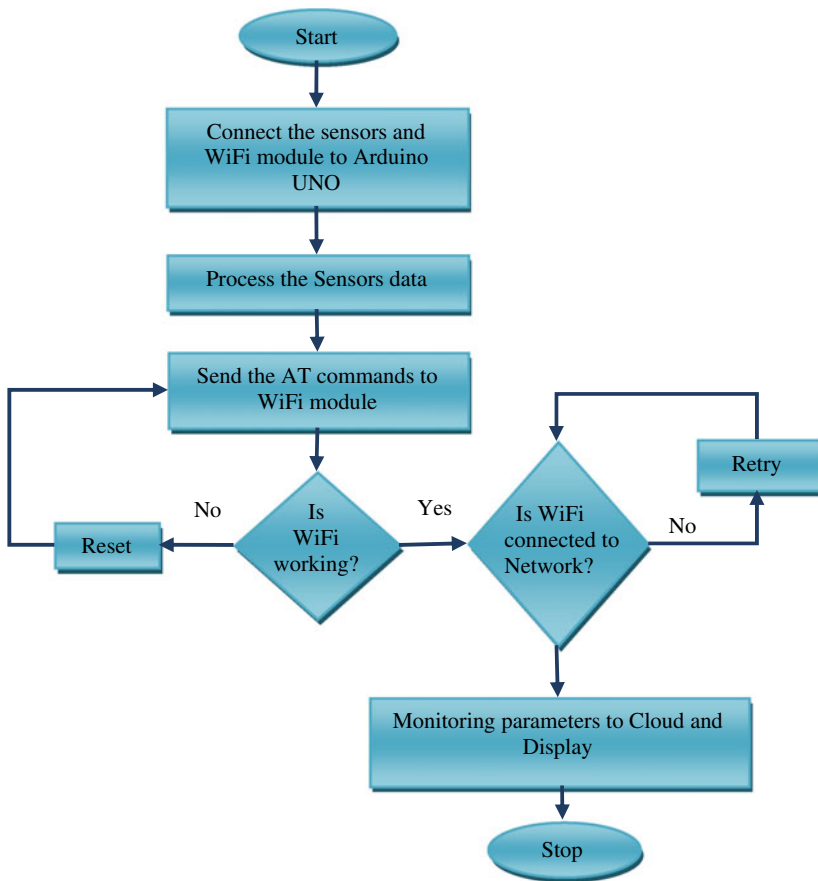


Fig. 2 Flowdiagram

Step:2 Process the sensor data using Arduino UNO with the help of Arduino programming.

Step:3 Establish TCP network connection to the WiFi module.

Step:4 Check whether the WiFi module is connected to the network or not?

Step:5 Once the WiFi connects send the sensor data to Cloud. If not go to step 3.

Step:6 The stored data can be viewed using Android app on the phone.

Step:7 A tweet alert has been sent whenever the data falls below or above the predefined value.

4 Hardware Implementation

The health parameters, for example, BP, beat rate and temperature level are estimated from the pulse and temperature sensors separately. The gathered medical information is stored in Arduino UNO through interfacing. This information is then sent to the ESP8266. The transmitted data is also transferred to the cloud. Here the pulse sensor is connected to the UART pins (0 and 1) of the ATmega328. Temperature sensor is associated with the pin 4 of Arduino UNO and Esp8266 WiFi module sequential pins (Tx&Rx) to the product sequential pins (8 and 9) of Arduino UNO. The arrangements of the proposed system has been given in Fig. 3. The measured values are shown in Fig. 4.

The measured data has been transmitted to IoT based cloud administration referred as ThingSpeak using the ESP8266 WiFi shield. Since ThingSpeak web service works in collaboration with MATH works, the transmitted data are stored in the graphical format. The stored values of Systolic (SC), Diastolic (DC), Pulse Rate (PR) and Temperature (T) on ThingSpeak have been shown in Fig. 5.

From the analysis of the above figure, the following parameters are obtained.
SC:92mmhg, DC:60mmhg.

Fig. 3 Real time implementation



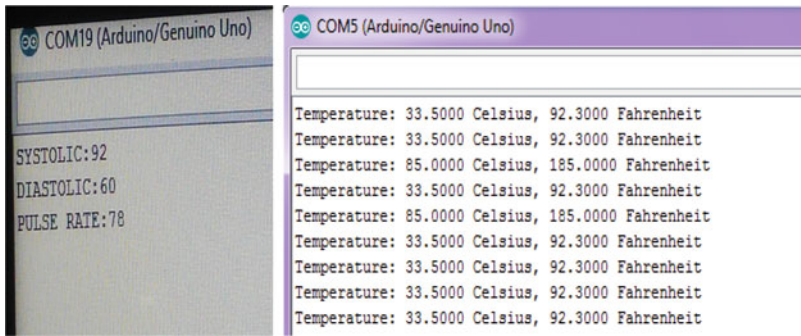


Fig. 4 Pressure and temperature sensor values on serial monitor

PR:78 bpm, T:33 °C.

The preserved data in the ThingSpeak cloud administration could be recovered from mobile application. It empowers individual to visualize their ThingSpeak directs in a simple manner by entering the channel ID. It has been appeared in Fig. 6. It shows the different project title. By clicking on the project title, it receives the values from thing speak server and displays the stored (sensors) values in graphical format.

The systolic, diastolic, body temperature and pulse values on Thing View app have been presented in Fig. 7. Based on the principles of the facts reviewed, the psychiatrist identified the illness and then estimated the state of the patient’s wellbeing. The main advantage of this method is that it is possible to access vital parameters of a particular person from anywhere in the world with the availability of Internet, IoT health care device and android mobile.

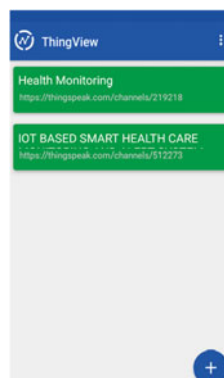
5 Conclusion

The IoT is becoming feasible solution in the field of health surveillance. Using remote patient surveillance to enable an individual’s wellbeing to be tracked and help a doctor to identify signs of disease. In this paper, IoT based healthcare system for patient monitoring has been proposed. Here, different types of sensors have been utilized to monitor the patient health parameters. They sense internal heat level and heartbeat rate separately as indicated by their capacity and afterward with the assistance of microcontroller then the data has been stored and processed further. These sensor values are then transmitted via wireless communication to a medical server. These data are subsequently received on an approved IoT framework smart phone. Based on the principles of the facts he reviewed, the psychiatrist identified the illness and then estimated the state of the patient’s wellbeing. In future, this framework can be actualized with insignificant advancement board with promptly fitting and play IoT and fundamental parameters can be observed in a pervasive strategy.



Fig. 5 SC, DC, PR and T values on ThingSpeak

Fig. 6 ThingView app window after addition of channels



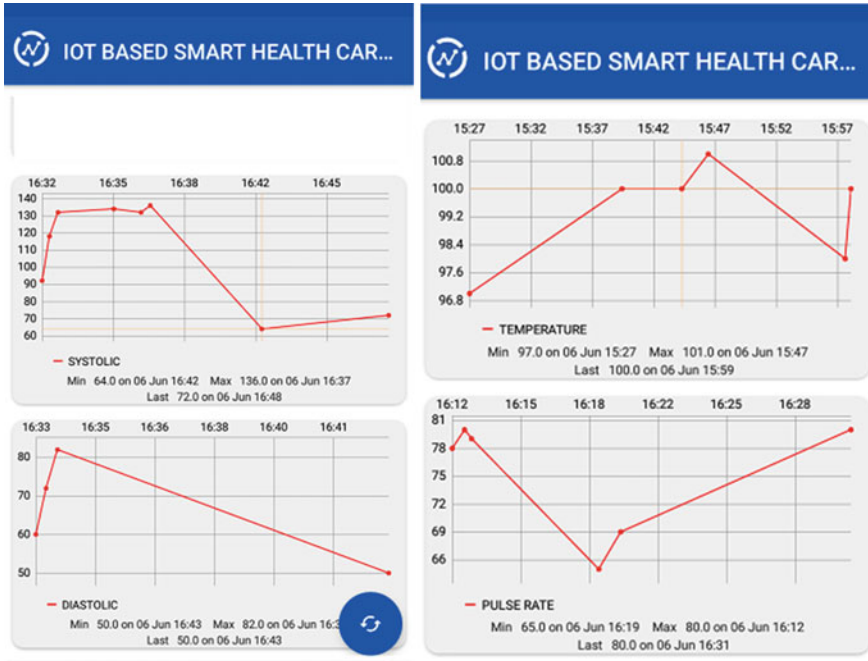


Fig. 7 Systolic and diastolic, body temperature and pulse values

References

1. Guadagni, Y. et al. (2017). RISK: a random optimization interactive system based on kernel learning for predicting breast cancer disease progression. In *Proceedings of the 5th International Work-Conference, IWBBIO 2017*, April 26–28 (pp. 189–196). Spain, Granada.
2. Ohta, S., Nakamoto, H., Shinagawa, Y., & Tanikawa, T. (2002). A Health monitoring system for elderly people living alone. *Journal of Telemedicine and Telecare*, 8(3), 151–156.
3. Pennacchiotti, M., & Zanzotto, F. M. (2008). *Natural language processing across time: An empirical investigation on Italian* (pp. 371–382). Berlin, Heidelberg: Springer.
4. Beccaceci, R., Fallucchi, F., Giannone, C. F., Spagnuolo, F., & Zanzotto, F. M. (2009). Education with ‘living artworks’ in museums. In *Proceedings of the 1st International Conference on Computer Supported Education*, (pp. 1–5).
5. Arcidiacono, G., De Luca, E.W., Fallucchi, F., & Pieroni, A., 2016. The use of lean six sigma methodology in digital curation. In *Proceedings of the CEUR Workshop*, (pp.1–7).
6. Ambrogio, A. D., Gaudio, P., Gelfusa, M., Luglio, M., Malizia, A., Roseti, C., & Marsella, S. (2017). Use of integrated technologies for fire monitoring and first alert. In *Proceedings of the International Conference of Information and Communication Technologies, AICT 2016* (pp. 1–5).
7. Pazienza, M. T., Scarpato, N., & Stellato, A. (2009). STIA: Experience of semantic annotation in Jurisprudence domain. *Artificial Intelligence and Applications*, 205, 156–161.
8. Bianchi, M., Draoli, M., Gambosi, G., Pazienza, M. T., Scarpato, N., & Stellato, A. (2009). ICT tools for the discovery of semantic relations in legal documents. In *Proceedings of the CEUR Workshop* (p. 582).

9. Boella, G., Di Caro, L., Humphreys, L., Robaldo, L., Rossi, P., & van der Torre, L. (2016). Eunomos, a legal document and knowledge management system for the Web to provide relevant, reliable and upto-date information on the law. *Artificial Intelligence Law*, 24(3), 245–283.
10. Morabito, V. (2015). Big data and analytics for government innovation. In *Big data and analytics* (pp. 23–45).
11. Zanella, A., et al. (2014). Internet of Things for Smart Cities. *IEEE Internet of things*, 1, 22–32.
12. Fallucchi, F., Alfonsi, E., Ligi, A., & Tarquini, M. (2014). Ontology-driven public administration web hosting monitoring system. *IJTRE*, 8842.
13. Bianchi, M., Draoli, M., Fallucchi, F., & Ligi, A. (2014). Service level agreement constraints into processes for document classification. In *Proceedings of the 16th International Conference on Enterprise Information Systems* (pp. 12–18).
14. Zhang, D., Zhou, L., & NunamakerJ, F. (2002). A Knowledge management framework for the support of decision making in humanitarian assistance/disaster relief. *Knowledge and Information systems.*, 4(3), 370–385.
15. Fallucchi, F., Tarquini, M., & De Luca, E.W. (2016). Knowledge management for the support of logistics during humanitarian assistance and disaster relief (HADR). *Lecture Notes in Business Information Processing* (pp. 1–7).
16. Fallucchi, F., Tarquini, M. and De Luca, E.W. (2016). Supporting humanitarian logistics with intelligent applications for disaster management. In *INTELLI 2016 : The Fifth International Conference on Intelligent Systems and Applications* (pp.51–56).
17. Pieroni, A., Scarpato, N., & Brilli, M. (2018). Industry 4. 0 revolution in autonomous and connected vehicle a non-conventional approach to manage big data. *Journal of Theoretical and Applied Information echnology*, 96(1), 10–18.
18. Colakovi, A., & Hadžialic, M. (2018). Internet of Things (IoT): a review of enabling technologies, challenges and open research issues. *Computer Networks*, 144, 17–39.
19. Wang, C., Daneshmand, M., Dohler, M., Mao, X., Hu, R. Q., & Wang, H. (2013). Guest editorial–special issue on internet of things (IoT): architecture, protocols and services. *IEEE Sensors Journal*, 13(10), 3505–3510.
20. Hao, Y., & Foster, J. (2008). Wireless sensor networks for health monitoring applications. *Physiological Measurements*, 29(11), R27–R56.
21. Gowthami, P., & Sathishkumar, P. (2016). An android based patient monitoring system. *International Journal of Innovative Research in Advanced Engineering*, 4(3), 54–57.

Detection and Classification of Intracranial Brain Hemorrhage



K. V. Sharada, Vempaty Prashanthi, and Srinivas Kanakala

Abstract Computer

-aided diagnosis systems (CAD), as their name suggests, utilize computers to assist doctors to obtain a quick and correct diagnosis. They focused on several scholars as they are built upon the concept of processing and examining pictures of various parts of the individual body meant for a fast and correct outcome. CAD systems are generally area specific because they are augmented for some certain kinds of infections, various parts of the individual body, diagnosis methods, etc. They analyze dissimilar types of inputs given, for example, signs, test center, result, health pictures, etc. varying on their territory. One of the maximum common kind of diagnosis depends on medical pictures. Our approach is to develop a model to identify either a brain hemorrhage is present or not in Computed Topography (CT) scan of the brain and also identify the kind of hemorrhage. The process of detecting and identifying hemorrhage contains many steps like image pre-processing, segmentation of image, extracting the features, and classifying the images.

Keywords CT scan · Brain haemorrhage · Image processing · Image segmentation

1 Introduction

Brain hemorrhage implies blood loss within brain. In brief, one quite stroke which causes draining round the tissues by an artery within the cerebrum is observed as brain hemorrhage. Draining can happen in between the cerebrum and also the layers that cover it. The irritation that is caused by the blood from trauma leads to increase of pressure on brain tissues; this results in reduced percentage of oxygen from reaching the brain cells. This brain hemorrhage is observed as medical emergency which needs prompt treatment. There are several factors which might cause or result in

K. V. Sharada · V. Prashanthi (✉)

Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

S. Kanakala

VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

cerebral hemorrhage. These components include head injury, blood vessel anomalies, liver disease, brain tumour, extreme high blood pressure, bleeding disorders, and utilization of illegal medications. Anomalies within the blood vessels are the most explanation in the majority of the intracerebral hemorrhages that out of nowhere happen in youngsters, anyway there could be other potential causes which incorporate blood sicknesses, cerebrum tumors, septicemia, or the work of liquor or illegal medications. Many people who experience a hemorrhage have side effects as though they are having a stroke and can create weakness on one side of their body or a feeling of numbness. Sometimes brain hemorrhage can cause a scope of various indications like sudden, serious migraine, difficulty in gulping, vision problems, loss of coordination with the body, confusion or trouble in understanding, difficulty in talking or stammering discourse, seizures, torpidity or stupor.

Brain hemorrhage often leads to several complications. Because of the draining nerve cells one cannot speak with different pieces of the body and in this manner stops ordinary working. Additionally, there are scarcely any basic issues that emerge after a brain discharge which incorporates development, discourse, or memory issues. A few difficulties might be lasting relying upon the area of drain and the harm that happens. These intricacies may incorporate—loss of motion, vision misfortune, and decreased capacity to talk or get words, disarray, or memory misfortune. These complications made brain hemorrhage an emergency condition that requires immediate treatment. Diagnosing a brain drain could be troublesome as certain individuals do not give any physical indications. Specialists need to do take the assistance of a CT scan or MRI scan so as to locate the specific area of the seeping in the mind. CT pictures are known to claim numerous points of interest over MRI. Along these lines, the standard of CT pictures is sufficiently high to precisely analyse Intracranial Brain Hemorrhage. PC supported determination frameworks (CAD), as their name proposes, use PCs to assist specialists with arriving at a quick and exact conclusion. They have been the focal point of numerous scientists since they depend on preparing and examining pictures of various pieces of the human body for brisk and exact outcomes. Computer aided design frameworks are normally area explicit as they are advanced for certain particular sorts of infections, portions of the body, analysis techniques, and so on. They investigate various assortments of information, for example, manifestations, research facility test results, clinical pictures, and so on depending on their space. One of the most generally perceived sorts of determination is the one that depends on clinical pictures. Such frameworks are helpful on the grounds that they can be incorporated with the product of the clinical imaging machine so as to deliver a brisk and precise finding. Then again, they can be trying since they consolidate the components of man-made reasoning and advanced picture preparing (Fig. 1).

The rest of the paper is organized as follows: In Session 2 Literature survey is discussed, in Session 3 proposed system is described, Session 4 describes the implementation, Session 5 explains the experimentation and results and finally in Session 6 conclusion is given of our main proceedings series.

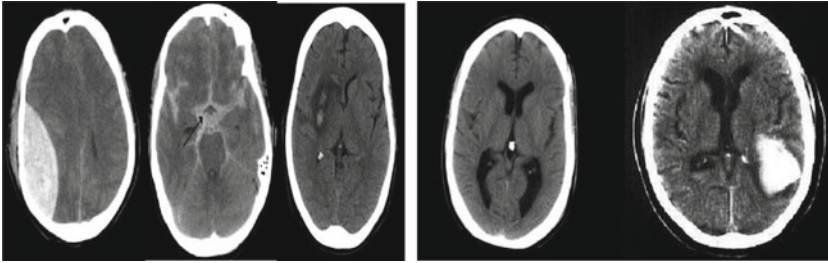


Fig. 1 CT scan images of five types of brain hemorrhages

2 Literature Survey

The paper [1] mainly discussed the pre-segmentation process. In this the techniques of dividing the input image into four quadrants using the method of splitting and merging are used. This is called as pre-segmentation of imaging. The pre-segmentation can be done with the following methods: (1) Thresholding: in thresholding the pixels of the input image is given certain threshold values and those pixels are mapped to related areas. (2) Region growing technique. (3) Supervised segmentation methods and unsupervised segmentation methods. Process of segmentation is as follows: as a first step the image is split into quadrants. Then histogram and pixel values are computed for each quadrant separately. The succeeding process for the above step is comparing each histogram value and pixel value with the predefined threshold criteria (or) histogram, pixel values. The final step is taking only the region of abnormality that is detected and feeding this as input to the segmentation process.

In [2] the method includes a combination of the machine with knowledge discovery techniques. During a CT scan the Intracranial pixel depth is noted. Based on the depth the pixel intensities are normalized using depth-dependent gray level normalization and the dense area of the intracranial region is segmented with the process of region growing and multi-resolution thresholding. The succeeding step is the construction of the decision tree by applying the algorithm. The first step in the method is pre-processing, and it involves the separation of pixels containing the extracranial part and the skull (intracranial pixels). This is used for recognition, measurement, and classification of a hematoma. In the next step the optimal threshold is chosen for hematoma segmentation by human experts and this differs from case to case which leads to instability. In the succeeding step the identification of the largest area containing Intracranial hematoma is done by considering the largest connected hyperdense region and excluding other smaller hyperdense regions. The final step is the construction of a decision tree using algorithm. Limitations: there is difficulty in recognizing subdural and epidural hematomas.

In [3] there is much focus on Cerebral Ischemia and had been said that if cerebral ischemia developed once there might not be a successful or complete cure for the abnormality. The method involved is based on the average thickness of the blood layer and based on thickness, and it is divided into grades such as good, recovery,

dead scaling them as 1, 2, and 3. Computed Tomography (CT) scanned images are given as input to the process. It is described in this process that a thick localized subarachnoid layer of blood or diffusion of blood might lead to delayed cerebral ischemia. This method only focuses on Cerebral Ischemia.

In [4] there is an automatic classification of images into two classes based on features. The two classes are abnormal and normal. This classification involves four phases. Pre-process includes the separation of pixels containing the extracranial part and intracranial part. In feature extraction the original dataset is restricted, and these extracted features serve as training data and saved into feature library. Next, they applied the SVM classifier and KNN classifier. While performing this method it was observed that the K-Nearest Neighbors classifier yielded better results in comparison with the SVM classifier. The postprocessing step is performed after image is classified as abnormal. In the post-processing step the image of the skull is removed and then the abnormal region is extracted. Generally, for removing skull we use brightest pixel cells as the brain matter appears in grey color. This approach categorizes only two classes (abnormal and normal) and gives generic results without any specific label.

In [5, 6] the cerebral microbleed is visualized by using susceptibility-weighted imaging (SWI). Here depending on susceptibility-weighted imaging they constructed rank based average pooling to identify cerebral microbleed. For detection of cerebral microbleed, Convolutional Neural Network is used which contains multiple layers. ReLU layer is used to map the convolution layer with SoftMax activation, tangent activation and Rectified linear unit activation. It is more effective than the convolution layer. So, it is more popular. The convolutional layer may have a greater number of elements in the feature set which will lead to dimension disaster termed as overfitting. In this situation pooling plays a major role that replaces the clusters of related elements in the feature set with statistic summary value. This is done using a pooling function. The last advance is rank-based normal pooling which incorporates a normal of non-zero negative actuations. Rank-based normal pooling can beat the issue of loss of helpful data which is brought about by normal pooling and most extreme pooling. This is a very complex method and has a computational burden.

In [7] the method proposed is a densely connected neural network which is called densenet. To detect cerebral microbleeds the algorithm used is densenet. A sliding window is utilized to cover the arrangement of unique pictures from left to right and through and through. The objective worth is chosen dependent on the focal pixel of the sub models. Later the cost matrix will be employed, based on the comparison between the entries in the cost matrix and target value final abnormality region is detected. Following are the steps included in this algorithm: (1) Traditional Convolutional Neural Network: the layer age of straight enactment, the capacity is done by means of convolutions [8]. The following layer, for example, ReLU layer, is utilized to outline convolution layer with SoftMax enactment, digression actuation, Rectified straight unit initiation, and so on. It is more effective than the convolution layer. So, it is more popular. (2) Densenet: It is used to establish connections between layers. These connections are used for feature maps. The layers that are between the blocks are called transition layers. Each layer receives information from its previous layer. (3) Transfer learning: it is based on fully convolutional multiscale residual

densenets. It considers the labeled samples to get high classification accuracy. It is meant to retrain the later layers of densenet. The usage of this transfer learning leads to an increase in accuracy rate.

In [9, 10] a complete CNN is trained with computed tomography (CT) scans. The algorithm resulted in high accuracy for the detection of acute Intracranial Hemorrhage. This algorithm has shown more accurate results than the measurements that are calculated by 2–4 radiologists.

The method used in [11, 12] is 3D quantitative analysis which performs 3D measurements of the parameters of the Intracerebral Brain Hemorrhage region based on computed tomography (CT) images. This resultant data is correlated with patient mortality. The image segmentation in this is done using a clustering algorithm whose strategy is based on fuzzy c-means which minimizes the objective function that represents the distance of feature vectors from the cluster centers. Based on the features of the clusters rule-based labelling is done whose components include fact-list, knowledgebase (or) rule-base, and inference engine. The labelling is done on the area, the color of the region. This is taken as input for the final step. This method not always perform correct segmentation.

In [13–15] the first step is pre-processing which includes division of the image into four quadrants called segmentation. Pre-processing is done using a tracking algorithm to separate skull images from the Gray dura matter in the CT scan image. Among these four regions the most vulnerable region which has the possibility of abnormality is considered and remaining regions are excluded. The next step is a grouping of all the identical homogenous regions to point out the abnormality. Following are the segmentation methods that are used: (1) Thresholding: it compares the pixel values with the threshold values. (2) Region growing techniques: it extracts useful or connected regions. (3) Supervised and unsupervised method: in this method there is the segmentation of the image in the training stage. Compared to supervised, unsupervised is efficient and less error sensitive. This method is not fully automatic. It requires a lot of human intervention.

3 Proposed Method

The proposed method includes CT scan images as the input datasets. Initially CT Scan images are converted into the format that a network/model can take as input. Images are converted into jpeg format. Later the images are processed so as to analyze the images accurately. Otsu's strategy is utilized with the end goal of segmentation. In general, it is a method to segment an advanced picture into numerous regions bolstered a few rules like arrangements of pixels, and so forth. Noise and unwanted pixels from the image are removed such as the skull part of the image which is of high intensity. The objective of the segment is to disentangle a picture to be increasingly significant and simpler to examine. Numerous methodologies for division exist, for example, thresholding and grouping. This stage sections the cerebrum picture into a few locales with the goal that we can segregate the ROI (the drain district). The result

of this stage is regularly used to distinguish the presence of hemorrhages with most extreme exactness. Later classification of images is done using the Weka tool. Otsu's method of segmentation also minimizes the within-class variance of the system by finding a threshold value automatically based upon the pixel values, probability values of a segment. Morphological operations and region growing techniques can be used to obtain further improvements on the segmented image and finally most filtered part of ROI is obtained and features like the area of ROI, the axis of ROI, the circumference of ROI, distance between skull and ROI are extracted, and they are fed into a neural network through a feature vector pattern. For the extraction of features accurately Region props tool of MATLAB is used. For the purpose of classification between various types of hemorrhages the shape of ROI plays a major role generally for Epidural and Intraparenchymal hemorrhages, there is a convex hull that is present in the region of Interest whereas for subdural hemorrhage it is concave in shape. The Region surrounding the ROI is called Bounding Box, which also plays an important role to extract features of region of interest (ROI). The extracted features from ROI will be stored as input to train the model. The kind of hemorrhage is then recognized based on the neural network which is trained. Once the set of input and output images are saved at that instance then there will be a saved network which is called a Network file. Later working of a network file is determined using a training percentage method.

3.1 Preprocessing

Preprocessing is the beginning step of the method that involves removal of high intensity pixel part of the image such as the skull part and removal of noise. Later the image is segmented into parts so that it becomes easy for the purpose of analyzing and it is done by the OTSU'S method which has 100% of accuracy and clustering of segments with similar properties is made so that identification abnormalities become easy.

3.2 Segmentation

Segmentation step is continued by the feature extraction step which is main step to group segments with similar properties together and then later feed into a neural network. The pixel intensity of the brightest part of the skull is about 250 pixels, and it must be removed in order to identify the part that has abnormality. It is done through MATLAB. Image based cad system is used for the purpose of image segmentation. Later based on the properties of segmented part the type of hemorrhage is detected using a classifier algorithm (Type detection algorithm).

Type Detection Algorithm (Segmentation)

```

1: BTD (Ss, p, q, Cc, W)  $\triangleleft$  Ss CT scan Data set. n count of feature. C classifier
W is vector.
2: Dd  $\leftarrow$  BUILDDATASET1 (Ss, p, q)
3: for c  $\in$  Cc do
4: Acu  $\leftarrow$  TESTINCLASSIFIER (Dd, c, wc)
5: end for
6: return c Acu
7: end
8: BUILDDATASET1 (Ss, p, q)  $\triangleleft$  Ss Scan Dataset. n count of feature
9: Assign Dd to empty
10: for j  $\leftarrow$  0,9 do
11: tt  $\leftarrow$  EXTRACINTFEATURE (j, n)
12: add tt to the end of Dd
13: end for
14: return Dd
15: end
16: EXTRACTINFEATURE (j, n)  $\triangleleft$  i brains CT scan
17: j is  $\geq$  251, j = 0; Apply steps 18 to 23
18: perform Otsu j
19: morphological j
20: regions growing j
21: region props
22: classifier
23: return vi
24: end
25: TESTINCLASSIFIER (Dd, c, wc)
26: partition Dd into 10 subset: Dd0..Dd9
27: for k  $\leftarrow$  0,9 do
28: Ta  $\leftarrow$  Dk
29: Tb  $\leftarrow$  Sk = 0,...,9
30: calculate accuracy
31: return avg of Aj's for j = 0 to 9
32: end for
33: end procedure.

```

3.3 Morphological Techniques

The followings are the morphological techniques used Opening-by-remaking, shutting-by-remaking, Supplement picture, Compute territorial maxima, Superimpose the picture, Process Background Markers, Watershed Transformation and the Segmentation, Imagine the Result, Evacuate Background Noise. Morphological operations basically remove the background and then dilates the inner segment region. An image

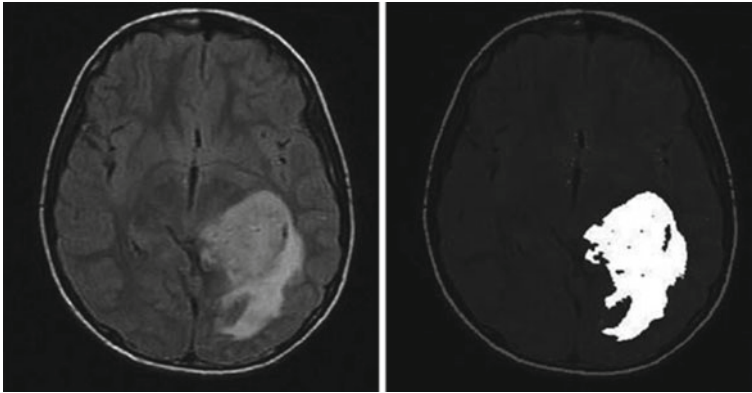


Fig. 2 Haemorrhagic region after applying segmentation and pre-processing techniques

containing potential regions with suspicious mass is the result of morphological operations.

3.4 Obtaining Region of Interest

Figure 2 highlighted images that indicate the hemorrhagic region after applying segmentation and pre-processing techniques. Region growing starts from a single point and then expand from that point by finding the region that is similar to that of the seed region. At the end of this step ‘Region of Interest’ is obtained which is utilized as contribution for ‘Neural Network’ and furthermore as training set. ROI is very crucial for the success of the proposed system. Resultant image after applying region growing techniques will be blank for normal brain image, whereas there will be some portion of image existing for brain image in abnormal condition after application of region growing technique.

4 Experimentation and Results

Experiments were done by changing the parameters given to the model: changing the way the data was split was tried by toggling the value of the shuffle argument—True or False as shown in Fig. 3. The change has not got any specific modifications to the output of the model. This has happened as little number of samples was utilized for performing segmentation. Next by changing the image size to 256X256 and the batch size to 64, the training process a lot more slower compared to when the values were set to 128×128 and 32 which trained the model approximately in 25 min. The training loss for the 256×256 images started learning later because the batch

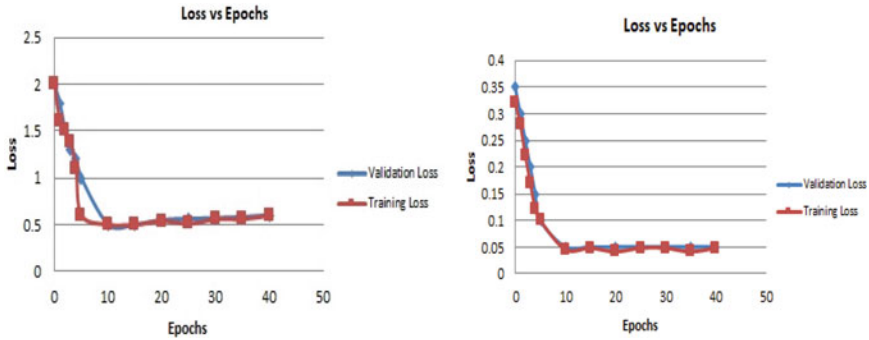


Fig. 3 (a) Loss vs Epoch: for 256X256 image size (b) Loss vs Epoch: for 128X128 image size

size was doubled and further we think it did not reach it. The model trained on 50 epochs but could have been appropriately trained for less number of epochs. This experiment mainly focused on segmentation of the data and localizing the affected area.

5 Conclusion

Automated frameworks for grouping medical images have increased a phenomenal degree of consideration recently. It has considerable impact in recognizing the existence of cerebral hemorrhage (the paired characterization issue) and in the event that it exists classification of hemorrhage is done and also problem of multiclass is also resolved. Although there are some algorithms which detect the hemorrhage efficiently and produce reasonable results, they are having few limitations like few algorithms find it difficult in recognizing subdural and epidural hematomas, few of them cannot detect subarachnoid hemorrhage, few consume more power and some are not suitable for large datasets, having longer computation time. Main cause for these limitations is existence of standardized procedures in less numbers. Since the diagnosis of Brain Hemorrhage is the very complicative and sensitive task, accuracy and reliability are given much priority. The investigations indicate that later pre-handling CT Scans, the double arrangement issue untraveled with 100% exactness. Additionally, the actualized framework accomplished over 92% exactness for the order issue of deciding the discharge type utilizing convolutional neural systems as a classifier. The outcomes are truly promising and more elevated levels of exactness for the characterization issue will be accomplished by getting a vastly improved dataset with high-level goals pictures held legitimately from the CT scanner. Also, unique element extraction and highlight choice procedures could be utilized to improve the presentation of the framework.

References

1. Li, Y., Hu, Q., Wu, J., & Chen, Z. (2009). A hybrid approach to detection of brain hemorrhage candidates from clinical head ct scans. In *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. vol. 1. IEEE.
2. Liao, CC., Xiao, F., Wong, J. M., & Chiang, I. J. (2008). A knowledge discovery approach to diagnosing intracranial hematomas on brain CT: recognition, measurement and classification. In *International conference on medical biometrics*. Berlin, Heidelberg, Springer.
3. Bhadauria, H. S., & Dewal, M. L. (2014). Intracranial hemorrhage detection using spatial fuzzy c-mean and region-based active contour on brain CT imaging. *Signal, Image and Video Processing*, 8(2), 357–364.
4. Mohsen, F., Pomonis, S., & Illingworth, R. (1984). Prediction of delayed cerebral ischaemia after subarachnoid haemorrhage by computed tomography. *Journal of Neurology, Neurosurgery & Psychiatry*, 47(11), 1197–1202.
5. Kyaw, M. M. (2013). Pre-segmentation for the computer aided diagnosis system. *International Journal of Computer Science & Information Technology*, 5(1), 79.
6. Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., & Heng, P. A. (2016). Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE transactions on medical imaging*, 35(5), 1182–1195.
7. Ramteke, R. J., & Monali, K. Y. (2012). Automatic medical image classification and abnormality detection k-nearest neighbour. *International Journal of Advanced Computer Research* 2(4), 190–196.
8. Wang, S., Jiang, Y., Hou, X., Cheng, H., & Du, S. (2017). Cerebral micro-bleed detection based on the convolution neural network with rank based average pooling. *IEEE Access*, 5, 16576–16583.
9. Kuo, W., Häne, C., Mukherjee, P., Malik, J., & Yuh, E. L. (2019). Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proceedings of the National Academy of Sciences*, 116(45), 22737–22745
10. Loncaric, S., Dhawan, A. P., Cosic, D., Kovacevic, D., Broderick, J., & Brott, T. (1999). Quantitative intracerebral brain hemorrhage analysis. In *Medical Imaging 1999: Image Processing* (Vol. 3661). International Society for Optics and Photonics.
11. Kyaw, M. M. (2013). Computer-Aided Detection system for Hemorrhage contained region. *International Journal of Computational Science and Information Technology*, 11–16.
12. Magoulas, G. D., & Prentza, A. (1999). Machine learning in medical applications. In *Advanced course on artificial intelligence*. Berlin, Heidelberg, Springer.
13. Prashanthi, V., & Srinivas, K. (2020). Plant disease detection using convolutional neural networks. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 2632–2637.
14. Hayward, R. D. (1977). Subarachnoid haemorrhage of unknown aetiology: A clinical and radiological study of 51 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, 40(9), 926–931.
15. Prashanthi, V., & Srinivas, K., Generating analytics from web log. *International Journal of Engineering and Advanced Technology*, 9(4), 161–165.

Implementation of Efficient Technique to Conduct DDoS Attack Using Client–Server Paradigm



Seema Rani and Ritu Nagpal

Abstract In modern times, every human being relies upon the internet for fulfilling their hefty needs as the internet offers a vast amount of information to users, so its availability to users is indispensable. Major objectives of security are availability, integrity, and confidentiality. DDoS (Distributed Denial of Service) is a universal cyber-attack that is a major intimidation for cyberspace. DDoS attack slows down network availability by overflowing illegal traffic over network bandwidth. Day by day, attackers improve upon their strategies by using new technologies and techniques. In this paper, a DDoS attack is proposed using python script. We focus on the volumetric DDoS attack effect on the performance of the server ultimately shutting it down. A DDoS attack is undertaken using a python script on the server in which multiple clients send multiple fake requests to the server to slow down the services/performance of the server.

Keywords Distributed denial of service (DDoS) · Denial of service (DoS) · Client · Server · Security · Volumetric DDoS attack · Zombies

1 Introduction

Cybersecurity is the most significant and crucial technology across the world, and social media, the internet, net banking, digital media create new challenges for cybersecurity [1]. Security industry experts, researchers, and related organization personnel are seen analyzing the cyber-attacks across the world and develop different techniques to prevent the cyber-attacks [1].

- Cyber-attack is a sort of outrageous action that exploits computer systems, computer networks, personal devices, and information systems.
- Attacks are being detected every moment, DoS and DDoS have grown over the internet like parasites [2].
- Government, private organizations, banks, schools, hospitals, and people across the world face financial crises due to cyber-attacks [1, 2].

S. Rani (✉) · R. Nagpal
Computer Science & Engineering, GJUS&T, Hisar, India

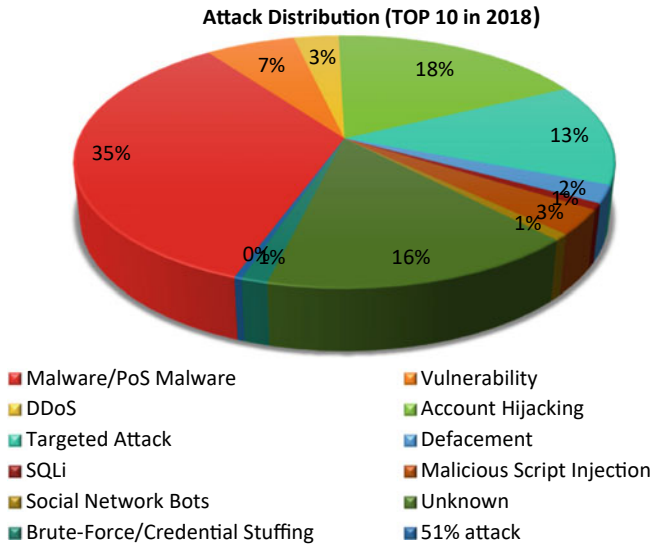


Fig. 1 Top internet attack distribution data

- The primary concern of the internet is security against repudiation, integrity, confidentiality, and availability of information.
- Distributed Denial of Service attack obstruct the legitimate user to access the network services and allows an attacker to access all leverage of services [3].
- It can be conducted by using a single machine or multiple machines named zombies [4].
- DDoS is a distributed version of DoS which is faster and difficult to detect and adverse than DoS [5, 6].

From the Fig. 1, DDoS is a measured attack used by cyber attackers to degrade the performance of the server. In this paper, efficient technique is used to attack the server which is described in the further sections.

2 DDoS: A Major Threat to the IspS

The effect of a fruitful DDoS assault on an ISP is boundless [7, 8]. Site execution is seriously undermined, bringing about baffled clients and different clients. Administration level arrangements (SLAs) [9] are disregarded, bringing about expensive help credits. The developing reliance on the Internet has the effect of fruitful DDoS assaults [10]. DDoS on ISPs [11] brings about the accompanying:

- Lost income
- Lost profitability

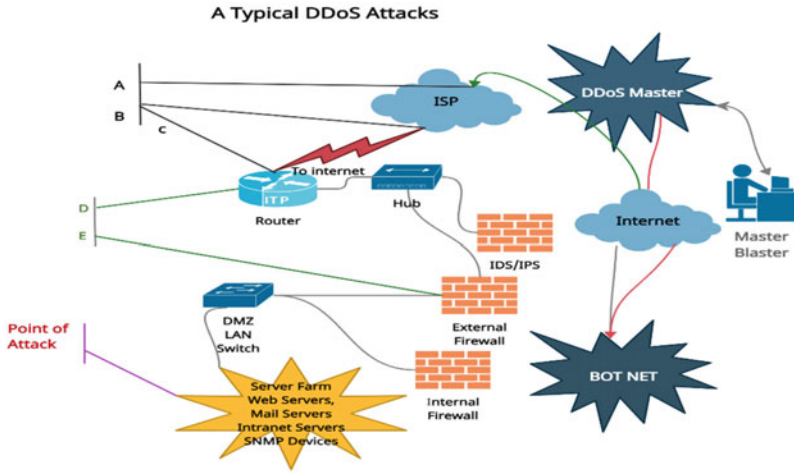


Fig. 2 DoS network

- Increased IT costs
- Mitigation costs
- Loss of clients
- Figure 2 shows how DDoS assault is extended on ISP.

Point A: This is the passage purpose of ISP.

Point B: This is the leave purpose of ISP.

Point C: This is the passage highlight your organization.

Point D and E: This is where Anti-DDoS or Firewalls or your IPS/IDS frameworks dwell.

From the above chart, it is very apparent that DDoS may assault a solitary point in your framework, yet the repercussions are felt from Point B onwards and can be ruined at Point B itself.

The graph in Fig. 3 shows the percentage of worldwide DDoS attack traffic between November 2017 and April 2018, sorted by originating countries. It is shown that during that period, 30% of DDoS attack traffic originated from the United States [12] (Table 1).

3 History of DDoS Attacks

Chowdhury [1], in this paper, discuss the recently occurred cyber-attacks and evaluate the loss of the economy due to increasing cyber-attacks. Denial of service (DoS), SQL injection, and decrypting RSA with Obsolete and Weakened Encryption (DROWN) attack have been discussed. The impact of these attacks on the business, organization, and individual has been defined. Other strategies are software patching,

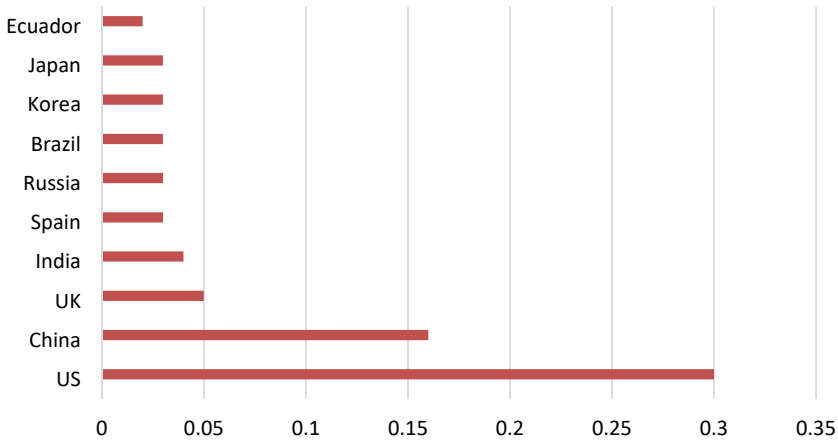


Fig. 3 Chart of various countries originating the DDoS attack

Table 1 Chart of DDoS attack of various countries

Countries	US	China	UK	India	Spain	Russia	Brazil	Korea	Japan	Ecuador
Total (100%)	30%	16%	5%	4%	3%	3%	3%	3%	3%	2%

the digital immune system by minimizing administration privilege using avoidance and detection system, using a firewall, and log file monitoring.

Mirkovic and Reiher [2] have discussed DDoS attack detection and mitigation taxonomy in this paper. The first DDoS attack is conducted using three-phase recruit, infect, and exploit. It proposed a degree of automation manually, semi-automatic, automatic. Describes various host scanning and vulnerability scanning strategies. Source address validity that exploited the weakness to deny service, attack rate dynamics, impact on the victim are explained in detail.

Purwanto et al. [3], in this paper, show how to mitigate and detect flooding DDoS attack. The researcher conducts a flooding attack and detects it by using signature-based and anomaly-based detection techniques. Statics detection, information theory, forecasting, soft computing, signal processing are categories of the anomaly that are discussed.

Sachdeva et al. [4], in this paper, analyze the impact of DDoS attack and defense. NS-2 simulation is used to evaluate DDoS attack conducted at internet-based web application. Quantitatively measure the performance using throughput, response rate, number of requests service to the server with or without attacks.

Li et al. [5], in this paper, discuss the volumetric DDoS attack in IoT (Internet of Things) devices. A DDoS detection method is proposed in three levels of the sliding window, single direction packet filter, and quintile deviation algorithm. An entropy calculator is used for the calculation of optimized slide window time which has a variation updating method and value counter dictionary and counter matrix.

4 DDoS Client/Server Attack

DDoS attacks jam resources at server making them unavailable to the users [13], slowing down the network [14], and making websites inaccessible [15]. We are sending a huge amount of traffic to the server to stop the response from the server. A server-side script response to the client causes a DDoS attack, many requests are forwarded to the server, so that the server stops responding [16–18]. To prevent the attack and allow only authorized users to access the server, a certain amount of time after that, the server is automatically closed.

The DDoS attack is executed using client scripts in which multiple clients send huge miscellaneous data per second and make the server stop working. During the DDoS attack, the server was on the network having IP address 192.56.101.1 on port 23. After the server started, client machines start their execution one by one sending data to the server. When clients send data to the server, it works properly, but as the number of users increases, server performance decreases gradually. Up to the eighth machine, server works and receives data, but at the ninth client machine server completely stopped working, and service was unreachable to the clients during a willful attack against the server and a denial of service attack takes place.

5 Implementation of Techniques to Generate DDoS Attack

5.1 *Number of Attacks Sent*

We generate an attack, the number of attack packets was sent to the victim while generating a DDoS attack. Denoted the time at which attack packets were sent to the server from different zombie [19, 20] machines and attack data saved into the text file for counting the number of attacks sent to the server. It is simply stored in the text file, and by using these files, we got the number of attacks sent to the server by counting the number of lines in the data file. Let us denote the number of attack packets sent to the server as Psuccess.

For experimental purposes, data is collected by sending packets to the server at the start time T_x by all the machines, and after the error time T_y , machines get an error message from the server and the server stops servicing them. Based upon the gathered information, we measure the performance of servers and clients. Approximately 60 packets per second are delivered to the server, and the number of attacks per second is evaluated using three attributes, start time of packet sent, error time, and a number of packet successful packet, after the total attack at the instant is computed using attack per second (Y).

5.2 Attacks Per Second

Attacks per second are evaluated using start time (Tx), error time (Ty), and number of attack packets sent (Psuccess).

Start time of packet sent = Tx, Error time = Ty

Number of a packet sent = Psuccess

Attack per second (y) = $P_{success}/T_y - T_x$

Using the above formula, attacks per second for the first client are calculated as follows: start time is 19:28:38 and error time is 19:29:43; thus, the difference between Ty – Tx is 65 seconds and the number of packets Psuccess is 4152. The value of attacks per second Y calculated is $4152/65 = 63.88s$, and so on.

5.3 Total attacks at instant

The value of total attacks at instant is computed using the value of attacks per second (Y), i.e., Y_n, Y_{n+1}, Y_{n+2} , where n is number of client machine.

$$\sum_{i=1}^n Y_n = X_i$$

where X = Total attacks at instant.

For the first machine, total attack packets/sec is 63.88, and for the second machine is obtained by adding the first and second machine data attacks per second, and for the third machine by adding machine first to third machine attacks per second, and so on.

5.4 Server Performance (Z)

At last, using the computed value of total attacks at instant and attacks per second, server performance is evaluated.

X = Total attacks at instant Y = Attacks per Second Z = Server Performance

$Z = \text{Maximum Attack per Second} / \text{Total Attack Instant} * 100$

$$Z = \frac{Y_i}{X_{initial}} * 100 \quad \text{where } i = 1, 2, 3 \dots n$$

Table 2 Attack traffic from clients to server

Client machine	Start time	Error time	Number of attacks sent	Attacks per second	Total attacks at instant (Server)	Server performance
Machine#01	19:28:38	19:29:43	4152	63.88	63.88	100
Machine#02	19:28:49	19:29:43	3404	63.03703704	126.92	98.6851495
Machine#03	19:29:01	19:29:43	2652	63.14285714	190.06	98.850812
Machine#04	19:29:06	19:29:43	2329	62.94594595	253.01	98.54254544
Machine#05	19:29:14	19:29:43	1806	62.27586207	315.28	97.49352203
Machine#06	19:29:22	19:29:43	1302	62	377.28	97.06165703
Machine#07	19:29:33	19:29:43	604	60.4	437.68	94.55684008
Machine#08	19:29:40	19:29:43	152	50.66666667	488.35	79.3192036
Machine#09	19:29:47	19:29:47	0	0	0.00	0
Attacks faced by server at the time of failure						488.35

Server performance is calculated by dividing the attacks per second by the maximum value of total attacks at an instant and multiply by 100. For example, for the first machine divide 63.88 by 63.88 and multiply by 100, second machine divide 63.037 by 63.88 and multiply by 100, and so on. The attacks faced by the server at the time of failure are 488.35. Table 2 shows the details of attacks by client machines.

6 Results of DDoS Attack Using Client/Server Paradigm

To measure the server performance, a graph was plotted between the server performance and the number of attacks per second. The graph below shows that server performance decreases gradually as the number of machines attacking the server increase, and at particular instant of time server completely stops working. The attacks faced by the server at the time of failure were 488 (Fig. 4).

The volumetric attacks [21] were launched to the server by sending a huge number of packets from multiple machines to the server [22], and this slows down its functioning and finally stops its services. Around 480 packets per second were sent to the server and made it stop working. The individual performance evaluation of each client machine was stored in a dataset as shown in Table 3.

Green, red, and yellow are used to represent different states of the machines. Green shows that the machine is up, red shows the machine is down, and yellow indicates that the machine has yet not started. At 15:48:25 machine #1 and server were running, at 15:48:32 machine #1, machine #2, and server were running. As time increased, at 15:49:02 eight machines were running and the server was down, and

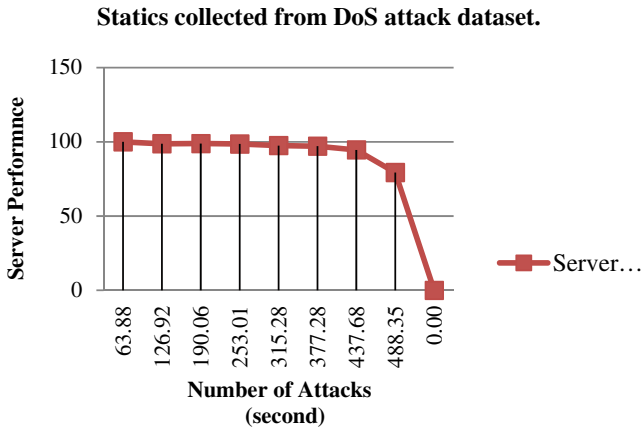


Fig. 4 Statics collected from DoS attack dataset

at 15:49:06 all machines were down. Based on this data, the performance evaluation graph was drawn and is shown in Fig. 5.

The light green line in the graph indicates the server, and the rest of the colored lines show machines, at time 15:49:02 server was down and after few microseconds, all machines were down, at 15:49:04 except machine #8 all machines were down, after two microseconds machine #8 also stopped working.

7 Conclusion and Future Work

In this paper, we computed the server performance that goes on decreasing when a number of zombie machines made bogus requests to the server and after a stipulated duration server completely stopped working. Our results will help in future courses of action to overcome the problem of rapidly increasing DDoS attacks. No doubt that attackers use advanced technologies and techniques to attack the machines, but as security practitioners, we need to think like attackers. To mitigate DDoS attacks, we can reduce the burden on the servers caused by attackers by load balancing and backups. As resources are limited, we must allow access only to legitimate users. We can grant time limits to the clients and after that duration, user will have to make the request again to the server for access to its resources. We can integrate this technique for the mitigation of DDoS attacks.

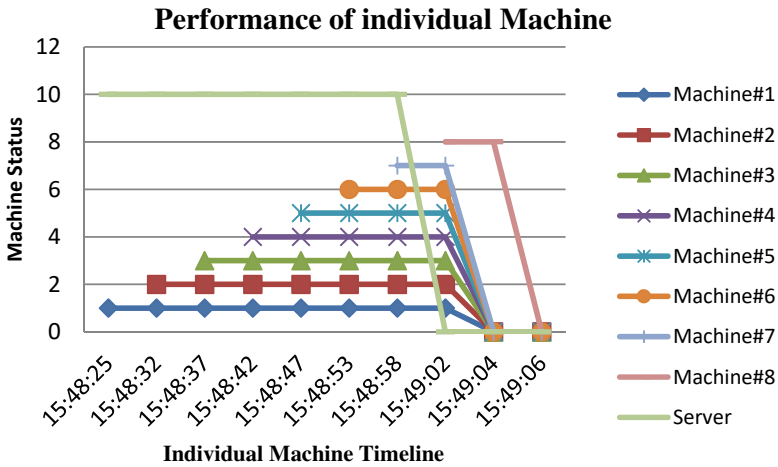


Fig. 5 Performance of individual machine

References

1. Chowdhury, A. (2016). In L. Batten & G. Li (Eds.), *Applications and techniques in information security* (Vol. 651, pp. 54–65).
2. Mirkovic, J., & Reiher, P. (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communication Review*, 34(2), 39.
3. Purwanto, Y., Kuspriyanto, Hendrawan, & Rahardjo, B. (2014). Traffic anomaly detection in DDoS flooding attack. In *8th International Conference on Telecommunication Systems Services and Applications* (pp. 1–6). Kuta, Bali, Indonesia: IEEE.
4. Sachdeva, M., Kumar, K., Singh, G., & Singh, K. (2009). Performance analysis of web service under DDoS attacks. In *IEEE International Advance Computing Conference* (pp. 1002–1007). Patiala, India: IEEE.
5. Li, J., Liu, M., Xue, Z., Fan, X., & He, X. (2020). RTVD: A real-time volumetric detection scheme for DDoS in the internet of things. *IEEE Access*, 8, 36191–36201.
6. Yihunie, F., Abdelfattah, E., & Odeh, A. (2018). Analysis of ping of death DoS and DDoS attacks. In *IEEE Long Island Systems, Applications and Technology Conference* (pp 1–4). Farmingdale, NY: IEEE.
7. Jamshed, M., & Brustoloni, J. (2010). In-network server-directed client authentication and packet classification. In *IEEE Local Computer Network Conference* (pp. 328–331). Denver, CO, USA: IEEE.
8. Guleria, A., Kalra, E., & Gupta, K. (2019). Detection and prevention of DoS attacks on network systems. In *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing* (pp. 544–548). Faridabad, India: IEEE.
9. Bajpai, D., Vardhan, M., Gupta, S., Kumar, R., & Kushwaha, D. S. (2012). In N. Meghanathan, D. Nagamalai, & N. Chaki (Eds.), *Advances in computing and information technology* (Vol. 176, pp. 719–728).
10. Tupakula, U. K., & Varadharajan, V. (2003). Counteracting DDoS attacks in multiple ISP domains using routing arbiter architecture. In *The 11th IEEE International Conference on Networks* (pp 455–460). Sydney, Australia: IEEE.
11. Kumar, K., Joshi, R. C., & Singh, K. (2007). In T. Sobh, K. Elleithy, A. Mahmood, & M. Karim (Eds.), *Innovative algorithms and techniques in automation, industrial electronics and telecommunications* (pp. 235–240).

12. Akamai (<https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/akamai-q2-2016-state-of-the-internet-security-report.pdf>)
13. Choi, J., Choi, C., Ko, B., & Kim, P. (2014). A method of DDoS attack detection using HTTP packet pattern and rule engine in cloud computing environment. *Soft Computing*, 18(9), 1697–1703.
14. Zhijun, W., Wenjing, L., Liang, L., & Meng, Y. (2020). Low-rate DoS attacks, detection, defense, and challenges: A survey. *IEEE Access*, 8, 43920–43943.
15. Carl, G., Kesidis, G., Brooks, R. R., & Rai, S. (2006). Denial-of-service attack-detection techniques. *IEEE Internet Computing*, 10(1), 82–89.
16. Tan, L., Pan, Y., Wu, J., Zhou, J., Jiang, H., & Deng, Y. (2020). A new framework for DDoS attack detection and defense in SDN environment. *IEEE Access*, 8, 161908–161919.
17. Raj, A. G. R., Sunitha, R., & Prasad, H. B. (2020). Mitigating DDoS flooding attacks with dynamic path identifiers in wireless network. In *Second International Conference on Inventive Research in Computing Applications Coimbatore* (pp. 869–874). Coimbatore, India: IEEE.
18. Agrawal, N., & Tapaswi, S. (2019). Defense mechanisms against DDoS attacks in a cloud computing environment: State-of-the-art and research challenges. *IEEE Communications Surveys & Tutorials*, 21(4), 3769–3795.
19. Munshi, A., Alqarni, N. A., & Abdullah Almalki, N. (2020). DDoS attack on IOT devices. In *2020 3rd International Conference on Computer Applications & Information Security* (pp. 1–5). Riyadh, Saudi Arabia: IEEE.
20. Udhayan, J., Hamsapriya, T., & Vasanthi, N. A. (2012). DDoS attack detection through flow analysis and traffic modeling. In V. V. Das, E. Ariwa, & S. B. Rahayu (Eds.), *Signal processing and information technology* (Vol. 62, pp. 89–94).
21. Salim, M. M., Rathore, S., & Park, J. H. (2020). Distributed denial of service attacks and its defenses in IoT: A survey. *The Journal of Supercomputing*, 76(7), 5320–5363.
22. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network anomaly detection: Methods, systems and tools. *IEEE Communications Surveys & Tutorials*, 16(1), 303–336.

Design and Development of Retrieval-Based Chatbot Using Sentence Similarity



Haritha Akkineni, P. V. S. Lakshmi, and Lasya Sarada

Abstract Chatbots or the well-known automated conversational agents have become a raging trend among all the sectors of businesses as a result of the rapid transition happening towards automation in processes. They are already being used extensively and will spread their wings to newer horizons shortly. The basic model of Chatbots is to interact with the user to answer their questions using various modes like text messages, voice replies, or any other predefined suitable interface. This paper discusses the development of a Chatbot for the college, Prasad V Potluri Siddhartha Institute of technology, to answer various questions related to the college like the facilities, procedures, policies, etc. This is a web-based software application implemented using Flask framework. This model is designed to capture text inputs from the user through a console and outputs the response in text format using machine learning concepts. A retrieval approach is implemented to process the input and to respond with an appropriate answer using logic adapters. The performance of this model is analyzed using a questionnaire that uses various parameters like performance, humanity, effect, and accessibility. This paper presents the overall approach used to design the Chatbot and compares the web application as-is study with the to-be website when the Chatbot is incorporated. The web application along with the Chatbot showed a 20% improvement in performance and 5% increase in accessibility by analyzing the performance metrics.

Keywords Chatbot · Machine learning · Retrieval approach · Flask framework

1 Introduction

The usage of Chatbots has skyrocketed in the recent times. They have a strong footing in the customer support industry which involves text-based dealings or support. A chatbot is basically a term for a conversational agent which in actual is a computer system software that can take natural language input and returns a chatty response instantaneously [1]. As observed, chatbots have been in great demand over the past

H. Akkineni (✉) · P. V. S. Lakshmi · L. Sarada
Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, India

few years. The number of internet users is increasing year-by-year. Proportionately, we can also observe a significant rise in the e-commerce sector which in turn demands increased customer support services for these online digital platforms. A delay of even 5 minutes in providing the customer support could reduce the chances of reaching out and selling to a customer [2]. A Chatbot of any particular organization is designed to perform certain tasks pertaining to their requirements. In reality, if the chatbots do not exist, the chances of getting the information across to each and every person of a particular organization (college in this case) are very meek. A Chatbot completes almost 25–30% of the tasks that are done by the organization to the end-user.

The College Enquiry Chatbot uses Machine Learning (ML) algorithms. The chatbot provides simple and trouble-free query clarification platform to students by addressing their queries in text format. The chatbot system makes it easy for the student to clarify his/her queries in lesser time. It is used to perform tasks like answering the questions related to the organization, supply the relevant links for the questions encountered by the user (if necessary), etc.

Basically, the initial chatbot named “ELIZA” was designed by Joseph Weizenbaum. With the innovation of the existing Chatbots, there came the evolution of the Chatbot technology. Chatbots can be implemented based on two approaches; Rule-Based Approach and Self-Learning Approach. In rule-based approach, there are some specific rules to be followed when training the bot. Here, the chatbot cannot answer complex questions triggered by the user. In self-learning approach, the chatbot follows the machine learning approach. This approach can further be divided into two models; Retrieval-Based Model (Chatbot retrieves the best response from the collection of responses) and Generative Model (Chatbot answers the question asked by the user from the set of answers). Chatbots that implement generative models are termed as intelligent chatbots.

The main sequence of flow for a functional chatbot involves three stages: (i) to get natural language user input, (ii) generate most appropriate automated response, and (iii) returning prudent natural language output. This College Enquiry Chatbot is developed using the Python Flask framework, where the front-end is written in HTML, CSS, and internal JavaScript. To implement the working of College Enquiry Chatbot, Chatterbot Library is used. Chatterbot is language independent and makes it easy to design a product that takes part in conversation. The inspiration to build this chatbot came when going through our college’s website. This Chatbot follows the generative approach, which answers the user from a predefined set of answers. This Chatbot functions similar to the initial chatbot “ELIZA.” Section 2 discusses the motivation behind developing this chatbot. Section 3 talks about the related work done in this area. Section 4 is about the proposed approach and Sect. 5 shows the implementation part and results. Section 6 is the conclusion part.

2 Motivation

For every student, to get the information related to the fee structure, due fees, events that are being organized, etc., is a lengthy and difficult process most of the times. Some of the events that are being organized by one department are not known to the other departments due to lack of effective communication channels. Sometimes, the changes in the fee structure are often not known to many students. We are living in a world where people are striving towards increased automation rather than manual intervention for repetitive tasks to arrive at faster and simpler solutions for the complex problems. This is the main reason behind building the College enquiry Chatbot where the user can ask the queries and a bot responds within no time. It reduces the users' time and can be easily available to any user with a mobile phone as well. With this Chatbot, the student no longer needs to take trips to notice boards or contact his/her friends or faculty to get timely information. This Chatbot in turn reduces the work of the organization. The user can get answers for any questions related to the college like address, admission process, and subsequent procedures which may be very difficult to find out for many students due to the pandemic situations.

3 Related Work

Harsh Pawar et al. developed College Enquiry Chatbot using Knowledge Database. This System is built using LUIS.ai, Microsoft Bot builder, and MongoDB for database. The user can interact with this system through web interface, in which the web application is connected to the bot through bot connector. This system will be a chatbot that responds to the questions by the user. In this system, they used LUIS.ai for training language model, which is used to identify the intention of the user and fetches the adequate response using pattern matching algorithm [3].

Nidhi Mishra et al. developed Dr. Vdoc, a medical Chatbot in 2018 which acts as a virtual doctor. The interaction can take place between a patient and a virtual chatbot, and this is made possible using natural language processing and pattern matching algorithms. It is more useful to provide awareness on the disease in case of unavailability of the doctors. The results showed that Dr. Vdoc is capable of returning the responses with 80% accuracy [4].

Pooja Prashanth et al. designed a college enquiry chatting system using knowledgeable database. In education system, few works can be very lengthy, time consuming, and require extra manpower. But in today's world, we see that almost everything is turning to digital format. This chatbot updates the students with college's cultural activities. It is based on client server architecture. All the information is kept in the database on the central server. By installing application in the smart phone, a user can access the information from the database. This system is

developed with the combined technologies of Artificial Intelligence, Knowledgeable Database, and Virtual Assistance, such that the chatbot can make conversations between humans and machine [5].

Sagar Pawar et al. developed a web-based college enquiry chatbot with results. The system uses bigram and some sentence similarity algorithms to provide answers to student or any other user queries. If the response is invalid or irrelevant to the user query, then user can mark the response as invalid. This is sent to the admin. The admin will be able to delete or modify the data [6].

Jayesh Gangrade et al. designed a Review on College Enquiry Chatbot in 2019. An intelligent voice recognition chatbot is designed and developed which replies to the user queries in a graphical user interface, which gives them a feel of talking to a person. In this system, the user must login into the web application to ask queries. This system also contains a notice board where the important notices are placed in the form of documents or portable document formats (pdf). Here, important keywords are fetched from the query, and those keywords are searched in the knowledge database using keyword matching algorithms. If the answer is not found, then the system displays the following—“Answer to this query is not available at the moment, please revert after some time” [7].

4 Proposed Approach

The proposed system is designed to solve the user queries in no time without the user approaching the college or contacting any other person. A retrieval-based model uses an archive of predefined responses and performs some analysis to pick a desired response based on the given input and text [8]. This analysis can be a simple expression match, or Machine Learning classifiers that work on the principle directed towards flows of graphs. The systems that are designed using this model choose an appropriate response from a fixed set instead of returning a new text. The chatbot is trained based on existing information in the database to provide best possible responses from the predefined responses. In case of retrieval-based chatbot [9], the selection of appropriate responses can be determined based on techniques like machine learning, deep learning, and keywords matching. This model cannot generate new output and can only provide predefined responses. Retrieval-based chatbot design is shown in Fig. 1.

User Query: The students can approach the Chatbot to resolve their queries related to the College. This Chatbot receives the queries that are of any type (Greetings, College related queries, General questions, etc.) by the user in the form of text. The user inputs his queries in the textbox.

Tokenizing the words: Tokenization is a process of breaking a sentence or the sequence into pieces called tokens, perhaps eliminating the special characters like punctuation marks. For example, if the input is “Hello, How are you?,” then the tokens obtained after tokenization are “Hello,” “How,” “are,” and “you.” After tokenizing,

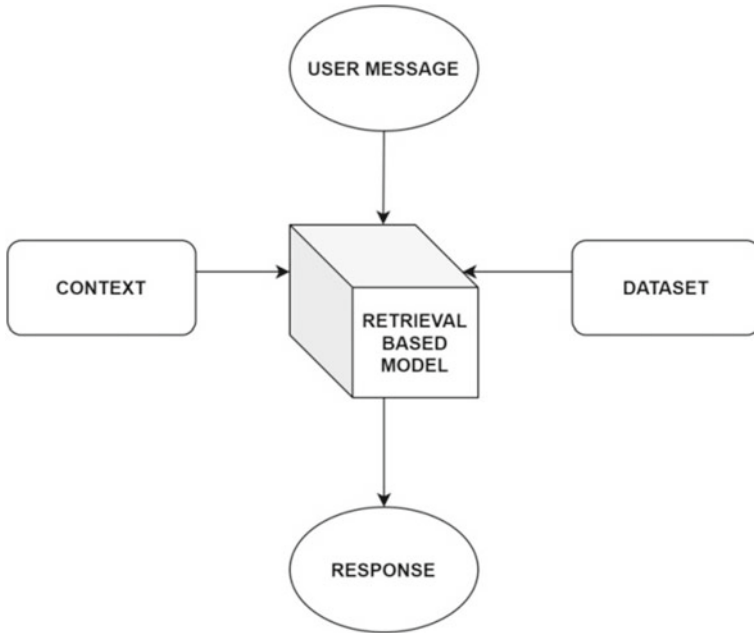


Fig. 1 Retrieval-based chatbot design

the tokens are processed by normalizer and matcher. Tables 1 and 2 depict how the normalizer and matcher process the tokens.

A normalizer processes the tokens by matching them with the patterns stored in the dataset. It matches the keywords of the input sentence with the patterns of the dataset and stores the matched keywords. A Matcher is used to find out the appropriate output

Table 1 Processing of the normalizer

S. No	Input	Output of the normalizer
1	I want	I want
2	to talk	to talk
3	in Hindi	in Hindi

Table 2 Processing of matcher

S. No	Input sentence	Pattern in database	Similarity threshold	Output sentence
1	I code in python	I prefer to code in python	0.5	The output will be “I prefer to code in python”
		I like coding in python	0.33	

for the user query from the existing dataset. The Matcher checks by converting text into ASCII and checks the sensitivity of the characters (lower/upper case).

Retrieving the matched response: Upon finding a sentence similar to the query from the predefined dataset, the chatbot retrieves an appropriate response from the same [10].

Calculating the similarity: The similarity threshold is calculated for the input query, out of many sentences of the dataset that match with the input. The higher threshold valued sentence is sent to the user.

For example, Input—"I code in python" and the dataset has two related sentences, Sentence 1—"I love to code in python" and Sentence 2—"I code in python language."

The similarity is checked by calculating the threshold of the input query with each of the two sentences in the dataset. The similarity threshold of the "Sentence 2" is higher than that of the "Sentence 1." So, the appropriate answer for the "Sentence 2" is returned to the user.

The intersection between two sentences, one which is the user query and the other is the pattern in database, gives the sentence similarity score. It is represented by $S1 \cap S2$ and $S2 \cap S1$, where $S1$ and $S2$ are two sentences.

The formula to calculate sentence similarity score is shown in Eq. (1).

$$Count(S1 \cap S2) \cup Count(S2 \cap S1) / Count(S1) \cup Count(S2) \quad (1)$$

Sample user input statement

$S1 =$ "I code in python." and the pattern in the database as: $S2 =$ "I prefer to code in python."

Using the similarity algorithm, we get:

$S1 = \{I \text{ code, code in, in python}\}$ $Count(S1) = 3$

$S2 = \{I \text{ prefer, prefer to, to code, code in, in python}\}$ $Count(S2) = 5$

Now, $S1 \cap S2 = 2$ and $S2 \cap S1 = 2$.

Calculated score for Sentence Similarity

Let $a = Count(S1 \cap S2) \cup Count(S1 \cap S2) = 2 \cup 2 = 4$.

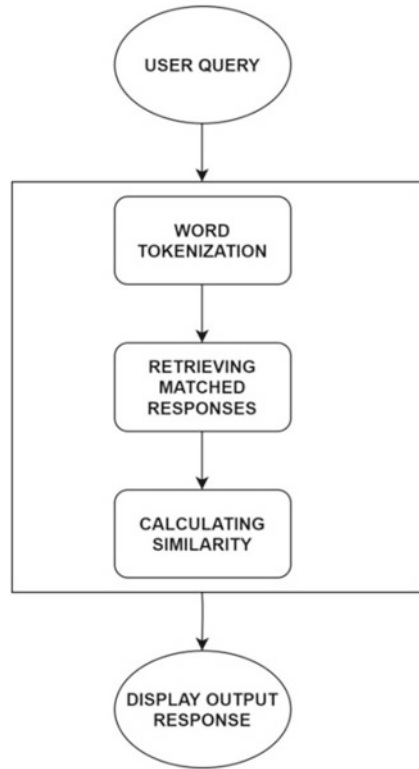
Let $b = Count(S1) \cup Count(S2) = 3 \cup 5 = 8$.

Similarity score is $a/b = 0.5$.

This procedure is adopted to calculate the sentence similarity score for each pattern stored in the database. The output for the users input query is displayed in the designed College enquiry Chatbot web app.

The proposed system design is shown in Fig. 2 which depicts the procedure of how the system is developed.

Fig. 2 College enquiry chatbot system design



5 Results and Discussion

This system follows three main steps as shown in Fig. 3, i.e., Get Input, Process Input, and Return Response.

Get Input: In the first step, the user asks the Chatbot a question from any source (sources can be screen, speech recognition, or API). In this project, the user uses a console to converse with the Chatbot [11]. The initial message is displayed as “Hi, Welcome to PVPSIT Chatbot. I am at your service. You can ask me questions!!”

Process Input: The input given by the user is processed by the logic adapters. The logic adapters are those which are used to decide the logic for the Chatterbot to select a response for the input asked by the user. Any number of logic adapters can be used in the project. There are two logic adapters used in this paper; “Mathematical Evaluation and “Best Match.” The bot gives back the response of the logic adapter that has higher confidence value. If any two logic adapters have the same threshold/confidence, then the first logic adapters’ response is returned to the user. One can set the parameters for the logic adapters like setting the default response, threshold, paths, etc. The “Best Match” logic adapter responds the user with the confidence level of 0.70, and when the bot does not find the best match for the input, it gives the default response

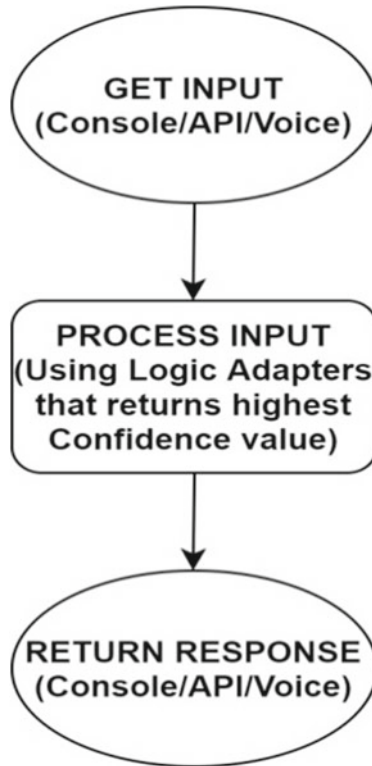


Fig. 3 Flow chart diagram of college enquiry chatbot

saying “I am sorry, but I do not understand. I am still learning.” Figure 4 gives the responses after processing the input data by the Chatbot.

Return Response: The bot returns the response through the console. The bot in this project is trained with the data related to the college and English language. This Chatbot can also respond to the input of various kinds like mathematical calculations, sports, emotions, and simple conversations. Figure 5 depicts the logic of how training is performed.

A questionnaire was constituted of various questions related to performance levels, humanity, effect, and accessibility, and was distributed among the students of the department. A sample of 240 responses was taken and analyzed, and the results conclude that the performance of the website improved by 20% when the chatbot is included. Considering humanity and affection, when compared to the existing one, it reached an average level and a 5% improvement over the previous version. The statistics are shown in Table 3.

Performance, Humanity, Effect, and Accessibility were the parameters that were considered to analyze the performance of the bot. A questionnaire was used for the comparative analysis [12]. This helps in designing more effective bots. To assess

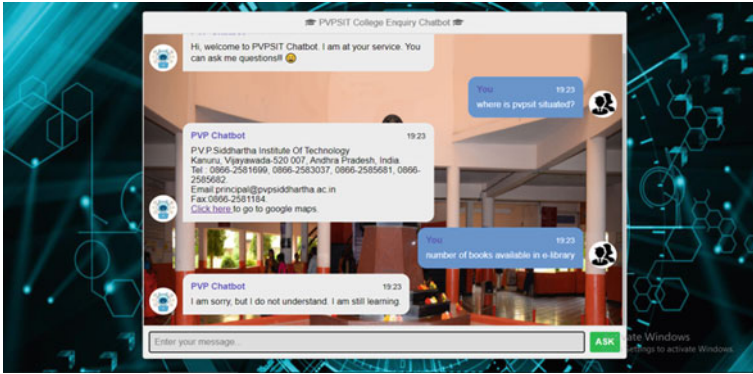


Fig. 4 Responses after processing the input data by the chatbot

```
# Prepare the training data
with open('training_data/personal_ques.txt', 'r') as fd:
    conv = []
    last = ''
    for line in fd:
        # Remove newline and whitespaces only from the end of the line
        line = line.rstrip()
        if line.startswith(' '):
            last += line
        else:
            if last:
                conv.append(last)
            last = line
    # Add the final string... last will be '' only if the file is empty
    if last:
        conv.append(last)

# Training with Personal Ques & Ans
trainer=ListTrainer(chatbot)
trainer.train(conv)

# Training with English Corpus Data
trainer_corpus = ChatterBotCorpusTrainer(chatbot)
trainer_corpus.train(
    'chatterbot.corpus.english'
)
```

Fig. 5 Training the chatbot

Table 3 ChatBot quality assessment

Category	Quality attribute	Metric	Old system	New system
Performance	Robustness to unexpected input	% of success	60%	80%
Humanity	Able to maintain discussion	Low to high	Low	Low
Affect	Provides greetings	Low to high	Low	Average
Accessibility	Can detect meaning and intent	% of success	70%	75%

the performance questions related to robustness, an unexpected input was questioned on. After analysis, it was evident that respondents stated that there is a clear improvement of 20% over the previous version. Coming to the category of humanity, questions based on maintaining discussion was assessed, and there was a significant improvement from average to high [13]. Coming to the effect, questions on providing greetings were assessed. There is no specific change noted in both the versions [14]. Coming to the accessibility, there was improvement by 5% which is not quite noteworthy. This paper presents the overall approach used to design the chatbot, and compares the web application as-is study with the to-be website when the chatbot is incorporated.

6 Conclusion

In this paper, a machine learning approach for College Enquiry Chatbot has been proposed. The responses of the Chatbot are obtained with the help of a “Retrieval-Based” approach. This Chatbot can perform almost 30% of the entire task of the college. This makes the work easy and provides results within seconds. This proposed chatbot is designed in python using “Flask Framework.” The chatbot mainly follows three steps: Getting input queries from the user, processing the input, and returning the responses to the user. The methodology of the Chatbot is that it takes the input from the user, tokenizes the input, retrieves the matched response, calculates the similarity score, and then returns the appropriate response to the user. The results of to-be website when Chatbot is incorporated and the as-is website are compared based on the performance, humanity, effect, and accessibility. Of all those tests, the website that incorporates the chatbot shows better performance, humanity, effect, and accessibility. Thus, implementing chatbot will reduce the work of the organization. Some of the limitations of the chatbot are domain knowledge, not possessing the personality.

References

1. Schachner, T., Keller, R., & von Wangenheim, F. (2020). Artificial intelligence-based conversational agents for chronic conditions: Systematic literature review. *Journal of Medical Internet Research*, 22(9), e20701. <https://doi.org/10.2196/20701>
2. Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems—CHI '17* (pp. 3506–3510).
3. Pawar, H., Prabhu, P., Yadav, A., Mendonca, V., & Lemos, J. (2018). College enquiry chatbot using knowledge in database. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. ISSN: 2321-9653.
4. Mishra, S. K., Bharti, D., & Mishra, N. (2017), Dr. Vdoc: A medical chatbot that acts as a virtual doctor. *Research & Reviews: Journal of Medical Science and Technology*, 6(3).

5. Prashant, B. P., Anil, M. S., Dilip, K. M. (2018). Online chatting system for college enquiry using knowledgeable database. *International Journal of Engineering Research in Electronics and Communication Engineering*, 5.
6. Pawar, S., Rane, O., Wankhade, O., Mehta, P. (2018). A web based college enquiry chatbot with results. *International Journal of Innovative Research in Science, Engineering and Technology*, 7(4).
7. Gangrade, J., Surme, S. S., Somu, S., Raskonda, S., Gupta, P. (2019). A review on college enquiry chatbot. *International Journal of Engineering Science and Computing*, 9(3).
8. Vamsi, G. K., Rasool, A., & Hajela, G. (2020). Chatbot: A deep neural network based human to machine conversation model. In *11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1–7). Kharagpur, India. <https://doi.org/10.1109/ICCCNT49239.2020.9225395>
9. Wu, Li, Z., Wu, W., & Zhou, M. (2018). Response selection with topic clues for retrieval-based chatbots. *Neurocomputing*, 316, 251–261.
10. Setiaji, B., & Wibowo, F. W. (2016). Chatbot using a knowledge in database: Human-to-machine conversation modeling. In *7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)* (pp. 72–77). Bangkok. <https://doi.org/10.1109/ISMS.2016.53>
11. Khan, R., & Das, A. (2018). Introduction to chatbots. In *Build better chatbots* (pp. 1–11). Berkeley, CA: Apress.
12. Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. In: *Proceedings of the 2019 7th International Conference on Information and Education Technology (ICIET 2019)* (pp. 111–119). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3323771.3323824>
13. Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*, 176(5), 619–625.
14. Seeger, A. M., & Heinzl, A. (2017). Human versus machine: Contingency factors of anthropomorphism as a trust-inducing design strategy for conversational agents. In *Lecture notes in information systems and organization* (Vol. 25, pp. 129–139). Cham: Springer.

A Pragmatic Study on Movie Recommender Systems Using Hybrid Collaborative Filtering



Akhil M. Nair and N. Preethi

Abstract The Movie Recommendation System (MRS) is part of a comprehensive class of recommendation systems, which categorizes information to predict user preferences. The sum of movies is increasing tremendously day by day, and a reliable recommender system should be developed to increase the user satisfaction. Most of the approaches are made to prevent cold-start, first-rater drawbacks, and gray sheep user problems, nevertheless, in order to recommend the related items, various methods are available in the literature. Firstly, content-based method has some drawbacks like data of similar user could not be achieved, and what category of these items the user likes or dislikes are also not known. Secondly, this paper discusses about collaborative filtering to find both user and item attributes that have been considered. Since there exist some issues pictured with collaborative filtering, so this paper further aims into hybrid collaborative filtering and deep learning with KNN algorithm of ratings of top K-nearest neighbors.

Keywords Collaborative filtering · Deep learning · Cold-start · First-rater · Gray sheep · KNN algorithm

1 Introduction

Nowadays, every online site provides Recommendation System for gaining users' interest in their websites or applications [1]. It is a very big challenge for the users to search for their kind of movies in this huge catalog. Recommendation Systems help in saving time and energy for both users and owners by creating codes to ensure user interest is always prioritized and they are offered similar products or items. These recommendation systems analyze our shopping, watching, listening patterns, and make predictions based on our past activities. The most basic model

A. M. Nair (✉) · N. Preethi
CHRIST (Deemed To Be University), Bengaluru, India
e-mail: akhil.nair@christuniversity.in

N. Preethi
e-mail: preethi.n@christuniversity.in

of the recommendation system is collaborative filtering which tends to assume that users may like items that are liked by similar users. This development has led to the collection of a humungous amount of data that may be difficult to be dealt with individually, so groups of similar interests and users are constructed in advance, to tackle individually [2]. This method of collaborative filtering helps to solve the problems and drawbacks of content-based recommendations.

2 Literature Review

The collaborative filtering method uses implicit or explicit feedback from the users [1]. Collaborative filtering involves two techniques: the memory-based approach and the model-based approach [3]. The memory-based approach finds user similarity using cosine similarity or Pearson correlation coefficient and takes weighted average ratings of users, whereas in the model-based approach machine learning is used to find user ratings of unrated item instances using Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Neural Network, and Deep-Matrix Factorization [4–6]. In this paper, we are using a memory-based approach which helps in the easy creation and elucidation of results. Figure 1 shows the architecture of a collaborative filtering-based recommender system. The system assumes a user to be x . It creates a set N of other similar users whose preferences match with the user x . Finally, the system predicts the x 's rating based on the ratings provided by x 's, similar users. Figure 1 shows a basic architecture of such recommender system.

The memory-based collaborative filtering techniques can be further classified into two: user-item filtering and item-item filtering [7]. In user-item filtering, a target user is chosen, and users similar to the target user are discovered based on the user preferences and feedback. Item-item filtering takes a target item and user ratings of

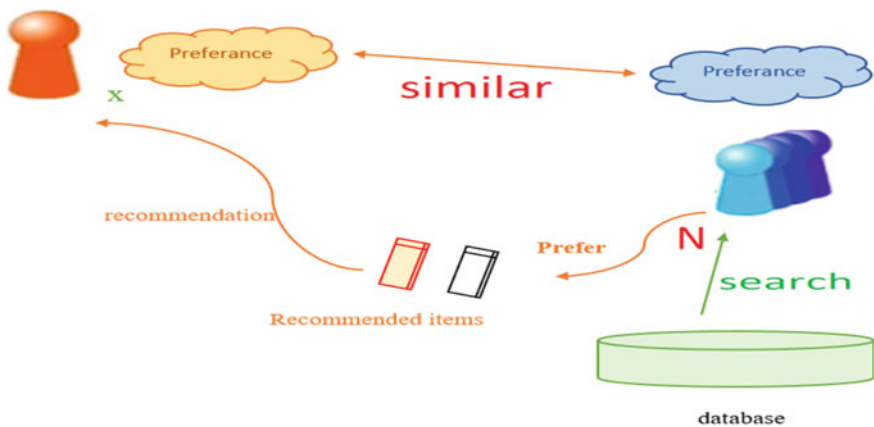


Fig. 1 Architecture of collaborative filtering system

that item are grouped together. There are certain disadvantages in the collaborative filtering method like Cold-Start, Sparsity Problems, First-Rater, Popularity Bias, etc. The complexity in finding K-most similar users or items is high and is given by $O(|U|)$, Where $|U|$ is the size of the utility matrix [8, 9]. To overcome these disadvantages in CF, hybrid CF-method and Deep Learning are considered in this paper.

To overcome the limitations of CF, a global baseline is combined with CF [6]. The negative predictions achieved by the system are corrected by averaging the distance from the baseline predictor. It also helps to extend collaborative filtering to large user bases. Since the difference between the baseline estimate and the rating matrix is evaluated and thus similarities can be applied to a normalized matrix [10]. While the ratings are limited to be non-negative, it is noted that certain ratings average together and compute the predictions to be negative. The baseline estimator for the weighted averaged is given by Eq. 1.

$$r_{xi} = b_{xi} + \frac{\sum_{j \in N(i:x)} S_{ij} * (r_{xj} - b_{xj})}{\sum_{j \in N(i:x)} S_{ij}} \tag{1}$$

- where $b_{xi} = \mu + b_x + b_i$
- $b_x =$ rating deviation of user $x = (avg \cdot rating \ of \ user) - \mu$
- $\mu =$ overall mean movie rating
- $b_i =$ rating deviation of movie.

Equation 1 specifies the baseline estimator for the prediction. To start with similarity s_{ij} of items i . This provides the most similar items I that are rated by the user x . The rating r_{xi} is then estimated as the weighted average. The predictions are evaluated using Root-Mean-Square-Error (RMSE).

Evaluating predictions: Root-Mean-Square-Error (RMSE) is calculated by Eq. 2.

$$\sqrt{\frac{\sum_{(x,i) \in T} (r_{xi} - r_{xi}^*)^2}{N}} \tag{2}$$

where $N = |T|$, r_{xi} is the predicted rating.

3 Proposed Methodology

This study uses the MovieLens10M dataset obtained from the movie lens database (<https://grouplens.org/datasets/movielens/latest/>). We exercised the two original files, namely, movies.csv and ratings.csv with records containing 9762 rows and 100,836 rows, respectively. The dataset consists of the ids of users, movies, and ratings of the movies provided by the users. The study evaluates the results of the proposed method with that of state-of-the-art collaborative filtering approaches. A cosine similarity metric is applied for one dimension, which is “ratings” to find

Table 1 Meta data about movie data set

S. No	Parameter	Numbers of instances
1	Users	610
2	Movies	9742
3	Ratings	1,00,836
4	Rating scale	5 Star (0.5 stars to 5 Star)

the similarity among the users. K-nearest neighbour (KNN) finds the n nearest and similar users as per the preferences and feedback. KNN builds a neighborhood of K similar users concerning the target users rendering to a similarity measure which is evaluated from the user ratings and other parameters. Item-item-based system is tested with KNN using both cosine and Pearson similarity measures. The value of k for the number of neighbors was set to 20 although the results were quite similar for both higher and lower values. The study was conducted on two types of recommendation algorithms that are collaborative filtering and neural networks using KNN. The proposed method is evaluated using the metrics shown in Table 1.

Neural networks are considered to be black-box algorithms, and the interpretation of their weights is done in an insightful way without the application of specific visualization techniques. The ratings and its frequency of occurrences is shown in Fig. 2.

Table 2 displays the KNN Methods for the number of movies which contains training set 6819 and testing set 2923 for k = 1, k = 3, and neural network (NN). The Accuracy, Recall, Precision, and F-measures are shown for various methods. For the size K = 1, the accuracy is 86.25%, Recall is 0.872, Precision is 0.853, and F-measures is 0.873. When k = 2, the accuracy is 91.2%, Recall is 0.926, Precision

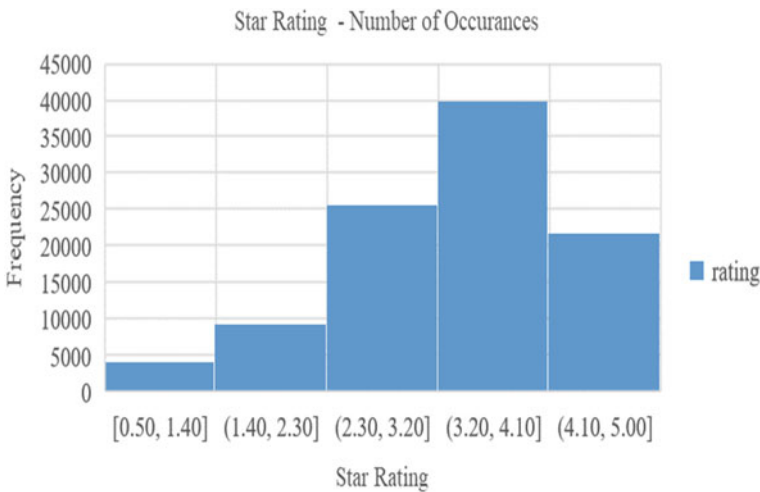


Fig. 2 Star rating—frequency of occurrences

Table 2 Performance of NN, KNN K = 1, and K = 3 classifiers

Method	Data set—number of movies		Accuracy (%)	Recall	Precision	F-Measure
	training set	Testing Set				
KNN K = 1	6819	2923	86.25	0.872	0.853	0.873
KNN K = 3	6819	2923	91.2	0.926	0.932	0.912
NN	6819	2923	94.6	0.943	0.932	0.964

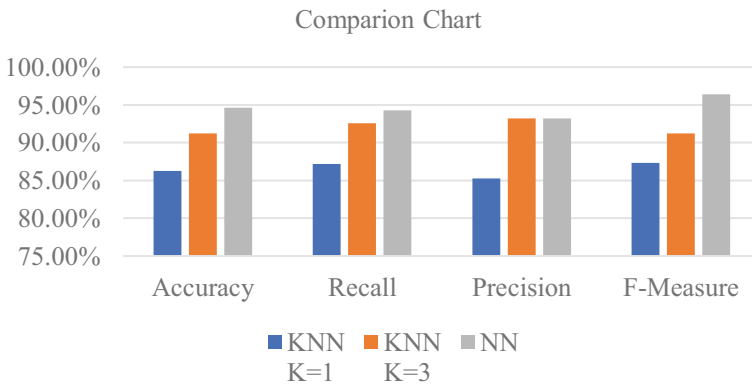


Fig. 3 Assessment graph for the NN, KNN K = 1, and 3 classifiers

is 0.932, and F-measures is 0.912, whereas in a neural network, the accuracy level is 94.6%, Recall is 0.943, Precision is 0.932, and F-measures is 0.964 which gives higher accuracy when compared with normal KNN method. The Accuracy, Recall, Precision, and F-measures values for the KNN and NN methods are shown clearly in the following Table 2 and Fig. 3.

4 Conclusion

Movie recommendation is one of the topics that is increasing research interest. Nevertheless, to convey relevant recommendations, it is required to deal with various recommendation methods to extract the information from movies. This study proposes a hybrid recommendation model which integrates collaborative filtering with deep learning to precise the cold-start problem, sparsity problems, first-rater, and popularity bias of collaborative filtering recommendation models. The proposed study evaluates the model with respect to RMSE on a movie dataset. We have implemented both the models with the specified dataset. It is clear from the results that

proposed model outperforms the simple model, which uses the average rating of users for rating prediction of various items. Also, the hybrid model outperforms the global baseline model for rating prediction of items with few ratings. The proposed model shows that the hybrid model is effective on rating predictions. A hybrid recommendation technique will eventually be adopted by the recommender systems community.

References

1. Sánchez-Moreno, D., Gil González, A., Muñoz Vicente, M., López Batista, V., & Moreno García, M. (2016). A collaborative filtering method for music recommendation using playing coefficients for artists and users. *Expert Systems with Applications*, 66, 234–244.
2. Arnold, A. N., & Vairamuthu S. (2019). Music recommendation using collaborative filtering and deep learning. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(7). ISSN: 2278-3075
3. Suganeshwari, G., & Syed Ibrahim S. P. (2016). A survey on collaborative filtering based recommendation system. In V. Vijayakumar & Neelamarayanan V. (Eds.), *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC—16')* (Vol. 49). Smart Innovation, Systems and Technologies. Cham: Springer.
4. Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, 69, 29–39.
5. Park, Y., Park, S., Jung, W., & Lee, S. (2015). Reversed CF: A fast collaborative filtering algorithm using a k-nearest neighbour graph. *Expert Systems with Applications*, 42(8), 4022–4028.
6. Sunitha, M., & Adilakshmi, T. (2018). Music recommendation system with user-based and item-based collaborative filtering technique. In G. Perez, K. Mishra, S. Tiwari, & M. Trivedi (Eds.), *Networking Communication and Data Knowledge Engineering* (Vol. 3). Lecture Notes on Data Engineering and Communications Technologies. Singapore: Springer.
7. Segaran, T. (2007). *Programming collective intelligence: Building smart web 2.0 applications*. O'Reilly Media.
8. Herrada, O. C. (2009). *Music recommendation and discovery in the long tail*. Ph.D. thesis, Universitat Pompeu Fabra.
9. Burke, R. (2007). *Hybrid web recommender systems*. The Adaptive Web (pp. 377–408). Springer.
10. Pottie, M., & Chabbert, M. (2009, August). *The pragmatic theory solution to the Netflix Grand Prize*. Netflix Prize Report. Retrieved February 5, 2017, from http://www.netflixprize.com/assets/GrandPrize2009_BPC_PragmaticTheory.pdf

An Anatomization of FPGA-Based Neural Networks



Anvit Negi, Devansh Saxena, Kunal, and Kriti Suneja

Abstract Ongoing advancements in the improvement of multilayer convolutional neural organizations have brought about upgrades in the precision of important recognition jobs, for example, huge category picture classification and cutting-edge automated recognition of speech. Custom hardware accelerators are crucial in improving their performance, given the large computational demands of Convolution Neural Networks (CNN). The Field-Programmable Gate Arrays (FPGAs) reconfigurability, computational abilities, and high energy efficacy makes it a propitious CNN hardware acceleration tool. CNN have demonstrated their value in picture identification and recognition applications; nonetheless, they require high CPU use and memory transmission capacity tasks that cause general CPUs to neglect to accomplish wanted execution levels. Consequently, to increase the throughput of CNNs, hardware accelerators using Application-Specific Integrated Circuits (ASICs), FPGAs, and Graphic Processing Units (GPUs) have been employed to improve CNN performance. To bring out their synonymity and dissimilarity, we group the works into many groups. Thus, it is anticipated that this review will lead to the upcoming development of successful hardware accelerators and be beneficial to researchers in deep learning.

Keywords FPGA · ASIC · Deep learning · Neural net

1 Introduction

In the past decade, artificial intelligence (AI) and machine learning (ML) instruments have gained considerable prominence due to advancements in computational structure which consists of area, power, and effectiveness. A range of programs have started using AI algorithms to improve overall efficiency than conventional methods. These applications include image processing [1], for example, face identification,

A. Negi · D. Saxena (✉) · Kunal · K. Suneja
Department of Electronics and Communication Engineering, Delhi Technological University,
New Delhi, India

K. Suneja
e-mail: kritisuneja@dtu.ac.in

banking and statistical surveying [2], mechanical arms in the robotized producing area [3], medical services applications [4], database administration and management [5], and face checking and investigation security applications [6].

The explosive development of big data over the past decade has inspired revolutionary approaches to obtain data from various sensors such as photos and voice samples. In reality, these Convolution Neural Networks (CNNs) [7] now are conceived as the standard method among the proposed methods by delivering “human-like” accuracy in various computer vision-related applications, such as classification [8], detection, segmentation [9], and speech recognition [10].

As CNNs need up to 38 GOP/s to identify a single frame [11], this output is obtained at the expense of a high computational price. Hence, to accelerate their execution, dedicated hardware is required. The most frequently used platform for implementing CNNs is Graphics Processing Units (GPUs), as they provide the highest performance with respect to computational throughput, hitting 11 TFLOP/s [12].

Nonetheless, FPGA systems are considered to be better energy efficient in terms of power consumption (vs GPUs). Thereby, several FPGA-based CNN accelerators were suggested, targeting both data centers for High Performance Computing (HPC) [13] and embedded applications [14].

Although GPU implementations have shown exceptional computational efficiency, for two reasons, CNN acceleration is heading momentarily towards FPGAs. First, recent advances in FPGA technology have brought about FPGA performance with a recorded performance of 9.2 TFLOP/s for the latter in striking distance to GPUs. Second, recent CNN production patterns are increasing the sparsity of CNNs and using extremely compact types of data.

The remaining paper is structured in the following way: Literature Review, summarizing the use of AI algorithms in the past decade. Explanation of CNN and challenges of FPGA-based implementation. A detailed analysis of FPGA, GPU, and ASIC-based implementation of AI networks which compares between the power, area performance, and efficiency parameters. Finally, we look into the optimizations available and scope of future research.

2 Literature Review

Lately, the application of AI algorithm as a replacement of conventional algorithms have become popular. With the advent of ML and AI, there is a worry to basically address computational cost and power-burning through AI calculations. This leads to the necessity for particular equipment with high capacity to perform AI estimations with large scope issues [15]. The aim of all the research is to achieve better and more capable handling of AI algorithmic calculations. Numerous researches have been conducted over the span of the latest decade discussing hardware and programming advancements and execution strategies in this field [1, 16–21].

Table 1 A comprehensive compendium of survey papers

References	Publication year	Description and scope
[16]	2010	This research addresses all the main architecture methods and models for the 1990–2010 “hardware neural network.” They exemplify numerous HNN implementations in a variety of “ANN models” like CNN, neural bursts, and so on. Digital, analog, composite, neuromorphic, “FPGA,” or optical implementations are used in these HNN
[19]	2013	The scope of the research work was if developers were designing customized versions or were using pre-tailored free platforms for the development. The “artificial neural networks (ANN)” software framework table was also included in the work
[1]	2017	The paper provides a brief analysis of past work on the main approaches of FPGA-based neural inference networks
[21]	2018	This survey paper gives insight about the Moore’s law in conjunction with the increasing use of CPUs and GPUs. Major applications of AI are also discussed
[22]	2020	This paper summarizes AI neural network’s hardware implementation, with emphasis on all three-hardware equipment: ASIC, GPU, and FPGA. A comprehensive flattening strategy has been used to pick the research articles to address issues of science

Here in Table 1, we discuss corresponding survey articles published roughly between 2009 and 2019 on the implementation of AI algorithms, more precisely neural nets for hardware. Few studies have covered hardware implementations of artificial neural networks [16, 19, 20]. While other studies have centered on FPGA-based deep learning neural network accelerators [1, 17, 18]. In [19], the FPGA implementation of coevolutionary neural networks was the subject of the authors (CNNs). The survey addressed GPU implementations in [21].

This examination distinguishes and addresses various papers. These papers were evaluated as per the hardware used. The bulk of the documents include FPGA-based implementations. With its high adaptability and solidness, the FPGA is viewed as a promising option for the use of these calculations. Likewise, recent FPGA design improvements have brought about making it more accessible because of which profound learning research has procured significant consideration [23].

One of the powerful “application-specific integrated circuit” (ASICs) is also used for AI algorithm implementation. ASICs are personalized chips that are conceived for a particular purpose. They save high-power and are speed-efficient, consume less silicon area, making them ideal solutions for accelerating AI algorithms [24]. To accelerate AI algorithms, graphical processing units (GPUs) are also used to pace up algorithms to hundreds of times the initial speed [25]. GPUs are made to carry out scalar and parallel intensive computing [26]. Unlike multicore CPUs, when accessing DRAM memories, GPUs try not to depend on shrouded latencies utilizing large cache

memories [27]. Such highlights have made GPUs increasingly more useful for AI acceleration.

Using small single-board processors, some AI algorithms are applied. For example, raspberry pi. As a result of their little size and low force prerequisites, single-board PCs are viewed as a decision for AI usage.

3 Overview of CNN

One of the classic profound learning networks is the deep convolutional neural network. They are widely used in these areas for the continued development of deep learning technologies, machine vision, and language recognition [28]. Earlier studies have indicated that the calculation of the cutting-edge CNNs is overwhelmed by the convolutional layers [29, 30] formation.

It contains numerous falling layers, pooled layers, and fully-connected complex layers. The convolutionary neural network distinguishes the image as the input, and obtains the outcome across several “convolutional layers, pooling layers and associated layers.”

3.1 Model of Convolutional Layer

The input f^{in} and convolution kernel composed of weights w_{ij} , comprises the convolution layer. The sampled function is set by balancing results to get the output f^{out} [30].

$$f_i^{out} = \sum_{i=1}^{n_{in}} f_i^{in} * w_{i,j} + b_i, 1 \leq i \leq n_{out} \quad (1)$$

3.2 Model of Pooling Layer

The pooling layer typically uses the maximum scan or core scan to minimize the size of the input matrix. The activity will viably lessen the following layer’s data processing ability while forestalling the loss of characteristic information.

3.3 Full Connection Layer

The layer changes over the input to straight space, hence, changing over the input to a direct space. Output is received.

$$f^{out} = \sum_{j=1}^n f_j^{in} * w_{i,j} + b \quad (2)$$

3.4 Activation Layer

A nonlinear change of the input is presented by the activation function excitation, and the output after each layer is regularly handled. Some common functions include nonlinear (Relu), trigonometric function (Tanh), shock response (Sigmoid), etc.

4 Challenges in FPGA

4.1 Tradeoff Between RTL and HLS Approaches

Via high-level abstractions, the HLS-based design approach enables fast growth. This demonstrates the conventional plan utilizing the OpenCL-based methodology [31]. Key features including partitioning, automatic pipelining of the CNN model, etc., can also be supported but for HW performance, however, the resulting design cannot be optimized.

4.2 Necessity of Diligent Design

Considering the particular characteristics of both FPGAs and Convolutional Neural Net, the optimization of throughput demands careful design. In general, two architectures for the same use of resources may have significantly different performances [30], and thus, the use of resources may not be a reliable overall performance measure. There are vastly different criteria for separate CNN layers. In addition, the use of FPGA relies heavily on the burst duration of memory access [32], so the access pattern of CNN memory must be cautiously configured. Furthermore, the compute-throughput ought to be adjusted with the memory or the memory can end up being the bottleneck [19].

4.3 FPGAs Over GPUs/ASICs/CPUs

Although the computer software environments are already mature for developing convolutional Neural Net designs for CPUs or GPUs, those for FPGAs are even now nascent. Furthermore, despite the HLS software, it can take several months for an experienced HW designer to implement the CNN model on FPGAs [33]. Therefore, in general, modeling efforts for GPUs is quite less when compared to FPGAs. Accounting for rapid changes growth related to neural nets, it may not be feasible to re-architect accelerators based on FPGA with every upcoming neural net algorithm. Therefore, the most recent NN algorithm cannot efficiently model an FPGA-based architecture. In addition, a Field-Programmable Gate Array (FPGA) modeling demands greater resource overhead than an ASIC design because of reconfiguration of logic blocks and switches. These factors give these custom boards a competitive disadvantage over other HW acceleration computing platforms for NNs.

4.4 A Comparative Study Between FPGA, ASIC, and GPU

Here, we bring the execution technique, GPUs/FPGAs/ASICs, in relation to regular merit figures. A distinction between the three methods is seen in Table 2. Since the ASIC specification is unaltered post-production, it becomes unsuitable for application where subsequent to installation, the model needs to be revised. ASIC is better used where low power and area are the goal. In comparison to ASICs, FPGAs are available for post-implementation upgrades. It is worth mentioning that even though the ultimate objective is ASIC execution, FPGAs are also used for prototyping and validation [22].

GPUs provide a solution at the software level, while FPGAs and ASICs have a solution at the hardware level. FPGAs and ASICs thus have greater stability compared to GPUs during the deployment process. As a result, the implementation of the GPU

Table 2 A comparison between the aforesaid

	FPGA (Hardware)	ASIC (Hardware)	GPU (Software)
Skills required	Verilog	Verilog	High level language
Implementation level	Medium	High	Medium
Versatility	High	High	Low
Post implementation change	High	Very low	High
Overall expenditure	Medium	High	Low
Area consumption	High	Low	High
Performance	Medium	High	Low/Medium
Power consumption	Medium	Low	High

is constrained by the current underlying hardware. Thus, in some situations, FPGAs can be quicker than GPUs.

5 A Compendium of Numerous Hardware Implementations

Implementation details of neural nets on various FPGA boards are shown in Table 3. Loop optimizations were first investigated in [30] to extract an FPGA-based CNN accelerator. This design was modeled utilizing HLS tools, therefore, it relies on the arithmetic of 32 floating points. Works in [34, 35] pursue the same unrolling scheme. In addition, [35] design has 16 bits of fixed-point arithmetic and RTL design, which results in an increase of performance by approximately two times. In recent works [36], the same unrolling and tiling scheme is used where authors report an improvement of $\times 13.4$ over their original works [30]. Consequently, unrolling and tiling loops can be proficient as it were for devices with large computational capabilities (i.e., DSP). This is often illustrated in works of Rahman et al. [34] which improves speed by 1.2 times over [30].

Conversely, all modeling factors looking for ideal loop unroll are fully explored in the works of Ma et al. [37, 40, 41]. More specifically, researchers show that using unroll function for single input arithmetic, the input FM and weights are optimally reused [38, 39, 42–46].

6 Conclusion and Future Work

In this paper, a comprehensive study related to development and deployment of FPGA in CNN accelerators has been provided. To highlight their similarities and distinctions, we categorized the works based on many parameters. We outlined the influential avenues for study and summarized the key themes of numerous works. Usage of more than one FPGA is important for computational acceleration of massive and extensive Neural Net models, considering the limited hardware resources available on an FPGA board. This allows the architecture to be partitioned across several FPGAs. Although a software for automatic partitioning will not be readily accessible, the manual approach to partitioning is vulnerable to error and unscalable. In addition, random splitting will increase the complexity of interconnections in a manner that the I/O pin count cannot complete the requirement of number of connections.

Lately, researchers have also reconnoitered models of the spike neural network (SNN) that model the biological neuron more closely [47]. This study concentrates on (ANN) which is an acronym for artificial neural network. SNNs and ANNs vary considerably. As a result, their training rules and network architectural structure

Table 3 Specifications of various hardware implementations

	Neural net	FPGA board	Frequency (MHz)	Performance (GOPs)	Consumed power (W)	LUT (K)	DSP	Memory (MB)
[30]	AlexNet-C	Virtex7 VX485T	100	62	19	186	2240	19
[14]	VGG16SVD-F	Zynq Z7045	150	137	10	183	780	18
[31]	AlexNet-C	Stratix5 GSD8	120	187	34	138	635	18
	AlexNet-F			72		272	752	30
	VGG16-F			118		524	1963	52
[34]	AlexNet-C	Virtex7 VX485T	100	75		28	2695	20
[36]	AlexNet-F	Virtex7 VX690T	150	826	126		14,400	
	VGG16-F			1280	160		21,600	
[35]	NIN-F	Stratix5	100	114	20	224	256	47
	AlexNet-F	GXA7		134	19	242	256	31
[40]	AlexNet-F	Virtex7 VX690T	156	566	30	274	2144	35
[41]	AlexNet-C	Virtex7 VX690T	100	62		273	2401	20
[37]	VGG16-F	Arria10 GX1150	150	645	50	322	1518	38
[42]	AlexNet-F	Arria10 GT1150	240	360		700	1290	47
	VGG-F		222	460		708	1340	49
	VGG-F		232	117		626	1500	33
[43]	AlexNet-C	Cyclone5 SEM	100	12		22	28	1
		Virtex7 VX485T	100	445			2800	
[39]	NiN	Stratix5 GXA7	150	283		453	256	30
	VGG16-F			352		424	256	44
	ResNet-50			251		347	256	39
	NiN	Arria10 GX1150	200	588		320	1518	31
	VGG16-F			720		263	1518	45
	ResNet-50			619		437	1518	39
[38]	AlexNet-F	Virtex7 VX690T	100	446	25	207	2872	37
	VGG16SVD-F			473	26	224	2950	47

vary drastically. Large-scale SNN modeling onto SOC/FPGA boards will offer an exhilarating and remunerating challenge for IT designers later on [45].

References

1. Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*.
2. Pau, L. F. (1991). Artificial intelligence and financial services. *IEEE Transactions on Knowledge and Data Engineering*.
3. Yao, X., Zhou, J., Zhang, J., & Boer, C. R. (2017). From intelligent manufacturing to smart manufacturing for industry 4.0 driven by next generation artificial intelligence and further on. In *Proceedings—2017 5th International Conference on Enterprise Systems ES*.
4. Bishnoi, L., & Narayan Singh, S. (2018). Artificial intelligence techniques used in medical sciences: A review. In *Proceedings of 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*.
5. Parker, D. S. (1989). Integrating AI and DBMS through stream processing.
6. Fraley, J. B., & Cannady, J. (2017). The promise of machine learning in cybersecurity. In *Conference of Proceedings—IEEE SOUTHEASTCON*.
7. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*.
8. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*.
9. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
10. Zhang, Y., Pezeshki, M., Brakel, P., Zhang, S., Laurent, C., Bengio, Y., et al. (2016). Towards end-to-end speech recognition with deep convolutional neural networks. In *Proceedings of Annual Conference of the International Speech Communication Association, INTERSPEECH*.
11. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*.
12. Nurvitadhi, E., Venkatesh, G., Sim, J., Marr, D., Huang, R., Ong, J. G. H., et al. (2017). Can FPGAs beat GPUs in accelerating next-generation deep neural networks? In *FPGA 2017—Proceedings 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*.
13. Ovtcharov, K., Ruwase, O., Kim, J., Fowers, J., Strauss, K., & Chung, E. S. (2015). Accelerating deep convolutional neural networks using specialized hardware. *Microsoft Research Whitepaper*.
14. Qiu, J., Wang, J., Yao, S., Guo, K., Li, B., Zhou, E., et al. (2016). Going deeper with embedded FPGA platform for convolutional neural network. In *FPGA 2016—Proceedings of the 2016 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*.
15. Rigos, S., Mariatos, V., & Voros, N. (2012). A hardware acceleration unit for face detection. In *2012 Mediterranean Conference on Embedded Computing*.
16. Misra, J., & Saha, I. (2010). Artificial neural networks in hardware: A survey of two decades of progress. *Neurocomputing*.
17. Baji, T. (2018). Evolution of the GPU device widely used in AI and massive parallel processing. In *2018 IEEE Electron Devices Technology and Manufacturing Conference EDTM 2018—Proceedings*.
18. Shawahna, A., Sait, S. M., & El-Maleh, A. (2019). FPGA-based accelerators of deep learning networks for learning and classification: A review.
19. Mittal, S. (2020). A survey of FPGA-based accelerators for convolutional neural networks. *Neural Computing & Applications*.

20. Guo, K., Zeng, S., Yu, J., Wang, Y., & Yang, H. (2017). [DL] A survey of FPGA-based neural network inference accelerator.
21. Blaiech, A. G., Ben Khalifa, K., Valderrama, C., Fernandes, M. A. C., & Bedoui, M. H. (2019). A survey and taxonomy of FPGA-based deep learning accelerators. *The Journal of Systems Architecture*.
22. Talib, M. A., Majzoub, S., Nasir, Q., & Jamal, D. (2020) A systematic literature review on hardware implementation of artificial intelligence algorithms. *The Journal of Supercomputing*.
23. Schneider, S., Taylor, G. W., Linqvist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*.
24. Faraone, J., Gambardella, G., Fraser, N., Blott, M., Leong, P., & Boland, D. (2018). Customizing low-precision deep neural networks for FPGAs. In *Proceedings—2018 International Conference on Field Programmable Logic and Applications FPL*.
25. Cheng, K. T., & Wang, Y. C. (2011). Using mobile GPU for general-purpose computing a case study of face recognition on smartphones. In *Proceedings of 2011 International Symposium on VLSI Design, Automation and Test VLSI-DAT 2011*.
26. Ouerhani, Y., Jridi, M., & AlFalou, A. (2010). Fast face recognition approach using a graphical processing unit “GPU.” In *2010 IEEE International Conference on Imaging Systems and Techniques IST 2010—Proceedings*.
27. Li, E., Wang, B., Yang, L., Peng, Y. T., Du, Y., Zhang, Y., et al. (2012). GPU and CPU cooperative acceleration for face detection on modern processors. In *Proceedings—IEEE International Conference on Multimedia and Expo*.
28. Lu, L., Liang, Y., Xiao, Q., & Yan, S. (2017). Evaluating fast algorithms for convolutional neural networks on FPGAs. In *Proceeding—IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines FCCM 2017*.
29. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
30. Zhang, C., Li, P., Sun, G., Guan, Y., Xiao, B., & Cong, J. Optimizing FPGA-based accelerator design for deep convolutional neural networks. In *FPGA 2015—2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*.
31. Suda, N., Chandra, V., Dasika, G., Mohanty, A., Ma, Y., Vrudhula, S., et al. (2016). Throughput-optimized openCL-based FPGA accelerator for large-scale convolutional neural networks. In *FPGA 2016—Proceedings of the 2016 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*.
32. Zhang, C., Fang, Z., Zhou, P., Pan, P., & Cong, J. (2016). Caffeine: Towards uniformed representation and acceleration for deep convolutional neural networks. In *IEEE/ACM International Conference on Computer-Aided Design Digital Technical Paper ICCAD*.
33. Guan, Y., Liang, H., Xu, N., Wang, W., Shi, S., Chen, X., et al. (2017). FP-DNN: An automated framework for mapping deep neural networks onto FPGAs with RTL-HLS hybrid templates. In *Proceedings—IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines FCCM 2017*.
34. Rahman, A., Lee, J., & Choi, K. (2016). Efficient FPGA acceleration of convolutional neural networks using logical-3D compute array. In *Proceedings of 2016 Design, Automation & Test in Europe Conference & Exhibition DATE 2016*.
35. Ma, Y., Suda, N., Cao, Y., Seo, J. S., & Vrudhula, S. (2016). Scalable and modularized RTL compilation of Convolutional Neural Networks onto FPGA. In *FPL 2016—26th International Conference on Field-Programmable Logic and Applications*.
36. Zhang, C., Wu, D., Sun, J., Sun, G., Luo, G., & Cong, J. (2016). Energy-efficient CNN implementation on a deeply pipelined FPGA cluster. In *Proceedings of International Symposium on Low Power Electronics and Design*.
37. Ma, Y., Cao, Y., Vrudhula, S., & Seo, J. S. (2017). Optimizing loop operation and dataflow in FPGA acceleration of deep convolutional neural networks. In *FPGA 2017—Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*.

38. Liu, Z., Dou, Y., Jiang, J., Xu, J., Li, S., Zhou, Y., et al. (2017). Throughput-optimized FPGA accelerator for deep convolutional neural networks. *ACM Transactions on Reconfigurable Technology and Systems*.
39. Ma, Y., Cao, Y., Vrudhula, S., & Seo, J. S. An automatic RTL compiler for high-throughput FPGA implementation of diverse deep convolutional neural networks. In *2017 27th International Conference on Field-Programmable Logic and Applications FPL*.
40. Li, H., Fan, X., Jiao, L., Cao, W., Zhou, X., & Wang, L. (2016). A high performance FPGA-based accelerator for large-scale convolutional neural networks. In *FPL 2016—26th International Conference on Field-Programmable Logic and Applications*.
41. Alwani, M., Chen, H., Ferdman, M., & Milder, P. (2016). Fused-layer CNN accelerators. In *2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*.
42. Wei, X., Yu, C. H., Zhang, P., Chen, Y., Wang, Y., Hu, H., et al. (2017). Automated systolic array architecture synthesis for high throughput CNN inference on FPGAs. In *Proceedings of the 54th Annual Design Automation Conference 2017*.
43. Motamedi, M., Gysel, P., & Ghiasi, S. (2017). PLACID: A platform for FPGA-based accelerator creation for DCNNs. *ACM Transactions on Multimedia Computing, Communications, and Applications*.
44. Ma, Y., Kim, M., Cao, Y., Vrudhula, S., & Seo, J. S. (2017). End-to-end scalable FPGA accelerator for deep residual networks. In *Proceedings—IEEE International Symposium on Circuits and Systems*.
45. Maguire, L. P., McGinnity, T. M., Glackin, B., Ghani, A., Belatreche, A., & Harkin, J. (2007). Challenges for large-scale implementations of spiking neural networks on FPGAs. *Neurocomputing*.
46. Negi, A., Saxena, D., & Suneja, K. (2020). High level synthesis of chaos based text encryption using modified Hill Cipher algorithm (pp. 3–7).
47. Thapa, S., Adhikari, S., Naseem, U., Singh, P., Bharathy, G., & Prasad, M. (2020). Detecting Alzheimer’s disease by exploiting linguistic information from Nepali transcript. *Communication in Computer and Information Science*.

Author Index

A

Ahmad, Faroze, 71
Ajitha, D., 61
Akhil, M. Nair, 489
Akkineni, Haritha, 477
Al-Obeidi, Ahmed S., 323
Ananth, A. G., 117
Annapurna, K., 371
Arora, Monika, 393

B

Babu, P. Ashok, 243
Bali, Malvinder Singh, 355
Bansal, Atul, 425
Bedekar, Pragati Patil, 433
Bhagya Lakshmi, K., 61
Bhatt, Chintan, 1

C

Cartagena, Jhonatan Fabricio Meza, 285
Chandani, Khushi, 393
Chand, Aryaman, 393
Chandru, R., 225
Chaouchi, Hakima, 1
Chitra Nair, S., 97
Choudhary, Tripti, 425
Chouhan, Dharamendra, 275

D

Das, Prodipto, 381
Deepalakshmi, P., 193
Dutonde, Abhimanyu, 433

E

Evangeline, L. Diana, 235

G

Garg, Tanya, 35
Gayathri, A. N., 415
Ghazo Al, Jaafar, 23
Goraya, Major Singh, 35
Goyal, Vishal, 425
Gul, Baseerat, 71
Gupta, Kamali, 355
Gupta, Rajendra, 333
Gupta, Surbhi, 23, 183
Gurdasani, Harsha, 117

H

Hamad, Abdulsattar Abdullah, 323
Hari Krishna, Ch., 161
Hasan, Moin, 35
Hegde, Sachinkumar, 275

J

Jayaram, Ramaprabha, 299
Jindal, Sonika, 345
Jose, Deepa, 285

K

Kadali, Kalyan Sagar, 445
Kadry, Seifedine, 183
Kakkasageri, Mahabaleshwar S., 171
Kalaiyarasi, M., 445
Kaleeswari, C., 127

Kamble, Sachin, 311
 Kanakala, Srinivas, 455
 Karthi, S., 445
 Karunanithi, K., 445
 Kaushal, Priyanka, 11, 183
 Khare, Ankur, 333
 Kumar, Mithilesh, 139
 Kumar, Neetesh, 151
 Kumar, S. M. Dilip, 275
 Kunal, 495
 Kuppusamy, K., 127
 Kushwaha, Alok Kumar Singh, 345

L

Lakhwani, Kamlesh, 255
 Lakshmi, K., 47
 Lakshmi, P. V. S., 477
 Let, G. Shine, 235

M

Mahadik, Yogesh, 311
 Majumder, Abhishek, 83
 Malik, Swati, 355
 Manimegalai, R., 225
 Manjula, S., 47
 Mohan, Neeraj, 183
 Morey, Amogh Manjunath Rao, 207
 Mudengudi, Shailaja S., 171

N

Nagaraju, V. Siva, 243
 Nagpal, Ritu, 465
 Nayak, Manoja Kumar, 107
 Nayak, Padmalaya, 11, 23
 Negi, Anvit, 495

O

Ojashwi, Ayushi, 139

P

Padhan, D. G., 11
 Pragaspathy, S., 445
 Prajapati, Keval B., 1
 Prasath, J. S., 285
 Prashanthi, Vempaty, 455
 Pratap, C. Benin, 235
 Pravalika, Neela, 23
 Preethi, N., 489

R

Rai, Newton, 207
 Rani, Seema, 465
 Raut, Atul, 433
 Roy, Sudipta, 83

S

Sachan, Anuj, 151
 Sachdeva, Aryan, 139
 Sachdeva, Monika, 345
 Saisindhutheja, Reddy, 403
 Sandhu, Ramandeep, 255
 Sarada, Lasya, 477
 Saravanan, S., 445
 Saxena, Devansh, 495
 Senthil Kumar, T., 299
 Shaik, Habibuddin, 207
 Sharada, K. V., 455
 Sharma, Madan Kumar, 139
 Shreyas, J., 275
 Shukla, Piyush Kumar, 333
 Shyam, Gopal K., 403
 Srinidhi, N. N., 275
 Srinivasaraju, P., 161
 Srinivasarao, G., 161
 Subbulakshmi, N., 225
 Sudar, K. Muthamil, 193
 Suneja, Kriti, 495
 Suresh, H. L., 207
 Swain, Prasanta Kumar, 107
 Swami, Rosy, 381
 Swetha, GK., 11

T

Tamilarasan, B., 323
 Thakre, Mohan, 311
 Thangadurai, N., 117
 Thivagar, M. Lellis, 323
 Tripathi, Anand, 265

V

Vallabhuni, Rajeev Ratna, 243
 Vamsi, T. Sairam, 161
 Varma, D. Prasanth, 371
 Varsha, M., 97
 Veena, S., 207
 Vijaya Lakshmi, K. N. V. S., 61
 Viji Rajendran, V., 415

Y

Yadav, Anjana, 265