

Guru Prasadh Venkataramani  
Karthik Sankaranarayanan  
Saswati Mukherjee  
Kannan Arputharaj  
Swamynathan Sankara Narayanan (Eds.)

Communications in Computer and Information Science

808

# Smart Secure Systems – IoT and Analytics Perspective

Second International Conference  
on Intelligent Information Technologies, ICIIT 2017  
Chennai, India, December 20–22, 2017  
Proceedings

# Communications in Computer and Information Science

808

*Commenced Publication in 2007*

Founding and Former Series Editors:

Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu, Dominik Ślęzak,  
and Xiaokang Yang

## Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),  
Rio de Janeiro, Brazil*

Phoebe Chen

*La Trobe University, Melbourne, Australia*

Joaquim Filipe

*Polytechnic Institute of Setúbal, Setúbal, Portugal*

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation of the Russian  
Academy of Sciences, St. Petersburg, Russia*

Krishna M. Sivalingam

*Indian Institute of Technology Madras, Chennai, India*

Takashi Washio

*Osaka University, Osaka, Japan*

Junsong Yuan

*Nanyang Technological University, Singapore, Singapore*

Lizhu Zhou

*Tsinghua University, Beijing, China*

More information about this series at <http://www.springer.com/series/7899>

Guru Prasad Venkataramani  
Karthik Sankaranarayanan  
Saswati Mukherjee  
Kannan Arputharaj  
Swamynathan Sankara Narayanan (Eds.)

# Smart Secure Systems – IoT and Analytics Perspective

Second International Conference  
on Intelligent Information Technologies, ICIIT 2017  
Chennai, India, December 20–22, 2017  
Proceedings

*Editors*

Guru Prasadh Venkataramani  
George Washington University  
Washington, DC  
USA

Kannan Arputharaj  
Anna University  
Chennai  
India

Karthik Sankaranarayanan  
Intel Microarchitecture Research Lab  
Santa Clara, CA  
USA

Swamynathan Sankara Narayanan  
Anna University  
Chennai  
India

Saswati Mukherjee  
Anna University  
Chennai  
India

ISSN 1865-0929                      ISSN 1865-0937 (electronic)  
Communications in Computer and Information Science  
ISBN 978-981-10-7634-3              ISBN 978-981-10-7635-0 (eBook)  
<https://doi.org/10.1007/978-981-10-7635-0>

Library of Congress Control Number: 2017961801

© Springer Nature Singapore Pte Ltd. 2018

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, express 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.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer Nature Singapore Pte Ltd.  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

On behalf of the Program Committee, it is my pleasure to present the proceedings of the International Conference on Intelligent Information Technologies (ICIIT 2017) held during December 20–22, 2017, at the College of Engineering Guindy, Anna University Chennai, India. ICIIT 2017 acted as a forum for researchers, scientists, academics, and industrialists to present their latest research results and research perspectives on the conference theme, “Internet of Things.”

We received 157 submissions from all over the world. After a rigorous peer-review process involving 351 reviews in total, 26 full-length articles were accepted for oral presentation and for inclusion in the CCIS proceedings. This corresponds to an acceptance rate of 23% and is intended for maintaining the high standards of the conference proceedings. The papers included in this CCIS volume cover a wide range of topics in IoT enabling technologies, IoT security, social IoT, Web of Things, and IoT services and applications.

The pre-conference tutorials on December 19, 2017, covered the thrust areas of IoT. The technical program started on December 20, 2017, and continued for next two days. Non-overlapping oral and poster sessions ensured that all attendees had opportunities to interact personally with presenters. The conference featured the following distinguished keynote speakers: Prof. Timothy A. Gonsalves of IIT Mandi, India, Prof. Roch H. Glitho of Concordia University, Canada, Prof. Selwyn Piramuthu of the University of Florida, USA, Dr. Balachandar Santhanam of IoT Group, Intel India, and Dr. Prateek Jain of Microsoft Research, India.

I take this opportunity to thank the authors of all submitted papers for their hard work, adherence to the deadlines, and patience with the review process. The quality of a refereed volume depends mainly on the expertise and dedication of the reviewers. I am thankful to the reviewers for their timely effort and help rendered to make this conference successful. I thank Prof. Marimuthu Palaniswami, Australia, Prof. Vijayan Sugumaran, USA, and Prof. Guru Prasad Venkataramani, USA, for providing valuable guidelines and inspiration to overcome various difficulties in the process of organizing this conference as general chairs. I would like to thank Prof. Saswati Mukherjee, Prof. Kannan A., and Prof. Swamynathan S. for their endless effort in all aspects as conference convener and conference chairs. I would like to thank Prof. Ranjani Parathasarthi, Prof. Uma G. V., Prof. Sridhar S., Prof. Geetha Ramani R., and the Program Committee members for their invaluable suggestions. I would also like to thank the PhD symposium chairs, poster/demo chairs, tutorial chairs, finance chair, registration chairs, publicity chairs, sponsorship chairs, liaison chairs, and Web chairs for their big support and contributions. For the publishing process at Springer, I would like to thank Leonie Kunz, Yeshmeena Bisht, Suvira Srivastav, and Nidhi Chandhoke for their constant help and cooperation.

My sincere and heartfelt thanks to Prof. Geetha T. V., Convener Committee member, Anna University, and Prof. Ganesan S., Registrar Anna University, for their

support of ICIIT 2017 and providing the infrastructure at CEG to organize the conference. I am indebted to the faculty, staff, and students of the Department of Information Science and Technology for their tireless efforts that made ICIIT 2017 at CEG possible. I would also like to thank the participants of this conference, who considered the conference above all hardships. In addition, I would like to express my appreciation and thanks to all the people whose efforts made this conference a grand success.

December 2017

Karthik Sankaranarayanan

# Organization

ICIIT 2017 was organized by the Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai, India.

## Chief Patron

Geetha T. V. Convener Committee Member, Anna University, India

## Patron

Ganesan S. Registrar, Anna University, India

## General Chairs

Marimuthu Palaniswami University of Melbourne, Australia  
Vijayan Sugumaran Oakland University, USA  
Guru Prasadh George Washington University, USA  
Venkataramani

## Program Chair

Karthik Sankaranarayanan Intel Research Lab (MRL), USA

## Steering Committee

Ahmed Abdelgawad Central Michigan University, USA  
Ajith Abraham MIR Labs, USA  
Atta ur Rehman Khan King Saud University, Saudi Arabia  
Markus Weinberger Aalen University, Germany  
Marimuthu Palaniswami University of Melbourne, Australia  
Rishi Bhatnagar Aeris Communications, India

## Convener and Proceedings Chair

Saswati Mukherjee Anna University, India

## Conference Chairs

Kannan A. Anna University, India  
Swamynathan S. Anna University, India



## **Conference Co-chairs**

Sridhar S. Anna University, India  
Deivamani Mallayya Anna University, India

## **Organizing Secretaries**

Thangaraj N. Anna University, India  
Selvi Ravindran Anna University, India

## **Panel Chairs**

Ranjani Parthasarathi Anna University, India  
Uma G. V. Anna University, India  
Geetha Ramani R. Anna University, India

## **Poster/Demo Chairs**

Ramachandran V. Anna University, India  
Yogesh P. Anna University, India

## **PhD Symposium Chairs**

Vani K. Anna University, India  
Indra Gandhi K. Anna University, India

## **Finance Chair**

Bama S. Anna University, India

## **Registration Chairs**

Vijayalakshmi M. Anna University, India  
Vidya K. Anna University, India

## **Publicity Chairs**

Indhumathi J. Anna University, India  
Muthuraj R. Anna University, India

## Sponsorship Chair

Sendhilkumar S. Anna University, India

## Tutorial Chairs

Mala T. Anna University, India

Geetha P. Anna University, India

Uma E. Anna University, India

## Local Liaison Chair

Kulothungan K. Anna University, India

## Web and Communication Chair

Abirami S. Anna University, India

## Technical Review Board

Sedat Akleylek	Ondokuz Mayıs University, Turkey
Ioannis Anagnostopoulos	University of Thessaly, Greece
Amjad Anvari-Moghaddam	Aalborg University, Denmark
Venkatalakshmi B.	Velammal Engineering College, India
Mohamad Badra	Zayed University, UAE
Thar Baker	Liverpool John Moores University, UK
Paolo Bellavista	Università di Bologna, Italy
Zorica Bogdanovic	University of Belgrade, Serbia
Rajendra Boppana	University of Texas, USA
Amrita Chaturvedi	IIT Kanpur, India
Pethuru Raj Chelliah	Reliance Jio Cloud Services, India
Simone Cirani	University of Parma, Italy
Luca Davoli	University of Parma, Italy
Ramesh Dharavath	IIT Dhanbad, India
Ke-Lin Du	Concordia University, Canada
Ali Emrouznejad	Aston University, Birmingham, UK
Ernesto Exposito	University of Pau and Pays Adour, France
Valerio Frascolla	Intel, Germany
Sudha Sadasivam G.	PSG College of Technology, India
Veena Goswami	KIIT University, India
Zhishan Guo	Missouri University of S&T, USA
Pao-Ann Hsiung	National Chung Cheng University, Taiwan
Mauro Iacono	University of Naples, Italy
Joe Louis Paul Ignatius	SSN Engineering College, Chennai, India
Nejmeddine Jouida	Eniso, Tunisia
Chandrasekaran K.	NITK, Surathkal, India

Muneeswaran K.	Mepco Schlenk Engineering College, India
Easwarakumar K. S.	Anna University, India
Selvakumar Kamalanathan	VIT University, Vellore, India
Joarder Kamruzzaman	Federation University, Australia
Thangavelu Kesavamurthy	PSG College of Technology, India
Manas Khatua	IIT, Jodhpur, India
Karthikeyan Krishnasamy	Coimbatore Institute of Technology, India
Vimal Kumar	University of Waikato, Hamilton, New Zealand
Aleksandra Labus	University of Belgrade, Serbia
Gyu Myoung Lee	Liverpool John Moores University, UK
Yiu-Wing Leung	Hong Kong Baptist University, SAR China
Yun Liu	Beijing Jiaotong University, China
Anand M.	VIT University, Vellore, India
Krishnamurthy M.	KCG College of Technology, India
Priyadharshini M.	KPR Institute of Technology, India
Sandhya M.	B. S. Abdur Rahman University, India
Wissam Mallouli	Montimage, France
Daisuke Mashima	ADSC Center of Illinois, Singapore
Barbara Masucci	Università di Salerno, Italy
Weizhi Meng	Technical University of Denmark, Denmark
Juan Pedro Muñoz-Gea	Polytechnic University of Cartagena, Spain
Venkatesh Narayanan	Microsoft, India
Yogesh P.	Anna University, India
Narayanasamy P. N.	PSG College of Technology, India
Alessandra De Paola	University of Palermo, Argentina
Zeeshan Pervez	University of West Scotland, UK
Selwyn Piramuthu	University of Florida, USA
Vincenzo Piuri	University of Milan, Italy
Tie Qiu	Dalian University of Technology, China
Baskaran R.	Anna University, India
Kanchana R.	SSN Engineering College, Chennai, India
Bhuvaneshwaran R. S.	Anna University, India
Rajinikumar Ramalingam	IIT Mandi, India
Virender Ranga	NIT Kurukshetra, India
Udai Pratap Rao	S. V. National Institute of Technology, India
Sangram Ray	National Institute of Technology Sikkim, India
Rukhsana Ruby	Shenzhen University, China
Mary Saira Bhanu S.	NIT Tiruchirappalli, India
Sasirekha S.	SSN College of Engineering, India
Elham Sahebkhorkhorasani	University of Illinois, Springfield, USA
Mohsen Amini Salehi	University of Louisiana, USA
Nickolas Savarimuthu	NIT Tiruchirappalli, India
Corinna Schmitt	University of Zurich, Switzerland
Chithra Selvaraj	SSN Engineering College, Chennai, India
Mercy Shalinie	Thiagarajar College of Engineering, India
Anshuman Singh	University of Central Missouri, USA

Houbing Song	Embry-Riddle Aeronautical University, USA
Jayashree Subramanian	PSG College of Technology, India
Mirnalinee T. T.	SSN Engineering College, Chennai, India
Javid Taheri	Karlstad University, Sweden
Latha Tamilselvan	B. S. Abdur Rahman University, India
Halina Tarasiuk	Warsaw University of Technology, Poland
Ali Yazici	Atilim University, Turkey
Kalidas Yeturu	IIT Tirupati, India

### **Additional Reviewers**

Bilge Say	Chavit Denninnart
Cigdem Turhan	Enrico Barbierato
Fannia Pacheco	Hossam Kasem
Laura Belli	Marco Viola
Matthieu Carre	Ramaprasad Vaddella
Razin Farhan Hussain	Sm Zobaed
Yan Li	

### **Organizing Committee**

L. Sairamesh	S. Kanimozhi
P. Chakradhar	T. Sindhu
V. Pandiyaraju	J. Sumathi
T. J. Vijaykumar	B. Senthilnayaki
D. Narashiman	R. L. Jasmine
P. Prabhavathy	H. Riasudheen
P. Shunmuga Perumal	B. R. Yuvaraj
V. Ezhilarasi	K. Mohana Bhindu
Tina Esther Trueman	G. Mahalakshmi

# Contents

## IoT Enabling Technologies

Comparative Study of Simulation Tools and Challenging Issues in Cloud Computing . . . . .	3
<i>S. R. Shishira, A. Kandasamy, and K. Chandrasekaran</i>	
Data Consumption Pattern of MQTT Protocol for IoT Applications. . . . .	12
<i>Hansa Lysander Manohar and T. Reuban Gnana Asir</i>	
Data Access in Heterogeneous Data Sources Using Object Relational Database . . . . .	23
<i>M. S. Hema, R. Maheshprabhu, and M. Nageswara Guptha</i>	
Optimization of UAV Video Frames for Tracking. . . . .	34
<i>A. Ancy Micheal and K. Vani</i>	
Failure Recovery Using Segment Protection in Software Defined Networks . . . . .	47
<i>V. Padma, Gayathri Santhosh, and Yogesh Palanichamy</i>	
Spectrum Sensing Based Heed Routing Performance Enhancement Strategy for Cognitive Radio Sensor Networks . . . . .	62
<i>S. Janani, M. Ramaswamy, and J. Samuel Manoharan</i>	

## IoT Security

Improved Recommendation Filtering Component Resilient to Trust Distortion Attacks in a MANET . . . . .	81
<i>Shirina Samreen and Akhil Jabbar Meerja</i>	
A Hybrid Group Key Management Scheme for UAV – MBN Network Environment Increasing Efficiency of Key Distribution in Joining Operation . . . . .	93
<i>R. Mahaveerakannan and C. Suresh Gnana Dhas</i>	
Lightweight Anonymity-Preserving Authentication and Key Agreement Protocol for the Internet of Things Environment . . . . .	108
<i>Ahmed Mohammed Ibrahim Alkuhlani and S. B. Thorat</i>	

ECC Based Proxy Signature Scheme with Forward Security . . . . .	126
<i>Aparna Bannore and Satish Devane</i>	
Efficient and Robust Secure In-Network Aggregation in Wireless Sensor Networks . . . . .	139
<i>Radhakrishnan Maivizhi and Palanichamy Yogesh</i>	
Context-Aware Conditional Probabilistic Hyper-exponential Reputation Technique for Mitigating Byzantine Attacks . . . . .	153
<i>Geetha Achuthan, Sreenath Niladuri, and A. G. Sareeka</i>	
<b>Social IoT</b>	
Illumination and Communication Using LED as Light Source in Underground Mines . . . . .	171
<i>B. Anitha Vijayalakshmi and M. Nesa Sudha</i>	
Leveraging Social Networks for Smart Cities: A Case-Study in Mitigation of Air Pollution . . . . .	179
<i>Nagarathna Ravi, Manoranjani R., Vimala Rani P., Mercy Shalinie S., and Karthick Seshadri</i>	
Smart Garbage Bin Systems – A Comprehensive Survey . . . . .	194
<i>Gulshan Soni and Selvaradjou Kandasamy</i>	
<b>Web of Things</b>	
Contextual Pattern Clustering for Ontology Based Activity Recognition in Smart Home . . . . .	209
<i>K. S. Gayathri, K. S. Easwarakumar, and Susan Elias</i>	
E-FPROMETHEE: An Entropy Based Fuzzy Multi Criteria Decision Making Service Ranking Approach for Cloud Service Selection . . . . .	224
<i>B. Akshya Kaveri, O. Gireesha, Nivethitha Somu, M. R. Gauthama Raman, and V. S. Shankar Sriram</i>	
Workflow Scheduling Using IOT Enabled Reputation of Service Providers in the Cloud . . . . .	239
<i>K. Kanagaraj and S. Swamynathan</i>	
SCICS: A Soft Computing Based Intelligent Communication System in VANET . . . . .	255
<i>Mamata Rath and Binod Kumar Pattanayak</i>	

Classification and Recommendation of Competitive Programming Problems Using CNN . . . . .	262
<i>S. Sudha, A. Arun Kumar, M. Muthu Nagappan, and R. Suresh</i>	
A Generic Context-Aware Service Discovery Architecture for IoT Services . . . . .	273
<i>S. Sasirekha, S. Swamynathan, and S. Keerthana</i>	
<b>IoT Services and Applications</b>	
Understanding How Adversarial Noise Affects Single Image Classification . . . . .	287
<i>Amit Adate, Rishabh Saxena, and Don.S</i>	
Top- <i>k</i> Category Search for an IP Address-Product Network . . . . .	296
<i>Ramalingeswara Rao Thottempudi, Pabitra Mitra, and Goswami Adrijit</i>	
Booking Based Smart Parking Management System . . . . .	312
<i>Jhanavi Jyothish, Mamatha, Srilakshmi Gorur, and M. Dakshayini</i>	
VIBI: A Braille Inspired Password Entry Model to Assist Person with Visual Impairments. . . . .	320
<i>V. Balaji, K. S. Kuppusamy, and Shaikh Afzal</i>	
A Rehabilitation Therapy for Autism Spectrum Disorder Using Virtual Reality. . . . .	328
<i>T. Manju, S. Padmavathi, and D. Tamilselvi</i>	
<b>Author Index</b> . . . . .	337

# **IoT Enabling Technologies**



# Comparative Study of Simulation Tools and Challenging Issues in Cloud Computing

S. R. Shishira<sup>1</sup>(✉), A. Kandasamy<sup>1</sup>, and K. Chandrasekaran<sup>2</sup>

<sup>1</sup> Department of MACS, NITK, Surathkal, India  
shishirasr@gmail.com, kandy@nitk.ac.in

<sup>2</sup> Department of CSE, NITK, Surathkal, India  
kchnitk@ieee.org

**Abstract.** Resource Scheduling lays a key role in large-scale cloud applications. It is difficult for the developers to do an extensive research on all the issues in real time as it requires infrastructure which is beyond the control, also network condition cannot be predicted. Hence simulations are used which imitates the real time environment. There are various simulators developed for the research as it is difficult to maintain the infrastructure on premise. Thus to understand the tools in deep, we focused on five open source tools such as Cloudsim, CloudAnalyst, iCancloud, Greencloud and CloudSched. The above mentioned tools are compared based on the respective architecture, the process of simulation, structural elements and performance parameters. In the paper, we have also discussed some of the challenging issues among the tools for further research.

**Keywords:** Data centres · Cloud simulator tools · Cloud computing  
Resource scheduling

## 1 Introduction

Cloud computing is a emerging platform which helps in retrieving the services from the remote server. It emerged based on the advancements of grid computing [1]. The benefits of cloud computing includes the services such as storage, network facility, and an efficient use of resources and balance the particular load on various datacenters [2]. Few examples include GoogleApp Engine, IBM cloud, Amazon EC2, Azure. Cloud computing helps in sharing of resources, network and storage for users as pay as u go model.

Cloud computing is composed on many nodes formed by CPU, network, bandwidth etc. Due to virtualisation, today cloud computing is an emerging field. One important technology is resource scheduling. When the demands are requested from the user, it has to be scheduled in an efficient manner based on the resources existed [5]. Efficient resource utilization is also majorly concerned. There is no available infrastructure to do the same as there are difficulties in

maintaining and also with the cost. Hence there are publicly available tools provided by Buyya and Murshed [3,4] for the researchers to proceed the work.

In this paper we have discussed five important tools which are openly available such as Cloudsim, CloudAnalyst, iCancloud, Greencloud and CloudSched. We have compared them based on the different criteria's. This paper is organized as follows: Sect. 2 gives the related work; Sects. 3 and 4 gives the details on the comparison based on different elements considered.

## 2 Related Work

This paper mainly focus on open-source simulators on cloud. This section briefs on the related work on the different simulators developed for cloud computing.

Gangsim tool is introduced by Howell and McNab [6] for grid computing. Gridsim toolkit is developed for modelling and simulation for grid computing by Buyya et al. [7]. Calheiros et al. Compared different scheduling algorithms at the application level on the proposed tool [8]. Sakellari and Loukas [9] provided a survey on various mathematical model approaches which is helpful for the researchers for further simulations and implementation of the particular modelling techniques. Youse et al. [10] designed a simulation-based cloud resource management model which focuses on dynamic service composition. Huu et al. [11] proposed a scheme constituting a model and an experimentation for energy data centres. In paper [12] Guérout et al. provided a survey on energy aware simulators and techniques with the help of Dynamic Voltage and Frequency Scaling. CloudSched tool has been developed by Tian et al. [15] which is cloud simulation tool for virtual machines in cloud data centres.

Based on the given configurations, CloudAnalyst gives the best scheduling results among the consumer groups. CloudAnalyst is an extension of cloudsim which has a improved GUI feature in it. Both Cloudsim and CloudAnalyst are implemented on Gridsim and SimJava which considers the cloud centre as a huge pool of resources with variety of workloads. GreenCloud has been developed by Zheng et al. [13] at a package level for energy-aware in the cloud data centres. Tian et al. [14] developed a tool called iCanCloud that is favoured for the cloud infrastructure which is been implemented in C++, and the proposed tool was compared with the Cloudsim for its performance.

## 3 Comparison of Cloud Simulator Tools

Cloud tools are divided into different categories based to their features. We have considered five open source simulation tools namely Cloudsim, iCancloud, Greencloud, CloudAnalyst, CloudSched. In this paper, we have studied the architecture, elements, process of simulation and performance metrics of the simulators.

### 3.1 Platform

Cloudsim, CloudSched, and CloudAnalyst are implemented in java, thus they can be executed on any machine. Green cloud is implemented in NS2 simulator and iCancloud in OMNET (Table 1).

**Table 1.** Comparison guideline for simulators

Items	Cloudsim	CloudAnalyst	iCancloud	Greencloud	CloudSched
Platform	Any	Any	OMNET	NS2	Any
Program language	Java	Java	C++	C++/OTCL	Java
Availability	Open source	Open source	Open source	Open source	Open source
Graphics	No	Yes (limited)	No	No	No
Parallel experiment	No	No	Yes	No	No
Energy consumption	Yes	No	No	Yes	Yes
Simulation time	Seconds	Seconds	Seconds	Tens of minutes	Seconds
Memory space	Small	Small	Medium	Large	Small

### 3.2 Programming Language

The Programming languages are meant to their respective platforms. Cloudsim and CloudSched are implemented in Java, whereas tools such as Greencloud is the combination of C++ and OTcl, and iCancloud is implemented in C++.

### 3.3 Availability and Graphical Support

All the simulators discussed in this paper are freely available for the users. Except CloudAnalyst all other tools do not support GUIs. However, there is no full support provided in the CloudAnalyst. Hence it is mentioned as limited during the comparison.

### 3.4 Parallel Experiments

It includes the combination of different machines working simultaneously to process the task. iCanCloud is one such tool which help in parallel experiments, and other simulator do not support this feature.

### 3.5 Energy Consumption Model

Energy consumption model is used to compare different scheduling strategies which is helpful and efficient. Except iCancloud and CloudAnalyst other simulators support energy consumption model.

### 3.6 Migration Algorithms

Migration algorithms are helpful when there is a load in on premise datacentre and needs to be offloaded to the public or private centres for saving the total energy consumption or to improve the utilization of resources. CloudSim, CloudSched and CloudAnalyst algorithms which are helpful in migration while others do not.

## 4 Comparison on Different Elements

### 4.1 Comparison 1: Architecture

Figure 1 shows the layered architecture of Cloudsim. At the fundamental layer, management of application, virtual machines, and hosts are provided. User code represents the entities, generates customer requests and implements applications. Basically in Greencloud, architecture is composed of different layers such as Access layers, where servers are placed, aggregation & cores which includes workloads. In iCancloud, the bottom layer consists of hardware layer which models the hardware part of the system. At the upper layer, there is a cloud hyper-visor which manages all the jobs. In CloudSched the top layer constitutes of interface which allows user to select the resources and scheduling process is done in the lower layer. Once the generation of user request is done, it is forwarded to the next layer (Figs. 2, 3, 4 and 5).

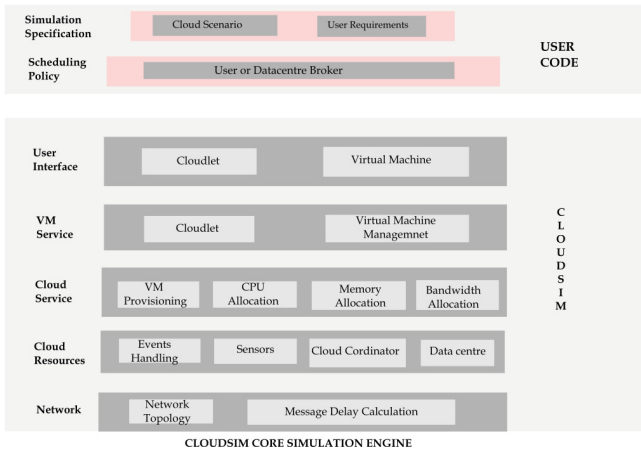


Fig. 1. Architecture of Cloudsim

### 4.2 Comparison 2: Simulator Elements

We discuss main building blocks of each simulator here.

1. Cloud datacenters modeling: CloudAnalyst and Cloudsim, the services rendered by the infrastructure level are simulated and it is carried out by modeling the data centers and each entity involved is a host or data centre. In Greencloud, server, links, workloads, are the basic elements. Here, server is used for task execution and workload for generation of user requests. In CloudSched, the data centre is composed of hosts for managing activities of virtual machines. In iCancloud, data centre represents certain set of virtual machines responsible for allocating and servicing jobs.

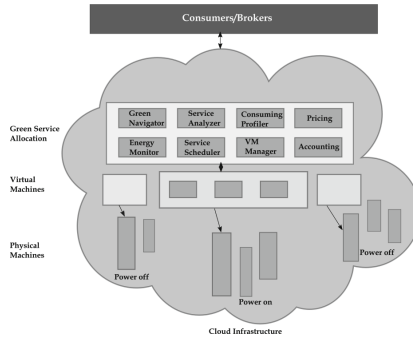


Fig. 2. Architecture of green cloud

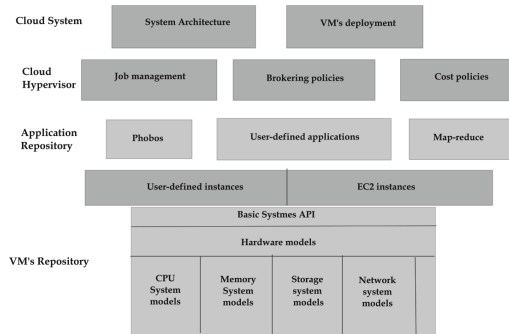


Fig. 3. Architecture of iCancloud

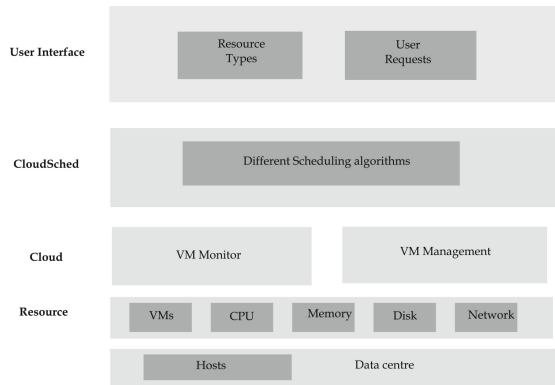
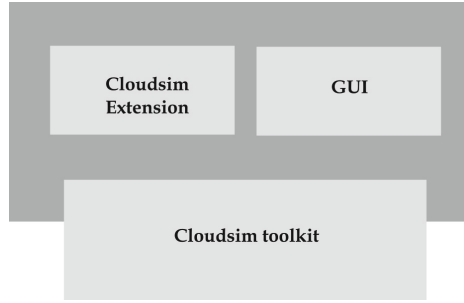


Fig. 4. Architecture of CloudSched



**Fig. 5.** Architecture of CloudAnalyst

2. Customer requirements: Cloudsim and cloud analyst models the requirement by adopting virtual machines by extending the object for implementing the services. iCancloud use virtual machines as the building block for creating the system. This model will be used for configuring the jobs. CloudSched randomly generates various types of virtual machines and allocates it on the basis of certain scheduling algorithms. Greencloud uses workloads arrival pattern to model the requirement. This helps user to adopt various choices for traffic load, network conditions etc. It generates the request in a log file. Overall in order to satisfy the ease of customer requirements, these tools provide interface as well as defined configurations.

### 4.3 Comparison 3: Simulation Procedure

We have divided the simulation process into four different categories.

1. Request generation
2. Datacentre Initiation
3. Allocation
4. Collection of output results

Simulators which we have discussed in this paper adopts all the four parts.

1. Generating requests: Generation of customer requests varies according to the simulator. Cloudsim, CloudSched, CloudAnalyst generates the requests as virtual machines instances and puts into waiting queue. Greencloud produces workloads and iCancloud uses jobs which is then added to a waiting queue which is to be executed.
2. Data centre initiation: Datacentre provides resources. All the five simulators discussed are similar in initializing the data centre and offer resources such as memory, storage etc.
3. Defining allocation: It describes scheduling which includes where and how the request is allocated and processed. Cloudsim, CloudAnalyst, iCancloud implement First Come First Serve allocation policy. Cloudsched adopts load balancing policies, whereas Greencloud implements Dynamic voltage frequency scaling to allocate the request.

4. Output the result: Results are gathered for evaluation of the performance. CloudAnalyst uses limited graphical user interface which enables user to setup the experiment quickly.

#### 4.4 Comparison 4: Performance Metrics

There are various different metrics for load balancing, energy efficient goals. Five simulators which we have discussed here use different metrics. Table 2 summarizes different metrics, objectives and the simulators which adopt these metrics.

**Table 2.** Comparison guideline based on metrics

Metrics	Optimization objectives	Simulator tools
Average resource utilization	Maximize resource	All five
Total number hosts needed	Maximize resource	All five
Average CPU utilization	Load balancing	All five
Make span	Load balancing	CloudSched

1. Resource utilization:
  - Average Resource utilization: Resources such as CPU, memory, harddisk can be computed and used.
  - Total number of hosts used: It describes the utilization of a cloud data-centre.
2. Metrics of load-balancing
  - Utilization of a single CPU: The observed time at which the average load is on a single CPU.
  - Makespan: It is defined as the duration of the execution time on all the hosts. In Cloudsched, it is defined as the maximum number of loads on the hosts.
3. Metrics on energy efficiency
  - Model: Energy consumption model depends on the disk storage, computation, processing and datacentre cooling system.

## 5 Conclusion

In this paper, we have compared five cloud open simulators such as Cloudsim, CloudAnalyst, Greencloud, iCancloud and CloudSched. Comparison is done based on different terms such as architecture, process, components and performance metrics. Overall we can see that none of the tools are perfect in all the aspects. Each tool is helpful in optimizing different objectives such as energy efficiency, load balancing, resource utilization. None of the tool optimize all the objectives.

We have listed few issues for cloud simulators which are as follows:

1. Model all cloud layers: Currently there are no such tools which can model all the cloud layers such as Infrastructure as a Service, Platform as a Service and Software as a Service.
2. Support for federated data centres: Simulation tool should model the federated data centres.
3. Ease of tool: Simulation tool must allow users to use it easily with the GUIs by accepting inputs from Text/Csv files and output to the same. Hence it is useful for the researchers to repeat the experiments and get efficient results.
4. Consideration of priorities: Different types of policies can be created for giving priorities for different virtual machines. Currently the tools which we have discussed in the paper do not consider it yet. Hence more real scenarios can be considered further.

## References

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., et al.: A view of cloud computing. *Commun. ACM* **53**(4), 50–58 (2010)
2. Tian, W.: Adaptive dimensioning of cloud data centers. In: Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC 2009, pp. 5–10. IEEE (2009)
3. Buyya, R., Murshed, M.: GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing. *Concurr. Comput.: Pract. Exp.* **14**(13–15), 1175–1220 (2002)
4. Luo, L., Wu, W., Tsai, W.-T., Di, D., Zhang, F.: Simulation of power consumption of cloud data centers. *Simul. Model. Pract. Theory* **39**, 152–171 (2013)
5. Tian, W., Zhao, Y., Xu, M., Zhong, Y., Sun, X.: A toolkit for modeling and simulation of real-time virtual machine allocation in a cloud data center. *IEEE Trans. Autom. Sci. Eng.* **12**(1), 153–161 (2015)
6. Howell, F., McNab, R.: SimJava: a discrete event simulation library for Java. *Simul. Ser.* **30**, 51–56 (1998)
7. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. *Future Gener. Comput. Syst.* **25**(6), 599–616 (2009)
8. Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A., Buyya, R.: CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Softw.: Pract. Exp.* **41**(1), 23–50 (2011)
9. Sakellari, G., Loukas, G.: A survey of mathematical models, simulation approaches and testbeds used for research in cloud computing. *Simul. Model. Pract. Theory* **39**, 92–103 (2013)
10. Youseff, L., Butrico, M., Da Silva, D.: Toward a unified ontology of cloud computing. In: Grid Computing Environments Workshop, GCE 2008, pp. 1–10. IEEE (2008)
11. Huu, T.N., Ngoc, N.P., Thu, H.T., Ngoc, T.T., Minh, D.N., Tai, H.N., Quynh, T.N., Hock, D., Schwartz, C., et al.: Modeling and experimenting combined smart sleep and power scaling algorithms in energy-aware data center networks. *Simul. Model. Pract. Theory* **39**, 20–40 (2013)



12. Guérout, T., Monteil, T., Da Costa, G., Calheiros, R.N., Buyya, R., Alexandru, M.: Energy-aware simulation with DVFS. *Simul. Model. Pract. Theory* **39**, 76–91 (2013)
13. Zheng, H., Zhou, L., Wu, J.: Design and implementation of load balancing in web server cluster system. *J. Nanjing Univ. Aeronaut. Astronaut.* **38**(3), 347 (2006)
14. Tian, W., Zhao, Y., Zhong, Y., Xu, M., Jing, C.: A dynamic and integrated load-balancing scheduling algorithm for cloud datacenters. In: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 311–315. IEEE (2011)
15. Tian, W., Liu, X., Jin, C., Zhong, Y.: LIF: a dynamic scheduling algorithm for cloud data centers considering multi-dimensional resources. *J. Inf. Comput. Sci.* **10**(12), 3925–3937 (2013)

# Data Consumption Pattern of MQTT Protocol for IoT Applications

Hansa Lysander Manohar<sup>1</sup> and T. Reuban Gnana Asir<sup>2</sup>(✉)

<sup>1</sup> College of Engineering, Guindy, Anna University, Chennai, India

<sup>2</sup> Video Business Unit, Nokia, Chennai, India

gnana\_asir.reuban@nokia.com

**Abstract.** In connecting the networks and people, the HTTP Protocol played a greater role and it is the most widely used protocol for data transfer in variety of applications. In IoT, we got to establish connections between “machines and things”. Their communication requirements were different from current needs of Internet and associated data communications. So, HTTP protocol looks quite heavy for Internet of Things (IoT) applications, due to the overheads of HTTP. Hence, we analysed the data consumption pattern of a light weight protocol Message Queue Telemetry Transport (MQTT) supported by literature study and practical validation using Orange-Pi controlled test bed. The test bed comprises PIR Motion sensor coupled with WeMos microcontroller that sends input to the Orange-Pi Gateway over MQTT Protocol. Along with the test results, this paper summarizes the benefits of using MQTT Protocol, in IoT Applications.

**Keywords:** IoT · Cloud systems · MQTT · HTTP · CoAP

## 1 Introduction

“We stand on the brink of a technological revolution that will fundamentally alter the way we live, work, and relate to one another”, says World Economic Forum, referring the fourth industrial revolution dominated by the recent technologies that include Internet of Things, Artificial Intelligence, etc. [1]. The Internet of Things is poised to dominate the internet technologies in upcoming years due to the enormous evolutions of ‘things’ which is set to cross beyond 28 billion users with internet connectivity by year 2020 [2], ranging from home appliances, connected cars, wearable’s, industrial automation gears, etc. The installed base is poised to grow and exceed 212 billion devices, including the connected devices of 30 billion growth in next 3 years, says the industry analyst firm IDC. It predicts the growth of intelligent systems driven by data collection, data analysis and decision making across both consumer and enterprise applications [3]. Thus the data transfer across the components plays a vital role in the success of IoT applications.

When it comes to the data transfer, HTTP hits the mind due to its successful data transportation in the TCP/IP network over the years. Is the requirement of IoT applications same as the HTTP applications? No. Hence we made a study on the protocols to identify the better suited protocol for the transport layer requirements of IoT

applications. One of the upcoming demand of IoT applications is to cope-up with the growth in the number of connections, growth in the number of devices, growth in the data transfer needs, wherever they are and whenever they are without choking up. This raises significance on the telemetry which allows things to get measured and monitored from a distance. With the improvement in telemetry technology it becomes possible to interconnect the monitoring and measuring devices from different locations.

Dependency on the smart devices and the ability to interact with other devices has raised the quest of smartness of individuals, corporate and Government organizations in every country. A woman shopping for groceries would like to get a view of what is there and what not in her kitchen. A man flying to Chennai wants to know if the flights going to that city are currently affected by weather or not. A doctor wants to know the patient's blood pressure ahead of his planned flight trip to abroad country so ensure his stability. The information that helps to take wise decisions may come from one or other forms of smart meters and equipments.

The challenge lies in the information transfer from the device to the person and to the application in a timely and effective way with increasing demand. The challenge goes to next stage based on the geographic distribution storage and computing power that shoots the cost as well which is a key factor for developing nations like India [4]. Fortunately, the advancement in the communication protocols and telemetry technologies makes it possible to receive and send the information over the internet reliably, despite network disturbance cases, little processing power of the monitoring devices, etc. using MQTT protocol [5].

This paper starts with an overview of MQTT protocol and its components and then tries to discuss the discriminating factors of MQTT Vs HTTP using the literature study. The main contribution of this paper is the MQTT protocol based application testbed and validation that is supported by the test results, observations and inference on the data consumption patterns.

## 2 MQTT Protocol

MQTT (Message Queue Telemetry Transport) Protocol is a light weight and an extremely simple protocol [6]. The publish/subscribe architecture of MQTT is designed with the characteristics of openness and easiness to implement, which can be scaled by single server to support up to thousands of clients accessible from remote. These characteristics of MQTT makes it ideal for usage in constrained environments where there is high latency or low network bandwidth and devices from remote sites that could have limited memory and processing capabilities [7].

The benefits of MQTT protocol includes the following:

- (a) It delivers the data relevant to any intelligent and decision-making asset that can utilize it.
- (b) It extends the connectivity range exceeding enterprise boundaries to reach to smart devices.
- (c) It provides the optimized connectivity options for remote devices and sensors.
- (d) It enables enormous scalability of management and deployment to IoT solutions.

## 2.1 MQTT Highlights

MQTT Protocol claims that, it minimizes the device resource requirements and network bandwidth with attempts to deliver with reliability. This characteristic is the validation goal that is explained in the later parts of this paper.

This approach of minimal resource requirement and reduced network bandwidth makes the MQTT protocol well-positioned for connecting machine to machine (M2M) communications, which is a critical aspect of the IoT.

The other key highlights include [5]:

- Open and royalty-free.
  - MQTT is easy to adopt, open to make and fit for variety of platforms, devices, and operating systems that are used at the network edge.
- Messaging model.
  - The publish/subscribe messaging model facilitates one-to-many distribution. Sender devices or applications need not know anything about the receiving device or applications, not even its address.
- Ideal for constrained networks.
  - MQTT message headers are retained as small as possible and ideal for fragile connections, low bandwidth, data limits, high latency networks. The fixed header is only two bytes, that too on demand, push-style message distribution keeping the network utilization low.
- Multiple service levels.
  - It gives the flexibility in handling various types of messages. For example, developers can design that the messages will be delivered exactly once, at least once, or at most once.
- Design.
  - Its designed to support remote devices with low processing power and minimal memory.
- Ease of use.
  - Usage and implementation is quite easy with simple set of command messages. Various applications of MQTT will be accomplished using CONNECT, DISCONNECT, SUBSCRIBE, UNSUBSCRIBE and PUBLISH methods.
- Built-in support for contact loss.
  - If the connection with client connection breaks abnormally, the information is sent to server facilitating the message either to get preserved for later delivery or to re-send.

## 3 Findings Over HTTP

In this section, let us list down the key factors of comparison between HTTP Protocol and MQTT Protocol for IoT applications based on the research done in this field by several experts in the past and the next section shares the observation based on validation results.

### 3.1 Comparison of MQTT vs HTTP

The below Table 1 gives a quick view on the comparison between MQTT and HTTP.

**Table 1.** Quick view of MQTT vs HTTP

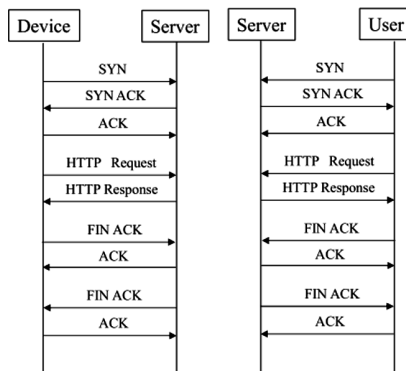
	MQTT	HTTP
Design orientation	Data centric	Document centric
Pattern	Publish/subscribe	Request/response
Complexity	Simple	More complex
Message size	Small, with a compact binary header just two bytes in size	Larger, partly because status detail is text-based
Service levels	Three quality of service settings	All messages get the same level of service
Extra libraries	Libraries for C (30 KB) and Java (100 KB)	Depends on the application (JSON, XML), but typically not small
Data distribution	Supports 1 to zero, 1 to 1, and 1 to $n$	1 to 1 only

- **High Power consumption by HTTP [8]:**

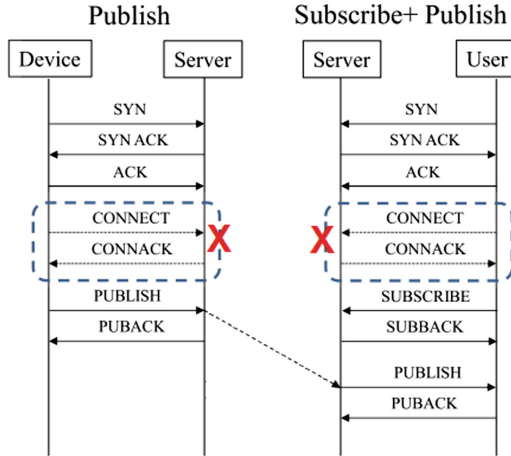
- In the dynamic data communication scenarios, HTTP is observed to be consuming more power. In the tests done by Hantrakul K et al., the HTTP protocol consumes 10 times higher power than MQTT protocol. They have witnessed MQTT sends 10 times more messages than HTTP in 1 h of operation.
- Tests done by Upadhyay et al. [9] reveals, power consumption of MQTT Protocol is lower and 30% faster performance than CoAP [10].

- **High Protocol overheads in HTTP:**

- IoT applications requires large number of information exchange with tiny packets. Hence the payload is quite less, whereas the overhead caused to transfer the payload is quite high.
- From the below Figs. 1 and 2 we see the elimination of CONNECT/CONNACK flow for MQTT cases, that reduces the overhead and latency, when compared to HTTP, leading faster data transfer as well [10].



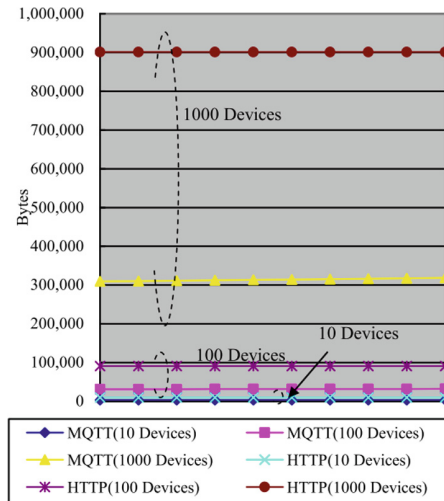
**Fig. 1.** Communication sequences of HTTP use case.



**Fig. 2.** Communication sequences of MQTT use case where CONNECT/CONNACK is eliminated

• **High Bandwidth consumption in HTTP:**

- From the research done by Yokotani and Sasaki [11] on the comparison of bandwidth usage between HTTP and MQTT on 2 different cases, with payload and without payload (where only topics exist, that is used to decide on the MQTT broker, which client receive which message).
- For MQTT topics cases, where zero payload exists and only the transmission bytes exists reveals, HTTP consumes 300% higher bandwidth as in Fig. 3.



**Fig. 3.** Characteristics with zero payload

- For MQTT message sharing cases, where pay load and transmission bytes exists, HTTP consumes 250% higher bandwidth as in Fig. 4.

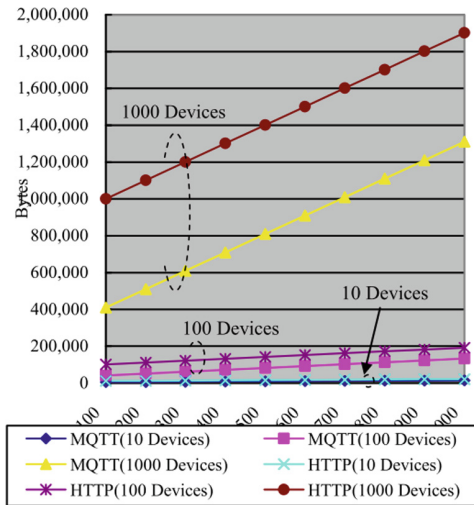


Fig. 4. Characteristics with payload and overhead.

When these studies reveals that MQTT is better choice that HTTP for IOT applications, we wished to get into it to detail to understand any additional behavior of MQTT and got an interesting observation, that will be explained in next chapter.

## 4 Experiment and Test Results

### 4.1 Test Environment

The test environment mainly comprises of following components as in Fig. 5.

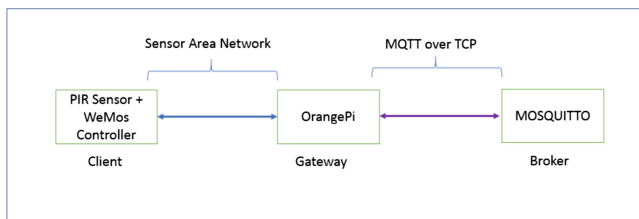
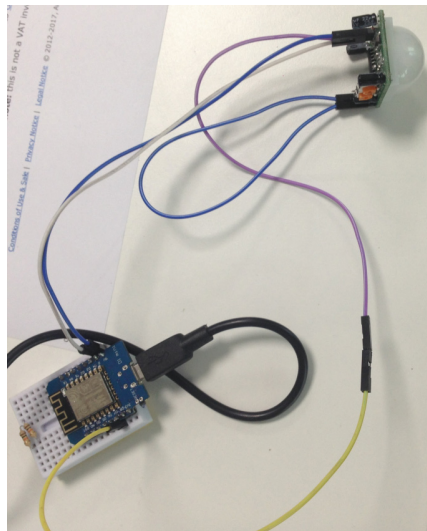


Fig. 5. Test bed overview

- PIR Sensor HC-SR501
- WeMos D1 mini ESP8266
- Orange Pi Zero Processor as MQTT Gateway
- Mosquitto MQTT Broker v3.1

### PIR Sensor HC-SR 501 + WeMoS ESP8266 Controller

The short distance communication is realized by the Wireless Sensor Networks (WSN) among the objects nearby. In this experiment, we've picked PIR motion sensor that contains low-cost Wi-Fi chip with full TCP/IP stack. The PIR motion sensor detects the presence of anyone coming closer or moving away and will send the signal. However, it's difficult to connect each other with the mobile communication networks, the Internet and WSN because there is not much standardization exists with respect to communication protocols and sensing technologies. The other restriction is from the data transmission from WSN in long distance due to the limitation of WSN's transmission protocols. Therefore, we house WeMos D1 mini controller as in Fig. 6, whose aim is to balance the heterogeneity between mobile communication, sensor network and Internet that strengthens the management of the terminal nodes, WSN and bridge [12].

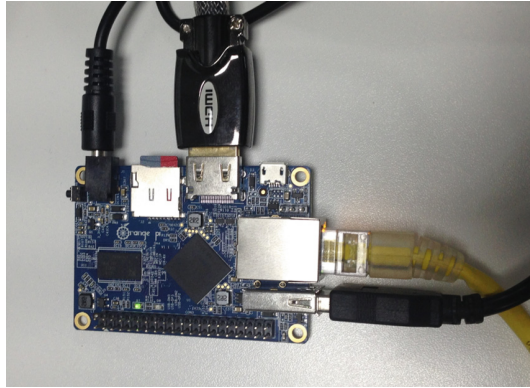


**Fig. 6.** PIR sensor and WeMos controller

### OrangePi Zero

Its an open source single board computer that can run on Operating Systems, that includes Android, Ubuntu, Debian, Armbian. This acts as IoT Gateway that converts the input from sensory network and passes to MQTT broker over TCP. In our tests, we tried to hook into these conversations using TCP dump and observed the pattern of MQTT communications as in Fig. 7.

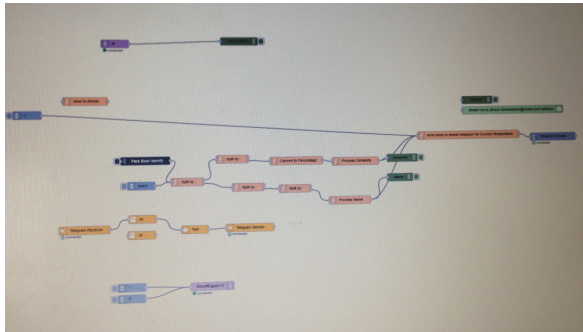




**Fig. 7.** MQTT Gateway with Orange-Pi

### Mosquitto Broker

This is an open source message broker that implements MQTT protocol versions 3.1 and 3.1.1. It carries out the Publish/Subscribe model that makes it suitable for messaging with low power sensors, mobile devices, embedded computers and Arduino micro controllers. To write the control logic, we used the Node-RED as in Fig. 8, that facilitates the flow based programming method with ease of use.



**Fig. 8.** Node red flow using Mosquitto MQTT Broker

## 4.2 Test Results

From the tests performed we tried to collect the TCP dump for the conversations between IoT Gateway and MQTT Broker as in Fig. 9.

The interesting observations are:

1. The sensor could sense either an object coming near or moving away and it sends the signal, which we see in the Fig. 10 as ON and OFF. Whenever the sensor senses a signal, it transmits to MQTT broker with TCP message of length 28 bytes.
2. When there is no signal change read by sensor, the TCP message with 0 bytes is transmitted, which is presumably the CONNACK.

```
root@OrangePi-mini:~# tcpdump -i eth0 -s 0 -n -v -e -X 'port 1883'
14:51:24.988731 IP 192.168.1.101.1883 > 192.168.1.74.18030: Flags [P.] seq 831, win
ack 830, win 14600, length 28
14:51:25.014693 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [.] ack 831, win
4868, length 0
14:51:27.721193 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [P.] seq 831:858
ack 831, win 4868, length 28
14:51:27.721772 IP 192.168.1.101.1883 > 192.168.1.74.18030: Flags [P.] seq 831:859
ack 858, win 14600, length 28
14:51:27.763282 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [.] ack 859, win
4840, length 0
14:51:34.987336 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [P.] seq 858:886
ack 859, win 4840, length 28
14:51:34.988001 IP 192.168.1.101.1883 > 192.168.1.74.18030: Flags [P.] seq 859:887
ack 886, win 14600, length 28
14:51:35.015302 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [.] ack 887, win
4812, length 0
14:51:37.714436 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [P.] seq 886:914
ack 887, win 4812, length 28
14:51:37.715023 IP 192.168.1.101.1883 > 192.168.1.74.18030: Flags [P.] seq 887:915
ack 914, win 14600, length 28
14:51:37.773745 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [.] ack 915, win
4784, length 0
14:51:45.492388 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [P.] seq 914:942
ack 915, win 4784, length 28
14:51:45.492817 IP 192.168.1.101.1883 > 192.168.1.74.18030: Flags [P.] seq 915:943
ack 942, win 14600, length 28
14:51:45.560836 IP 192.168.1.74.18030 > 192.168.1.101.1883: Flags [.] ack 943, win
4756, length 0
```

Fig. 9. TCP dump collected from MQTT Gateway

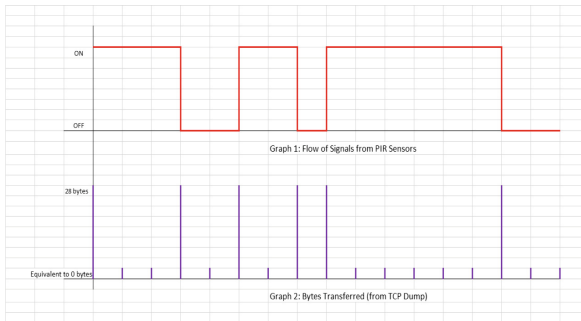


Fig. 10. Graphical representation of TCP message length

This pattern is graphically represented in Fig. 10 and we realize MQTT consumes 28 bytes data, only when there is signal ON/OFF to be transmitted. On all other cases, it remains very minimal with byte length as zero. This pattern is another feather on MQTT’s effective data consumption trend.

## 5 Future Work

In the coming days, we wish to extend the test bed integrated with actuators like SONOFF switches and observe the data consumption pattern between the MQTT broker and the actuators. With additional efforts, this test bed will get integrated with multiple types of IoT applications, other protocols used in IoT applications like CoAP and make a study on the data consumption patterns in both simple cases and under traffic situations.

## 6 Conclusion

With the findings of the MQTT Data consumption pattern, it is evident that MQTT Protocol is essential for effective unitization of network bandwidth, to reach out to devices at remote locations, low power enabled devices. Usage of HTTP as the transport layer for IoT Applications will help for initial stages only. With the increasing trend of number of IoT applications it is a value add to switch to MQTT to match various usecases that pops-up like IaaS [13]. The REST API support of MQTT to Push or Pull the data as and when required, is getting leveraged by industry champions in Cloud Platforms include Aercloud, IBM IoT [14].

### Abbreviations and Acronyms

CoAP - Constrained Application Protocol  
HTTP - Hyper Text Transfer Protocol  
IaaS - IoT As A Service  
IoT - Internet of Things  
MQTT - Message Queue Telemetry Transport  
TCP - Transmission Control Protocol  
WSN - Wireless Sensory Networks

**Acknowledgment.** We would like to thank Nokia, College of Engineering Guindy, for giving us such an opportunity to carry out this research work and for providing us the requisite resources and infrastructure for carrying out the research. Special thanks to Mr. Dinesh Birlasekaran, Mr. Wilson Anandaraj from Nokia for being a mentor and continuous inspiration towards the research on IoT.

## References

1. <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
2. The Internet of Things: Making sense of the next mega-trend. Goldman Sachs Global Investment Research, IoT Primer, September 2014
3. IDC, Worldwide Internet of Things (IoT) 2013–2020 Forecast: Billions of Things, Trillions of Dollars, Doc # 243661, October 2013
4. Reuban Gnana Asir, T., Anandaraj, W., Naga Sivaranjani, K.: Internet of things and India's readiness. In: International Conference on Computing Paradigms (ICCP2015), pp. 274–279, July 2015
5. Lampkin, V., Leong, W.T., et al.: Building smarter planet solutions with MQTT and IBM WebSphere MQ telemetry, September 2012
6. <https://en.wikipedia.org/wiki/MQTT>
7. Barata, D., Louzada, G., Carreiro, A., Damasceno, A.: System of acquisition, transmission, storage and visualization of pulse oximeter and ECG data using android and MQTT. *Procedia Technol. J.* **9**, 1265–1272 (2013)

8. Hantrakul, K., Sitti, S., Tantitharanukul, N.: Parking lot guidance software based on MQTT protocol. In: 2017 International Conference on Digital Arts, Media and Technology (ICDAMT) (2017)
9. Upadhyay, Y., Borole, A., Dileepan, D.: MQTT based secured home automation system. In: 2016 IEEE Symposium on Colossal Data Analysis and Networking, CDAN 2016 (2016)
10. Amaran, M.H., Noh, N.A.M., Rohmad, M.S., Hashim, H.: A comparison of lightweight communication protocols in robotic applications. *Procedia Comput. Sci.* **76**, 400–405 (2015)
11. Yokotani, T., Sasaki, Y.: Comparison with HTTP and MQTT on required network resources for IoT. In: International Conference on Control, Electronics, Renewable Energy, and Communications 2016, Conference Proceedings, CCEREC 2016 (2016)
12. Zhu, Q., Wang, R., et al.: IOT gateway: bridging wireless sensor networks into internet of things. In: IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, pp. 347–352 (2010)
13. Reuban Gnana Asir, T., Manohar, H.L., Anandaraj, W., Naga Sivaranjani, K.: “IoT as a Service” Internet of things and India’s Readiness. In: International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS 2016), pp. 1093–1096 (2016)
14. Ray, P.P.: A survey of IoT cloud platforms. *Future Comput. Inf. J.* **1**(1–2), 35–46 (2017)

# Data Access in Heterogeneous Data Sources Using Object Relational Database

M. S. Hema<sup>1</sup>(✉), R. Maheshprabhu<sup>1</sup>, and M. Nageswara Gupta<sup>2</sup>

<sup>1</sup> Aurora's Scientific Technological and Research Academy, Hyderabad, India  
ghema\_shri@yahoo.co.in

<sup>2</sup> Sri Venkateshwara College of Engineering, Bangalore, India

**Abstract.** Data sourcing or integration is inevitable in current business scenario. Major issues of data sourcing from heterogeneous data sources are lack of semantic richness and deprived querying. To overcome these issues, an Ontology Based Data federation using Object Relational Database (OBDF-ORDB) architecture has been proposed and implemented. This OBDF-ORDB architecture consists of semantic layer and transformation & query layer. In semantic layer the ontology used to create local and global schema to enrich the semantics. In transformation & querying layer, Object Relational Database (ORDB) is used for storing the local ontology, global ontology to improve storage, maintenance and retrieval. The transformation rule engine proposed for the architecture converts and stores the local ontology and global ontology from flat OWL file to ORDB. The user queries are passed to ORDB for result extraction. To analyze the performance of the OBDF-ORDB architecture E-shopping application is selected. Experimental results shows that the proposed OBDF-ORDB architecture is relatively better than the traditional data access and ontology based data access in recall and response time. It is observed that the recall mechanism in OBDF-ORDB architecture has been improved by 25% compared to traditional data federation and response time is reduced by 15% compared to ontology based data federation.

**Keywords:** Data integration · Federation · Ontology · Heterogeneous data  
Object Relational Database

## 1 Introduction

Now a days, the need for accessing heterogeneous database information available in multiple and distributed sources are increasingly higher. Particularly in the context of decision making applications exploration of data is required. The data integration is a possible solution to address the above issue. The data integration is a process of combining data from the heterogeneous data sources [1]. The three types of data integration methods are 1. Data propagation 2. Data consolidation and 3. Data federation [2]. Data consolidation collects the data and create consolidated data warehouse, data propagation replicates data in the target data source, whereas the data federation provides a virtual view for two or more data sources. The user submits the queries against virtual view and gets the result. The creation of virtual view among different

data sources is a challenging task due to their heterogeneity nature. Various types of heterogeneities are data model heterogeneity, syntactical heterogeneity and logical heterogeneity. The logical heterogeneities are further divided into semantic heterogeneity, schematic heterogeneity and structural heterogeneity [3, 4]. Among these heterogeneities, the semantic heterogeneity is very difficult to resolve. The semantic heterogeneity is caused by different interpretation or meaning of data [5]. The semantic heterogeneities are divided into structural level heterogeneity and data level heterogeneity [6]. Ontology is used to resolve the semantic heterogeneities. Ontology is a formal specification of conceptualization [5, 7]. It is used in sharing of data or information and reusability of domain knowledge. The ontology is used to represent domain knowledge to resolve semantic heterogeneities in data federation.

The ontology architectures for data federation are single ontology approach, multiple ontology approach and hybrid ontology approach [4]. Flat files and database are widely used to store ontology. If flat file size is larger, the storage, maintenance and retrieval become complex task. A possible remedy is to store ontology in database that enhances easy maintenance and retrieval [8, 9].

In this paper to achieve improved recall and response time of data federation, ontology based data federation by using ORDB is implemented. The components and steps for building OBDF-ORDB architecture is illustrated in detail.

#### Contributions

1. Despite of numerous works on ontology based data federation a smooth extension from traditional ontology based data federation is proposed and implemented. The strengths and weakness of the proposed has been investigated.
2. This paper suggests ontology as a solution for resolving semantic heterogeneities and to store in ORDB. It improves recall and response time of ontology based data federation to a considerable level.

In addition to the introduction section, there are five more sections in this article. The remainder of the paper is organized as follows. The related work in both data federation and ontology based data federation is presented in Sect. 2. Section 3 describes OBDF-ORDB architecture. It contains ontology construction for data federation, storing ontology in ORDB, query processing and result integration. Section 4 uses a demonstration case to illustrate the architecture. Section 5 discusses the contributions of the applications in the proposed architecture. The conclusion is presented in Sect. 5.

## 2 Related Work

Surveys on data federation based ontology are found in [3, 10–12]. An extensive semantic search model was anticipated. The keyword search was used for syntactic matching and semantic search was used for interpretation of meaning of the term. The benefits of both keyword and semantic based search were explained [13]. Ontology based data federation using mediator based architectures was proposed. The mediator architecture has interface layer to integrate all local schema and abstracts the semantic complexity [7, 14, 15].

The global ontology was created via Local As View [LAV] [4] or Global As View [GAV] [6, 16, 17] to retrieve the data from the respective local ontology. For generating global ontology from local ontology, the shared vocabulary [18] was provided. A framework was proposed and implemented for mapping local source and global ontology [19]. For computing similarities among various ontology specifications a method was employed for ensuring the performance such as reusability and accuracy [20, 21]. An algorithm was proposed and implemented for ontology classification based on their domain [22]. Efforts were also made to store ontology in the database as an independent entity [8]. A method was proposed for automatic data migration from data intensive application to semantic web [23, 24] and OWL ontology [25, 26] was designed using mapping rule engine. Ontologies were stored in relational databases for rapid query processing [27–29]. Later the Object Relational Databases were used to realize real time entities and mapping of ontology to ORDB was implemented [30].

A tool was proposed to combine the intelligent techniques to support domain expert for constructing ontologies [31]. A method was proposed to convert SPARQL query to SQL query [32]. A conceptual model was proposed for mapping various data integration applications [33]. The overview of Ontology Based Data Access is proposed and the main challenges that are yet to be addressed were also discussed [34].

The Limitations of ontology based data integration are

If the ontology is stored in flat files, it is difficult to maintain and retrieve when file size increases. If the ontology is stored in a relational database, it is efficient for storage, maintenance and retrieval but it has some limitations such as (1) there is no direct relationship between relational database and real world objects. (2) There is no provision to store methods internally and (3) If ontology is stored in a relational database, it may lose some of the relationship among attributes and tables.

### 3 Proposed Methodology

The main objectives of this proposed methodology are design OBDF-ORDB architecture and to enhance the recall and response time of ontology based data integration. The OBDA-ORDB architecture is shown in Fig. 1. This architecture consists of three layers. The bottom most layer is data source layer that contains heterogeneous data sources. The middle layer is wrapper mediator layer, which creates local and global ontology and maps local and global ontology in the semantic layer. The semantic layer ensures semantic richness by resolving heterogeneities such as semantic, schematic, schema isomorphism and structural discrepancies using ontology. The semantic layer is further divided into local ontology service, mapping service and global ontology service, which are discussed in Sects. 3.1, 3.2 and 3.3 respectively. The transformation and querying layer converts OWL ontologies stored in flat files into ORDB relations by using mapping rules. ORDB maps local ORDB relation with global ORDB relation and store all records as instances in the local ontology ORDB. This layer improves data retrieval time using ORDB. This layer is further classified into transformation rule engine service, local ORDB service, local ORDB to global ORDB mapping service and global ORDB service that are discussed in the Sects. 3.4 and 3.5. The user request in SQL3 is presented in the top layer for data extraction. The extracted results are integrated using result integration service and send integrated result to user for decision support and analysis.

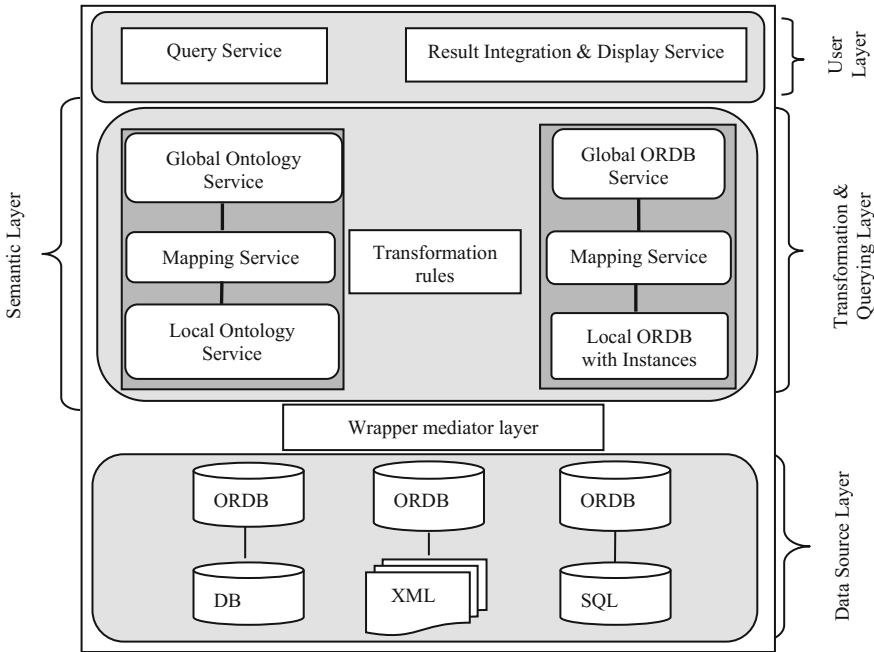


Fig. 1. OBDF-ORDB architecture

### 3.1 Creation of Local Ontology

The local ontology is created from the local schema of the respective data sources. It consists of two steps: first step is to complete analysis of data sources which means that what data is stored, how it is stored and the meaning of the data (semantics). The second step creates local ontology for the respective data sources. It is a semi-automatic process which analyze and extracts the data manually, also creates local ontology automatically by using protégé tool. The rules to create local ontology are shown in Table 1 (Subset of these rules is adopted from [7]).

### 3.2 Building Mapping Rules

Mapping rules are defined to resolve the semantic heterogeneities among local ontologies. Due to independent construction of data sources their own local ontology is used to represent their entity, attribute and the relation.

#### Rule – I

If the class properties of C1 and C2 are semantically equivalent then the local ontology are mapped with single global ontology class to resolve naming conflict.

#### Rule – II

A few local ontology properties of a class are to be concatenated before mapping with global ontology to resolve structural discrepancies.



**Table 1.** Local schema to local ontology mapping

S. No	Data source schema elements	Equivalent OWL statement
1	Table	Class
2	Data type	Data property
3	Foreign key	Object property
4	Primary key, unique Key	Inverse functional property
5	Check constraints	Value restriction
6	NOT NULL	Required property

**Rule – III**

Some classes have unique properties that are appended in global ontology.

**Rule – IV**

Some classes have common properties but they are placed at different positions. In global ontology these properties are mapped together to resolve schema isomorphism.

**3.3 Global Ontology Creation**

The main objective of data federation is to provide a virtual view for the users to access data, regardless of its actual organization and location. This is accomplished by creating a global ontology, which has been created using Hybrid ontology approach [19]. The global ontology is built by analyzing the local ontology in which each class and its associated properties in the local ontologies should be mapped to the global ontology. The mapping between local ontologies and global ontology has been done by using mapping rules which was explained in Sect. 3.1.

**3.4 Creation of Transformation Rule Engine Service**

In this service, set of rules has been defined to convert global and local ontologies that are stored in OWL flat file into ORDB relations.

**Axiom – I**

Ontology class map to ORDB relation, each property in an ontology class maps to each column in the ORDB relation and an additional unique column ID is created. This is used for retrieving the contents uniquely from the ORDB relation.

**Axiom – II**

If an inherited class in an ontology maps to ORDB relation, an ID different from its base class is appended along with the properties of the inherited class.

**Axiom – III**

The properties defined in ontology are of two types namely scalar and vector. The vector type properties are converted to array of rows in ORDB.

**Axiom – IV**

If a scalar property of one class is referring to scalar property of different class in ontology then the referential key relation is created in ORDB between the two relations.

**Axiom – V**

The scalar properties in ontology are subjected to value restriction. These scalar properties are converted to attributes with check constraints in ORDB.

**Axiom – VI**

Similar to value restriction this is added as constraint to the attribute in relation or ORDB.

**Axiom – VII**

Inverse functional property is stored as NOTNULL constrains along with the attributes.

**3.5 Creation of Local ORDB, Global ORDB and Mapping**

The local ORDB and global ORDB are created from the local ontology OWL file and global ontology OWL file respectively by using transformation rule that is defined in Sect. 3.3. After creating local ORDB, the records of the native data sources are converted into ORDB format and are attached as instances of the respective local ORDB. The mapping between local ORDB and global ORDB is defined.

**3.6 Query Processing**

The query processing is performed in transformation and querying layer. In this layer, query is received from the user. The received query is posted against the global ORDB. The query is divided into subqueries based on the mapping rules and directed to the local ontologies. The results are retrieved from the databases and it is directed for integration.

**3.7 Result Integration and Display**

It integrates the results from the transformation and query layer. The integration is performed using union operation. It automatically eliminates duplication.

$$R = r_1 \cup r_2 \cup r_3 \cup \dots \cup r_n. \tag{1}$$

R-integrated result.

r<sub>1</sub>, r<sub>2</sub>, r<sub>3</sub>.....r<sub>n</sub> - results from the different local ORDB.

**4 Results and Discussions**

Few electronic gadget e-shopping enterprises have been chosen for experimentation. Chosen enterprises have autonomously developed heterogeneous databases. The following tables are selected from those enterprises and implemented the prototype.

Item\_Category (category\_identifier, category\_cname, category\_desc)

Customer\_detail (customer\_identifier, Customer\_cname, Customer\_address, customer\_cphone\_no, Cutomer\_cemail\_id)

Products\_details (product\_identifier, category\_identifier, product\_desc, pbrand, pprice)

Order\_details (order\_identifier, product\_identifier, customer\_identifier, no.of\_products).

Here three databases are designed using heterogeneous DBMS (ORACLE, MYSQL, SQL SERVER). In these databases the table and attributes have different names and it is structurally organized in a different mode. These data source with 4000 records in each are selected for experimentation.

Local and Global ontology have been constructed using protégé 4.2 tool. Local ontology is created from a local schema of the data sources by using rules that is described in Sect. 3.1. It includes 4 classes, and 18 properties (4 object properties and 14 data properties) and the following OWL constructs: 'inverseOf', functional property [22]. The local ontology and data source mapping has been implemented by using Portege 'ontop' plug in. Extraction of classes, properties from local OWL ontology file and OWL global ontology file are performed using java, OWL-API and Jena. The java wrapper class datum method is used for mapping ontology class and Object-Relational objects in ORDB.

Example for information extraction from product OWL file

```
Class = products
Object property = product_id
Data property = brand
Data property = product_description
Data property = category_id
Data property = price
```

Example for transformation from extracted information into ORDB relation.

```
CREATE TYPE product_id AS VARRAY (12) OF REF order_items.
CREATE TYPE products AS OBJECT (Id ref (products), product_idproduct_id,
brand VARCHAR, product_description VARCHAR, category_id NUMBER, price
NUMBER).
```

```
CREATE TABLE products_table OF products.
```

#### 4.1 Recall

The ontology based data federation and traditional data federation recall is calculated for combination of 20 query sets for simple, aggregated functions and sub query set. The recall has been calculated by using following formula

$$Recall = \frac{((No. of relevant records) \cap (No. of records retrieved))}{(No. of records retrieved)} \quad (2)$$

The recall of the above two methodologies is compared and shown in Fig. 2.

The ontology based data federation solves all semantic heterogeneities and it improves the recall by 20.8%.

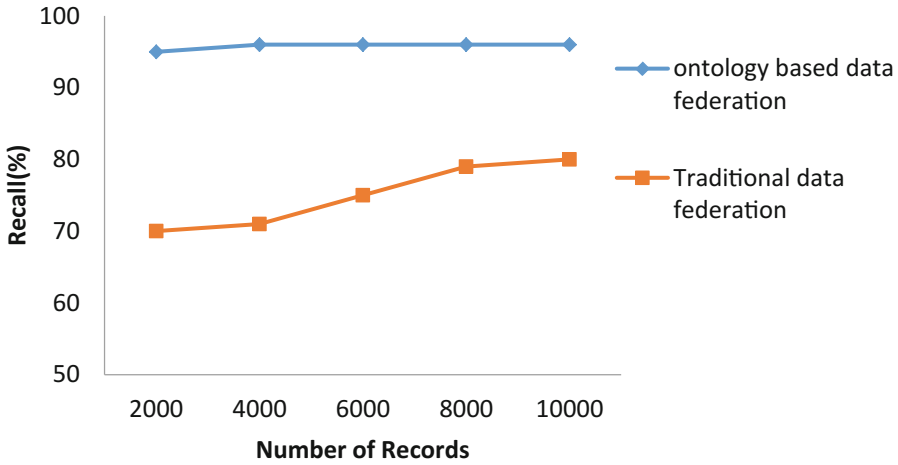


Fig. 2. Recall rate

### 4.2 Execution Time

The response time of the query has been calculated for hybrid ontology based data federation and ontology based data federation by using ORDB. In the former one the ontology is stored in flat files and in the later one the ontology is stored in ORDB. The query response time has been measured with the help of java function ‘*System.currentTimeMillis ()*’. The comparisons for simple queries, aggregated queries and sub queries are shown in Figs. 3, 4 and 5 respectively.

Experimental results conclude that the response time for simple query, aggregate query and sub query has been improved by 20%, 10% and 10% respectively in ORDB approach than the hybrid ontology method.

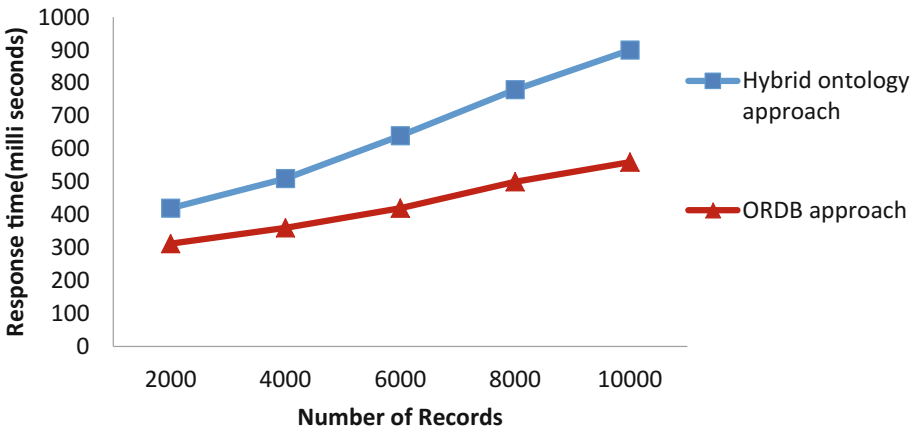


Fig. 3. Average response time for simple query sets

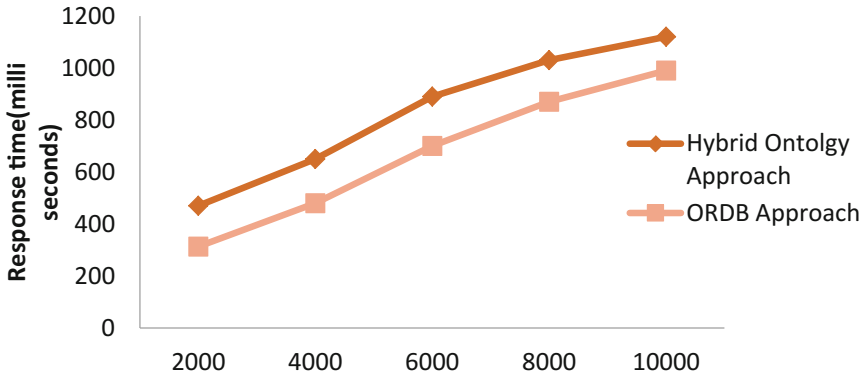


Fig. 4. Average response time for aggregate query sets

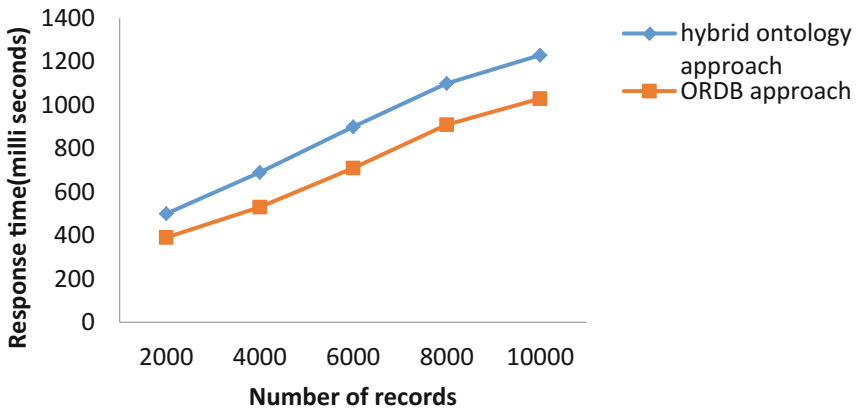


Fig. 5. Average response time for sub query sets

## 5 Conclusion

The proposed OBD-ORDB architecture has been implemented for ontology based data federation from heterogeneous data sources. The findings are 1. The local ontology has been created and hybrid ontology approach is used to create global ontology. 2. The ontology OWL flat file has been transformed and stored in ORDB using transformation rule engine. This improves the semantic richness and querying capability. The results show that the proposed system improved the recall by 20% when applied to the E-shopping example and response time by 15% respectively when compared to traditional data federation system. In future, the cache may be included for retrieval of results and it may be compressed to improve the space utilization.

## References

1. Lenzerini, M.: Data integration a theoretical perspective. In: Proceedings of the Twenty-First Symposium on Principles of Database Systems, pp. 233–246. ACM SIGMOD-SIGACT-SIGART, New York (2002)
2. Hema, M.S., Chandramathi, S.: Federated query processing service in service oriented business intelligence. In: Das, V.V., Stephen, J., Chaba, Y. (eds.) CNC 2011. CCIS, vol. 142, pp. 337–340. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-19542-6\\_62](https://doi.org/10.1007/978-3-642-19542-6_62)
3. Busse, S., Kutsche, R.D., Leser, U., Weber, H.: Federated information systems: concepts, terminology and architectures. *Forschungsberichte des Fachbereichs Informatik* **99**(9), 1–38 (1999)
4. Gagnon, M.: Ontology-based integration of data sources. In: 10th International Conference on Information Fusion, pp. 1–8. IEEE (2007)
5. Xiao, H.: Query processing for heterogeneous data integration using ontologies, Ph.D. thesis, University of Illinois, Chicago (2006)
6. Hu, G.: Global schema as an inversed view of local schemas for integration. In: International Conference on Software Engineering Research, pp. 206–212. SERA (2006)
7. Song, F., Zacharewicz, G., Chen, D.: An ontology-driven framework towards building enterprise semantic information layer. *J. Adv. Eng. Inform.* **27**(1), 38–50 (2013). Elsevier
8. Konstantinou, N., Spanos, D.E., Chalas, M., Solidakis, E., Mitrou, N.: VisAVis: an approach to an intermediate layer between ontologies and relational database contents. In: WISM, p. 239 (2006)
9. Vysniauskas, E., Nemuraite, L., Paradauskas, B.: Hybrid method for storing and querying ontologies in databases. *J. Electron. Electr. Eng.* **9**, 67–72 (2011)
10. Sheth, A.P., Larson, J.A.: Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Comput. Surv.* **22**(3), 183–236 (1990). ACM
11. Kashyap, V., Sheth, A.: Semantic and schematic similarities between: a context-based approach. *Int. J. Very Large Data Bases* **5**(4), 276–304 (1996)
12. Hull, R., King, R.: Semantic database modeling: survey, applications, and research issues. *ACM Comput. Surv.* **19**, 202–260 (1987)
13. Fernandez, M., Cantador, I., Lopez, V.: Semantically enhanced information retrieval: an ontology-based approach. *J. Web Semant.: Sci. Serv. Agents World Wide Web* **9**(4), 434–452 (2011)
14. Wiederhold, G.: Mediators in the architecture of future information systems. *IEEE Comput.* **25**(3), 38–49 (1992). IEEE
15. Langegger, A.: Virtual data integration on the web-novel methods for accessing heterogeneous and distributed data with rich semantics. In: International Conference on Information Integration and Web based Integration System, WAS 2008, pp. 559–562. ACM (2008)
16. Hua, Z., Ban, J.: Ontology-based integration and interoperation of XML data. In: Sixth International Conference on Semantics, Knowledge and Grids, Beijing, pp. 422–423. IEEE (2010)
17. Pinheiro, J.C., Vidal, V.M., Macêdo, J.A., Sacramento, E.R., Casanova, M.A., Porto, F.A.: Query processing in a three-level ontology-based data integration system. In: Proceedings of the 12th International Conference on Information Integration and Web-Based Applications & Services, pp. 283–290. ACM (2010)
18. Zhang, L., Ma, Y., Wang, G.: An extended hybrid ontology approach to data integration. In: International Conference on Biomedical Engineering and Informatics, BMEI 2009, pp. 1–4 (2009)

19. Zhao, Y., Zhang, S., Yan, Z.: Ontology – based model for resolving the data-level and semantic-level conflict. In: International Conference on Information and Automation. IEEE (2009)
20. Rodriguez, M.A., Egenhofer, M.J.: Determining semantic similarity among entity classes from different ontologies. *IEEE Trans. Knowl. Data Eng.* **15**(2), 442–456 (2003). IEEE
21. Harrison, R., Chan, C.: Distributed ontology management system. In: Proceedings of 18th Annual Canadian Conference on Electrical and Computer Engineering, Saskatoon, Canada, pp. 661–664 (2005)
22. Glimm, B., Horrocks, I., Motik, B., Shearer, R., Stoilos, G.: A novel approach to ontology classification. *Web Semant.: Sci. Serv. Agents World Wide Web* **14**, 84–101 (2012)
23. Stojanovic, L., Stojanovic, N., Volz, R.: Migrating data-intensive web sites into the semantic web. In: Proceedings of the 2002 ACM Symposium on Applied Computing, pp. 1100–1107. ACM (2002)
24. Dou, D., LePendu, P., Kim, S., Qi, P.: Integrating databases into the semantic web through an ontology-based framework. In: 22nd International Conference on Data Engineering Workshops Proceedings, p. 54. IEEE (2006)
25. Ghawi, R., Cullot, N.: Database-to-ontology mapping generation for semantic interoperability. In: Third International Workshop on Database Interoperability (InterDB 2007), vol. 91 (2007)
26. Xu, Z., Zhang, S., Dong, Y.: Mapping between relational database schema and OWL ontology for deep annotation. In: International Conference on Web Intelligence, IEEE/WIC/ACM, pp. 548–552. IEEE, December 2006
27. Wang, S., Zhang, X.: A high efficiency ontology storage and query based on relational database. In: International conference on Electrical and Control Engineering, pp. 4253–4256 (2011)
28. Al-Jadir, L., Parent, C., Spaccapietra, S.: Reasoning with large ontologies stored in relational databases: the OntoMinD approach. *Data Knowl. Eng.* **69**(11), 1158–1180 (2010)
29. Astrova, I., Kalja, A., Korda, N.: Automatic transformation of OWL ontologies to SQL relational databases. In: IADIS European Conference on Data Mining (MCCSIS), pp. 5–7 (2007)
30. Jia, C., Yue, W.: Rules-based object-relational databases ontology construction. *J. Syst. Eng. Electron.* **20**(1), 211–215 (2009)
31. Denaux, R., Dolbear, C., Hart, G., Dimitrova, V., Cohn, A.G.: Supporting domain experts to construct conceptual ontologies: a holistic approach. *Web Semant.: Sci. Serv. Agents World Wide Web* **9**(2), 113–127 (2011)
32. Wang, J., Zhang, Y., Miao, Z., Lu, J.: Query transformation in ontology-based relational data integration. In: Asia-Pacific Conference on Wearable Computing Systems (APWCS), pp. 303–306. IEEE (2010)
33. Calhau, R.F., de Almeida Falbo, R.: An ontology-based approach for semantic integration. In: 14th IEEE International Conference on Enterprise Distributed Object Computing (EDOC), pp. 111–120. IEEE (2010)
34. De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rosati, R.: Using ontologies for semantic data integration. In: Flesca, S., Greco, S., Masciari, E., Saccà, D. (eds.) *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*. SBD, vol. 31, pp. 187–202. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-61893-7\\_11](https://doi.org/10.1007/978-3-319-61893-7_11)

# Optimization of UAV Video Frames for Tracking

A. Ancy Micheal<sup>(✉)</sup> and K. Vani

Department of Information Science and Technology,  
Anna University, Chennai 600025, India  
ncysus17@gmail.com

**Abstract.** In this digital era, UAV is becoming a trend setter in surveillance and gathering traffic information. Shortage of time to view the entire video necessitates video optimization. In this paper, a novel method has been proposed for optimizing the video frames without variation in the tracking path of the object. The keyframes are extracted using absolute difference of histogram of consecutive frames with mean as threshold. Then principal keyframes are selected at regular interval and finally compiled into an optimized video. In the tracking session, the region of interest is obtained from the first frame of the video and the SURF features are extracted and tracked with KLT tracker. A SURF feature is tracked along the video and position is tabulated. The tracking path is represented graphically to evaluate the tracking deviation from original and optimized video. The proposed method had successfully achieved the average time saved as 90.68% with negligible tracking deviation.

**Keywords:** Unmanned Aerial Vehicle (UAV) · Keyframes extraction  
Histogram · Speeded Up Robust Feature (SURF)  
Kanade Lucas Tomasi tracker (KLT)

## 1 Introduction

The dominance of Unmanned Aerial Vehicle in this new media age is aggrandizing. Its potentiality of reaching the areas unvisited or unaccessed by human increases its popularity. Small size, low cost and large landscape coverage pertains unmanned aerial vehicle in a favored position than stationary cameras. The evolution of drones are drastic from world war age to current digitized era. The application of drones are limited due to climatic conditions, battery life, limited UAV payload and physical obstacles [1]. Though there are disadvantages, the informations of unaccessed areas and its flexibility overshadows all the downside of drones. Even though the UAVs were initially developed for armed forces, it is gaining worldwide demand in commercial use.

The implementation of drones have widened in areas such as oceanography, forest ecology, traffic monitoring, criminal investigation, military surveillance and traffic analysis. Drones are used to monitor traffic and accidental control on Expressway, the mapping of landslide affect area, large scale town mapping and 3-Dimensional model construction, crop damage assessment. The data protection laws and privacy acts limits the civilian usage of drones in many countries. UAV is much effective in real time



implementation due to its high frequency time series data. Although ground based cameras provide high resolution data, it cannot cover large area. Satellites provide large coverage information but provide low resolution data. UAV provides data at low cost, large area coverage and high resolution due to its adjustable flight height. It functions as eye in the sky to gather information. The yielded videos are processed for analysis and object detection and tracking. Continuous recording of information leads larger video files. Video optimization is essential for video storing, video indexing and effective information gaining in shorter time rather than playing the entire video. Optimizing the video should conserve the highly informative images without deviation in the tracking. Lack of tracking consistency would to misleading surveillance analysis.

The two specific objectives of this paper are optimizing the UAV video and evaluating the tracking consistency. The proposed methodology involves the combining technique of keyframe extraction and selecting principal keyframe [2]. Further on, tracking deviation is compared with the original video and the optimized video. The paper is structured as follows: Section “Related works” forefront the video summarization techniques, object detection and tracking in UAV; Section “Video Optimization” presents the techniques of keyframe extraction and selecting principal keyframes; Section “Object detection and tracking” explores the object detection and tracking in UAV video; Section “Experiments and Discussion” illustrates the results of the above discussed methodology in two datasets; Section “Conclusion” wrap up this papers performance.

## 2 Related Work

The role of UAV in transportation management and military surveillance is nonpareil. With advancement in technology, UAV cover large area with high definition camera. In 2000, drones was used to identify track and monitor vehicle, later on by 2015 drones were used for route planning optimization, detect road boundaries and extract factors such as acceleration and trajectories [3]. The application of UAV are categorized into commercial, safety management and research purpose. NGOs in Japan use drones to inspect illegal whaling. In Nepal, drones are used to protect wildlife habitation. Implementation of UAV ranges from agriculture to civil industry [4]. UAV images are used for 3-D reconstruction of agricultural area. In agriculture, analysing individual plant or tree is impractical. Such limitations are overcome by usage of UAV. The high spatial resolution images are further enhanced by superresolution algorithm based on sparse representation. Classification of edge orientation is done by adaption of multiple pair of dictionaries. The details are preserved both qualitatively and quantitatively with reduction in edge ringing and blurring [5].

Collecting aerial information is strenuous during bad weather condition and varying illumination conditions. Image enhancement, noise reduction, illumination correction and video stabilization is done to obtain unambiguous data [6]. Large video data leads to complexity in video storage and time required to watch the entire content. Keyframe extraction is done by selecting the first frame of the shot segment. As other frames was not inspected lack of continuity persisted. The colour histogram and curve segmentation is used to determine the sharp corners, and the frame related to sharp corners are

extracted as keyframe [7]. In [8], the global and local feature are combined for video segmentation. To extract the global feature, SIFT and SIFT-point distribution histogram is obtained. Distance between pair of images are computed from SIFT-PDH. With an adaptive threshold, the shot boundaries are detected. In [2], the temporal order of the frames is retained by adopting histogram on consecutive frames. Keyframes are extracted based on absolute difference of histogram of consecutive frames. The threshold is obtained from the mean and standard deviation of absolute difference of histogram of consecutive frames.

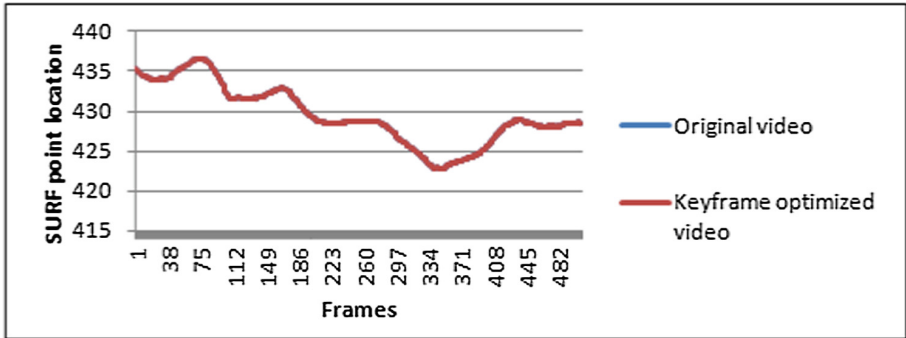
Keyframes extraction is performed through object based event. The input video is initially segmented into shots. The important frame selection is based on presence and absence of objects. The dynamics of frames are obtained and temporal contented are integrated [9]. In [10], the visual feature such as color, spatial frequency and spatial resolution are obtained. The entire video is splitted into three variants: Keyframe groups, equal size frame groups, unequal size frame group using Eratosthenes Sieve Theory and then optimal set of clusters are obtained from Davies-Boulding Index.

### 3 Video Optimization

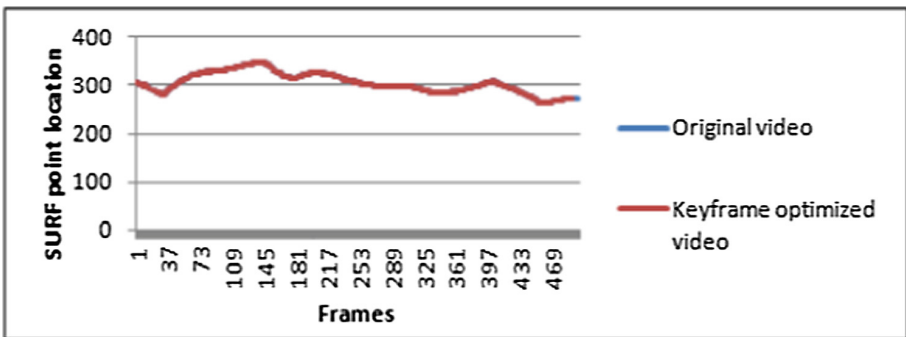
To gather the content of a video, the entire video has to be viewed. Shortage of time to view the entire video necessitates optimization of the video. The optimization has to be done without deteriorating the trajectory of the objects in the video. Object path provides more cue about the informations related to the respective object. Optimization of video with false object trajectory misleads the surveillance. In this paper, video optimization methodology has been proposed to reduce the duration of the video and maintaining the trajectory structure of the video. The methodology is tested on 6 videos from UAV123 dataset and 2 videos form VIVID dataset. The average duration of the video is 20.37 s. The redundant frames present in the video is reduced by keyframe extraction.

#### 3.1 Keyframe Extraction

Video optimization based on keyframes can be done using sampling, scene segmentation and shot segmentation. Scene segmentation and shot segmentation does not take temporal position of frames into account. In this paper, keyframe extraction is done based on sampling. The keyframe extraction is done in two section. In the first section, the video is converted to frames. The absolute difference of histogram and its sum is calculated for consecutive frames. The mean of the sum of the absolute difference is assigned as threshold. In second section, the threshold is compared against sum of the absolute difference. The object path of the original video overlaps with the respective object path of the keyframe optimized video resulting in no deviation (See Figs. 1 and 2). The average time required to view the keyframe video is 9.23 s. The keyframes extraction saves 54.66% of viewing time.



**Fig. 1.** Tracking deviation between original video and keyframe optimized video of CAR10 video from UAV123 dataset



**Fig. 2.** Tracking deviation between original video and keyframe optimized video of VIVID2 video from VIVID dataset

The steps to extract keyframes are as follows:

- Step 1: Video is converted to frames
- Step 2: Consecutive frames are converted into grayscale
- Step 3: Histogram is obtained for the consecutive frames
- Step 4: Absolute difference of the histogram of the consecutive frames is obtained
- Step 5: Sum of the absolute difference (sum) is calculated
- Step 6: Mean of the sum of absolute difference of histogram of consecutive frames is assigned as threshold
- Step 7: Compare the sum with threshold, If  $\text{sum} > \text{threshold}$  then the frame is selected as keyframe else goto Step 2. The extracted keyframes of the video are shown in Fig. 3.

Although the obtained keyframes gives the summary of the original video, redundant keyframes still exist which paves way to further optimization without deviation in tracking.



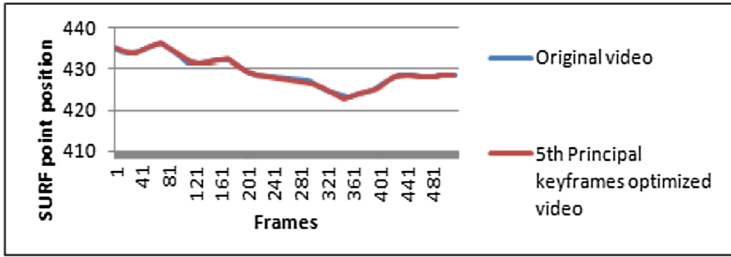
**Fig. 3.** Samples of extracted keyframes of CAR10 video from UAV123 dataset

### 3.2 Principal Keyframes Selection

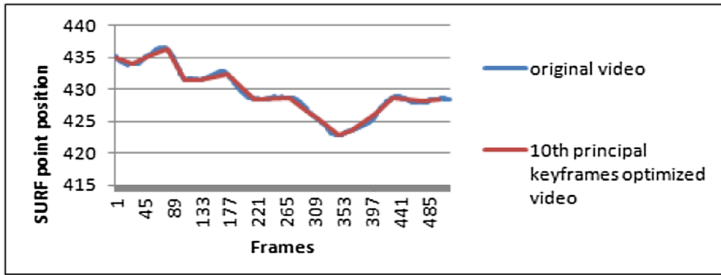
The keyframes are extracted from the above steps (Sect. 3.1). Principal keyframe range is set to select the primary keyframes at regular interval, thereby discarding redundant keyframes. Initially, the first keyframe is retained and the principal keyframe range is selected as 2. Every 2<sup>nd</sup> Keyframe is selected as principal keyframe and the selected principal keyframes are compiled into an optimized video. Tracking analysis exhibits no deviation between the original video and the optimized video. Hence the principal keyframe range increases to 3. Likewise, every 3<sup>rd</sup> keyframe is selected and subjected to tracking analysis. With no deviation in tracking, the principal keyframe range increases to 4. The procedure continues till higher tracking deviation occurs between the original video and the optimized principal keyframe video (See Figs. 4(a), (b), (c), (d) and 5(a), (b), (c), (d)).

It has been found that with every 23<sup>rd</sup> keyframe as principal keyframe, tracking is consistent with negligible deviation. Furthermore, increase in principal keyframe range as 24, leads to apparent tracking variation and tracking loss (See in Figs. 4(d) and 5(d)). Henceforth, retaining every 23<sup>rd</sup> keyframe as principal keyframe, provides precise information about the original video with less tracking variation. The average time duration of 23<sup>rd</sup> principal keyframe optimized video is 1.93 s and average time saved is 90.68%. The steps involved in principal keyframe selection are as follows:

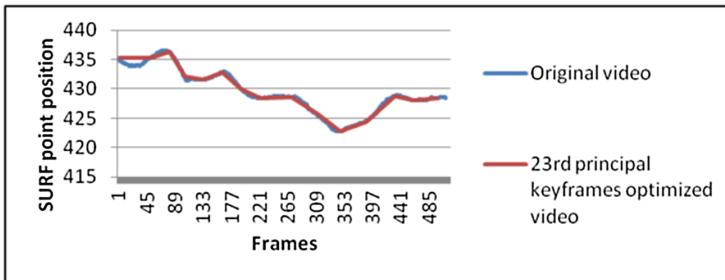
- Step 1: Retain the first keyframe and every 23<sup>rd</sup> keyframe
- Step 2: Continue Step 1 till the last keyframe. The principal keyframes are obtained (See Fig. 6).
- Step 3: Compile the retained principal keyframes into optimized video.



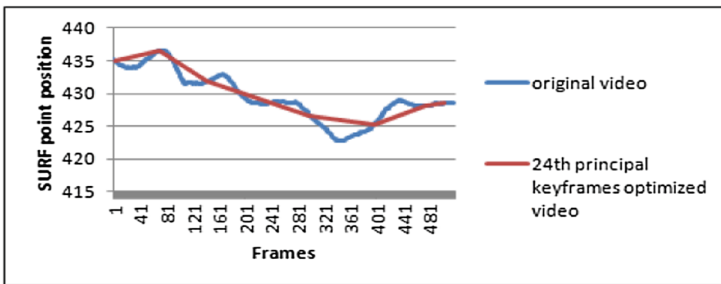
(a)



(b)

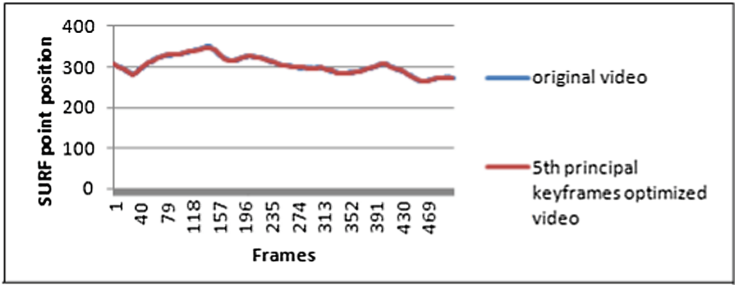


(c)

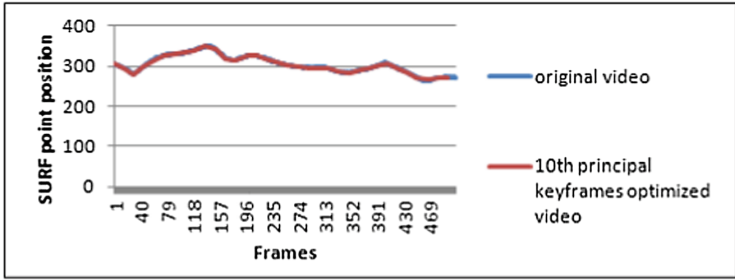


(d)

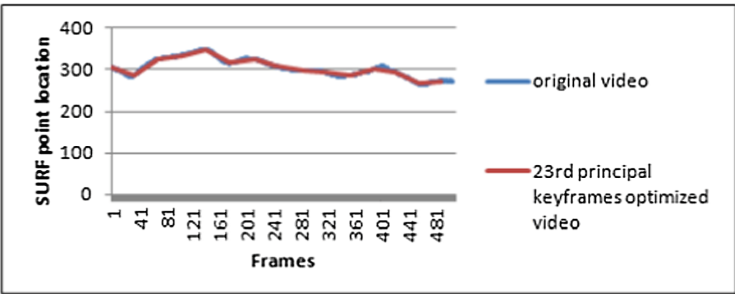
**Fig. 4.** Tracking deviation between original video and optimized videos of CAR10 video from UAV123 dataset: (a) original video and 5<sup>th</sup> principal keyframes optimized video (b) original video and 10<sup>th</sup> principal keyframes optimized video (c) original video and 23<sup>rd</sup> principal keyframes optimized video (d) original video and 24<sup>th</sup> principal keyframes optimized video.



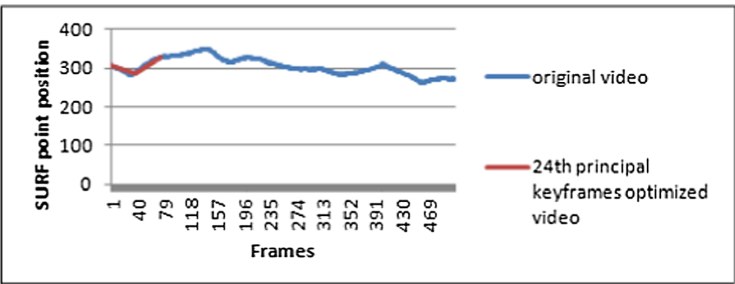
(a)



(b)

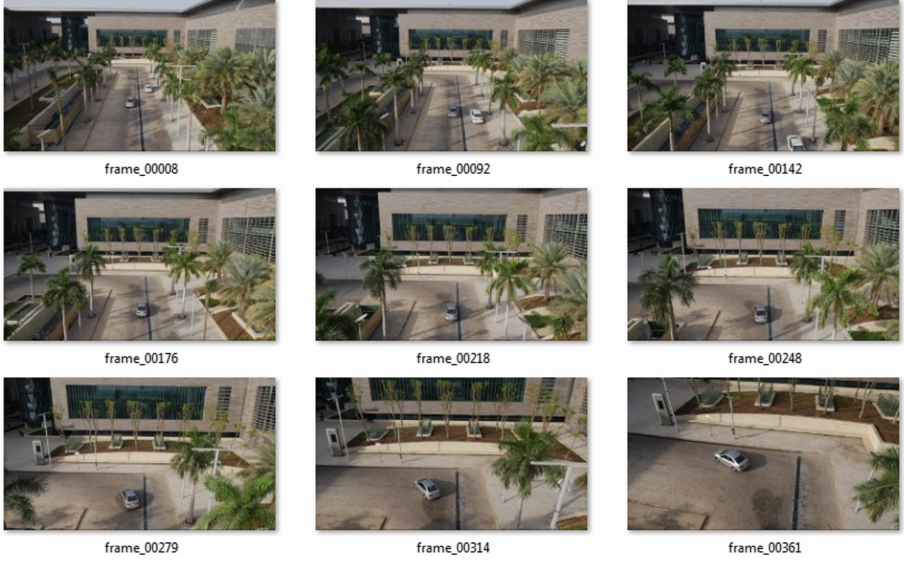


(c)



(d)

**Fig. 5.** Tracking evaluation between the original and optimized videos of VIVID2 video from VIVID dataset: (a) original video and 5<sup>th</sup> principal keyframes optimized video (b) original video and 10<sup>th</sup> principal keyframes optimized video (c) original video and 23<sup>rd</sup> principal keyframes optimized video (d) original video and 24<sup>th</sup> principal keyframes optimized video.



**Fig. 6.** Samples of 23<sup>rd</sup> keyframes as principal keyframes of CAR10 from UAV123 dataset.

## 4 Object Detection and Tracking

The object detection and tracking is done for the original and the optimized video to analyze the tracking variation. In the original video, a region of interest is selected and SURF descriptors are obtained and the object is tracked with KLT tracker [11]. A SURF descriptor location is tabulated for every frame throughout the video till the last frame. The same approach is followed for tracking the optimized video. In order to compare the tracking difference between the original video and the optimized video, the tracked object path is compared.

### 4.1 Object Detection – Speeded Up Robust Feature

SIFT feature descriptor is commonly used, due to its invariance to scale and rotation and partial invariance to illumination and 3D viewpoint. The heavy mathematical computation and complications of SIFT leads way to SURF. In 2006, Bay et al. proposed Speeded up Robust Features. The SURF is compiled as follows:

Step 1: Laplace of Gaussian is approximated with Box filter. Image filtering is faster in an Integral image. To obtain the integral image of an image, the sum of all pixels of a rectangular region is calculated. Using box filters, the values are approximated.

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (1)$$

Step 2: SIFT uses Hessian and DOG to select scale and location interest point, whereas SURF uses determinant to find both. The determinant elements must be weighted to obtain a good approximation.



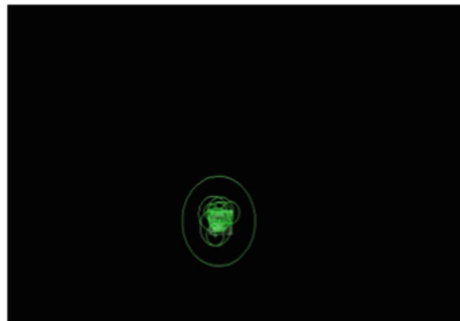
(a)



(b)



(c)



(d)

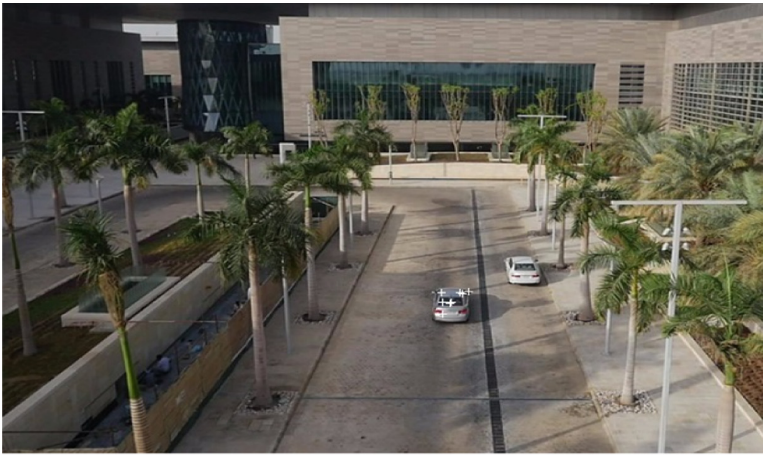
**Fig. 7.** (a) First frame of the video (b) Select the region of interest to track (c) Grayscale conversion (d) SURF descriptors



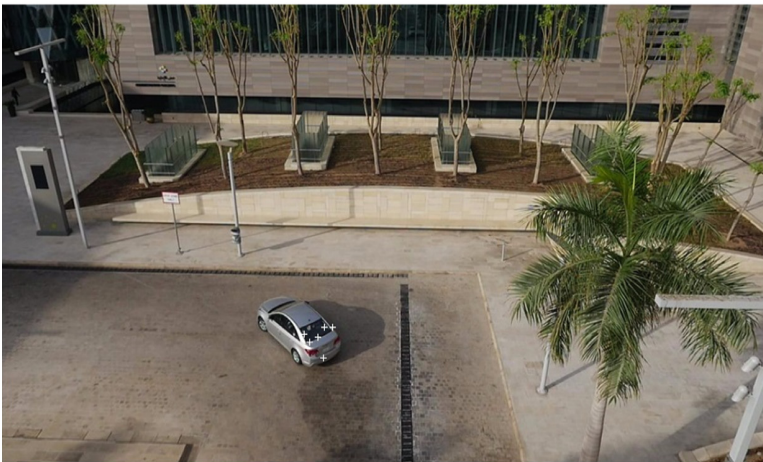
$$\det(H_{approx}) = D_{xx}D_{yy} - (WD_{xy})^2 \quad (2)$$

The Eqs. (1) and (2) are referred by Bay in [12].

Step 3: Once the interest points have been localized in both speed and scale, the orientation assignment and keypoint descriptor is determined. The Haar wavelet is obtained along the x and y direction. Four dimensional descriptor represents each subregion, which includes the sum of absolute values and sum of wavelet response. The SURF descriptors are obtained from the selected region of interest (See Fig. 7 (a), (b), (c) and (d)).



(a)



(b)

**Fig. 8.** (a) and (b): The SURF descriptors are tracked along the car throughout the video

## 4.2 Object Tracking – Kanade Lucas Tomasi Tracker

Kanade Lucas Tomasi tracker exhibits frame to frame tracking. It generates object feature trajectory between consecutive images [11]. To generate a continuous trajectory, current displacement vector is added to previous. In this paper, the SURF points are tracked along the video with the KLT (See in Fig. 8(a) and (b)). In order to tabulate the path of the object trajectory, one SURF point position is tracked throughout the video.

## 5 Experiments and Discussion

The experiment is performed in UAV123 dataset [13] and VIVID dataset [14]. UAV123 dataset is a low altitude HD UAV dataset. It has resolution of  $1280 \times 720$  px and the altitude of the object varies between 5–25 m. The videos are downsampled to  $832 \times 468$  px. The VIVID dataset focus on high altitude coverage, with low resolution of  $640 \times 480$  px. The implementation is performed in MATLAB 2013. The object path of original video and keyframe video exhibits no deviation, hence leading to further optimization by removing redundant keyframe at regular interval. The comparative analysis of object path between original video and optimized principal keyframes video of CAR10 is consistent till principal keyframe range as 23 (See Fig. 4(a), (b), (c)). Increase in the principal keyframe range as 24 leads to higher tracking deviation (See Fig. 4(d)). In CAR14 video, tracking deviation is negligible till principle keyframe range as 23, further increase in the principal keyframe range as 24 leads to more tracking deviation, resulting in loss of object path information. CAR6 video exhibit no tracking deviation till 23<sup>rd</sup> principal keyframe range, whereas at 24<sup>th</sup> principal keyframe range tracking deviation begins which is not negligible.

In VIVID2 video, tracking is consistent till 23<sup>rd</sup> principal keyframe range (See Fig. 5(a), (b), (c)). The point tracking is lost when the principal keyframe selection increases to 24 (See Fig. 5(d)). The object path is tracked only till 75<sup>th</sup> frame, further on, the SURF feature points disappears. In VIVID3 video, the tracking is lost in the 49<sup>th</sup> frame. Contradictions such as loss of point tracking, higher tracking deviation between original video and optimized video and higher loss of information occurs. These contradictions may lead to imprecise analysis of the video. Henceforth, in terms of interest point tracking both the dataset respond well with principal keyframe range as 23. The time optimization of the proposed work is given in Table 1. The average time saved in the keyframe extraction is 54.71%. The proposed principal keyframe selection methodology achieves average saving time as 90.68%.

$$\text{Time saved} = \text{original video duration} - \text{optimized video duration} \quad (3)$$

$$\text{Percentage of time saved} = \frac{\text{Time saved}}{\text{Duration of original video}} \times 100 \quad (4)$$

$$\text{Average time saved} = \frac{\sum \text{Percentage of time saved}}{\text{total no. of videos}} \quad (5)$$

**Table 1.** Time computation of original and 23<sup>rd</sup> Principal keyframes optimized video

Video	Computational time - original video (sec)	Computational time - 23 <sup>rd</sup> principal keyframes optimized video (sec)	Time saved	Percentage of time saved
Car6	20.36	2.23	18.13	89.05
Car6turn	20.35	1.89	18.46	90.713
Wakeboard8	20.14	1.81	18.33	91.10
Car10	20.7	2.01	18.69	90.3
Boat1	20.17	1.56	18.61	92.26
Car14	20.06	1.8	18.5	92.22
Boat2	20.12	1.44	18.68	92.84
Car1	20.73	2.14	18.59	89.67
Vivid2	20.52	2.1	18.42	89.76
Vivid3	20.69	2.06	18.63	90.04

## 6 Conclusion

In this research paper, an optimized method have been described and discussed for UAV video summarization for object tracking. It is essential to summarize the video to save time. In the proposed method, the video is optimized in such a way that tracking deviation is negligible in both high and low altitude UAV videos. The redundant frames in video is reduced by keyframe extraction. The extracted keyframes has still more redundant keyframes which can be reduced further. Henceforth, the principal keyframes are selected by retaining first frame and every 23<sup>rd</sup> keyframe. The sustained keyframes are compiled into optimized video. The compiled video and the original video is subjected to object detection and tracking. The position of the object is tabulated which exhibit negligible deviation between the original and optimized video. The keyframe extracted video saves 54.71% of time whereas the optimized 23<sup>rd</sup> principal keyframes video saves 90.68%. The average video duration is 20.37 s, which is reduced to 9.23 s by keyframe extraction, which is further reduced to 1.93 s without object trajectory deviation. The optimized video supports the interest point tracking efficiency and more duration of the time is saved.

## References

1. Guido, G., Gallelli, V., Rogano, D., Vitale, A.: Evaluating the accuracy of vehicle tracking data obtained from Unmanned Aerial Vehicle. *Int. J. Transp. Sci. Technol.* **5**, 136–151 (2016)
2. Sheena, C.V., Narayanan, N.K.: Key-frame extraction by analysis of histograms of video frames using statistical methods. In: 4th International Conference on Eco-Friendly Computing and Communication Systems, vol. 70, pp. 36–40. Elsevier (2015)
3. Emmanouil Barmponakis, N., Eleni Vlahogianni, I., John Golias, C.: Unmanned aerial aircraft systems for transporation engineering. *Int. J. Transp. Sci. Technol.* **5**(3), 111–122 (2016)

4. Idries, A., Mohamed, N., Jawhar, I., Mohamed, F., Al-Jaroodi, J.: Challenges of developing UAV applications: a project management view. In: Proceedings of International Conference on Industrial Engineering and Operations Management. IEEE (2015)
5. Haris, M., Watanabe, T., Fan, L., Widyanto, M.R., Nobuhara, H.: Superresolution for UAV images via adaptive multiple sparse representation and its application to 3-D reconstruction. *IEEE Trans. Geosci. Remote Sens.* **55**(7), 4047–4058 (2017)
6. Balasubramanian, A., Kamate, S., Yilmazer, N.: Utilization of robust video processing techniques to aid efficient object detection and tracking. *Procedia Comput. Sci.* **36**, 579–586 (2014)
7. Zhao, L., Qi, W., Li, S.Z., Zhang, H.J.: Key-frame extraction and shot retrieval using nearest feature line (NFL). In: Proceedings of the ACM Multimedia 2000 Workshops, Los Angeles, pp. 217–220 (2000)
8. Hannane, R., Elboushaki, A., Afdel, K., Naghabhushan, P., Javed, M.: An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram. *Int. J. Multimedia Inf. Retrieval* **5**, 89–104 (2016)
9. Barhoumi, W., Zagrouba, E.: On-the-fly extraction of keyframes for efficient video summarization. In: AASRI Conference on Intelligent Systems and Control, Canada, vol. 4, pp. 78–84. Elsevier (2013)
10. Kumar, K., Shrimankar, D.D., Singh, N.: Eratosthenes sieve based key-frame extraction technique for event summarization in videos. *Multimedia Tools Appl.* 1–22 (2017). <https://doi.org/10.1007/s11042-017-4642-9>. Springer
11. Jabar, F., Farokhi, S., Sheikh, U.U.: Object tracking using SIFT and KLT tracker for UAV-based applications. In: IEEE International Symposium on Robotics and Intelligent Sensors, Malaysia (2015)
12. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). *Comput. Vis. Image Underst. (CVIU)* **110**(3), 346–359 (2008)
13. Mueller, M., Smith, N., Ghanem, B.: A benchmark and simulator for UAV tracking. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 445–461. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46448-0\\_27](https://doi.org/10.1007/978-3-319-46448-0_27)
14. Collins, R., Zhou, X., Teh, S.K.: An open source tracking testbed and evaluation web site. In: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (2005)

# Failure Recovery Using Segment Protection in Software Defined Networks

V. Padma<sup>2</sup>, Gayathri Santhosh<sup>1</sup>(✉), and Yogesh Palanichamy<sup>1</sup>

<sup>1</sup> Department of Information Science and Technology, College of Engineering, Anna University, Chennai, India

m.gayath@yahoo.com, yogesh@annauniv.edu

<sup>2</sup> Department of Computer Science and Engineering,

G.Narayananamma Institute of Technology and Science, Hyderabad, India

padma.js80@gmail.com

**Abstract.** Software Defined Networking (SDN) is a network architecture that decouples the control and data planes. SDN enables network control to make its programmable directly. Software defined networking is the OpenFlow protocol and its architecture is designed for Local Area Networks (LAN). It does not include effective mechanisms for fast resiliency. Fast resiliency is a major requirement in metro and carrier-grade Ethernet network. The proposed scheme aims to reduce the recovery time during single link network failures. The controller calculates the backup path proactively using segment protection scheme. The switches identify link failures in segments using Bidirectional Forwarding Detection (BFD) protocol and reroute the traffic. As the link is recovered the switches will start using the best path. The controller deletes the backup segment entries when the corresponding working path entries expire. This paper experiments the link protection scheme that aims to enhance the OpenFlow architecture by adding fast recovery mechanisms in the switch and the controller. This is achieved by enabling the controller to add backup paths proactively along with the working paths and enabling the switches to perform the recovery actions locally. Recovery time is less compared to the switch-controller and round trip time which provides better results. The system performance is evaluated by finding the packet loss and switch over time by comparing it with the current OpenFlow implementations. The system performs reasonably better than the existing ones in terms of switch over time.

**Keywords:** SDN · OpenFlow · LAN

## 1 Introduction

### 1.1 Software Defined Network (SDN)

Software Defined Networking (SDN) is a new networking paradigm in which the forwarding hardware is decoupled from control decisions [16]. In traditional networks, the control and data planes are combined together. The control plane is for configuring the node and programming the paths. Once these paths have been determined, they are pushed down to the data plane. Data forwarding at the hardware level has been done by using control information. In this traditional approach, once the flow management

(forwarding policy) has been decided, adjustment to the policy has been done by changing the configuration of the devices. Hence scalability is difficult because of changing traffic demands, increasing use of mobile devices, and the impact of “big data.”

Figure 1 clearly depicts the traditional and SDN network views. In SDN, control is moved out of the individual network nodes and pushed into the separate, centralized controller [16]. SDN switches are controlled by a network operating system that collects information using the API and manipulates their forwarding plane which provides an abstract model of the network topology. The controller can therefore exploit complete knowledge of the network to optimize flow management and support service-user requirements of scalability and flexibility.

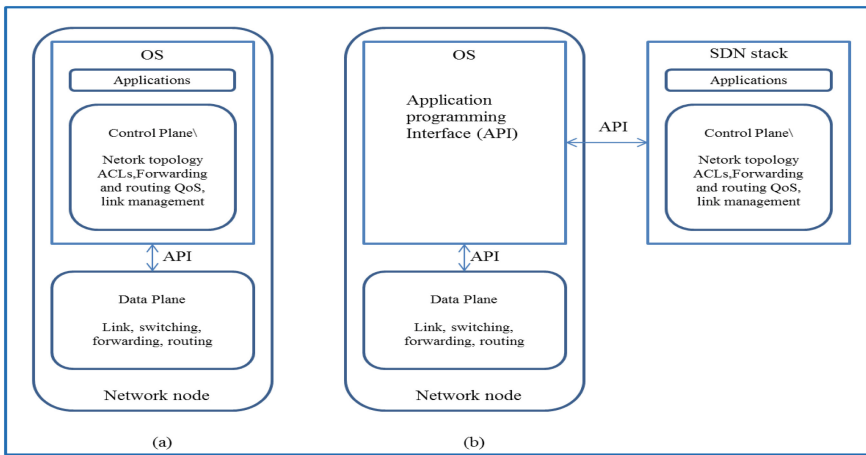


Fig. 1. (a) Traditional network view and (b) SDN network view

**OpenFlow:** OpenFlow is driven by the SDN principle of decoupling the control and data forwarding planes. This standardizes information exchange between the two planes [16]. OpenFlow networks consist of following three components which is shown in Fig. 2.

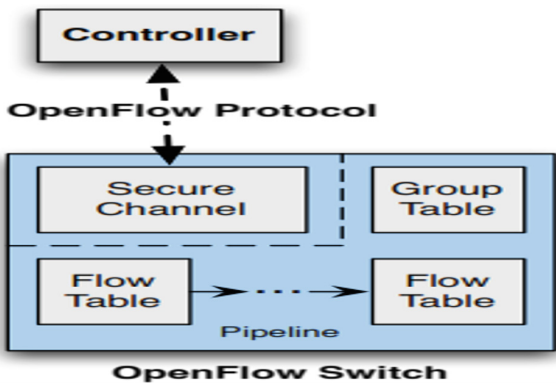


Fig. 2. Components of an OpenFlow switch

*OpenFlow Switch:* OpenFlow switch consists of one or more flow tables, a group table and a secure channel to an external controller [16]. A flow table consists of a list of flow entries. Each entry has match fields, counters and instructions. Incoming packets are compared with the match fields of each entry and matching starts at the first flow table and may continue to additional flow tables [16]. If there is a match, the packet is processed according to the action contained by that entry. When there is no matching flow entry, the outcome depends on the configuration of the table-miss flow entry. Counters are used to keep statistics about packets. The group table contains group entries, where each group entry contains a list of action buckets with specific semantics dependent on group type. The actions in one or more action buckets are applied to packets sent to the group. The main function of the switch is flow table lookup and packet forwarding.

*OpenFlow Controller:* The controller is a software program responsible for manipulating the switch's flow table, using the OpenFlow protocol [16]. It makes a decision on how to handle the packet. It can drop the packet, or it can add a flow entry directing the switch on how to forward similar packets in the future. Thus the controller essentially centralizes the network intelligence, while the network maintains a distributed forwarding plane through OpenFlow switches and routers.

*OpenFlow Protocol:* The OpenFlow protocol deals with defining the format of the messages passed between the control plane and the OpenFlow switch through the secure channel [16]. The format of the messages has to be understood as well as generated by both the entities. This standard format of message passing is defined in the OpenFlow protocol.

Fast resiliency is a major requirement in metro and carrier-grade Ethernet network. Hence this project aims to reduce the recovery time during single link network failures. The controller calculates the backup path proactively using segment protection scheme [16]. The switches identify link failures in segments using Bidirectional Forwarding Detection (BFD) protocol and reroute the traffic. As the link is recovered the switches will start using the best path. The controller deletes the backup segment entries when the corresponding working path entries expire. This scheme is proposed to provide fast resiliency in OpenFlow networks. This proposed scheme employs segment protection scheme to compute the backup paths proactively. This considers single link failures of the network. The recovery is performed locally by the switch. The controller intervention is avoided during failure recovery. Once the link is up, the switches start forwarding using the best path. The first and last switches of each segment are enabled to send Bidirectional Forwarding Detection (BFD) packets to identify the failure of any links in the segment. When a link failure is detected by the switch, it reroutes the packets through backup segments. To avoid expiration, flow entries for backup segments are installed permanently and a novel mechanism is introduced for deleting these entries when the corresponding working path flow entries are removed.

The remainder of the paper is organized as follows. Section 2 provides a literature survey in terms of various approaches in providing resiliency in OpenFlow networks. Section 3 highlights the system requirements and presents the overall system architecture. Section 4 discusses the results and Sect. 5 evaluates the performance of the proposed system. Section 6 gives the conclusions of this work and shows the directions for future enhancements.

## 2 Existing Schemes

Initially OpenFlow was designed for Local Area Networks (LANs). Hence it doesn't include fast resiliency mechanisms. Many schemes had been proposed in the past for reducing the restoration time in OpenFlow networks [5]. This survey discusses the existing approaches used to reduce the failure recovery time and the limitation of the same.

Staessens et al. [2] and Sharma et al. [3] proposed a restoration scheme for OpenFlow carrier grade Ethernet networks. In this scheme, switch connected to the disrupted link directly notifies the controller about the topology change. Upon notification, the controller identifies the disrupted flows, computes the backup paths, and updates the data plane flow tables considering the failure. In other words, the method followed in this scheme is reactive. It is also recognized that in big networks, these full-state controllers could be overloaded by recovery requests.

Yu et al. [7] considered OpenFlow resiliency in IP networks. This scheme is also based on a full-state controller that is notified by the switch upon link failure occurrence. In data plane approach, when a link fails the switch attached to that link will send notification to all other switches that sends traffic through this failed link.

Desai and Nandagopal [4] proposed a data plane approach in OpenFlow networks. This also overloads the Controller. Kempf et al. [8] proposed a similar kind of approach in transport networks based on OpenFlow. In this, each established flow is monitored by sending frequent probe messages. Hence the failure is quickly detected. But the backup path computation is done by the controller only [5]. In path protection approach, backup paths are pre computed along with the working paths. Both are installed in the switches with different priorities. Hence when a link fails, the switch can use the backup path without the intervention of the controller.

Sharma et al. [1] proposed this approach using the fast-failover groups functionality of OpenFlow specification. This approach monitored the working path aliveness by the ingress switch using a similar mechanism as the one proposed in [8]. Here the controller intervention is totally avoided.

Nguyen et al. [13] introduced a novel method for fast switchover by using the select group instead of failover group. This method has better bandwidth utilization than the standard one. Link protection mechanism also pre computes the backup path as that of path protection approach. Instead of computing a single backup path between the source and destination, this computes backup paths from every node in the working path to the destination.

Sgambelluri et al. [6] used this scheme in protection approach. This achieves lesser fail over time than path protection scheme, as here every switch in the working path can directly divert the traffic through the back up path. But this increases the number of flow entries in the flow table.

Fonseca et al. [9] proposed a backup control mechanism which involves a backup controller in an OpenFlow based network. This avoids the problem of having a single point of failure. Various algorithms have been developed and studied for dividing the given path into a set of segments.



Xu et al. [10] proposed a scheme called PROMISE, which divide an active path into several possible overlapping active segments and protect each of them with a detour called backup segment. In the worst case this algorithm takes exponential time to compute the backup paths.

Todimala and Ramamurthy [11] presented a dynamic partitioning sub path protection routing technique. Here primary path is partitioned into sub paths and then backup paths for these sub paths are computed dynamically. Tewari and Ramamurthy [12] based on uniform fixed-length segment protection method. Primary path is divided into fixed-length segments with the exception of the last segment.

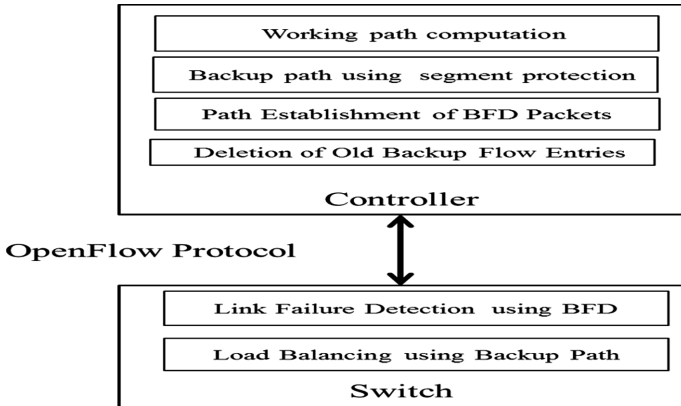
Table 1 summarizes the major approaches and their drawbacks. Earlier works in OpenFlow deal with the problem of path failure reactively. Once the OpenFlow switch identifies the path failure it informs the OpenFlow controller and the OpenFlow controller identifies the alternate or backup paths and inform the OpenFlow switch. In this reactive approach, packet loss is inevitable. Recent work in this area attempts a proactive approach. In the proposed scheme, a proactive approach using segment protection is attempted to achieve moderate switch over time, bandwidth efficiency and moderate number of flow entries

**Table 1.** Comparison of various recovery approaches

Approach	Concept	Drawback
Restoration	<ul style="list-style-type: none"> <li>• Switch notifies the controller about the change</li> <li>• Controller identifies disrupted flows, computes backup path and updates flow tables</li> <li>• Recovery time is around 200 ms</li> </ul>	Controller is overloaded by recovery requests
Data plane mechanism	<ul style="list-style-type: none"> <li>• On failure, switches that send traffic through disrupted links are notified</li> <li>• Recovery performed by controller</li> </ul>	Controller is overloaded by recovery requests
Path protection	<ul style="list-style-type: none"> <li>• Controller precomputes backup paths</li> <li>• Installs it in switches using fast fail over groups</li> <li>• Recovery around 50 ms</li> </ul>	<ul style="list-style-type: none"> <li>• Switchover time in path protection is more compared to link and segment protection</li> <li>• Bandwidth efficiency is more</li> </ul>
Link protection	<ul style="list-style-type: none"> <li>• Controller precomputes backup paths using link protection</li> <li>• Installs it in switches</li> <li>• Recovery around 20–50 ms</li> </ul>	<ul style="list-style-type: none"> <li>• No of flow entries in the switch are more</li> <li>• Bandwidth efficiency is less</li> </ul>

### 3 System Architecture

The system architecture is represented in Fig. 3 and it gives the details of the proposed scheme in a diagrammatic way. The system can be classified into two sections such as Controller part and the Switch part which is discussed later and the system works as follows.



**Fig. 3.** Modified OpenFlow scheme

- Step 1: The controller computes the working path for the packet in message.
- Step 2: Then it divides the working path into a set of working segments of fixed length and finds the backup segment for each. Working and backup segment entries are installed in the switch.
- Step 3: It also enables the segment start and end switches to send BFD packets periodically. These packets are routed through the working segment.
- Step 4: When a link fails, the end switches of that segment identify the failure using BFD protocol and reroute the packets through the backup segment.
- Step 5: When the working path expires, the backup path entries are also deleted by the controller.
- Step 6: Backup segments are also used to distribute the traffic when there is no failure.

### 3.1 Controller Part

**Working Path Computation:** When a packet is redirected to the controller, shortest path is calculated between the source and destination with the help of link discovery. Controller constructs a graph structure  $G$  by identifying the links and nodes from the discovery module. Then it finds the shortest path between the source and destination in the graph structure using bidirectional Breadth First Search (BFS).

**Backup Path Computation Using Segment Protection:** When a new flow arrives, backup path is calculated by the controller. This involves two steps. They are

- Identification of Segments
- Installation of Backup Flow entries

*Identification of Segments:* The backup path is calculated using segment protection method. In this scheme, working path is divided into number of working segments. For each of the working segment, the backup segment has to be found. For this a new graph

$G_p$  is constructed from the original graph  $G$  by removing the edges of working path from  $G$ . The minimum number of hops ( $min\_hop$ ) of a working segment is decided according to the size of the network. Now initial working segment ( $S_w^i$ ) is formed by taking the first  $min\_hop + 1$  nodes from the working path. If a backup path ( $S_b^i$ ) exists for this working segment. Working segment and backup segment are added into the list of working ( $S_w$ ) and backup segments ( $S_b$ ) respectively. Then the next working segment is formed by starting from last node of previous working segment and adding next  $min\_hop$  nodes from the working path. If the number of left over nodes in the working path is lesser than  $min\_hop$  then these remaining nodes form the last segment. This process is repeated till we cover all the nodes of working path. If the backup path doesn't exist for a segment then one more node is added to the working segment and the process of identifying backup segment is repeated.

*Installation of Backup Flow Entries:* After identifying working and backup segments, the in port and out port for each of the node in the working and backup segments are identified. The first and last element of the working segment and the corresponding backup segment will be same. These are the nodes which divert the traffic through alternate path during failure. Hence group table entries have to be added with two buckets for these nodes. First bucket will be associated with the port to be used for working path. Second bucket will be associated with the port to be used for backup path. The Fast fail over group type is used which sends the traffic through first active bucket. For intermediate nodes in working and backup segments separate flow entries are installed with different priority value. Backup segment entries priority will be lesser than the priority of working segment entries. Figure 4 describes the segment identification.

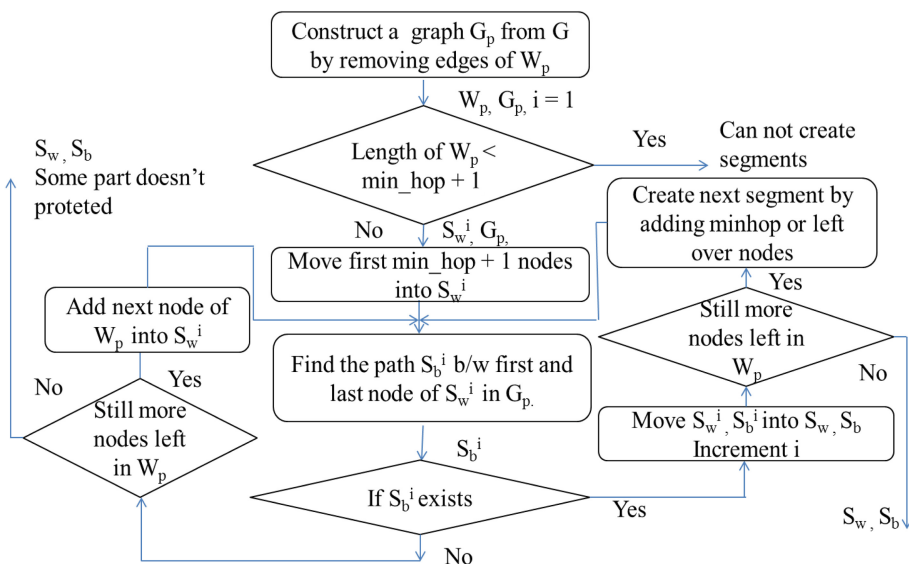


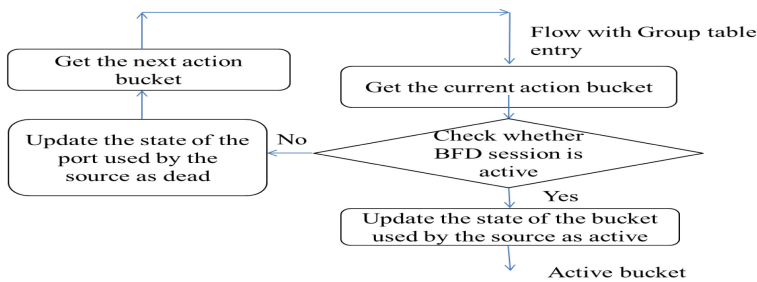
Fig. 4. Flow diagram of segment identification

**Path Establishment for BFD Packets:** This module is incorporated at the OpenFlow Controller. After identifying the working and backup segments, the controller enables segment start and end switches to send BFD packets periodically to identify any link failures in that segment. It also installs flow entries for BFD packets in such a way that the switches can route these packets through the corresponding working segments.

**Deletion of Old Backup Flow Entries:** While installing the flow entries, backup flow entries are added as permanent entries and the working path entries are installed with ‘Flow Removed’ flag. Hence when the working flows are deleted the switches send ‘FlowRemoved’ message to the controller. At that time the controller retrieves the source, destination and the switch from which the message is sent. If it receives the ‘Flow Removed’ message from source or destination of any protected flow then it sends ‘FlowMod’ messages to the switches to delete the backup flow entries for that source and destination.

### 3.2 Switch Part

**Link Failure Detection using BFD:** OpenFlow switch has to detect the failure using BFD packets. These packets are sent periodically between the first and last nodes of working segments. Hence when a link fails these packets won’t be received by the first node of the working segments. If the time limit exceeds, then the start node assumes that the failure of a link has occurred and updates the aliveness of the corresponding port. After that the packets are sent through the next alive port (bucket). Link failure detection is shown in Fig. 5.



**Fig. 5.** Flow diagram for link failure detection

**Load Balancing using Backup Path:** Backup path is utilized to distribute traffic when there is no failure. Distribution of traffic through working and backup path is done in a round robin fashion. This is achieved using ‘select’ group type of group tables. When both action buckets of the group are active the flow of packets are directed through one of them in a round robin fashion. During failure of working path link, the flow is routed through another active bucket which follows the backup path. This improves the bandwidth utilization in normal condition.

## 4 Results and Discussions

Two servers have been used for emulation. One server is run as a virtual machine using virtual box. Mininet is used here to create a realistic virtual network, running real kernel, switch and application code, on a single machine in seconds, with a single command. **Ryu** is a framework to write OpenFlow Controllers.

### 4.1 Controller Part

**Working Path Computation:** In the OpenFlow controller a handler for packet\_in message is added. In this handler function, NetworkX modules are used to identify the shortest path. This path computation uses bidirectional forwarding detection algorithm. Figure 7 shows the working path computed by the controller for the ping request between h1 and h2.

**Backup Path Computation Using Segment Protection.**

**Identification of Segments:** Once the working path is computed, the controller has to split the working path into a set of segments except the last one. Then it finds the backup segment for each of the working segments. If some of the segments doesn't have backup segment then that part alone is left as unprotected. The working and backup segments for the working path (S12-S11-S10-S9-S8-S7-S6-S5-S4-S3) of h1 and h2 is shown in Fig. 6.

```

padma@padma:~/ryu
segment creation completed
path
12      11      10      9      8      7      6      5      4      3
working segments
[(1, (12, 9, [12, 11, 10, 9])), (2, (9, 6, [9, 8, 7, 6])), (3, (6, 3, [6, 5, 4, 3]))]
backup segments
[(1, (12, 9, [12, 22, 21, 20, 9])), (2, (9, 6, [9, 18, 17, 16, 6])), (3, (6, 3, [6, 15, 14, 13, 3]))]
calling install path
bfd session
s12-eth1
success
s9-eth2
success
s9-eth1
success
s6-eth2
success
s6-eth1
success
s3-eth2
success

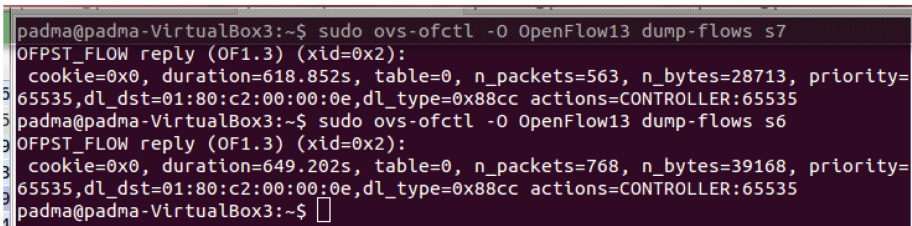
```

Fig. 6. Working path, working segments, backup segments and BFD sessions

**Installation of Backup Flow entries:** After identifying the working and backup segments, the controller has to frame the flow entries for the same and send FlowMod messages [14] to the corresponding switches to add those flow entries. First and last switch of the segments will have group table entries as they have to reroute the traffic in case of link failure.

**Path Establishment for BFD Packets:** Once the working segments are identified, the controller enables BFD sessions between the first and last switch of each segment. The default destination Medium Access Control (MAC) address of the BFD packets generated by BFD sessions is ‘00:23:20:00:00:01’. Hence this function installs flow entries with destination MAC as ‘00:23:20:00:00:01’ in all other switches (middle) of the working segments in order to route the BFD packets through the working segments.

**Deletion of Old Backup Flow Entries:** The controller installs the backup path entries as permanent entries to avoid the expiration of these entries. Hence when the working path expires due to expiration of idle timeout period, we have to delete the corresponding backup path entries. For this a ‘Flow Removed’ message handler is added in the controller. For deletion purpose the controller maintains a list of source and destination hosts tuple between which protected path is installed already. After deleting the backup path entries it updates this list by deleting the corresponding source and destination tuple. Figure 7 shows the flow entries of S6 and S7 after the expiration of working path between h1 and h2.



```

padma@padma-VirtualBox3:~$ sudo ovs-ofctl -O OpenFlow13 dump-flows s7
OFPST_FLOW reply (OF1.3) (xid=0x2):
  cookie=0x0, duration=618.852s, table=0, n_packets=563, n_bytes=28713, priority=
6 65535,dl_dst=01:80:c2:00:00:0e,dl_type=0x88cc actions=CONTROLLER:65535
padma@padma-VirtualBox3:~$ sudo ovs-ofctl -O OpenFlow13 dump-flows s6
OFPST_FLOW reply (OF1.3) (xid=0x2):
3  cookie=0x0, duration=649.202s, table=0, n_packets=768, n_bytes=39168, priority=
3 65535,dl_dst=01:80:c2:00:00:0e,dl_type=0x88cc actions=CONTROLLER:65535
padma@padma-VirtualBox3:~$

```

Fig. 7. Flow entries of S6 and S7 switches after the expiration of working path

## 4.2 Switch Part

**Link Failure Detection Using BFD:** Whenever a link in the working path fails, the BFD session status of that corresponding session will go down. This status is used to update the action bucket aliveness of the group entries. If the first action bucket is not alive then the packets will be forwarded through next alive action bucket. Figure 8 shows the BFD session state change of the BFD session between S6 and S9 during the failure of S7–S8 link.

**Load Balancing Using Backup Path:** To utilize the bandwidth allocated to the backup paths efficiently, the functions are modified to install the group entries and select the active action buckets in round robin fashion. This enables the switches to route the alternate flows through working and backup segments respectively. The request of the ping uses routed through one action bucket (s3-eth3 port). The reply of the ping is routed through another action bucket (s3-eth2 port).

Filter: bfd		Expression... Clear Apply			
	Source	Destination	Protocol	Length	Info
30	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Up, Flags:
39	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Up, Flags:
74	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Up, Flags:
84	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Up, Flags:
20	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
15	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
51	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
51	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
24	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
94	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
99	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
877	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
810	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
608	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
130	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
811	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
871	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag
574	169.254.1.0	169.254.1.1	BFD Cont	66	Diag: Control Detection Time Expired, State: Down, Flag

Fig. 8. State change of the BFD session during link failure

## 5 Performance Evaluation

The performance of the system is evaluated using the following metrics [16].

1. Number of packets lost
2. Switch over time
3. Number of flow entries

**Number of Packets Lost:** To find the number of packets lost, ping request is given between the two hosts of the emulated network with a count option. The flow of the ping request is observed before and after failure using Wireshark. The ping statistics of the flow displays the number of packets transmitted and the number of packets received. Using these values we can find the number of packets lost during the transmission. This ping test is performed several times by specifying different values for ping intervals and the observations are recorded. The results are plotted in a graph as shown in Fig. 9.

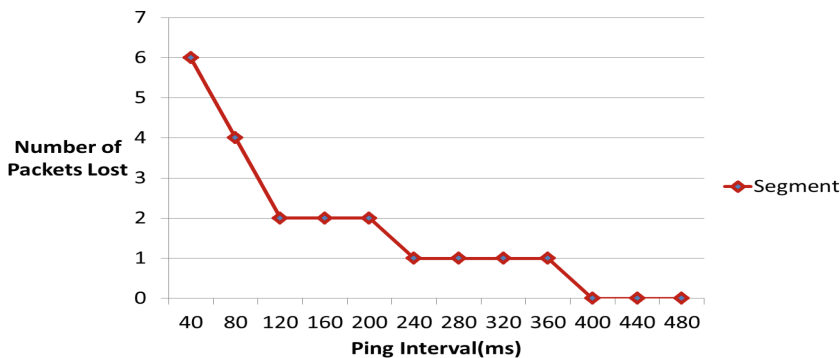


Fig. 9. Ping interval (milliseconds) vs packet loss

From the graph we can observe that the number of packets lost becomes zero after 400 ms. The reason behind this is, our system generates BFD packets at 100 ms intervals and the BFD status is updated only after 3 continuous loss of reply. Hence the system needs at least 300 ms to identify the failure and reroute the traffic. The packets generated within this interval will be lost. If we take the ping interval of 80 ms, then the packets generated within 300 ms is 3 ( $= 300/80$ ). But the link failure may occur in between two successive generations of packets. It may result in one more or less packet loss. Hence the number of packets lost for this interval can have a value from 2 to 4. The observed value recorded in the graph is 3.

**Switch Over Time:** Switch over time is the time required by the switch to identify the link failure and reroute the traffic through the backup path. Our system with BFD interval of 100 ms requires at least 300 ms to identify the link failure. There are systems which support up to 50 ms for BFD transmission interval [15]. In that case we can reduce the switch over time to 150 ms.

Networks of real world will have delays. Hence a delay of 100 ms is set for the links created in Mininet and the ping test is performed. Theoretically the switch over time can be computed using the Eqs. (1) and (2). The value of Multiplier (M) will vary from 0 to 3 according to the position of the failed link in the segment. If the failed link is the 3<sup>rd</sup> hop of 4 hop segment, then all the three BFD packets which will be lost would have been generated already. Hence the value of M in the Eq. (1) will become 0. If the failed link is the first hop of the segment, then the value of M in the Eq. (1) will become 3. For the segment size of 3 and BFD interval of 100 ms this will give the value as 600 to 900 which can be considered as minimum and maximum value of estimated switch over time. To get the value of switch over time from the observations, we have to find the minimum ping interval at which the number of packets lost becomes zero. For various segment sizes, we found the value of switch over time and plotted the results in a graph as shown in Fig. 10. For segment size 3, in the emulated topology, the link S4–S5 is failed and the switch over time is found. These values fall in the range of minimum (M = 0) and maximum (M = 3) value of estimated switch over time.

$$\text{Switch over time} = M * \text{BFD\_min\_tx} + \text{Seg\_RTT} \quad (1)$$

$$\text{Seg\_RTT} = 2 * \text{No\_of\_hops} * \text{Link\_delay} \quad (2)$$

where

M – Multiplier

BFD\_min\_tx – Transmission Interval for BFD Packets

Seg\_RTT – Round trip time of the segment

No\_of\_hops – Number of hops in the segment

Link\_delay – Delay of the links in the network

**Comparison with Path and Link Protection Systems:** To compare the segment protection with other schemes, the ping test is performed with various ping intervals on path, segment and link protection systems where the link delay of emulated network is



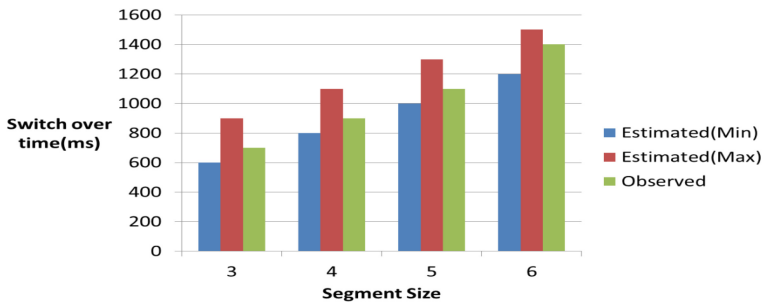


Fig. 10. Switch over time for various segment size

set as 100 ms. As there were some slight variations in the number of packet lost for the same set of inputs, the highest value among them is used to plot the graph which is shown in Fig. 11. This variation is due to the time difference between the last ping request sent and the failure of the link. From the graph it is very clear that the number of packets lost in segment protection is lesser than that of path protection system. The Fig. 12 shows the number of flow entries needed to install a single protected flow between two hosts. Here the size of the segment is taken as 3.

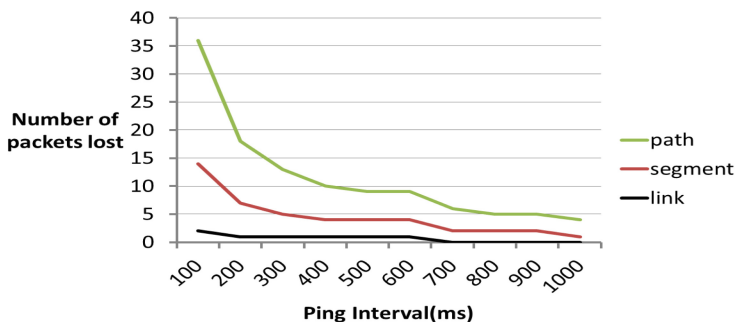


Fig. 11. Comparisons of path, link and segment protection based on number of packets lost

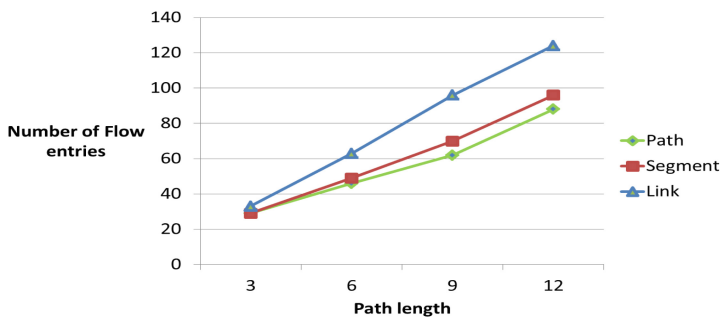


Fig. 12. Comparisons of path, link and segment protection based on number of flow entries

Further the bandwidth efficiency is very low in link protection as we have to allocate bandwidth for all backup paths. Moreover, all networks won't have a backup path from every node of working path to the destination. Hence segment protection is the better option among the three as it provides moderate switch over time with moderate bandwidth efficiency. So the segment protection is the better option among the three.

## 6 Conclusion

Segment protection scheme is implemented using fast fail over group type of openFlow in openFlow networks. The switches identify the link failure of a segment using BFD protocol and reroute the traffic using next available action buckets. Once the failure is rectified, the switches are reverted back to the working path. Here the achieved recovery time is determined only by the segment failure detection time. This depends on the segment size which has to be selected according to the network and working path size. This gives better result than path protection system. Though the link protection gives lesser switch over time, the number of backup flow entries is more in that which leads to poor bandwidth efficiency. Hence segment protection gives moderate switch over time with moderate bandwidth efficiency. In future, we try to improve the bandwidth efficiency by calculating the available bandwidth of working and backup segments periodically and assigning weights to them accordingly. We can perform dynamic load balancing by splitting the traffic through working and backup paths according to their weights. The system can be enhanced to handle node failures.

## References

1. Sharma, S., Staessens, D., Colle, D., Pickavet, M., Demeester, P.: Enabling fast failure recovery in OpenFlow networks. In: Proceedings of 8th International Workshop on the Design of Reliable Communication Networks (DRCN), pp. 164–171, October 2011
2. Staessens, D., Sharma, S., Colle, D., Pickavet, M., Demeester, P.: Software defined networking: meeting carrier grade requirements. In: Proceedings of 18th IEEE Workshop on Local Metropolitan Area Networks (LANMAN), pp. 1–6, October 2011
3. Sharma, S., Staessens, D., Colle, D., Pickavet, M., Demeester, P.: OpenFlow: meeting carrier-grade recovery requirements. *J. Comput. Commun.* **36**(6), 656–665 (2013)
4. Desai, M., Nandagopal, T.: Coping with link failures in centralized control plane architectures. In: Proceedings of 2nd International Conference on Communication Systems and Networks (COMSNETS), pp. 1–10, January 2010
5. Abujassar, R.S., Ghanbari, M.: Efficient algorithms to enhance recovery schema in link state protocols. *Int. J. UbiComp. (IJU)* **2**(3), 53–58 (2011)
6. Sgambelluri, A., Giorgetti, A., Cugini, F., Paolucci, F., Castoldi, P.: OpenFlow-based segment protection in ethernet networks. *J. Opt. Commun. Netw.* **5**(9), 1066–1075 (2013)
7. Yu, Y., Xin, L., Shanzhi, C., Yan, W.: A framework of using OpenFlow to handle transient link failure. In: Proceedings of International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), pp. 2050–2053, December 2011

8. Kempf, J., Bellagamba, E., Kern, A., Jocha, D., Takacs, A. and Skoldstrom, P.: Scalable fault management for OpenFlow. In: Proceedings of IEEE International Conference on Communications (ICC), pp. 6606–6610, June 2011
9. Fonseca, P., Bennessy, R., Mota, E., Passito, A.: A replication component for resilient OpenFlow-based networking. In: Proceedings of IEEE Network Operations and Management Symposium (NOMS), pp. 933–939, April 2010
10. Xu, D., Xiong, Y., Qiao, C.: Novel algorithms for shared segment protection. *IEEE J. Sel. Areas Commun.* **21**, 1320–1331 (2003)
11. Todimala, A., Ramamurthy, B.: A dynamic partitioning sub-path protection routing technique in WDM mesh networks. *J. Clust. Comput.* **7**, 259–269 (2004)
12. Tewari, R., Ramamurthy, B.: Optimal segment size for fixed-sized segment protection in wavelength-routed optical networks. In: Proceedings of IEEE 3rd International Symposium on Advanced Networks and Telecommunication Systems (ANTS), pp. 1–3, December 2009
13. Nguyen, K., Minh, Q.T., Yamada, S.: Novel fast switchover on OpenFlow switch. In: Proceedings of Conference on Consumer Communications and (CCNC), pp. 543–544 (2014)
14. OpenFlow Switch Specification 1.3.0. <https://www.opennetworking.org>
15. [www.cisco.com/c/en/us/td/docs/switches/datacenter/sw/5\\_x/nxos/interfaces/configuration/guide/if\\_cli/if\\_bfd.html](http://www.cisco.com/c/en/us/td/docs/switches/datacenter/sw/5_x/nxos/interfaces/configuration/guide/if_cli/if_bfd.html)
16. Padma, V., Yogesh, P.: Proactive failure recovery in OpenFlow based software defined networks. In: 3rd International Conference on Signal Processing Communication and Networking (ICSCN) (2015)

# Spectrum Sensing Based Heed Routing Performance Enhancement Strategy for Cognitive Radio Sensor Networks

S. Janani<sup>1</sup>(✉), M. Ramaswamy<sup>2</sup>, and J. Samuel Manoharan<sup>3</sup>

<sup>1</sup> Department of ECE, A. V. C College of Engineering,  
Mannampandal, Mayiladuthurai 609305, Tamil Nadu, India  
jananiphdl5@gmail.com

<sup>2</sup> Department of Electrical Engineering, Annamalai University,  
Annamalai Nagar, Chidambaram 608002, Tamil Nadu, India

<sup>3</sup> Department of ECE, Bharathiyar College of Engineering and Technology,  
Karaikal 609609, Tamil Nadu, India

**Abstract.** The paper formulates the cluster based (Hybrid Energy Efficient Distribution) HEED routing strategy for cognitive radio sensor networks (CRSN) using the principles of spectrum sensing. The theory of spectrum sensing enables the secondary users to effectively use the vacant channels and ensure a sense of protection for the primary users. Though heterogeneous co-operative spectrum sensing periodically increases the throughput of the network, still it necessitates measures for improving the other performance indices. The focus relates to reducing the spectrum sensing duration with a view of maximizing the time for data transmission. The attempts reiterate to extract an efficient solution through hybrid sequential and parallel channel sensing (CS) technique for routing the information. The efforts bring out the benefits in terms of the parameters that include throughput, energy, packet loss, delay and overhead through a comparative study with CS- LEACH and CS-AODV routing schemes. The results obtained in the NS2 platform allow it to claim a place for the use of CS-HEED in real world applications.

**Keywords:** Routing · Cognitive radio sensor networks · CS-HEED  
Spectrum scheduling · Hybrid sequential-parallel channel sensing

## 1 Introduction

The cognitive radio (CR) forms a technique for sensing the vacant bands and enabling the use of the available bands to transmit data. It can operate in the licensed spectrum band, where the CR refers to the secondary user (SU) and acquires access when not used by the primary user (PU). The philosophy of dynamic access gathers interest to foster the cognitive radio enabled devices engage the spectrum efficiently with the available bands.

A cognitive radio sensor network (CRSN) constitutes a multichannel network capable of transferring data between a source and a destination. It inherits two main differences from a traditional wireless sensor network (WSN) in the sense that the

number of available channels differ from time to time and the set of available channels differ for each node in the CRSN. The nodes of a single network in the WSN usually use the same set of the available channels.

While the challenges in WSN that include low energy and hardware limitation increase the complexity of spectrum management, the CRSN does not consider the energy and hardware limitations as constraints. Though the wireless sensor networks (WSNs) offer a fault tolerant nature, serve to be flexible, and find wide spread use, it experiences the constraints of link connectivity, limited bandwidth and processing capability. Besides the theory of clustering allows an efficient way to design efficient network architecture and facilitate effective routing schemes. It reduces the communication overhead, increases reliability and being a structured topology serves to increase the system capacity and stability.

However when the channel experiences fading and shadowing, a *cooperative spectrum sensing* approach gathers strength to augment the degraded performance. It encompasses two successive stages, where in the sensing stage, every cognitive user performs spectrum sensing individually and in the reporting stage it communicates the local sensing observations to a common receiver and allows the latter to make a final decision either on the absence or the presence of the primary user. It selects the most favourable user (*cluster head*) with the largest reporting channel gain to collect the sensing results from the other users in the same cluster and forward them to the common receiver.

## 1.1 Related Works

The CR enabled WSN has been seen to reduce congestion, increase the throughput of the network and offer a reliable performance [1, 2]. It has been operated with fixed spectrum allocation characterized by resource constraints in terms of communication and processing capabilities. The CR enabled sensor nodes have been endowed with the potential ability to access the multiple alternative channels.

The clustered CRSNs have been formed with a number of clusters and periodically transmit their sensed data to the sink through hierarchical routing [3]. The sensor nodes have been equipped with sense and switch facility with the licensed channel by dynamic channel access to reduce the energy consumption.

The sensing strategies have been related to the sensing order optimization and acquiring the stopping time in sequential manner in the event of the channels being sensed one after the other [4]. A time schedule has been assigned to each secondary user for sensing each particular channel at a particular instant. A co-operative spectrum sensing approach has been implemented to increase the efficiency of sensing and reduce the sensing time [5] for allowing multiple SU's sense the same channel at the same time.

A cluster based cooperative spectrum sensing has been proposed in cognitive radio systems for reducing the reporting errors by the fading channels. The decision fusion and energy fusion schemes have been orchestrated to circumvent the drawbacks [6].

An online decision scheduling algorithm has been suggested to determine the sensing period together with a sequential detection for spectrum sensing, suitable for short term channel change [7].

Two problems have been witnessed due to spectrum sensing and channel state estimation and augur efforts for maximizing the secondary user (SU) throughput. It has been solved by means of the secondary user (SU) reporting their sufficient statistics to a fusion centre (FC) and enabling a level triggered sampling [8].

The primary user has been facilitated to transmit its information to the PU's receiver directly or assisted by the SU depending on maximization of the throughput of the secondary user and primary user in each time slot [9]. A game based spectrum allocation mechanism has been proposed for the different number of channels and the dynamic bidding game based spectrum allocation strategy developed [10].

A co-operative sensing scheduling embedded in partially observable Markov decision process has been analyzed as an efficient method of spectrum sensing for decreasing the transmission time of the secondary user for exploiting the other channels effectively [11]. An energy efficient spectrum sensing technique has been outlined for reducing the sensing duration for each user [12].

In group based co-operative spectrum sensing, the secondary users have been grouped such that different groups become responsible for performing different sensing rounds. Three efficient adaptive assignment heuristics have been explained to perform the assignment of users to the group and the assignment of groups to the sensing rounds in a way that the throughput efficiency remains maximized [13].

A partially myopic access strategy has been articulated to prove that it allocates SU traffic to idle spectral bands on an energy efficient framework [14]. The TDMA/round-robin fashion has been used to ensure that the secondary station efficiently shares the specific resources and exhibits perfect coordination. [15].

The cooperative spectrum sensing embedded with energy harvesting secondary user has been showcased to reduce the sensing time and decrease the energy efficiency with increased throughput [16]. A prioritized ordering heuristic has been developed to order channels under the spectrum and a scheduling assignment included for achieving optimal solution [17].

Two heuristic algorithms have been put forth for spectrum sensing and compared with the optimal one-convex concept as applied in the design of an algorithm to solve the heterogeneous scenario optimally [18]. An ant colony based energy efficient sensor scheduling algorithm has been elucidated to provide the required sensing performance and increase the overall secondary system throughput [19]. The CSMA-CA has been laid to achieve fair and efficient throughputs in multi-hop networks by characterizing the worst case bounds for CSMA-CA in one-hop neighbourhood topology [20].

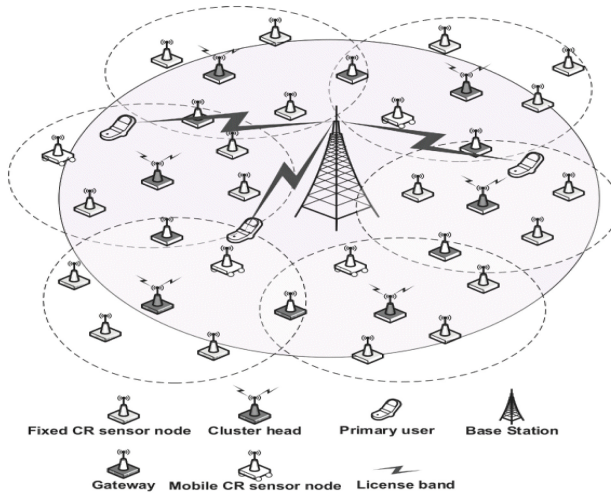
Despite the continuous efforts, still the influence of a cluster of the routing scheme and its subsequent benefits on the performance invites attention.

## 2 Problem Formulation

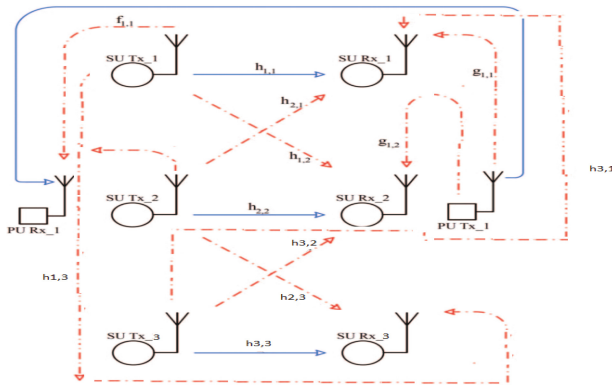
The emphasis orients to evolve a cluster based HEED methodology for routing the data through a hybrid sequential and parallel multi channel strategy in a CRSN. The exercise owes to evaluate its performance on a NS2 portal and establish its merits in terms of indices through a comparative study with CS-HEED, CS-LEACH and CS-AODV methods.

### 2.1 System Model

The procedure involves a clustered CRSN with a primary user and three secondary TXs and three secondary RXs. It operates on the assumption that the PU’s activity in the channel remains independent which allows it to transmit information in the time-slotted fashion with slot duration equal to  $T$ . The Figs. 1 and 2 show the network and system model with one primary trans-receiver and three secondary trans-receivers respectively.



**Fig. 1.** System model for clustered cognitive radio sensor networks



**Fig. 2.** System model with a primary user transmitter (Tx), three secondary transmitters

The process augurs a sensing sequence that includes the time schedule and the priority for the users to sense the channel. If the channel gain operating on secondary transceiver pair be  $s_{x,i}$  (here  $x = 1, 2, 3$ ) and the channel gain operating from the PU transmitter be  $p_{x,y}$ ,

Then the secondary user transmission capacity on the  $m^{\text{th}}$  channel turns out to be as in Eq. 1

$$C_{x,y} = B_x \log_2(1 + \sigma_{x,y}) \quad (1)$$

where  $B_x$  refers to the bandwidth of the  $x^{\text{th}}$  channel,  $1 \leq x \leq 3$

The received SNR of the  $y^{\text{th}}$  SU on  $x^{\text{th}}$  channel can be expressed from Eq. 2 as

$$\sigma_{x,y} = \frac{p_y \alpha_{x,y}^2}{\alpha_x^2} \quad (2)$$

where  $p_y$  is the transmission power for  $y^{\text{th}}$  secondary user

$\alpha_x^2$  is the noise power on  $x^{\text{th}}$  channel

If  $\sigma_{x,y}$  follows to be the same for all the SUs, then the SNR can be related as in Eq. 3

$$\vartheta_{x,y} = \frac{\gamma_x \rho_{x,y}^2}{\alpha_x^2} \quad (3)$$

where  $\gamma_x$  governs the transmission power of the primary user at channel  $x$ .

The collective decisions arrive from several individual sensing results and therefore a node which engages to send packet to another node, transmits a series of short preamble messages on the common channel control (CCC) that contains the destination ID and location of transmitter node. The data transmission includes the acknowledgement (ack), negative acknowledgement (nack) and end of data which begins on the selected channel. The sensing time however can be reduced based on the requirements of the individual users. In time slotted fashion since the status of channel does not change during each time slot, thus by reducing the sensing time, the throughput of the network can be increased.

The collective decisions in co-operative spectrum sensing, however depends on the several individual sensing results. The sensing time can be reduced based on requirements of individual users and in time slotted fashion since the status of channel does not change during each time slot, thus by reducing the sensing time, the throughput of the network can be increased.

If the PUs, whose pilot signals remain detected by the secondary receiver and  $\beta$  be the minimum sensing time(MST), required to satisfy the given detection quality under additive white Gaussian noise (AWGN) channel by the optimal detector, then the matched filter becomes equal to as observed from Eq. 4

$$MST, \beta = \frac{(Q^{-1}(P_f) - Q^{-1}(P_d))^2}{\vartheta f_s} \quad (4)$$

where  $\vartheta$  is the detected SNR



$f_s$  is the receiver sampling frequency

$P_d$  is the Probability of detection

$P_f$  is the Probability of false alarm

The sensing results provided by the different users can be combined to evolve a decision based on AND and OR fusion rules where they pave the way for deciding k out of N.

If the N users continue to be heterogeneous, then the cumulative probability detection can be written as in Eq. 5

$$Q_d = 1 - (1 - P_d)^N \text{ [i.e. OR fusion rule]} \quad (5)$$

The cumulative Probability of false alarm may be given by Eq. 6

$$Q_f = 1 - (1 - P_f)^N \text{ [i.e. OR fusion rule]} \quad (6)$$

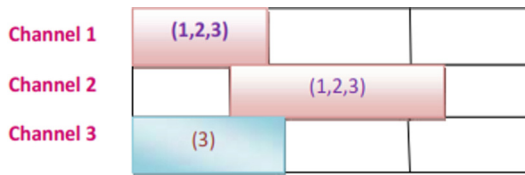
The Eqs. 7 and 8 explain the cumulative probability detection and that of the false alarm respectively

$$Q_d = (P_d)^N \text{ [i.e. AND fusion rule]} \quad (7)$$

$$Q_f = (P_f)^N \text{ [i.e. AND fusion rule]} \quad (8)$$

### 3 Proposed Methodology

The scheme evolves a hybrid sequential-parallel strategy shown in Fig. 3, where the channels divide into several subsets and within each subset, a subset of users adopt the sequential cooperative sensing while the different channel subsets orient to sense in parallel. It necessitates finding the channel subsets, the assignment of users to each subset and the sequential sensing order with each subset.



**Fig. 3.** Sequential-parallel channel sensing strategy, channel (1&2) sensing sequentially and channel (3) sensing in parallel

It echoes to define a function  $R_S(CS_m, V)$  as the maximum expected throughput obtained from the sequential co-operative sensing by  $V$  users with the channel subset  $CS_m$ . The secondary network controller decides the spectrum opportunities for transmission at the beginning of each time slot. The exercise involves the use of greedy heuristic algorithm to maximize the throughput

The greedy approach starts with the knowledge of the channel subset and the number of users assigned to the subset and the algorithm follows the steps in Fig. 4 for the remaining users and channels.

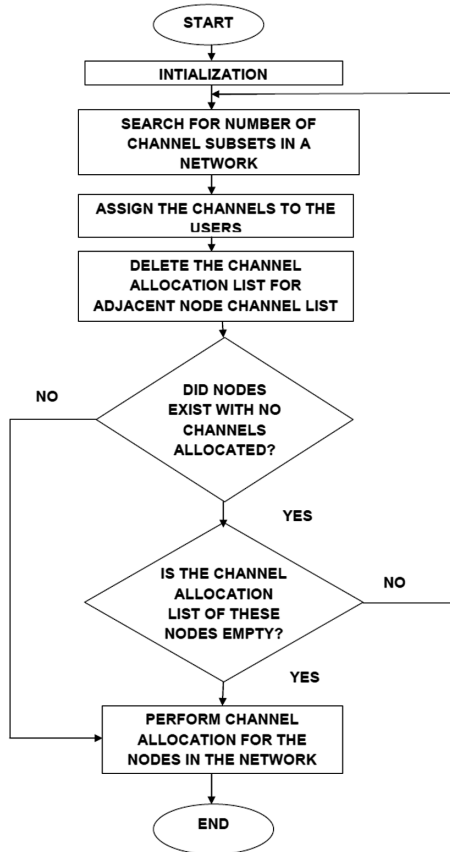
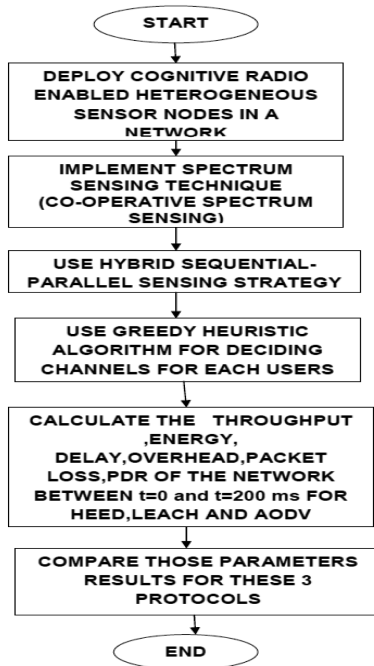


Fig. 4. Flowchart for Greedy heuristic approach

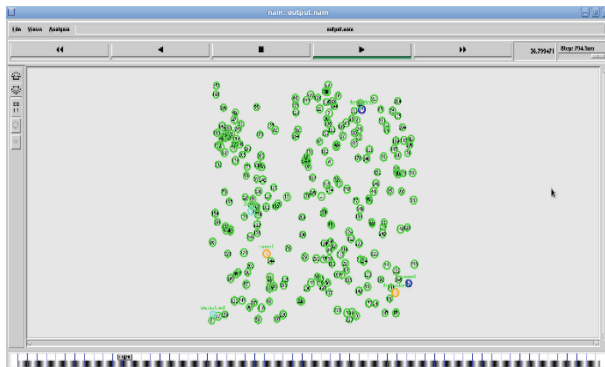
The flowchart in Fig. 5 explains the various steps involved in creating the cluster based CS-HEED and enabling the transfer of data.



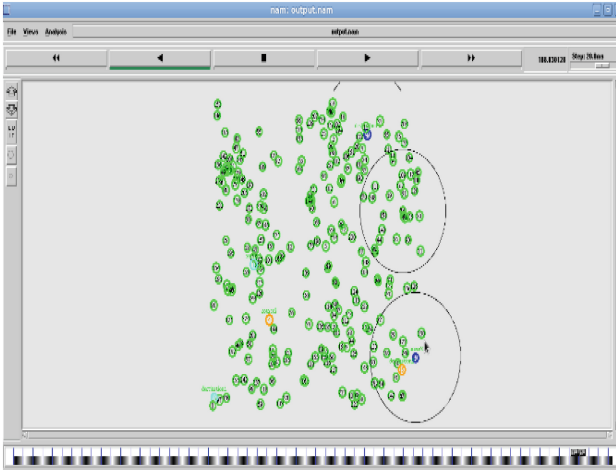
**Fig. 5.** Flowchart for proposed approach

## 4 Simulation Results

The Figs. 6 and 7 below show the network model using NS2 simulator and routing of data in the network respectively. It evaluates the performance with the parameters for the chosen network configured as in Table 1. The methodology investigates the routing of data using CS-HEED, CS-LEACH and CS-AODV protocol. The Table 1 displays the simulation parameters used for modelling the network.



**Fig. 6.** Network model



**Fig. 7.** Routing of data using NS2

**Table 1.** Simulation parameters

Simulation parameters	Value
Network size	1000 × 1000 m
Number of sensor nodes	250
Communication range	250 m
Initial energy of sink	100 J
Packet size	1000 to 5000
Transmission power	0.8 W
Reception power	0.6 W
Simulation time	200 s

It assumes the beginning of time slot as  $t = 0$  (i.e. reference point) and the elapsed time when it completes the sensing process for channel as  $T_I^{(i)}$ . The throughput  $T$  from all channel spectrum opportunities can be determined from Eq. 9

$$\max_A E\{R(A)\} = \sum_{i=1}^M \frac{(T - T_I^{(i)}(A))CS_i(1 - V_i)}{T} \tag{9}$$

where  $(T - T_I^{(i)})CS_i(1 - V_i)$  is the expected throughput

- R the expected normalized throughput T
- T the elapsed time for a channel which remain sensed (no throughput gain)
- $E\{.\}$  the expectation operation
- A the sensing strategy
- $T_I^{(i)}(.)$  be a function of A

The energy efficiency metric may be defined as the ratio of actual energy spent in transmitting one packet over the total energy spent to transmit or receive a packet as in Eq. 10.

$$R(L, CH) = \frac{Energy_{pkt}(\pi_S(L, CH)P)}{Energy_{pkt} + E_{Control}} \quad (10)$$

where

$E_{Control}$  relates to the energy incurred in transmitting control information and monotonically increases as a function of the packet size

$\pi_S(L, CH)$  the steady state probability of node being in transmission state

$P$  the probability of successful transmission

$Energy_{pkt}$  the energy consumed in transmitting successful packet

The packet transmission delay can be expressed as in Eq. 11

Delay = The time taken to wait in the queue when the node is asleep

+ The time spent in community control information

+ The time taken to send the data over data channel.

$$Delay = T_W + T_{CI} + T_d \quad (11)$$

The control overhead for the network may be determined from Eq. 12.

$$C_{OH} = l_{ack} + l_{nack} + l_{endd} \quad (12)$$

where

$C_{oH}$  refers to the control over head

$l_{ack}$  the length of acknowledgement

$l_{endd}$  the length of end of data.

The packet delivery ratio (PDR) can be expressed as in Eq. 13

$$PDR\% = \frac{Number\ of\ Packets\ successfully\ transmitted}{Total\ Number\ of\ Packets} \times 100 \quad (13)$$

The Figs. 8, 9, 10, 11, 12 and 13 bring out the improved performance when the network operates with the hybrid sequential-parallel channel sensing technique. The Figs. 14, 15, 16, 17, 18 and 19 show the CS-HEED protocol to be the best when compared with other two protocols in terms of the performance indices.

The graph depicted in Fig. 8 bring out the influence of channel sensing in its attempt to achieve a higher PDR. Though PDR values decrease with increasing size of the packets transmitted, still they turn out better than that obtained without the channels being sensed.

It is significant to observe that there is minimum loss of packets for the CS-HEED protocol as portrayed in Fig. 9, in addition to decreasing the overhead as seen from Fig. 10 and there from support an increase in the network efficiency.

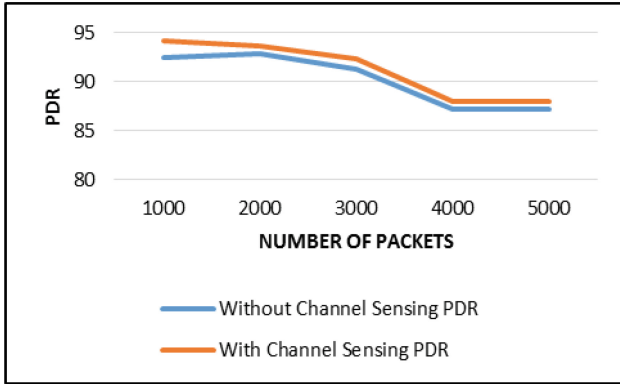


Fig. 8. Performance of PDR comparison with and without hybrid channel sensing

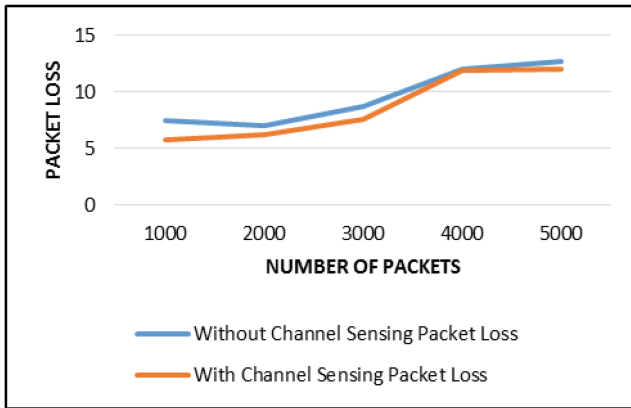


Fig. 9. Performance of packet loss comparison with and without hybrid channel sensing

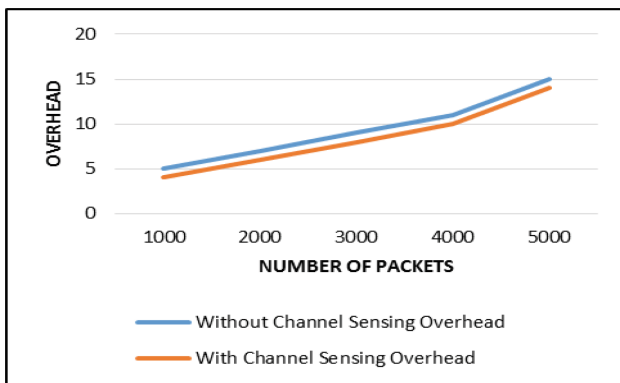
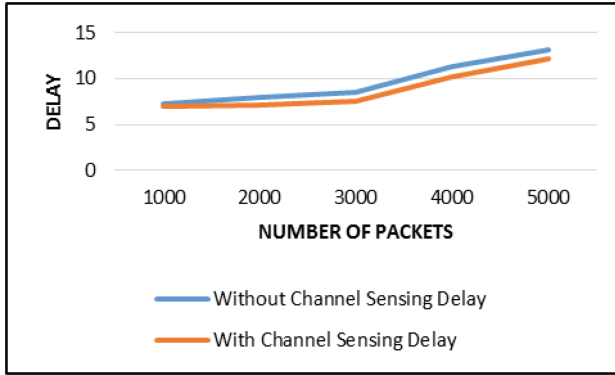
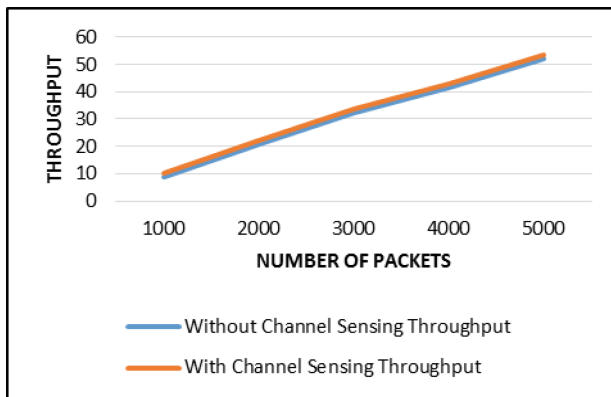


Fig. 10. Performance of overhead comparison with and without hybrid channel sensing



**Fig. 11.** Performance of delay comparison with and without hybrid channel sensing



**Fig. 12.** Performance of throughput comparison with and without hybrid channel sensing

It follows from Figs. 11 and 12 that the CS-HEED routing enjoys a minimum delay for the data transmission between the source and destination nodes and enable the highest throughput respectively, which in turn enhances the performance of the network.

The Fig. 13 explains that the nodes consume the least energy when CS-HEED transfers the data, thereby increasing the lifetime of the network.

The bar diagram in Fig. 14 articulates a higher PDR for CS-HEED over the other two similar routing schemes for increasing sizes of the data and facilitates a significant improvement in the performance of the network. The Packet Loss shown for varying packet sizes in Fig. 15 projects a lower value for the CS-HEED protocol and in turn fosters a faster data transfer rate. The bars in Fig. 16 reflect a smaller decrease in the overhead across increasing packet sizes for CS-HEED when compared with the other two routing schemes.

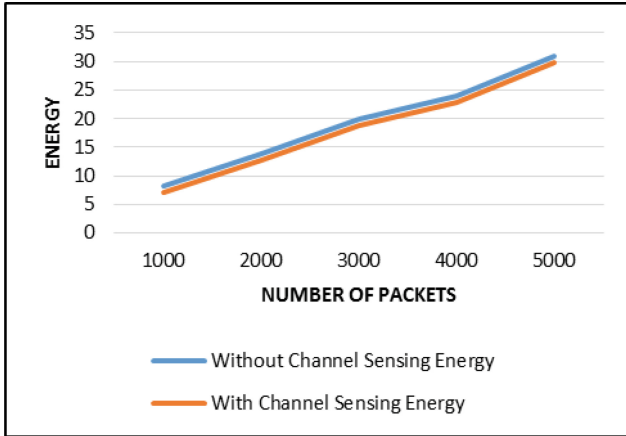


Fig. 13. Performance of energy comparison with and without hybrid channel sensing

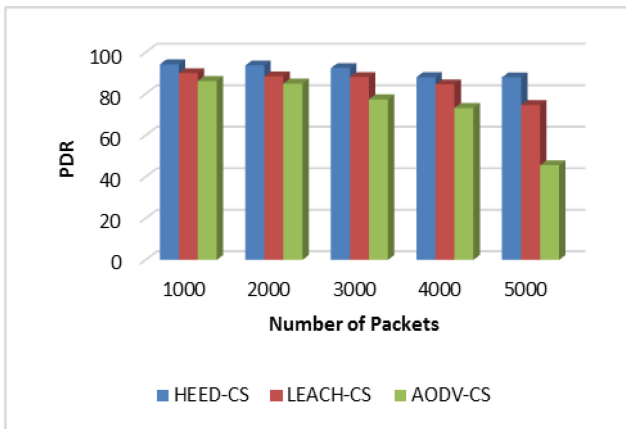


Fig. 14. PDR vs number of packets

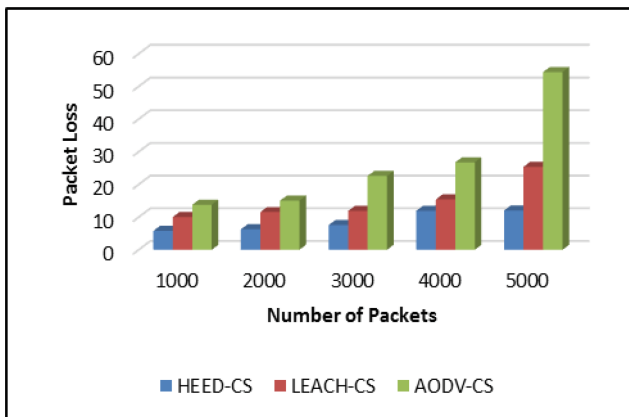
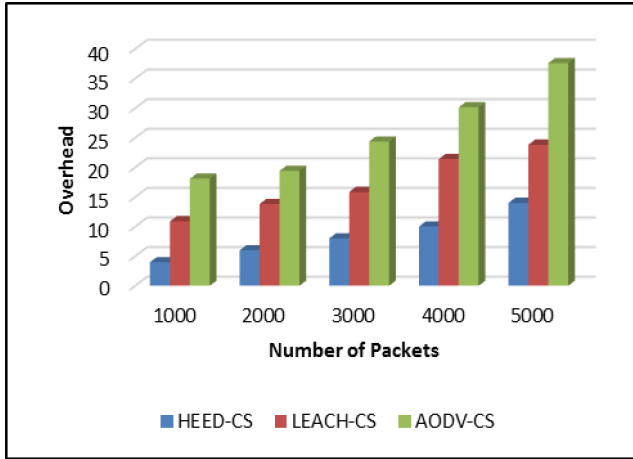


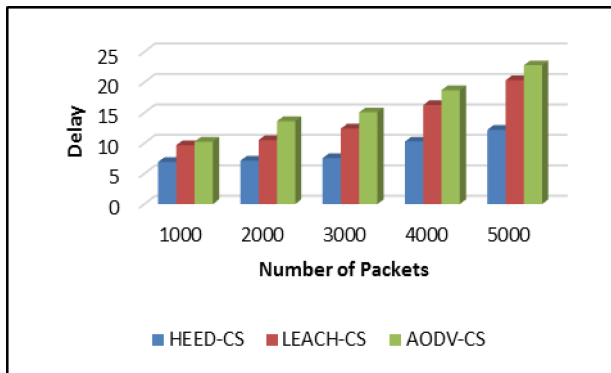
Fig. 15. Packet loss vs number of packets



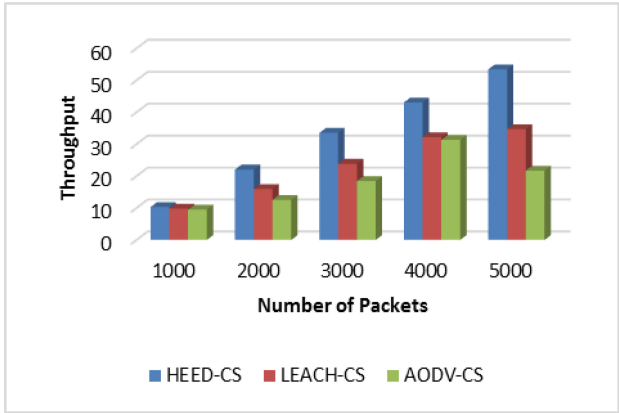


**Fig. 16.** Overhead vs number of packets

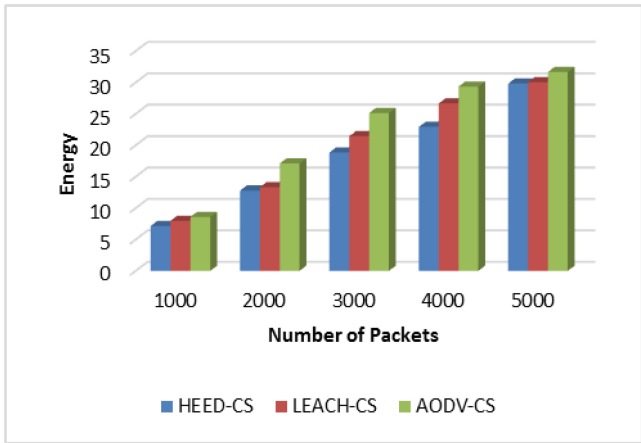
The CS-HEED based data transfer scheme facilitates a smaller delay for increasing sizes of the packet transmission between source and destination node as observed from Fig. 17 thus enabling a higher performance of a network. It follows from the bar diagrammatic representation in Fig. 18 that CS-HEED protocol exhibits a linear increase in throughput over large sizes of data transmission. It evinces from the bar chart in Fig. 19 that the network requires the least energy to transmit the data of same size through the use of the CS-HEED routing scheme and allows it for varying packet sizes.



**Fig. 17.** Delay vs number of packets



**Fig. 18.** Throughput vs number of packets



**Fig. 19.** Energy vs number of packets

## 5 Conclusion

A cluster based data transfer strategy for cognitive radio sensor network has been introduced through the use of spectrum sensing technique. The hybrid sequential-parallel channel sensing technique has been evaluated for three routing methods that include HEED (CS-HEED), LEACH (CS-LEACH) and AODV (CS-AODV) protocol. The performance has been measured using the indices such as throughput, PDR, packet loss, delay, overhead and energy both before and after channel sensing to establish the benefits of sensing process. The results have been presented to claim the best performance for the cluster based CS-HEED over the other two approaches and allows it to claim a place for its use in the utility world.

**Acknowledgment.** The authors acknowledge with thanks the authorities of Annamalai University for providing the facilities to carry out this work.

## References

1. Akan, O., Karli, O., Ergul, O.: Cognitive radio sensor networks. *IEEE Netw.* **23**(4), 34–40 (2009)
2. Vijay, G., Ben, E., Bdira, A., Ibnkahla, M.: Cognition in wireless sensor networks: a perspective. *IEEE Sens. J.* **11**(3), 582–592 (2011)
3. Ren, J., Zhang, Y., Zhang, N., Zhang, D., Shen, X.: Dynamic channel access to improve energy efficiency in cognitive radio sensor networks. *IEEE Trans. Wirel. Commun.* **15**(5), 3143–3156 (2016)
4. Kim, H., Shin, K.G.: Efficient discovery of spectrum opportunities with MAC layer sensing in cognitive radio networks. In: *IEEE Transactions on Mobile Communication Conference*, vol. 7, no. 5, pp. 533–545, May 2008
5. Cabric, D., Tkachenko, A., Roderson, R.W.B.: Spectrum sensing measurements of pilot, energy and collaborative detection. In: *Proceedings of IEEE Mobile Communication Conference*, pp. 2342–2348, October 2006
6. Sun, C., Zhang, W., Letaief, K.B.: Cluster based cooperative spectrum sensing in cognitive radio systems. In: *Proceedings of ICC 2007*, pp. 2511–2515. IEEE Communication Society, July 2007
7. Liu, Q., Wang, X., Cui, Y.: Robust and adaptive scheduling of sequential periodic sensing for cognitive radios. *IEEE Trans. Sel. Areas Commun.* **32**(3), 503–515 (2014)
8. Yilmaz, Y., Guo, Z., Wang, X.: Sequential joint spectrum sensing and channel estimation for dynamic spectrum access. *IEEE Trans. Sel. Areas Commun.* **32**(11), 2000–2012 (2014)
9. Gao, H.Y., Ejaz, W., Jao, M.: Co-operative wireless energy harvesting and spectrum sharing in 5G networks. *IEEE Access* **4**, 2790–2796 (2016)
10. Jing, C., Sheng, W.J., Wencho, Y.: Spectrum allocation strategy for heterogeneous wireless service based on bidding game. *KSII Trans. Internet Inf. Syst.* **11**(3), 1336–1356 (2017)
11. Zhang, T., Wu, Y., Lang, K., Tsang, D.H.K.: Optimal scheduling of co-operative spectrum sensing in cognitive radio networks. *IEEE Syst. J.* **4**(4), 535–549 (2010)
12. Eryigit, S., Bayhan, S., Tugcu, T.: Energy efficient multi channel co-operative sensing scheduling with heterogeneous channel conditions for cognitive radio networks. *IEEE Trans. Veh. Technol.* **62**(6), 2690–2699 (2013)
13. Khalid, L., Anpalagan, A.: Adaptive assignment of heterogeneous users for group based co-operative spectrum sensing. *IEEE Trans. Wirel. Commun.* **15**(1), 232–246 (2016)
14. Michelusi, N., Mitra, U.: Cross-layer estimation and control for cognitive radio : exploiting sparse network dynamics. *IEEE Trans. Cogn. Commun. Netw.* **1**(1), 128–145 (2015)
15. Zame, W., Xu, J., Van Der Schaar, M.: Co-operative multi-agent learning and co-ordination for cognitive radio networks. *IEEE J. Sel. Areas Commun.* **32**(3), 464–477 (2014)
16. Abdulkadir, C., Aisharoa, A., Kamal, A.E.: Hybrid energy harvesting based co-operative spectrum sensing and access in heterogeneous cognitive radio networks. *IEEE Trans. Cogn. Commun. Netw.* **3**(1), 37–48 (2017)
17. Ceik, A., Kamal, A.E.: Green co-operative spectrum sensing and scheduling in heterogeneous cognitive radio networks. *IEEE Trans. Cogn. Commun. Netw.* **2**(3), 238–248 (2016)

18. Zhang, T., Tsang, D.H.K.: Co-operative sensing scheduling for energy efficient cognitive radio networks. *IEEE Trans. Veh. Technol.* **64**(6), 2648–2662 (2015)
19. Liu, X., Evans, B.G., Moessner, K.: Energy efficient sensor scheduling algorithm in cognitive radio network employing heterogeneous sensor. *IEEE Trans. Veh. Technol.* **64**(3), 1243–1249 (2015)
20. Jindal, A., Psounis, K.: On the efficiency of CSMA-CA scheduling in wireless multi-hop networks. *IEEE/ACM Trans. Netw.* **21**(5), 1392–1406 (2013)

# **IoT Security**

# Improved Recommendation Filtering Component Resilient to Trust Distortion Attacks in a MANET

Shirina Samreen<sup>1(✉)</sup> and Akhil Jabbar Meerja<sup>2</sup>

<sup>1</sup> Anurag Group of Institutions, Hyderabad, Telangana, India  
shirina.samreen@gmail.com

<sup>2</sup> Vardhaman College of Engineering, Hyderabad, Telangana, India

**Abstract.** The resilience of a mobile ad hoc network (MANET) to malicious packet droppers is directly related to the design of a trust management framework (TMF) which assigns a trust metric value to each node through behavioral monitoring. As the network topology in a MANET changes dynamically, a node may be uncertain about the trust metric associated with some other node with which it never had interacted. Under such circumstances, recommendations from other trusted peers may help in the computation of trust which is known as indirect trust component. Even though recommendations facilitate the trust computation of a TMF, malicious nodes may launch trust distortion attacks which intend to misguide the nodes about the trustworthiness of other nodes. Dishonest recommenders form a category of malicious nodes which launch the trust distortion attacks. To cope up with such attackers, a TMF should encompass a recommendations filtering component. The current work serves to improvise upon the recommendations filtering scheme used in an uncertainty reasoning based TMF proposed in our earlier work.

**Keywords:** Mobile ad hoc networks · Trust management framework  
Dishonest recommenders · Uncertainty reasoning · Trust distortion attacks

## 1 Introduction

Trust management frameworks facilitate in the formation of a reliable route based upon the trust metric computed by assessing the node behavioral characteristics. Due to limited battery power and dynamically changing network topology, the behavioral monitoring is limited to the neighborhood. This results in nodes having complete uncertainty about the behavior of other nodes with which the former never had an interaction. Hence the node has to depend upon recommendations received from other trusted nodes about the behavior of those nodes.

The current paper proposes an enhancement to one of the modules of a novel trust management framework (TMF) robust to dishonest recommenders [1]. The TMF proposed in earlier works [2] is based upon uncertainty reasoning, wherein the trust associated with a node is represented by a triple comprising the components of belief, disbelief and uncertainty. It is used in the design of a reliable routing protocol for a MANET called as Path Allegiance Metric based routing protocol which requires a node

to have an assessment of trust metric upon its upstream and downstream nodes which fall upon a source to destination path. When a node's trust assessment upon another node is complete uncertainty, then it depends upon the recommendations received from its neighborhood to compute the indirect trust.

The existence of dishonest recommenders requires a recommendation filtering module to filter out the recommendations received from dishonest recommenders so as to have accurate computation of a composite indirect trust obtained from honest recommenders. The proposed filtering mechanism employs a clustering-based algorithm similar to the algorithm proposed by Yu et al. [3]. It involves two phases: first stage clustering based upon the least inter-cluster distance and second stage clustering through merging based upon inter-cluster distance threshold.

The proposed approach is an improvement over the recommendations filtering proposed in earlier works [1] and is distinct in the following ways:

- Computation of inter-cluster distance based upon weighted mean centroid rather than the mean of the distances between individual pair-wise recommendations (data points) between two clusters.
- Additional merging of clusters based upon inter-cluster distance threshold.
- Computation of indirect trust using the cluster with the largest recommendations count and least uncertainty.
- Usage of weighted mean centroid to compute the indirect trust from the selected.

The above two modifications to the recommendation filtering scheme proposed earlier [1] act as enhancements to the recommendations filtering scheme by filtering out most of the dishonest recommenders' recommendations.

## 2 Related Work

When the trust management framework involves an indirect trust computation component, the recommendations received from multiple recommenders have to be appropriately aggregated to generate the final indirect trust. A malicious node intending to perform a trust distortion attack can compromise the indirect trust evaluation component by acting as a dishonest recommender. The various forms of dishonest recommender attacks can be categorized as follows:

**Slandering attack:** The strategy comprises propagation of fake negative recommendations so as to reduce the overall trust of a victim node.

**Self-Promoting attack:** The strategy comprises the propagation of fake positive recommendations so as to increase the overall trust of a node. The dishonest recommender and the recommendee can mutually cooperate so as to upgrade each others trust by misleading the trust management framework.

**Collusion attack:** It involves a combination of slandering and self-promoting attack wherein malicious nodes collaboratively work to upgrade their own trust and lower other nodes trust through fake recommendations.

**Time based On-Off dishonest recommenders attack:** It involves an intelligent malicious node providing fake recommendations as well as correct recommendations

at random periods of time so as to evade detection/ filtering out by the indirect trust computation component.

**Node based On-Off dishonest recommenders attack:** It involves an intelligent malicious node providing fake or incorrect recommendations to some nodes and correct recommendations to some other nodes so as to perform trust distortion and launch a conflicting behavior attack.

The counter-measures to overcome the above attacks involve a recommendation filtering mechanism to eliminate the recommendations from dishonest recommenders. The various approaches which can be employed are as follows: (i) Majority-rule based approaches (ii) Personal-experienced approaches and (iii) Service-reputation based approaches. In majority-rule based approaches, the recommendations which are most consistent with each other and which form the majority in the set of total number of recommendations received are considered to compute the overall indirect trust. Yu et al. [3] proposed an approach wherein the recommendations are clustered based upon a similarity measure and the cluster with largest recommendations count is selected to compute the indirect trust.

In personal-experienced based approaches [4], the recommenders are chosen based upon the personal experience of the evaluating node with the recommendee. The approach can fail when the deviation between the interaction history of the evaluating node and the recommender node with the recommendee is quite large. Especially, the approach can fail, when the malicious node exhibits an on-off attack pattern wherein the recommendee can exhibit a malicious behavior with evaluating node and normal behavior with the recommender node. Such behavior can result in drastic rise in the number of false positives and false negatives.

In service-reputation based approaches [5], the selection of a recommender is based upon the reputation metric value assigned to each node based upon its network services. The approach fails, when the malicious node exhibits intelligent behavior like time-based on-off attack pattern or node-based on-off attack pattern. The failure of the approach is also attributed to the fact, that a generalized metric value computed for service-reputation cannot be used to assess the credibility of a recommender, since a node may be benign with respect to packet forwarding but malicious with respect to providing recommendations.

### 3 Proposed Scheme

The recommender filtering scheme proposed is a part of the indirect trust component of the trust management framework which facilitates the design of a reliable routing protocol called as Path Allegiance Metric based routing protocol. The trust management framework is based upon uncertainty reasoning wherein the trust is represented as a tuple comprising belief, disbelief and uncertainty. The TMF employs beta probability distribution Beta ( $\alpha, \beta$ ) which involves two variables  $\alpha$  and  $\beta$  to represent a measure of packet forwarding and packet dropping behavior respectively. The variables  $\alpha$  and  $\beta$  are then used to compute the values of belief, disbelief and uncertainty. The TMF involves direct trust component and indirect trust component and the final trust



associated with a node can be solely based upon direct observations or may depend upon the recommendations received from the peers based upon the uncertainty associated with the direct trust. The indirect trust is computed by aggregating the received recommendations using a recommendations filtering mechanism so as to, filter out the dishonest recommendations. Hence the accuracy of the overall trust computed using indirect trust depends upon the efficiency of the recommendations filtering mechanism.

The proposed recommendations filtering mechanism is based upon clustering technique proposed by Yu et al. [3]. The formation of clusters is based upon the least dissimilarity computed using Jousselmé’s distance [6]. The following phases are associated with the working of the proposed mechanism.

- Clusters Formation based upon least inter-cluster distance
- Clusters Formation based upon inter-cluster distance threshold.

### 3.1 Inter-cluster Distance Computation

The computation of inter-cluster distance involves two different scenarios listed below:

- (i) Computation of inter-cluster distance between clusters having only one recommendation.
- (ii) Computation of inter-cluster distance between clusters having multiple recommendations.

The former scenario involves the usage of Jousselmé’s distance and the latter scenario involves the usage of weighted centroids of the clusters and also the Jous-selmé’s distance.

In the first phase, initially, each recommendation individually is treated as a cluster, followed by the computation of distance between each pair of clusters. The clusters with smallest distance are merged together into one cluster which constitutes the output of the first step. The output of the first step is given as input the second one wherein the inter-cluster distance between the given set of clusters is computed followed by the merging of least distance clusters. The process is repeated until the number of clusters is reduced to (R/2, where R is the number of recommendations received initially). At this stage, the second phase takes over.

When the clusters are of size is 1, the Jousselmé’s distance between the two recommendations pertaining to each of the clusters can be used to compute the dis-similarity measure. Let p1 and p2 represent two basic probability assignments comprising the tuple <b, d, u>, then,  $0 \leq \Delta(p_1, p_2) \leq 1$

$$\Delta(p_1, p_2) = \sqrt{\frac{1}{2} \left( \|\vec{p}_1\|^2 + \|\vec{p}_2\|^2 - 2\langle \vec{p}_1, \vec{p}_2 \rangle \right)}$$

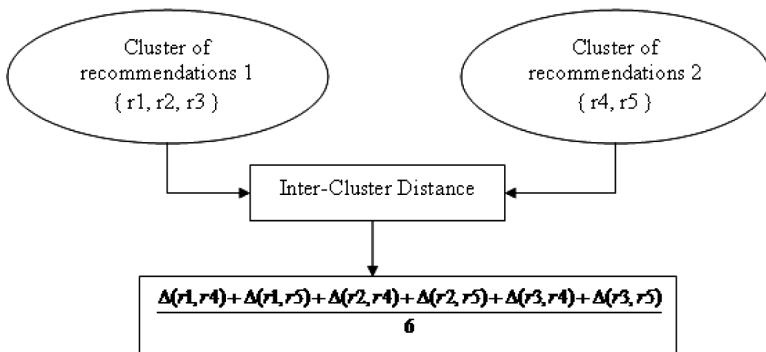
Here  $\|\vec{p}_1\|^2 = \langle \vec{p}_1, \vec{p}_1 \rangle$ ,  $\|\vec{p}_2\|^2 = \langle \vec{p}_2, \vec{p}_2 \rangle$ , and  $\langle \vec{p}_1, \vec{p}_2 \rangle$  is scalar product defined by  $\langle \vec{p}_1, \vec{p}_2 \rangle = \sum_{i=1}^{2^\Omega} \sum_{j=1}^{2^\Omega} p_1(A_i) p_2(A_j) \frac{|A_i \cap A_j|}{|A_i \cup A_j|}$  where for i, j = 1, 2, . . . . .,  $2^\Omega$ .

The main contribution/novelty of the paper lies with the computation of inter-cluster distance, when the clusters contain more than one recommendation. In our previous work [1], the computation involves the summation upon the pair-wise distances upon all possible pair of recommendations such that, the first member in the pair is drawn from the first cluster and the second member is drawn from second cluster. Mathematically the idea is represented as follows:

$$OSM_{c_1,c_2} = \frac{\sum_{\forall(m_1 \in c_1 \wedge m_2 \in c_2)} (1 - \Delta(m_1, m_2))}{|c_1| \times |c_2|}$$

where,  $c_1$  and  $c_2$  represent two clusters with multiple recommendations, represent the opinion similarity measure between clusters  $c_1$  and  $c_2$ ,  $|c_1|$  and  $|c_2|$  represent the size of the two clusters equal to the number of recommendations within each.  $\Delta(m_1, m_2)$  is the Jousseme’s distance between two recommendations represented by  $m_1$  and  $m_2$ .

The difference between our earlier work [1] and the proposed mechanism with respect to inter-cluster distance computation based on which the merging of clusters takes place is depicted through an example in Figs. 1 and 2. In the current proposal, when the clusters are of size greater than 1, then the distance between them is computed using the distance between their weighted mean centroids. The computation of the weighted mean centroid of a cluster is done as follows:



**Fig. 1.** Inter-cluster distance computation using the average of pair-wise distances between the members of cluster

Let  $m_1, m_2, m_3, \dots, m_n$  represent the recommendations which form the members of a cluster. To compute the weighted centroid, the weights assigned to each member have to be computed as follows:

$$w(m_i) = \frac{\sum_{1 \leq j \leq n, i \neq j} \Delta(m_i, m_j)}{n - 1}$$

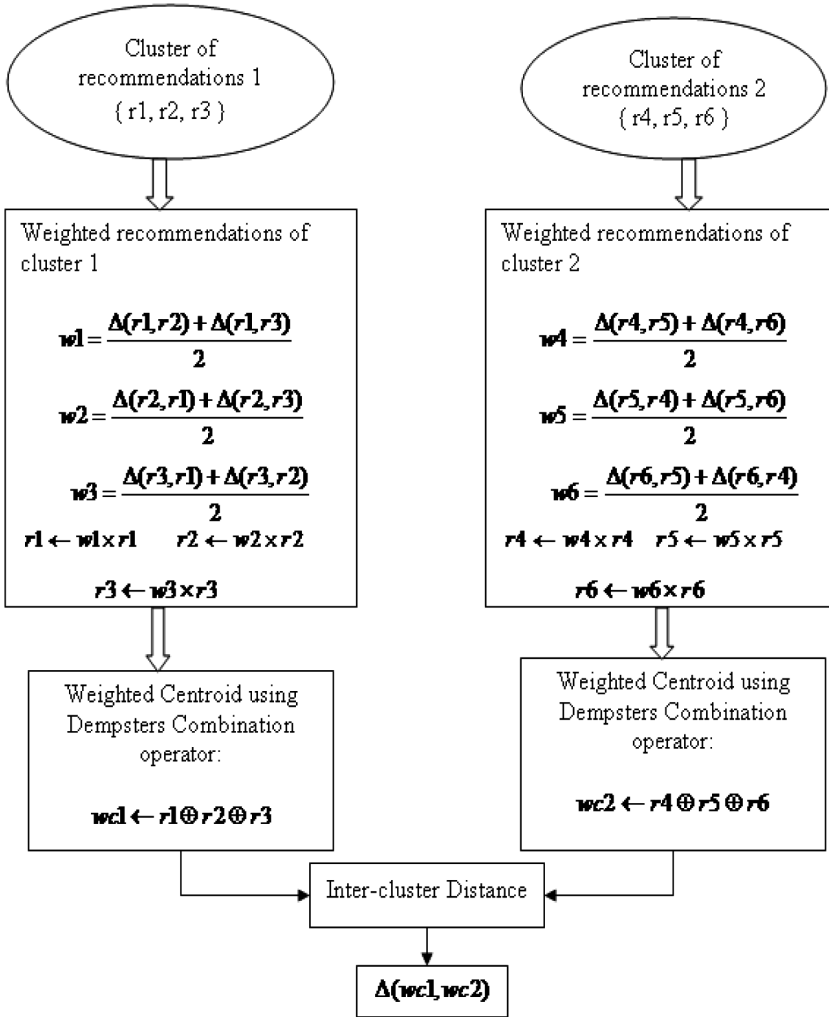


Fig. 2. Inter-cluster distance computation using the weighted centroids of the clusters

Then each of the weighted recommendations is computed as follows: Let the recommendation defined by  $p_i, 1 \leq i \leq n$  be represented as follows in the form of a basic probability assignment:

$$p_i(\{T\}) = b_i, p_i(\{-T\}) = d_i \text{ and } p_i(\{T, -T\}) = u_i$$

then, weighted recommendation  $\hat{p}_i$  is computed as follows:

$$\hat{p}_i(\{T\}) = b_i \times w(p_i) \quad \hat{p}_i(\{-T\}) = d_i \times w(p_i) \quad \text{and} \\ \hat{p}_i(\{T, -T\}) = u_i \times w(p_i)$$

After the computation of weighted recommendations, the centroid is computed using the Dempster's rule of combination.

### **Algorithm to perform recommendations filtering**

Recommendations\_Filtering (  $r_1, r_2, \dots, r_n$  )

Let  $R = \{ r_1, r_2, \dots, r_n \}$  be the set of recommendations

Let  $T = \{ \}$  be an empty set initially

Initialize the  $n$  recommendations  $r_1, r_2, \dots, r_n$  to clusters

$c_1, c_2, \dots, c_n$  and set  $S = \{ c_1, c_2, \dots, c_n \}$

Initialize the weighted centroid of each cluster to the  $\langle b, d, u \rangle$  tuple of each recommendation

For each cluster  $c_i$  in  $S$  such that  $i = 1$  to  $n$  do

  For  $j = i+1$  to  $n$  do

    Compute inter-cluster distance between  $c_i$  and  $c_j$

    If  $|c_i|=1$  and  $|c_j|=1$  then

      Compute the Jousselmes distance between  $r_i$  and  $r_j$

$\Delta(c_i, c_j) \leftarrow \Delta(r_i, r_j)$

    Else

      Compute the weighted centroids of  $c_i$  and  $c_j$

$wc_i \leftarrow \text{weighted\_centroid}(c_i)$

$wc_j \leftarrow \text{weighted\_centroid}(c_j)$

$\Delta(c_i, c_j) \leftarrow \Delta(wc_i, wc_j)$

      If  $\min > \Delta(c_i, c_j)$  then

$\min \leftarrow \Delta(c_i, c_j)$

      fi

    EndFor

EndFor

For  $i = 1$  to  $n$  do

  For  $j=i+1$  to  $n$  do

    if  $\min == \Delta(c_i, c_j)$  then

$c_k \leftarrow \{c_i\} \cup \{c_j\}$

$c_i \leftarrow c_k$

$T \leftarrow T \cup \{c_j\}$

    fi

  EndFor

EndFor

$S \leftarrow S - T$

**Algorithm to compute weighted centroid**weighted\_centroid( $c_x$ )For each recommendation  $r_i \in c_x$  such that  $i=1$  to  $n$  do    For each recommendation  $r_j \in c_x$  such that  $j=1$  to  $n$  do        If  $j=i$  then

continue

Else

 $w_i = w_i + \Delta(r_i, r_j)$ 

EndFor

 $w_i = w_i / (n-1)$      $wr_i < b, d, u > <- r_i < w_i.b, w_i.d, w_i.u >$ 

EndFor

 $weighted\_centroid(c_x) \leftarrow wr_1 \oplus wr_2 \oplus wr_3 \oplus \dots \oplus wr_n$     where  $\oplus$  is the Dempster's combination operator.**3.2 Clusters Formation Based upon Inter-cluster Distance Threshold**

After the number of clusters are reduced to  $R/2$  (where  $R$  is the number of recommendations received), further reduction of the number of clusters is done through merging the clusters based upon inter-cluster distance threshold (set as 0.3). Through experimental analysis, the value of 0.3 as inter-cluster distance threshold or opinion similarity measure was observed to have the most similar <belief, disbelief, uncertainty> components forming the trust metric. For the clusters, with single recommendation, the weighted centroid by default is the recommendation itself and for clusters with multiple recommendations, the computation of weighted centroid is as explained in Sect. 3.1. The second stage of merging involves merging of those clusters whose inter-cluster distance is less than or equal to the threshold.

**Algorithm to perform second stage cluster merging**Second stage cluster merging  $S = \{c_1, c_2, \dots, c_k\}$ For each cluster  $c_i$  in set  $S$  such that  $i = 1$  to  $k$  do    For  $j=i+1$  to  $n$  do        Compute the inter-cluster distance between  $c_i$  and  $c_j$             If  $DIST\_THRESH > \Delta(c_i, c_j)$  then                 $c_k <- \{c_i\} \cup \{c_j\}$                  $c_i <- c_k$                  $T <- T \cup \{c_j\}$ 

fi

EndFor

EndFor

 $S <- S - T$

### 3.3 Indirect Trust in the Trust Management Framework Using the Recommendations Filtering Mechanism

After the second stage of cluster merging, the existing set of clusters is considered and the one with highest number of recommendations is chosen. In case, more than one such cluster is available, then the one with least uncertainty is chosen. From the chosen cluster, the weighted mean centroid is computed which forms the final indirect trust used by the TMF to compute the overall trust of a node.

## 4 Performance Evaluation

The efficiency of the proposed improved RecommFilter scheme is determined through a performance analysis through ns-2 network simulator [7]. The network topology used for simulation comprises 50 nodes which follow a random way point mobility model. The routing protocol used is the Path Allegiance Metric Routing Protocol (PAMRP) proposed in our earlier work [2] which is designed based upon the proposed uncertainty reasoning based trust management framework. It forms the most reliable path by excluding the nodes with lower trust values from the path. The attack model used for simulation comprises two types of malicious nodes: Dishonest recommenders and Packet droppers. Data transmission is performed by ten pairs of nodes each sending 8 kb UDP-CBR (Constant Bit Rate) per second. Table 1. below lists the ns-2 experimental parameters an each data value refers to an average of 20 simulation runs. The following are the metrics involved in the performance analysis:

**Trust Convergence** is defined as the gradual change in the average trust value of a node towards the actual (true) trust metric value which reflects the benign or malicious nature of a node. In this context, average trust is computed using all the received recommendations after dishonest recommenders are filtered out.

**Packet Delivery Fraction** is defined as the ratio of number of packets received by the destination to the number of packets sent by the source. As the recommendations filtering scheme, efficiently filters out the dishonest recommenders, malicious

**Table 1.** Experimental parameters

Parameter	Value
Coverage area	800 m × 800 m
Number of nodes	50
Malicious packet droppers	30%
Transmission range	100 m
Simulation time	1000 s
Mobility	Random way point model
Traffic type	UDP – CBR (Constant Bit Rate)
Packet size	512 bytes
Speed	50 m/s
Pause time	1 s

packet droppers are eliminated from the source to destination path thereby increasing the packet delivery ratio.

The performance analysis of the proposed improved RecommFilter scheme involves a comparison with the RecommFilter scheme [1], RecommVerifier scheme [8], and E-Hermes scheme [4]. Figures 3 and 4 show the trust convergence in the presence of 40% dishonest recommenders and 30% malicious packet droppers. Two attack scenarios involving slandering attack and self-promoting attack are considered and the convergence of trust value to its actual value reflecting the true nature of a node is illustrated in the figures. It can be observed that the trust value converges faster to its original value in the case of improved RecommFilter as the computation of inter-cluster distance is based upon weighted centroid employing the shafers aggregation operator whereas the RecommFilter scheme computes the inter-cluster distance based upon the summation of distances between the individual recommendations within each cluster. Hence the former approach precedes the latter in efficiently filtering out of dishonest recommendations. The RecommVerifier scheme has a majority rule based approach employing two novel mechanisms termed as Time-Verifying and Proof-Verifying. It has a lower performance compared to RecommFilter as the trust model is based upon Bayesian inference and the approach creates avenues for a trust distortion attack, when the evaluating node provides fake recommendations in the Proof-Verifying mechanism. The E-Hermes scheme has the lowest performance as it involves a personal-experience based approach which filters out recommendations based upon the deviation from its own experiences which may not work out in a dynamic environment like MANET.

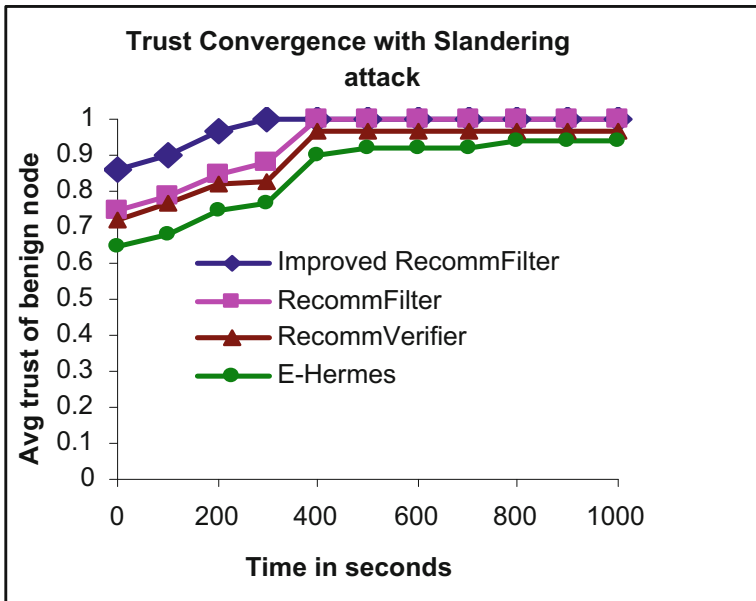


Fig. 3. Trust convergence with slandering attack

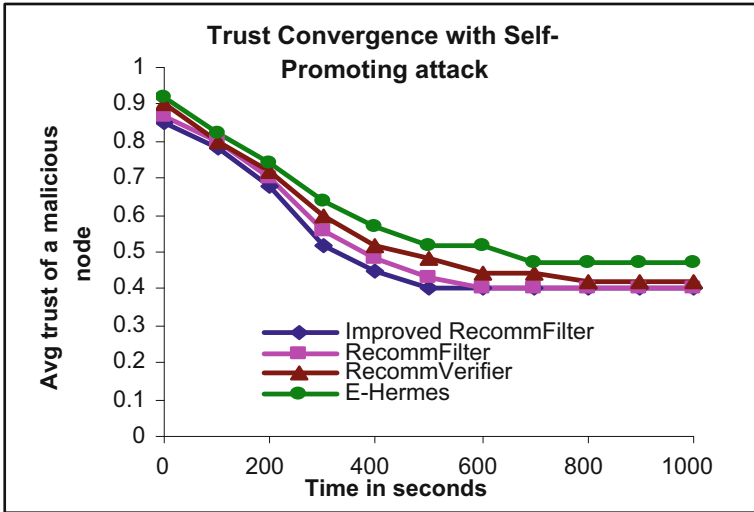


Fig. 4. Trust convergence with self-promoting attack

The packet delivery fraction is directly proportional to the efficiency of the recommendations filtering mechanism. Since the dishonest recommendations are filtered out, the accuracy of trust evaluation is going to be higher and the malicious packet droppers can be avoided while forming a source to destination path. Hence, the PDF of the improved RecommFilter is highest followed by the RecommFilter, RecommVerifier and E-Hermes schemes as can be observed in Fig. 5.

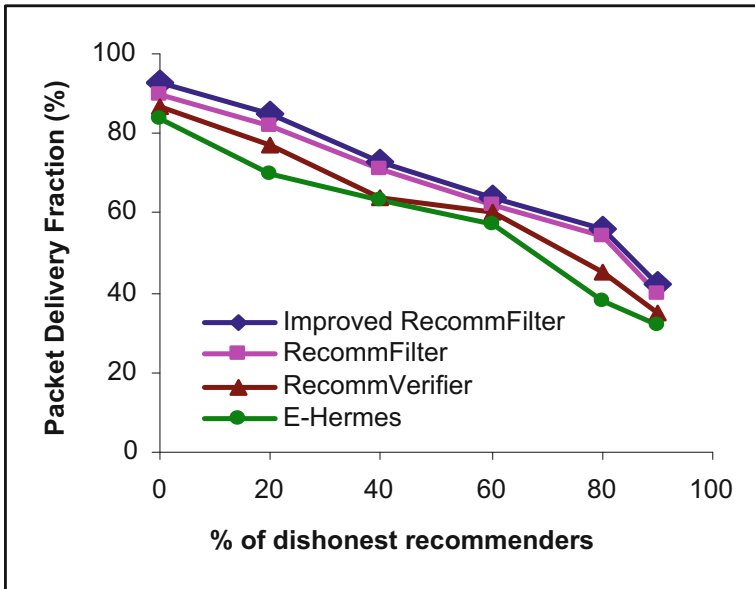


Fig. 5. Packet delivery fraction with collusion attack.



## 5 Conclusion and Future Work

The proposed work is an improvement over the RecommFilter scheme which intends to refine the recommendations filtering mechanism using the clustering mechanism based upon Joussemme's distance between the weighted centroids of the individual recommendations within a cluster. The efficiency of the proposed scheme is also attributed to the usage of weighted centroid computed using the Shafer's aggregation operator upon the weighted recommendations wherein, the recommendations with higher distance with other recommendations within the cluster are discounted proportionally in computing the integrated recommendation of the cluster. The future work intends to integrate the proposed scheme with another scheme which intends to filter out the recommenders initially based upon a credibility measure of the recommender which reflects the referral characteristics and the consistency of the referral behavior to counter the dishonest recommenders with intelligent attack patterns involving time based on-off attack and conflicting behavior attacks.

## References

1. Samreen, S., Narsimha, G.: Dynamically adaptive recommender filtering scheme to defend against dishonest recommenders in a MANET. *Int. J. Sci. Res.* **4**(5), 388–398 (2015). ISSN (Online): 2319-7064
2. Samreen, S., Narsimha, G.: A novel approach to secure routing through path allegiance metric in a mobile ad hoc network. In: *IEEE 2015 International Conference on Electronics, Computing and Communication Technologies (CONECCT 2015)*, July 2015
3. Yu, H., Liu, S., Kot, A.C., Miao, C., Leung, C.: Dynamic witness selection for trustworthy distributed cooperative sensing in cognitive radio networks. In: *13th IEEE International Conference on Communication Technology (ICCT)*, pp. 1–6, September 2011
4. Zouridaki, C., Mark, B.L., Hejmo, M., Thomas, R.K.: E-Hermes: a robust cooperative trust establishment scheme for mobile ad hoc networks. *Ad Hoc Netw.* **7**, 1156–1168 (2009)
5. Zhou, R., Hwang, K.: Powertrust: a robust and scalable reputation system for trusted peer-to-peer computing. *IEEE Trans. Parallel Distrib. Syst.* **18**, 460–473 (2007)
6. Joussemme, A.L., Grenier, D., Bosse, E.: A new distance between two bodies of evidence. *Inf. Fusion* **2**(2), 91–101 (2001)
7. The Network Simulator - ns-2. <http://www.isi.edu/nsnam/>
8. Chen, S., Zhang, Y., Liu, Q., Feng, J.: Dealing with dishonest recommendation: the trials in reputation management court. *Ad Hoc Netw.* **10**, 1603–1618 (2012)

# A Hybrid Group Key Management Scheme for UAV – MBN Network Environment Increasing Efficiency of Key Distribution in Joining Operation

R. Mahaveerakannan<sup>1</sup>(✉) and C. Suresh Gnana Dhas<sup>2</sup>

<sup>1</sup> Information Technology, St. Peter's University, Chennai, India  
mahaveerakannan10@gmail.com

<sup>2</sup> Vivekananda College of Engineering for Women, Namakkal, India  
sureshc.me@gmail.com

**Abstract.** The advent of wireless Ad-Hoc networks has seen an associated emergence of group applications. In order to secure group applications and prevent unauthorized users from accessing communication data that cannot be protected by wireless Ad-Hoc networks and IP multicast alone, group communication content needs to be encrypted by a shared group key. Key management is necessary to ensure the safety of this group key and to protect the group communication. However, problems arise when group key management schemes are applied to the wireless environment relates to three issues: performance, security and network compatibility. In this paper, the focus is made on UAV-MBN, a military network with a proposal to group key management architecture for the UAV-MBN. In order to tackle two particular wireless group key management problems, operational efficiency and multiple-membership changes, a hybrid group key management approach - which operates within each theatre of UAV-MBN is included. By performing micro-key management, this approach can reduce the operational costs associated with key management and improve the operational efficiency of wireless group key management. A membership-oriented group key management scheme is concentrated to tackle the performance problem of multiple-membership changes. This approach, compared to traditional application-oriented group key management approaches, offers more effective management of multiple-membership changes.

**Keywords:** Centralized control · Decentralized control  
Distributed management · Mobile ad hoc networks · Security management  
Unmanned autonomous vehicles

## 1 Introduction

Group key management needs to identify a way to efficiently distribute keying materials to its group members and raise the operational efficiency, which is to be dealt with high priority. This becomes more critical when a group key management approach is applied in the wireless environment due to the resource limitations of both UAV-MBN networks and mobile devices. Several group key management approaches [7–10] exists

and are investigated. However, these encounter serious operational efficiency problems when they are applied in the wireless theatre based architectures similar to that of UAV-MBN. In this paper, a new wireless group key management approach named as hybrid group key management (HGKM) is planned, to be implemented in UAV-MBN network based on the proposed wireless group key management architecture. This new group key management approach, HGKM for UAV-MBN is designed to address the operational efficiency issues with the following assumptions:

- A small key management structure that enables micro-key management operations to achieve efficiency for both the key controller and group members;
- A combination of centralized and decentralized key management approaches that allows members to involve in key management in order to reduce the overhead for the key controller during rekeying;
- A simple message delivery scheme that supports reliable transmission of rekeying messages; and
- Key controllers that co-operate to facilitate rekeying during the handoff.

## 2 Existing Group Key Management Approaches

In the distributed group key management approaches such as Group Diffie-Hellman Key Exchange [1, 6] and Tree-based DH Key Management [8–10], there is no explicit key controller that manages the group key and supporting keys. The group key is generated in such a way that each member contributes its own parts to calculate the shared group key. A number of expensive exponential computations are required in the key generation, which cannot be afforded by mobile devices participating in the UAV-MBN. Furthermore, in order to calculate the group key, each member needs to broadcast its contribution within the group following a one-by-one transmission pattern, which is a lengthy pattern for large groups. Distributed group key management [11, 12] approaches consequently have a slow key converging speed during key generation. Two serious performance issues namely the expensive computation requirement and the slow speed of key generation prevents the distributed key management [13, 14] approaches from being applied to the UAV-MBN network. The centralized group key management schemes such as LKH [3–5] and OFT [2] also face operational efficiency problems when they are deployed in the UAV-MBN network. In centralized group key management, all group members are organized into a single hierarchical key tree. Due to operational efficiency issues, distributed and centralized group key management approaches are not suitable for application in the UAV-MBN network. A more efficient group key management approach is required for the UAV-MBN network.

## 3 Hybrid Group Key Management (HGKM)

Based on the analysis of various centralized group key management approaches, it has become apparent that the reason for operational inefficiency in centralized approaches is rooted in the organization of all group members in one single and large hierarchical

structure. In order to minimize the rekeying impact on the group members and to reduce the overhead of key management, micro-key management operations are proposed to be performed on small management structures as theatres called as operation units. An operation unit is a small fixed-sized key management structure. Each operation unit contains a hierarchical structure for the purpose of key management. Under the HGKM approach, all group members within a theatre form a key management group called theatre-key-management-group for the purpose of key distribution. The theatre-key-management-group is composed of operation units. There are two types of operation units in HGKM: *leader unit and member unit*.

The upper level, called the leader units level, is formed by leader unit(s). There are two roles in the leader unit: *leader and leadership candidate*. Leader refers to a group member in a leader unit who is designated as a leader of a member unit on the lower level. The responsibility of the leader is to assist the theatre key controller (TKC) to distribute the keying materials to the group members within its member unit. On the other hand, leadership candidate refers to a member in a leader unit who has not been assigned as a leader yet. Its responsibility is to work as a backup to the current leader. When a leader leaves the group, a candidate can be designated as the new leader of the affected member unit immediately. The purpose of having leadership candidates is to reduce the operational cost of rekeying when the leader leaves the group.

### 3.1 Generation of Leader and Member Units

In the initial stage of HGKM, a TKC first creates the leader unit. After the TKC completes the generation of leader unit(s), the TKC creates member units for the new incoming group members and designates a member from the leader unit to be the leader of the newly-generated member unit. During the generation of operation units, the TKC continually monitors the changes in the ratio of leadership candidates.

#### The Join Operation

Based on the proposed wireless group key management architecture when a user joins a group, the TKC needs to enforce backward secrecy to prevent the new joining member from decrypting previous group data by updating the theatre traffic encryption key (CTEK). Therefore, in HGKM, the join operation starts with the user sending the join request to the group key controller (GKC) for authentication and authorization. In addition, the incoming member also sends a request via Internet Group Management Protocol (IGMP) to the TKC requesting group communication data from the base station.

$$user \rightarrow GKC : \{group\ join\ request\}$$

$$user \rightarrow TKC : \{request\ for\ receiving\ the\ group\ communication\ data\}$$

After receiving the join request, the GKC validates the user. If authentication is successful, the GKC sends the new incoming member the group traffic encryption key (GTEK) encrypted with a pair-wise key known only to the GKC and the user, while the GKC informs the TKC that the user is authenticated and authorized.

$$\begin{aligned} GKC &\rightarrow user : \{k_{GTEK}\}k_{GKC} - user \\ GKC &\rightarrow TKC : \{user \text{ is authenticated}\} \end{aligned}$$

After receiving this control message from the GKC, the TKC contacts the new joining member and invokes the join procedure. There are three steps to perform the joint action in the theatre.

- Step1: the TKC assigns the new joining member into an operation unit, where the leader unit has priority over the member unit.
- Step 2: the TKC sends the new member a set of keys including the theatre traffic encryption key (CTEK) and the supporting keys it is entitled to know.
- Step 3: the TKC invokes the join rekeying procedure to update the keys for the remaining members of the theatre.

In HGKM, there are two types of join operations within the theatre: the joining leader unit and the joining member unit. The TKC assigns the new member into an operation unit according to the key management policy, where the leader unit is given priority over the member unit. The purpose of doing this is to ensure there are enough members working as leaders and leadership candidates in the leader units' level for key management because the responsibility of a leader is to assist the TKC to distribute rekeying messages within its member unit.

### The Joining Leader Unit

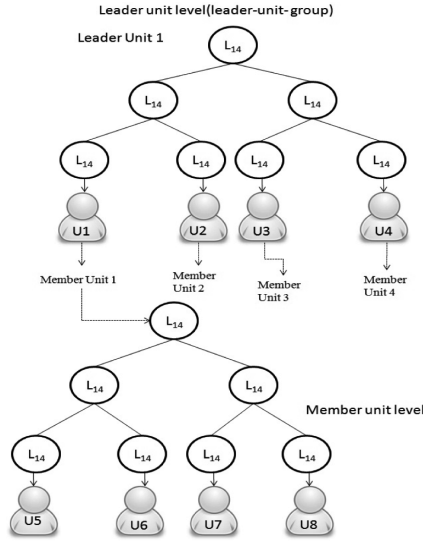
Once the TKC finds an available empty slot in the leader unit, the TKC assigns the joining member into that slot. In order to ensure backward secrecy, the TKC needs to generate new keys to replace the affected current keys. After sending the newly-generated keys to the new member, the TKC invokes the rekeying procedure to update the affected keys for the remaining group members. This rekeying process follows the bottom-up procedure and is divided into two steps.

- Step 1: the TKC updates the keys for the directly-affected leader unit where the new member resides. This rekeying procedure is the same as that in LKH.
- Step 2: the TKC multicasts a rekeying message in the theatre-key-management group where all remaining members reside to update the CTEK for the group members within the theatre.

In order to better illustrate the joining leader unit, we provide an example, shown in Fig. 1, to further describe this rekeying procedure. In Fig. 1, when user 8 joins the group, the TKC assigns user 8 into leader unit 2 as a leadership candidate. The TKC generates the new keys ( $k'_{CTEK}, k'_{78}, k'_{58}$ ) to replace the current keys ( $k_{CTEK}, k_{78}, k_{58}$ ) to ensure backward secrecy. The TKC sends these new keys to user 8.

$$TKC \rightarrow user8 : \{k'_{CTEK}, k'_{78}, k'_{58}\}k_8$$

In this rekeying message, the total of  $h_{unit} + 1$  keys is encrypted by the TKC, where  $h_{unit}$  is the height of the key tree for the operation unit. After rekeying the new joining user, the TKC invokes the rekeying procedure to update the keys for the remaining group members within the theatre. Following the rekeying procedure, in step 1, the



**Fig. 1.** Logical structure of HGKM

TKC updates the affected keys in leader unit 2 where the new member, user 8, has been assigned. The TKC generates two rekeying messages for user 7 and users 5 and 6 respectively to update the affected keys.

$$\begin{aligned}
 \text{User 7} &: \{k'_{78}, k'_{58}\}k_7 \\
 \text{User 5, 6} &: \{k'_{58}\}k_{56}
 \end{aligned}$$

Due to the small size of these rekeying messages, the TKC places them into a single integrated rekeying message and multicasts it directly to the affected leader unit 2.

$$\begin{aligned}
 &TKC \rightarrow \{leader\ unit\ 2\} \\
 &: \{for\{user\ 7\} : \{k'_{78}, k'_{58}\}k_7, for\ \{user\ 5, 6 : \{k'_{58}\} k_{58}\}
 \end{aligned}$$

When the members in leader unit 2 receive this integrated rekeying message, each member can gain the latest keys from the corresponding section. In this step, the TKC sends a single rekeying message and the number of keys encrypted is:

$$1 + 2 + \dots + h_{unit} = \frac{h_{unit}(h_{unit} + 1)}{2} \tag{1}$$

where  $h_{unit}$  is the height of the key tree for the operation unit.

After updating keys in leader unit 2, the TKC generates a rekeying message that contains the latest theatre TEK ( $k'_{CTEK}$ ) encrypted by the current CTEK ( $k_{CTEK}$ ) and multicasts it within the theatre-key-management-group that accommodates all the group members within the theatre.

$$TKC \Rightarrow \{theatre - key - management - group\} : \{k'_{CTEK}\}k_{CTEK}$$

When the remaining group members receive this message, they can gain the new CTEK for the future group communication. During the rekeying procedure of the joining leader unit, it can be observed that the TKC sends 3 rekeying messages. One is for the new joining member, one is for the directly-affected leader unit where the new member resides and the last is for all remaining members. In these three rekeying messages, the total number of keys encrypted by the TKC is

$$(h_{unit} + 1) + (1 + 2 + \dots + h_{unit}) + 1 = \frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 1 \quad (2)$$

where  $h_{unit}$  is the height of the key tree for the operation unit.

During the rekeying, the members in the directly-affected leader unit receive 2 rekeying messages; the first is the integrated message containing all the keying materials for leader unit 2 while the other contains the new CTEK for all remaining group members. The members outside the directly-affected leader unit only receive one single rekeying message to update the CTEK.

**The Joining Member Unit**

If the TKC cannot find an available slot in the leader unit(s) for the new joining member, then the incoming user is assigned to a member unit. The joining procedure is similar to the joining leader unit. The TKC generates new keys to replace the affected current keys in the directly-affected member unit and sends these new keys to the new joining user. After this, the TKC invokes the rekeying procedure to update keys for the remaining group members. Two steps similar to those performed for the joining leader unit are involved.

- Step 1: the TKC updates the keys for the directly-affected member unit where the new joining member has been assigned. The TKC generates and multicasts an integrated rekeying message which contains all the rekeying messages required by the directly-affected members unit.
- Step 2: the TKC multicasts a rekeying message that contains the new CTEK encrypted by the current CTEK to the theatre-key-management group to update CTEK for all remaining group members.

Considering a scenario Fig. 2, when user 9 joins the group, the TKC assigns it into member unit 2. The TKC generates the new keys ( $k'_{CTEK}$ ,  $k'_{910}$ ,  $k'_{912}$ ) and sends them to new joining member, user 9.

$$TKC \rightarrow user9 : \{k'_{CTEK}, k'_{910}, k'_{912}\}k_9$$

In this rekeying message, it can be observed that the number of keys encrypted by the TKC is  $h_{unit} + 1$ , where  $h_{unit}$  is the height of the key tree for the operation unit. After rekeying user 9, the TKC invokes the rekeying procedure to update the keys for the remaining group members. First, in step 1, the TKC updates the keys for the directly-affected operation unit, member unit 2. The TKC refreshes the current keys for users 10, 11, 12 and user 2, the leader of a member unit 2 by multicasting an integrated rekeying message.

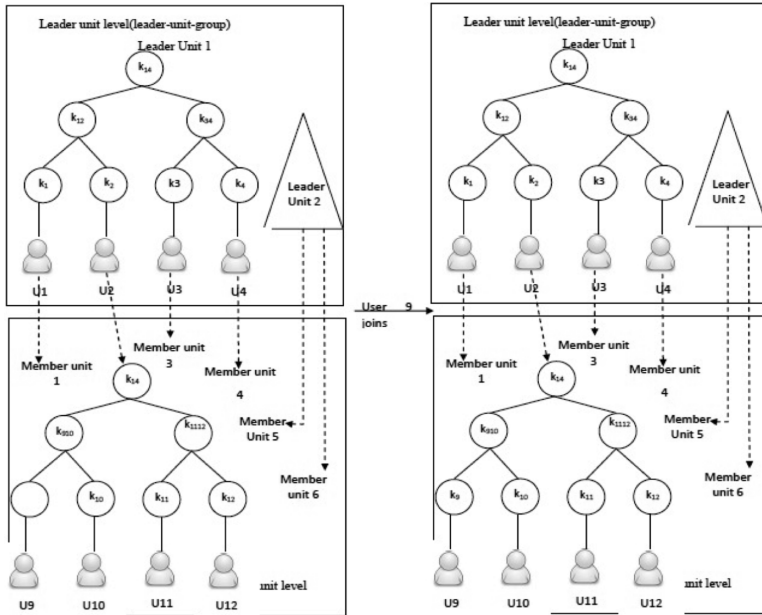


Fig. 2. New user 9 joins the group

$$TKC \rightarrow \{member\ unit\ 2\} : \{for\ \{user\ 10\} : \{k910', k912'\} k10 \\ for\ \{user\ 11, 12\} : \{k'_{912}\} k_{1112}\} \\ for\ \{user\ 2\} : \{k'_{912}\} k_2$$

After receiving this message, users 2, 10, 11, 12 can update their keys from the corresponding section. For the purpose of reliable delivery, the leader of the member unit 2, user 2, stores this rekeying message for a period of time until it receives new keying materials from the TKC. In this process, the TKC sends just one rekeying message and the number of keys encrypted by the TKC during this step is:

$$(1 + 1 + 2 + \dots + h_{unit}) = \frac{h_{unit}(h_{unit} + 1)}{2} + 1 \tag{3}$$

where  $h_{unit}$  is the height of the key tree for the operation unit.

After rekeying the directly-affected member unit, in step 2, the TKC multicasts a rekeying message, which contains the new CTEK,  $k'_{CTEK}$ , to the theatre-key-management-group to update the CTEK for all remaining group members within the theatre.

$$TKC \Rightarrow \{theatre - key - management - group\} : \{k'_{CTEK}\} k_{CTEK}$$

Three rekeying messages are sent by the TKC during the whole rekeying process. One message is for the new joining group member, one is for the directly-affected



member unit and the final message is for the remaining members. The number of keys encrypted by the TKC is:

$$(h_{unit} + 1)(1 + 1 + 2 + \dots + h_{unit}) + 1 = \frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 2 \quad (4)$$

where  $h_{unit}$  is the height of the key tree for the operation unit.

During this rekeying, the leader and the members of the directly-affected member unit receive two rekeying messages: the integrated rekeying message and a message updating the CTEK for all remaining group members. The group members outside the directly-affected member unit only receive one rekeying message updating the CTEK.

From Table 1, The number of messages sent by the TKC and the number of keys encrypted by the TKC during the rekeying process in two different types of join operations is given as;  $h_{unit}$ : the height of the key tree for operation unit in HGKM

**Table 1.** Joining a leader and member unit

	Number of messages sent out by TKC	Number of keys encrypted by TKC
Joining leader unit	3	$\frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 1$
Joining member unit	3	$\frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 2$

## 4 Performance Analysis

Operational efficiency is the highest priority in any wireless group key management scheme due to the resource limitations of both UAV-MBN networks and mobile devices. A wireless group key management approach cannot be recognized as efficient and practical if it cannot meet the requirements of operational efficiency. Therefore, in this section, we analyse and evaluate the operational costs of key management in HGKM to demonstrate that HGKM is an efficient and practical wireless group key management approach suitable for deployment in the UAV-MBN network. In order to evaluate the performance of HGKM, we set the centralized group key management approaches LKH and OFT as the benchmarks. The reasons for selecting these as benchmarks are: (i) LKH and OFT can be applied within the theatre; (ii) LKH and OFT also apply a hierarchical key structure to manage keying materials which are similar to HGKM; and (iii) LKH and OFT are considered to be two of the most widely-used and efficient group key management approaches because the operational costs of these are  $O(\log_d n)$  where  $d$  is the degree of the key tree and  $n$  is the total number of group members.

#### 4.1 Communication Cost

The theatre key controller (TKC) is the main key management entity in HGKM. Consequently, the communication cost generally refers to the communication overhead of the TKC during the rekeying procedure caused by the join and leave procedures. The communication cost can be measured by the number of rekeying messages transmitted by the TKC during rekeying. Without loss of generality, we apply a binary tree to build the key management structure in HGKM, LKH, and OFT, as a binary tree is easy to create, manage and maintain.

##### The Communication Cost of the Join Operation

In HGKM, there are two kinds of join actions: joining the member unit and joining the leader unit. From Table 2, it can be observed that the communication cost of the join action in LKH and OFT is proportional to the size of the whole group. Along with the increased group size, the communication cost of the join procedure in LKH and OFT also becomes larger. In contrast, the communication cost of the join procedure for the TKC in HGKM is a constant value that is achieved by applying an integrated rekeying message to contain all the rekeying information for the affected operation unit. In HGKM, micro-key management is performed within a small fixed-sized operation unit. The number and the size of rekeying messages sent to the affected operation unit are therefore minimal. Because HGKM has the ability to place the rekeying messages into one single integrated message, it can utilize the capacity of multicast transmission to improve the throughput and efficiency of network transmission. We apply the theory of expectation value to calculate the average communication cost of a join action in HGKM.

**Table 2.** The communication cost of the join action for TKC

Group key management algorithm		Communication cost
HGKM	Join member unit	3
	Join Leader unit	3
LKH		$h + 1$
OFT		$h + 1$

$$Cost_{communication}(join) = Cost_{joining\_leader\_unit} \times P_{joining\_leader\_unit} + Cost_{joining\_member\_unit} \times P_{joining\_member\_unit} \quad (5)$$

where  $Cost_{joining\_leader\_unit}$  and  $Cost_{joining\_member\_unit}$  are the communication costs of joining leader unit and joining member unit respectively.

$P_{joining\_leader\_unit}$  and  $P_{joining\_member\_unit}$  present the probability of joining leader unit and joining member unit correspondingly.

In HGKM, the probability of joining leader unit and joining member unit join are:

$$P_{joining\_leader\_unit} = \frac{n_{members\_in\_leader\_units}}{n_{total\_group\_members}}$$

$$P_{joining\_member\_unit} = \frac{n_{members\_in\_member\_units}}{n_{total\_group\_members}}$$

where  $n_{members\_in\_leader\_units}$  is the number of members in the leader units,  $n_{members\_in\_member\_units}$  is the number of members in the member units and  $n_{total\_group\_members}$  is the total number of members in the whole group.

- The communication cost of the join action in HGKM is a constant value independent of the size of the group;
- The communication cost of the join action in LKH and OFT is three to five times higher than that of HGKM. Moreover, the communication cost of the join action in LKH and OFT increases with the growth of the number of group members;
- The gap in the communication cost between HGKM and LKH and OFT becomes increasingly wider as the group size increases.

These three features ensure that HGKM can achieve much better communication efficiency in key management during the join procedure than the LKH and OFT approaches.

### 4.2 Computation Cost

Computation cost is another important parameter that can be used to evaluate the operational efficiency of group key management. The task of encryption and decryption is the heaviest work done by the TKC and group members during rekeying, making it an appropriate criterion for assessment. For the TKC, this parameter can be measured by the number of keys encrypted during the rekeying procedure. For group members, this parameter can be approximately assessed by ascertaining the number of rekeying messages received by members. During rekeying, every rekeying message is multicasted to all group members to improve communication efficiency. Each member needs to process every single received message to determine if it is the intended recipient. The computation cost for the member is therefore directly related to the number of received messages. In the next two sections, we analyse the computation costs of the join and leave operation respectively. We set the costs of LKH and OFT as the benchmarks for the purpose of comparison.

#### The Computation Cost of the Join Operation

- (i) The computation cost of the join operation for the TKC Table 3 summarizes the computation cost of a single join operation in HGKM, LKH, and OFT, where the formulas are from the Sect. 3.

**Table 3.** The computation cost of the join operation for TKC

Group key management	Computation cost of join action for TKC	
	Leader unit	Member unit
HGKM	$\frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 1$	$\frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 2$
LKH	$\frac{(h + 1)(h + 2)}{2} - 1$	
OFT	$2 h$	

$h_{unit}$ : the height of the key tree for operation unit in HGKM  
 $h$ : the height of the group key tree in LKH and OFT

In Table 3, it can be observed that the computation cost of the join action in HGKM is proportional to the power of the height of the key tree for the operation unit,  $2O(h_{unit})$ . This cost is independent of the size of the group members. Once the size of the operation unit is determined, the computation cost of the join operation for HGKM becomes a constant value. In contrast, in terms of LKH, the computation cost of the join operation is proportional to the power of the height of the group key tree,  $O(h^2)$ , while the computation cost of the join action for OFT is in ratio to the height of the group key tree,  $O(h)$ , which is due to the dependent key generation bypassing the old key through a one-way function to obtain a new key.

In Sect. 3, we introduced two join scenarios in HGKM: joining leader unit and joining member unit. Based on the theory of expectation value, for a single join action, the average computation cost is:

Where  $p_{leader\_unit}(join)$  and  $p_{member\_unit}(join)$  are the probability of joining leader unit and the probability of joining member unit respectively. In HGKM, the probability of joining leader unit and joining member unit are:

$$p_{leader\_unit}(join) = \frac{n_{member\_in\_leader\_units}}{n_{total\_group\_members}}$$

$$p_{member\_unit}(join) = 1 - p_{leader\_unit}(join)$$

where  $n_{members\_in\_leader\_units}$  is the number of members in the leader units and  $n_{total\_group\_members}$  is the total number of member in the group. Therefore, the average computation cost of the join action for HGKM can be simplified as follows:

$$\begin{aligned} Cost_{computation}(join) &= \left( \frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 1 \right) \times p_{leader\_unit}(join) \\ &\quad + \left( \frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 2 \right) \times p_{member\_unit}(join) \\ &= \frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 2 - \frac{n_{member\_in\_leader\_units}}{n_{total\_group\_members}} \end{aligned}$$

The comparison of computation cost of join for the TKC in HGKM, LKH, and OFT is shown in Fig. 3. In this example, we assume that the size of the operation unit is 32.

In Fig. 3, three features of the computation cost of the join action for the TKC can be observed.

- Due to the application of dependent key generation, the computation cost of the join action for the TKC in OFT is low and proportional to the height of the group key tree,  $O(h)$ .
- HGKM can achieve the same level efficiency as that of OFT because there keying process is confined within the small-sized operation unit. The computation cost of the join action for HGKM is insensitive to increasing group size. Despite the growth in the number of group members, the computation cost of the join operation for HGKM is little changed. The reason for this is the application of micro-key management within the scope of the operation unit;

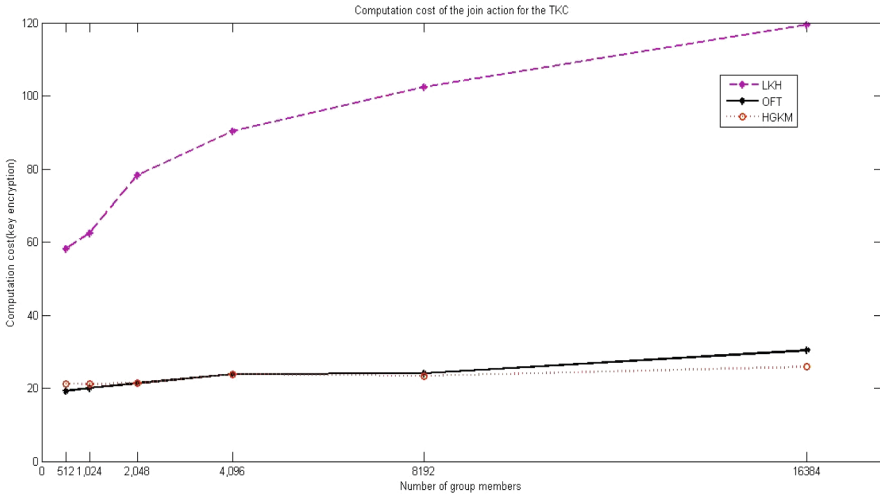


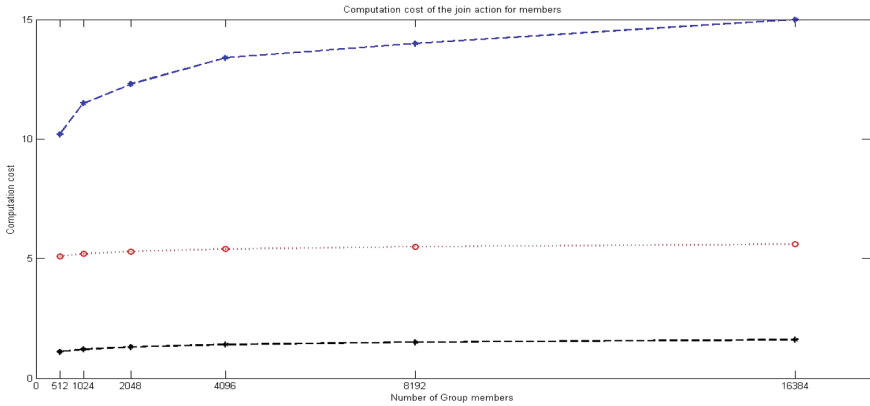
Fig. 3. The computation cost of the join action for the TKC

- LKH has the highest computation cost of the three approaches, with a computation cost of the join action two to six times higher than that of HGKM.

Moreover, the computation cost of the join action for LKH increases logarithmically as the group grows. The gap in the computation cost between LKH and HGKM consequently becomes increasingly wider. In conclusion, HGKM has an advantage over LKH in the computation cost of the join action, with the cost being similar to that of OFT. Moreover, the advantage increases with the growth of the group size.

- (ii) The computation cost of the join operation for members For group members, the computation cost can be measured by the number of received rekeying messages. As discussed in Sect. 3, in the centralized group key management approaches, each rekeying message is multi-casted to the whole group. Furthermore, due to the small number and size of these rekeying messages, they can be placed into an integrated message for more efficient transmission. As a result of this transmission, members in the directly-affected operation unit only need to process this integrated message to find the relevant information. For the purpose of comparison, in HGKM, we assume that the computation cost of the join action for the members in the directly-affected operation unit is the number of rekeying messages contained in the integrated package. Table 4 summarizes the computation cost of the join action for members in HGKM, LKH, and OFT, where the formula is from the previous Sect. 3.

Figure 4, illustrates the computation cost of join on the members’ side for HGKM, LKH, and OFT. In this example, we suppose that the size of the operation unit is 32.



**Fig. 4.** The computation cost of the join action for members

**Table 4.** The computation cost of the join operation for group members

Group key management algorithm	Joining leader unit		Joining member unit	
	Members of the directly affected operation unit	Members outside the directly - affected operation unit	Members of the directly affected operation unit	Members outside the directly-affected operation unit
HGKM	$h_{unit} + 1$	1	$h_{unit} + 1$	1
LKH	$H$			
OFT	$H$			

$h_{unit}$ : the height of the key tree for operation unit in HGKM

$h$ : the height of the group key tree, which equals  $2 \log n$ ,  $n$  is the total number of group members.

From Table 4, it can be observed that the computation cost of join for the members in the directly-affected operation unit is proportional to the height of the key tree for the operation unit in both join scenarios. In contrast, the computation cost of the join action for LKH, and OFT equals the height of the group key tree. Compared to the size of the key tree for LKH, and OFT, HGKM has small operation units. HGKM consequently has better computation efficiency than LKH and OFT. In addition, the most important improvement in the computation cost for HGKM is that the computation cost of the join action for group members outside the directly-affected operation unit is just one. This achievement minimizes the impact of rekeying on the remaining members and drastically improves the computation efficiency for members.

From Table 5, it is observed that the computation cost of the join action for the members in the directly-affected operation unit in HGKM is a constant value and independent of the group size. The reason for this is that the height of the key tree for the operation unit is determined by the size of an operation unit.

**Table 5.** The operation cost for HGKM

HGKM		Operation Cost
Communication	Join(TKC)	3
Computation	Join (TKC)	$\frac{(h_{unit} + 1)(h_{unit} + 2)}{2} + 2 - \frac{n_{member\_in\_leader\_units}}{n_{total\_group\_members}}$
	Join (member)	Members in the directly –affected operation unit: $h_{unit} + 1$ Members outside the directly –affected operation unit: 1

$h_{unit}$ : height of key tree for the operation unit in HGKM

$n_{member\_in\_leader\_units}$ : total number of members in leader units

$n_{total\_group\_members}$ : total number of members in HGKM

## 5 Conclusion

In order to tackle the problems of operational efficiency in wireless group key management, we have proposed a novel group key management approach: hybrid group key management (HGKM) in this chapter. This approach is specifically designed for the UAV-MBN network. The major contribution of HGKM is that HGKM performs micro-key management within a small area known as an operation unit. In HGKM, group members are organized into a number of operation units for the purpose of key management. Based on these operation units, micro-key management is performed. The features that allow HGKM to achieve operational efficiency can be summarized as follows: (1) key management can be restricted within the small scope of the operation unit to reduce operational costs from the perspectives of communication, computation and key storage; (2) HGKM combines the features of centralized and distributed key management approaches to allow members to involve in the key management in order to reduce the operational cost of the TKC during the rekeying process; and (3) HGKM has a simple and built-in reliable message delivery scheme based on the operation unit to provide reliable transmission of keying materials. Based on the analysis, evaluation, and comparison presented in this chapter, we conclude that HGKM can improve operational performance from the perspectives of communication, computation and key storage. HGKM offers an efficient and practical wireless group key management approach for the UAV-MBN network.

## References

1. Steiner, M., Tsudik, G., Waidner, M.: Diffie-Hellman key distribution extended to group communication. In: Proceedings of 3rd ACM Conference on Computer and Communications Security, pp. 31–37, March 1996
2. McGrew, D.A., Sherman, A.T.: Key establishment in large dynamic groups using one-way function trees: TIS. IEEE Trans. Softw. Eng. **29**(5), 444–458 (2003)
3. Wallner, D., Harder, E., Agee, R.: Key management for multicast: issues and architectures. RFC 2627, June 1999
4. Wong, C.K., Gouda, M., Lam, S.S.: Secure group communications using key graphs. In: Proceedings of ACM SIGCOMM, pp. 68–79, October 1998

5. Wong, C.K., Gouda, M., Lam, S.S.: Secure group communications using key graphs. *IEEE/ACM Trans. Netw.* **8**(1), 16–30 (2000)
6. Steiner, M., Tsudik, G., Waidner, M.: Key agreement in dynamic peer groups. *IEEE Trans. Parallel Distrib. Syst.* **11**(8), 769–780 (2000)
7. Rafaeli, S., Hutchison, D.: A survey of key management for secure group communication. *ACM Comput. Surv.* **35**(3), 309–329 (2003)
8. Kim, Y., Perrig, A., Tsudik, G.: Communication-efficient group key agreement. In: *Proceedings of 17th International Information Security Conference*, pp. 229–244, July 2004
9. Kim, Y., Perrig, A., Tsudik, G.: Group key agreement efficient in communication. *IEEE Trans. Comput.* **53**(7), 905–921 (2004)
10. Kim, Y., Perrig, A., Tsudik, G.: Tree-based group key agreement. *ACM Trans. Inf. Syst. Secur.* **7**(1), 60–96 (2004)
11. Sun, Y., Trappe, W., Ray Liu, K.J.: A scalable multicast key management scheme for heterogeneous wireless networks. *IEEE/ACM Trans. Netw.* **12**(4), 53–666 (2004)
12. Kim, H., Chung, B., Lee, Y., Park, Y., Yoon, H.: Weakness of the synchro-difference LKH scheme for secure multicast. *IEEE Commun. Lett.* **11**(9), 765–767 (2007)
13. Park, M.-H., Park, Y.-H., Jeong, H.-Y., Seo, S.-W.: Key management for multiple multicast groups in wireless networks. *IEEE Trans. Mob. Comput.* **12**(9), 1712–1723 (2013)
14. Mapoka, T.T., Shepherd, S.J., Abd-Alhameed, R.A.: A new multiple service key management scheme for secure wireless mobile multicast. *IEEE Trans. Mob. Comput.* **14**(8), 1545–1559 (2015)



# Lightweight Anonymity-Preserving Authentication and Key Agreement Protocol for the Internet of Things Environment

Ahmed Mohammed Ibrahim Alkuhlani<sup>1(✉)</sup> and S. B. Thorat<sup>2</sup>

<sup>1</sup> S.R.T.M University, Nanded-Waghala 431606, India  
Alkohlanyl@gmail.com

<sup>2</sup> I.T.M College, Nanded-Waghala 431602, India  
suryakant\_thorat@yahoo.com

**Abstract.** Internet of things (IoT) creates a world-wide network of interconnected objects or things that will have an active role in the Future Internet (FI). Such things will be readable, recognizable, locatable, addressable, and/or controllable via the Internet; in order for the IoT to expand there should be a trust in the IoT security infrastructure. The number of applications and services expected to be numerous so in order to access these applications and services a secure and robust authentication protocol is required. In this paper, we propose a robust and lightweight mutual authentication and key agreement protocol for the IoT environment. We have used lightweight computational cryptographic functions to maintain low computational, memory and energy consumption. The security analysis and performance evaluation prove the protocol is lightweight and resist most of known security related attacks. Moreover, formal security verification was conducted using AVISPA tool. The result shows that the proposed protocol is secure and safe.

**Keywords:** IoT · IoT authentication · Biometric-based authentication  
Constraints network · Lightweight authentication

## 1 Introduction

Nowadays, more IoT applications have been implemented, such as smart home systems [1], healthcare systems, connected cars, surveillance devices, environmental monitoring, and smart wearable devices [2–4]. Huge amounts of sensitive and personal information are exchanged.

It is very important to define how the IoT things could efficiently and securely communicate and exchange information among themselves and with remote servers. Security and privacy are a key challenge to IoT [5].

Things in IoT have limited computational capability, limited energy, and small memory. They communicate using low rate and low power wireless technologies such as IEEE 802.15.4 BLE ZigBee etc. [6, 7] meanwhile; existing traditional security techniques require a considerable amount of energy for processing. Therefore, we require efficient and robust security mechanisms that provide a similar level of security of the existing traditional techniques with the limited resources of the IoT devices. In IoT, we require authentication and key agreement techniques that allow two remote

entities to mutually authenticate and negotiate secret keys that are used to protect the sensor data against various types of active and passive attacks [8].

Therefore, in this paper we proposed a secure and lightweight mutual authentication and key agreement protocol for IoT environment. We have used lightweight computational cryptographic functions such as hash function and XOR operator which is suitable to use on constrained platforms such as IoT and wireless sensor network (WSN) [9]. Also deep Security analysis and performance evaluation are conducted to prove the protocol is lightweight and robust.

The rest of the paper is organized into six sections; Sect. 2 presents a literature review of related schemes. In Sect. 3 preliminaries related to IoT authentication are discussed. In Sect. 4 we present our proposed protocol. In Sect. 5 we provide security and performance analysis. In Sect. 6 the formal security analysis using AVISPA software is conducted, and the paper is wrapped up with the conclusion in Sect. 7.

## 2 Literature Review

In 2012 Das et al. [10] proposed a new authentication scheme for hierarchical WSNs that support the feature of dynamic node. At the same year Liu et al.'s [11] proposed user authentication and access control scheme for IoT. The scheme uses RBAC access control. In 2013, Turkanović and Hölbl [12] and Xue et al. [13] claimed that the protocol of [10] is impractical, and proposed enhanced protocols to overcome its drawbacks. Li et al. [14] proved that the scheme of Xue et al. is prone to problems such as stolen-verifier attack, off-line password guessing attack. In 2014, Turkanović et al. [15] proposed a lightweight authentication protocol for heterogeneous WSN based on the notion of IOT. The scheme proved to be computationally lightweight and consumes less memory and energy. At the same year, Ndibanje et al. [16] found some security weakness in [11] scheme; therefore, they propose an enhanced protocol that offers user anonymity and mutual authentication. In 2015, He et al. [17] showed that the Xue et al. protocol is susceptible to off-line password guessing attacks, and user and sensor node impersonation attacks. Amin and Biswas [18] claimed that the scheme of Turkanović is not efficient in terms of energy consumption, and proposed a user authentication and key agreement scheme in multi-gateway based on WSN. In 2016, Farash et al. [19] found some security weaknesses in Turkanović et al. [15] such as off-line password guessing attacks, and man-in-the-middle attacks. Then, they proposed an enhanced user authentication and key agreement scheme for heterogeneous WSN for the IoT concept. In the same year, Amin et al. [20] revealed that the scheme of Farash et al. is insecure and susceptible to stolen-smartcard attacks, off-line password-guessing attacks, user-impersonation attacks, and fails to preserve user-anonymity. Afterwards, Arasteh et al. [21] claimed that the scheme proposed by Amin et al. in [20] has security weaknesses and is prone to Replay attacks and DoS attacks, and proposed an enhanced protocol to overcome these drawbacks. Recently in 2017, Dhillon and Kalra [22] proposed an enhanced three-factor biometric authentication protocol for IoT network based on Turkanović et al. scheme. Jiang et al. [23] proposed Lightweight Three-factor Authentication and Key Agreement Protocol for Internet-integrated WSNs based on the idea of public key primitive Rabin cryptosystem.

### 3 Preliminaries

#### 3.1 One-Way Hash and Bio-Hash Function

Hash function takes arbitrary input data and returns a string with a fixed size, which is referred to as a hash value or (a message digest). One of the important properties of one-way hash function is that it is very sensitive: any small change to the input data results in a totally different output hash value. Biometric is not always a constant value; it may change with time and environment. So, the general one-way hash function is not the proper choice for hashing biometric. To resolve this issue, researchers in [25, 26] have suggested Bio-hash function which proved its accuracy and flexibility with biometrics.

Bio-Hash function refers to a special type of one-way hash function that can be used to hash different types of biometrics such as (Fingerprint, iris, retina, and voice). The input data of biometric may vary a little bit, but the result hash value of Bio-hash function remains the same. In the contrary, if the variation is significant, the output becomes different.

#### 3.2 Network Model

IoT is the concept of connecting smart devices to the global network (the Internet) which allows users to access the IoT services remotely. As depicted in Fig. 1 through an application on the remote user smartphone the user can directly connects to a specific IoT device inside the network (smart home). In order to lower the processing burden for the sensor node, the protocol uses the gateway node as a mediator for the authentication process [24].

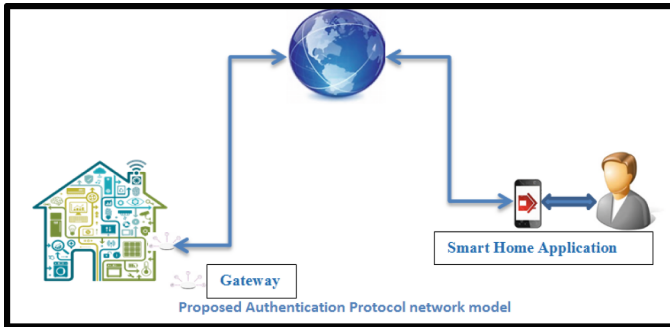


Fig. 1. Network model

### 4 Proposed Authentication and Key Agreement Protocol

In the following sections, we propose an authentication and key agreement protocol for IoT network. Our proposed scheme has a pre-deployment phase (system setup phase), registration phase, login phase, authentication and key agreement phase, and password change/update phase. In Table 1, there is a brief description of notations used within the protocol.

**Table 1.** List of notations used throughout the protocol

Symbol	Description
$U_i$	User
GWN	Gateway
$N_j$	IoT node
$ID_i$	Unique identity for each user $U_i$
$PW_i$	Password of the user $U_i$
$B_i$	Biometric key of $f_{ngi}$ , where $B_i = BK(H(f_{ngi}))$
$f_{ngi}$	Biometric template of user $U_i$
$X_{gui}$	Unique Shared secret key between each $U_i$ and GWN
SP	Smartphone
$K_{sg}$	Shared secret key between $N_j$ and GWN
$X_{gn}$	Master secret key and GWN secret password
$N$	High entropy Nonce generated by GWN to mask its secret Key
$S$	Used to masked $U_i$ identity during communication
$Y$	Used to masked $N_j$ identity during communication
$ID_j$	Unique identity for each node $N_j$
$CR_j$	Password of IoT node $N_j$
$K_i$	Random Nonce generated by $U_i$ to construct the session key
$K_j$	Random Nonce generated by $N_j$ to construct the session key
SK	Session key to encrypt communication between $U_i$ and $N_j$
$T_{s1}, T_{s2}, T_1-T_4$	Timestamp used throughout the Scheme
$\Delta T, T_c$	Time Range of allowed transmission delay, Current time
$h(\cdot), H(\cdot)$	One-way hash function, Bio-hashing function
$\parallel, \oplus$	Concatenate operation, X-OR operation
$g_i, f_i, e_i$	Values used to protect the identity and password of the user

#### 4.1 System Setup Phase

This is the pre-deployment phase in which each embedded device/sensor of IoT network has to be configured with certain parameters prior to authentication. This phase is executed by the system administrator (SA) in offline mode as follows:

- **Step 1.** SA assigns a master secret key ( $X_{gn}$ ) for the gateway (GWN), the master secret key  $X_{gn}$  is known only to SA and the GWN.
- **Step 2.** SA assigns unique identity  $ID_j$  for each IoT node  $N_j$  in the IoT network and also computes the password  $CR_j$   $CR_j = h(ID_j \parallel X_{gn})$ . Therefore, each node will have a unique secret key  $CR_j$ .
- **Step 3.** SA chooses a random secret number  $K_{sg}$  that is shared between the GWN and the  $N_j$ .
- **Step 4.** SA embeds ( $ID_j, CR_j, K_{sg}$ ) into node's tamper-proof memory and ( $X_{gn}, K_{sg}, ID_j$ ) to GWN memory.

### 4.2 Registration Phase

Registration phase is divided into two phases. The first one is for the registration of nodes of IoT network, and the second is for the registration of the outside/remote users.

**IoT Node Registration Phase.** This phase is performed between the  $N_j$  and the GWN. Details of this phase are depicted in Fig. 2b.

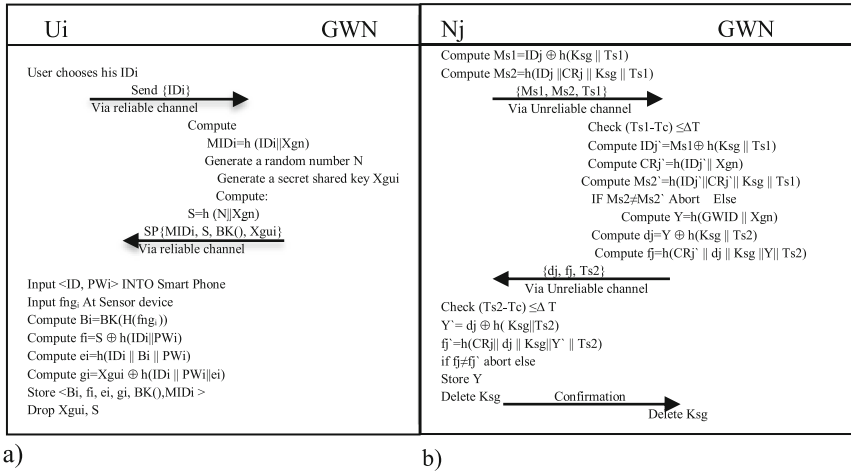


Fig. 2. Registration phase (a) user (b) IoT node

- Step 1.** In order to provide ID anonymity, IOT node  $N_j$  computes  $Ms1 = IDj \oplus h(Ksg||Ts1)$  and for message verification computes  $Ms2 = h(IDj||CRj||Ksg||Ts1)$ . In which T1 is a fresh timestamp and sends to GWN ( $Ms1, Ms2, Ts1$ ) through an unreliable channel.
- Step 2.** Upon the reception of the message from the  $N_j$  the GWN, first, verifies whether or not the time received T is within the allowed time span to avoid replay attack  $(Ts1 - Tc) < \Delta T$ . If it is not within the allowed time, the GWN refuses to accept the  $N_j$ ; otherwise, if the verification holds, GWN computes  $IDj` = Ms1 \oplus h(Ksg||Ts1)$ ,  $CRj` = h(IDj`||Xgn)$ ,  $Ms2` = h(IDj`||CRj`||Ksg||Ts1)$ , and checks whether  $Ms2` \neq Ms2$  then the  $N_j$  is not legitimate and session is aborted. If not, the GWN authenticates the  $N_j$ . The GWN continues and computes  $Y = h(GWID||Xgn)$ ,  $dj = Y \oplus h(Ksg||Ts2)$  and  $fj = h(Y||dj||Ksg||Ts2)$  then the GWN sends to  $N_j\{dj, fj, Ts2\}$  through unreliable channel. When  $N_j$  received the message from the GWN, it first verifies the time for replay attacks if the time T is within the allowed time span. Then it continues with the registration process, or else, it rejects the message. If the verification holds, the  $N_j$  computes  $Y` = dj \oplus h(Ksg||Ts2)$  and very fies if  $fj` = fj$  holds then the GWN is legitimate and then the  $N_j$  stores Y and deletes the shared key Ksg from the device memory.
- Step 3.** In the last step, the  $N_j$  sends a confirmation message to the GWN and deletes the shared key Ksg and IDj from the GWN memory.

**User Registration Phase.** The second phase of registration is done with the user  $U_i$ . At the end of this phase, the user will be authorized and registered with the GWN. Details of this phase are depicted in Fig. 2a.

- **Step 1.**  $U_i$  sends his identity  $ID_i$  to the GWN via a reliable/secure channel. Upon the reception of message sent from the  $U_i$ , the GWN computes masked  $ID_i$  with the GWN master secret key  $X_{gn}$ ,  $MID_i = h(ID_i||X_{gn})$ ; then, the GWN generates a secret random key  $X_{gui}$  that will be shared between the  $U_i$  and the GWN for further secure communication.

GWN also generates a random number  $N$  with high entropy, then computes  $S = h(X_{gn}||N)$  and customizes the user's smartphone (SP) with  $\{X_{gui}, BK(), S, MID_i\}$  where  $BK()$  refers to the biometric key generation and extraction function.

- **Step 2.** Upon the reception of the message sent from the GWN, the  $U_i$  inputs his  $ID_i$  and credentials password  $PW_i$  and fingerprint  $fng_i$  using smartphone sensor device. Using the  $BK()$ , the user computes  $B_i = BK(H(fng_i))$ , then computes  $fi = S \oplus h(ID_i||PW_i)$ ,  $ei = h(ID_i||B_i||PW_i)$ , and  $gi = X_{gui} \oplus h(ID_i||PW_i||ei)$ .
- **Step 3.** Finally user stores  $\{B_i, MID_i, fi, ei, gi, BK()\}$  in the SP and deletes  $X_{gui}$  and  $S$  from the SP memory. Note that  $X_{gui}$  is the secret key shared between the  $U_i$  and the GWN, and the value  $CR_i$  needs it to be computed at the login phase  $CR_i = h(PW_i||X_{gui})$ . Hence, to be safe from smartphone breach/stolen attacks and offline password guessing attacks,  $X_{gui}$  is deleted from the SP and will be recomputed at login phase. Furthermore, the value  $S$  is used to preserve the identity anonymity of the  $U_i$  when the message is exchanged in the authentication phase.

### 4.3 Login Phase

This Phase is done between the  $U_i$  and the  $N_j$ . After the registration phase is completed, the user logs in to initiate a request to access the desired device in the IoT network. Our proposed protocol uses the user fingerprint, username and password for login. A detailed description of this phase is as follows:

- **Step 1.**  $U_i$  opens the IoT application (smart home App) on his smartphone (SP) then inputs his fingerprint ( $fng_i$ ) on the smartphone device sensor to compute  $B_i' = BK(H(fng_i))$ , then compares the calculated  $B_i'$  with the stored  $B_i$  if ( $B_i' \neq B_i$ ). Then the user is rejected. Otherwise, the user is asked to enter his identity  $ID_i$  and the password  $PW_i$ . Afterwards,  $U_i$  Computes  $ei' = h(ID_i||PW_i||B_i)$  and checks whether ( $ei' \neq ei$ ). If so, the session is aborted as the user is not a legitimate user. Once the user is proved to be legitimate and his  $fng_i$ ,  $ID_i$  and  $PW_i$  are correct, user proceeds to step 2.
- **Step 2.**  $U_i$  Computes  $X_{gui}' = gi \oplus h(ID_i||PW_i||ei')$ ,  $CR_i = h(PW_i||X_{gui}')$  and  $S = fi \oplus h(ID_i||PW_i)$ , the  $U_i$  generates a random nonce  $K_i$  which is the user part of the session key to be used to encrypt the data. Also generates a fresh timestamp  $T1$  to be used to avoid a reply attack. After generating  $K_i$  and  $T1$ , the user starts to prepare the authentication messages that are to be sent to the IoT node  $N_j$  that  $U_i$  wants to access. To provide identity anonymity and avoid user traceability attack for the  $U_i$ 's  $ID_i$ , the  $U_i$  computes  $M1 = ID_i \oplus h(S||T1)$ . The identity of the user  $U_i$  is kept secret.  $S$  is a highly secure value; it is a combination of the GWN master secret

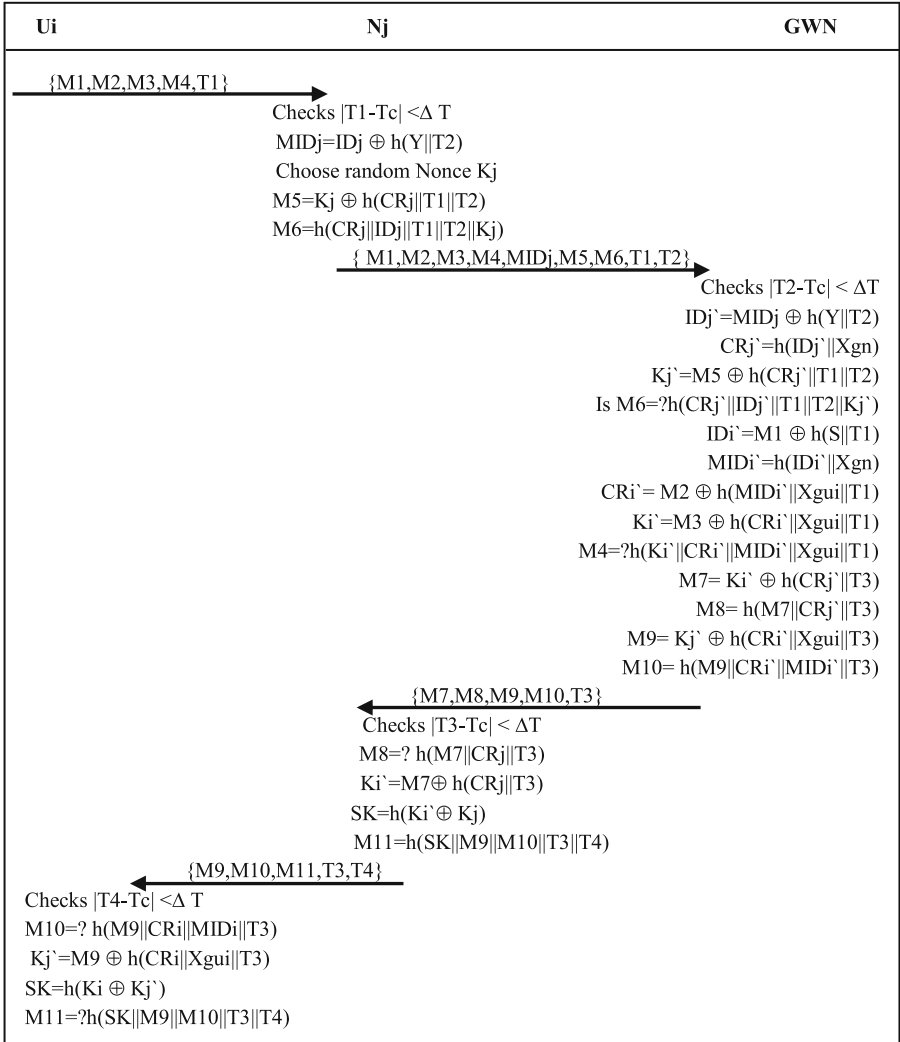
key  $X_{gn}$  and a high entropy random number  $N$  which makes it difficult for an attacker to break. In  $M2 = CR_i \oplus h(MID_i || X_{gui} || T1)$ , the user  $CR_i$  is safely protected from man in the middle attack and replay attack by using the shared password  $X_{gui}$  and the fresh time  $T1$ . Note that these messages are sent through an unreliable channel and the one-way hash function  $h$  maintains the integrity of these messages, and any tiny change to the hash value is discovered. The third message  $M3 = K_i \oplus h(CR_i || X_{gui} || T1)$  carries the  $U_i$  part of the session key  $K_i$ , and eventually  $M4 = h(K_i || CR_i || MID_i || X_{gui} || T1)$  verifies that the previously sent values  $M1, M2, M3$  are not changed, modified, or deleted by any attacker.

- **Step 3.**  $U_i$  chooses the IoT node  $N_j$  he wishes to access and send  $\{M1, M2, M3, M4, T1\}$  to it via an unreliable channel.

#### 4.4 Authentication Phase

After the deployment of the IoT network and registration of both users and IoT nodes, the user logs in and chooses the desired node he wants to access. The authentication phase comes to mutually authenticate a user with chosen node and the gateway. Moreover, manages a secure key agreement by securely exchanging key parts of the session key between the  $U_i$  and  $N_j$ . authentication phase is completed in 4 messages handshakes, a user who wants to access data from IoT network can directly access a specific IoT device without the need to access the gateway first. The gateway works as an authenticator for both the IoT node and user. Details of this phase is depicted in Fig. 3. Authentication steps are as follows:

- **Step 1.** Upon the reception of the login message  $\{M1, M2, M3, M4, T1\}$  from  $U_i$ ,  $N_j$  verifies the time  $|T1 - T_c| < \Delta T$ . If  $T$  is within the allowed time span, then  $N_j$  proceeds with the authentication. Otherwise, the user is considered illegitimate and the session is aborted.
- **Step 2.** After the verification of the freshness of  $T1$  passes,  $N_j$  computes  $MID_j = ID_j \oplus h(Y || T2)$ . The identity of the node  $ID_j$  is masked with the value  $Y = h(GWID || X_{gn})$ , and fresh time stamp  $T2$  to avoid any replay attack and to preserve the identity anonymity of  $N_j$ . Then Next  $N_j$  generates a random nonce  $K_j$  which is the  $N_j$  part of the session key to be used to encrypt the data in further communication with the  $U_i$ .
- **Step 3.**  $N_j$  continues to prepare the necessary values for authentication, and computes  $M5 = K_j \oplus h(CR_j || T1 || T2)$ , and the verification message  $M6 = h(CR_j || ID_j || T1 || T2 || K_j)$ .
- **Step 4.** As the node  $N_j$  is a constrained device, it delegates the authentication of the  $U_i$  to the GWN by sending the message received from  $U_i \{M1, M2, M3, M4, T1\}$ , along with its own message  $\{MID_j, M5, M6, T2\}$ .
- **Step 5.** After receiving the message sent from the  $N_j$ , the GWN checks the time freshness of the received messages  $|T2 - T_c| < \Delta T$ . If the time difference between the sent time  $T2$  and the current time of the GWN  $T_c$  is within the allowed time span, the GWN continues with the Authentication of the  $N_j$ , or else it aborts the session and sends a rejection message to the  $N_j$ .



**Fig. 3.** Authentication and key agreement phase of the proposed protocol

- **Step 6.** After the time verification passes, the GWN first checks the legitimacy of  $N_j$ . The GWN computes the  $ID_j' = MID_j \oplus h(Y || T2)$  using the secret value which was previously stored by the GWN in the  $N_j$  memory. It should be noted that only GWN can compute the value of  $Y$  user, as it is the only part that has the hashed value of  $Y$ . Using the newly computed  $ID_j'$ , the GWN computes  $CR_j' = h(ID_j' || X_{gn})$ .
- **Step 7.** Using the newly computed  $ID_j'$  and  $CR_j'$ , the GWN extracts the  $N_j$  session key part by computing  $K_j' = M5 \oplus h(CR_j' || T1 || T2)$ .  $T1$  and  $T2$  are used to avoid the replay attack.



- **Step 8.** The GWN checks if the received value  $M6 = h(CR_j || ID_j || T1 || T2 || K_j)$  is equal to the GWN version of  $M6 = h(CR_j || ID_j || T1 || T2 || K_j)$ , if so, then the  $N_j$  is authenticated and considered legitimate. Therefore, GWN proceeds to check the authenticity of  $U_i$ ; otherwise, GWN rejects  $N_j$  and aborts any further transaction.
- **Step 9.** After the GWN verifies the legitimacy of the  $N_j$ , it has to check the authenticity of the  $U_i$ . The GWN extracts the identity of the  $U_i$  by computing  $ID_i = M1 \oplus h(S || T1)$ . The identity of the user  $U_i$  is kept secret to maintain the ID anonymity and avoid user traceability attacks.  $S$  is a highly secure value; it is a combination of the GWN master secret key and the high entropy random number  $N$ . Using the newly computed  $ID_i$ , the GWN computes the masked identity of the  $U_i$   $MID_i = h(ID_i || X_{gn})$  using the GWN secret key  $X_{gn}$ . It should be noted that  $S$  can be computed only by the GWN and stored in a hash format in the  $U_i$ 's Smartphone memory during registration.
- **Step 10.** Using the newly computed  $MID_i$  and the GWN- $U_i$  shared password  $X_{gui}$ , the GWN extract  $CR_i = M2 \oplus h(MID_i || X_{gui} || T1)$ . Then using newly computed  $CR_i$  and the shared password  $X_{gui}$ , GWN extracts the  $U_i$  session key part  $K_i = M3 \oplus h(CR_i || X_{gui} || T1)$ .
- **Step 11.** The GWN checks if its version of  $M4 = h(K_i || CR_i || MID_i || X_{gui} || T1)$  is equal to the  $M4$  sent from  $U_i$ . If so, the user  $U_i$  is legitimate; if not, GWN declines the  $U_i$  and sends a message to  $N_j$  stating that  $U_i$  is not a legitimate user, then session is aborted.
- **Step 12.** After GWN verifies the authenticity of both  $U_i$  and  $N_j$  and extracts their session key parts  $K_i$  and  $K_j$ , GWN prepares the messages  $\{M7, M8, M9, M10\}$  and sends to the  $N_j$ , then to the  $U_i$  so that both the  $U_i$  and the  $N_j$  mutually authenticate with the GWN. Therefore, the  $N_j$  and the  $U_i$  can compute the session key (SK) and start encrypting their communication.
- **Step 13.** GWN computes  $M7 = K_i \oplus h(CR_j || T3)$ ,  $M8 = h(M7 || CR_j || T3)$ ,  $M9 = K_j \oplus h(CR_i || X_{gui} || T3)$ ,  $M10 = h(M9 || CR_i || MID_i || T3)$ ,  $M7$  and  $M8$  are used by the  $N_j$ ,  $M7$  is used to mask the user part of the session key  $K_i$ , and  $M8$  to ensure the legitimacy of the GWN. The same applies to  $M9$  and  $M10$ . They are used by the user  $U_i$  in which  $M9$  is used to mask the  $N_j$  part of the session key  $K_j$ , and  $M10$  to ensure the legitimacy of the GWN. The message  $\{M7, M8, M9, M10, T3\}$  is sent to  $N_j$ .
- **Step 14.** Upon the receipt of the message sent from the GWN, the  $N_j$  checks the time  $|T3 - Tc| < \Delta T$ . If  $T$  is within the allowed time span,  $N_j$  proceeds with the authentication; if not, the GWN is considered illegitimate and the session is aborted.
- **Step 15.** After the time verification passes, using the stored value of  $CR_j = h(ID_j || X_{gn})$  and the lately received  $M7$ ,  $N_j$  verifies if the received value of  $M8 = h(M7 || CR_j || T3)$ . If the verification holds, then GWN is legitimate, and thus  $N_j$  and GWN are mutually authenticated; otherwise, the message is intercepted and changed by an attacker, and the session is aborted, and a rejection message is sent to the GWN.
- **Step 16.** If the verification of the legitimacy of the GWN holds,  $N_j$  computes  $K_i = M7 \oplus h(CR_j || T3)$  to extract the  $U_i$  session key part  $K_i$ , and then construct the session key (SK) using its own session key part  $K_j$  and the newly computed  $K_i$ .

- **Step 17.** The  $N_j$  computes the session key  $SK = h(K_i \oplus K_j)$  and  $M11 = h(SK || M9 || M10 || T3 || T4)$ , and sends  $\{M9, M10, M11, T3, T4\}$  to  $U_i$ .  $M10$  is used by the  $U_i$  to verify the legitimacy of the GWN, and  $M11$  to verify the legitimacy of the  $N_j$ .
- **Step 18.** Upon the receipt of the message sent from the  $N_j$ , the  $U_i$  checks the time  $|T4 - Tc| < \Delta T$ . If  $T$  is within the allowed time span, the  $U_i$  proceeds with the authentication. Otherwise, the  $N_j$  is considered illegitimate and session is aborted.
- **Step 19.** If the time verification holds, then using the values  $CR_i$  and  $MID_i$  the  $U_i$  checks whether the received  $M10 = h(M9 || CR_i || MID_i || T3)$ . If correct, then the GWN is legitimate; if not, the GWN is impersonated and session is aborted.
- **Step 20.**  $U_i$  extracts the session key part of the  $N_j$  using its secret values  $CR_i$  and  $X_{gui} K_j = M9 \oplus h(CR_i || X_{gui} || T3)$  and using its stored session key  $K_i$  and newly computed  $K_j$   $U_i$  constructs its version of the session key  $SK = h(K_i \oplus K_j)$ .
- **Step 21.** Finally the  $U_i$  check if the received  $M11 = h(SK || M9 || M10 || T3 || T4)$  then, the  $N_j$  is legitimate. So,  $U_i$  authenticates  $N_j$  and GWN and starts using  $SK$  for further messages encryption between the user  $U_i$  and the IoT node  $N_j$ . Otherwise, the  $U_i$  rejects the  $N_j$  and considers it a malicious attacker.

#### 4.5 Password Change/Update Phase

For reliability and security purposes, the facility of changing/updating the password should be considered when designing any authentication protocol in the case of IoT and constrained networks. It is preferred to keep messages exchanged and communication at minimum so, this phase is executed locally at the user side without interfering with SA or GWN.

- **Step 1.** The user opens the smart home application on his SP and using the password change form. He is asked to inputs his fingerprint on the SP's sensor device then verifies his fingerprint. If the verification passes, the user then is asked to enter his IDi and Password PWi and verifies if stored  $e_i = h(ID_i || PW_i || B_i)$ . If verification holds, go to step 2.
- **Step 2.** The user is asked to enter his new password PWinew, in order to extract the values  $S$  and  $X_{gui}$  SP compute  $S = f_i \oplus h(ID_i || PW_i)$  and  $X_{gui} = g_i \oplus h(ID_i || PW_i || B_i)$ . Then  $U_i$  computes  $e_{inew} = h((ID_i || PW_{inew} || B_i))$   $f_{inew} = S \oplus h(ID_i || PW_{inew})$ ,  $g_{inew} = X_{gui} \oplus h(ID_i || PW_{inew} || B_i)$ .
- **Step 3.** Replace the old values of  $e_i$ ,  $f_i$ ,  $g_i$ , with the new values  $e_{inew}$ ,  $f_{inew}$ ,  $g_{inew}$ .

## 5 Security Analysis and Performance Evaluation of the Proposed Protocol

In this section, we illustrate the security features and detailed security evaluation of the proposed protocol. The evaluation is conducted by two different methods. The first one proves the high security of the protocol through theoretical analysis and a comparison with some other related protocols. The second method of the evaluation conducted a formal security analysis using AVISPA simulation software.

### 5.1 Security Analysis of the Proposed Protocol

Security features and comparison with the related protocol is presented in Table 2.

**Table 2.** Security features comparison with other protocols

Security feature	Farash [19]	Yeh [30]	Amin [20]	Proposed scheme
Mutual authentication	Yes	Yes	Yes	<b>Yes</b>
Key agreement	Yes	Yes	Yes	<b>Yes</b>
Password protection	No	Yes	Yes	<b>Yes</b>
Password-change	Yes	Yes	Yes	<b>Yes</b>
Dynamic node addition	Yes	No	Yes	<b>Yes</b>
User anonymity	No	No	No	<b>Yes</b>
Node anonymity	Yes	No	Yes	<b>Yes</b>
Stolen SP&SC breach attack resilience	Yes	No	Yes	<b>Yes</b>
Traceability attack resilience	No	Yes	No	<b>Yes</b>
Replay attack resilience	Yes	No	No	<b>Yes</b>
Privileged-insider attack resilience	No	Yes	Yes	<b>Yes</b>
Stolen verifier attack resilience	No	Yes	yes	<b>Yes</b>
Impersonation attack resilience	Yes	No	Yes	<b>Yes</b>
Many logged-in with same id attack resilience	Yes	–	Yes	<b>Yes</b>
Password change attack resilience	Yes	–	Yes	<b>Yes</b>

**Mutual Authentication.** In the proposed protocol the  $U_i$ , the  $N_j$  and the GWN all of them authenticate each other. The GWN authenticates  $U_i$  and  $N_j$  by computing  $M_4$  and  $M_6$  respectively. In contrary  $N_j$  and the  $U_i$  both authenticate the GWN by computing  $M_8$  and  $M_{10}$  respectively and finally  $U_i$  receives  $M_{11}$  and authenticate the  $N_j$  and GWN.

**Key Agreement.** The  $U_i$  and the  $N_j$  contribute individually to produce a secure session key, in login phase, the  $U_i$  generate a nonce  $K_i$  and computes  $M_3 = K_i \oplus h(CR_i || X_{gui} || T_1)$ ,  $K_i$  is securely protected by the shared password  $X_{gui}$  and the one-way hash function. The IoT node  $N_j$  also generates a nonce  $K_j$ , its part of the session key and computes  $M_5 = K_j \oplus h(CR_j || T_1 || T_2)$   $K_j$  is securely protected by the password  $CR_j = h(ID_j || X_{gn})$  and the one-way hash function. Both the  $U_i$  and the  $N_j$  successfully compute  $SK = h(K_i || K_j)$ .

**User Anonymity.** User anonymity means hiding the identity of the communicating parties during the authentication and key agreement process. The proposed scheme never transmits the identity of the user  $ID_i$  without protection, and never saves inside the smartphone unmasked.

When  $U_i$  sends a message  $\{M_1, M_2, M_3, M_4, T_1\}$  to the  $N_j$ ,  $M_1 = ID_i \oplus h(S || T_1)$ . The identity of the user  $ID_i$  is protected with one way hash function  $h(S || T)$  where  $S = h(X_{gn} || N)$  and  $T_1$  is the fresh time sent by  $U_i$ ,  $X_{gn}$  is GWN secret key which is

known only by GWN, and  $N$  is a high entropy random number generated by the GWN to mask its secret key. The combination of both values with one-way hash function keeps them secure, and also keeps the identity of the  $U_i$  secure. On other messages, the identity is masked and sent only inside one way hash function  $h(\text{MID}_i || \text{X}_{\text{gui}} || \text{T}_1)$ ,  $h(\text{K}_i || \text{CR}_i || \text{MID}_i || \text{X}_{\text{gui}} || \text{T}_1)$  which makes it infeasible to retrieve  $U_i$  identity by any attacker.  $\text{ID}_i$  is sent unmasked only one time during the registration through a secure channel.

**Security Against Smartphone Stolen/Breach Attack.** According to [27] a good hacker might use some power analyzing techniques to get the data inside the smart device. The proposed protocol is resistant to such attacks as we are going to explain.

*Password Off-Line Guessing Attack.* In the proposed protocol each value has the password ( $e_i$ ,  $f_i$ ,  $g_i$ ) is combined with other values and hashed by one way hash function making it hard to break or get the password. The values  $S = h(\text{X}_{\text{gn}} || N)$  which are sent from the GWN to the  $U_i$  during registration is combined with two values;  $\text{X}_{\text{gn}}$  which is the secret key of the GWN, known only by him, and  $N$  which is a highly entropy random number known only by the GWN. After computing  $f_i$  and  $g_i$ , both variables  $S$  and the secret shared key  $\text{X}_{\text{gui}}$  (which is known by the GWN and the  $U_i$ ) will be deleted from the smartphone.

*Identity Off-Line Guessing Attack.* The Identity of the user is securely stored inside the smartphone, and each value has  $\text{ID}_i(\text{MID}_i, e_i, f_i, g_i)$  is secure with one way hash function. So, to get  $\text{ID}_i$  we need to know  $\text{PW}_i$ ,  $S$ ,  $\text{X}_{\text{gui}}$ ,  $\text{X}_{\text{gn}}$  and  $\text{B}_i$ .

**User/Node Impersonate Attack.** Impersonating a legitimate user/node happens when an attacker uses the private values of a legitimate user/node such as identity or password or intercepts and forges a message sent from the  $U_i/N_j$  to other participants.  $\text{ID}_i$  and  $\text{PW}_i$  are secured as we mentioned in phone/card breach attacks. When  $U_i$  sends the login message to  $N_j$   $\{\text{M}_1, \text{M}_2, \text{M}_3, \text{M}_4, \text{T}_1\}$  the attacker needs to have  $\text{ID}_i$ ,  $\text{CR}_i$ ,  $S$ ,  $\text{X}_{\text{gui}}$ , and  $\text{K}_i$  to compute  $(\text{M}_1 - \text{M}_4)$ . Each message in the login is hashed using different secret keys. Therefore, to calculate  $\text{M}_1$  the attacker needs to know  $\text{ID}_i$  and  $S$  which both are known only by the  $U_i$  and the GWN. In  $\text{M}_2$  also, the attacker needs to know the shared secret key  $\text{X}_{\text{gui}}$  and  $\text{MID}_i$  which are known only by the  $U_i$  and GWN. The same applies to  $\text{M}_3$  in which the attacker needs to know  $\text{K}_i$ ,  $\text{CR}_i$ , and  $\text{X}_{\text{gui}}$  which are all kept secret from attackers also when the node  $N_j$  sent  $\text{MID}_j$ ,  $\text{M}_5$ ,  $\text{M}_6$ ,  $\text{T}_2$  to the GWN the attacker doesn't know  $\text{ID}_j$ ,  $\text{CR}_j$ ,  $Y$  and  $\text{K}_j$  and is computationally infeasible to compute way hash function.

**User Traceability.** The attacker can trace user  $U_i$  when sending a login message. The attacker compares two different login messages and finds constant values, and hence can differentiate between users. In the proposed protocol, the user sends  $\text{M}_1 = \text{ID}_i \oplus h(S || \text{T}_1)$  where the user  $\text{ID}_i$  is hidden and also  $\text{M}_1$  value is dynamically changed because of the time  $\text{T}_1$  which is different in every login.

**Node Traceability.** The same with the  $N_j$ , when sending the masked identity  $\text{MID}_j = \text{ID}_j \oplus h(Y || \text{T}_2)$ , the value of masked identity of the node is changeable in every login by the timestamp  $\text{T}_2$  so, The proposed protocol is safe against tractability attacks.

**Privileged Insider and Stolen-Verifier Attacks.** In the proposed scheme, GWN does not store user password  $PW_i$  in any tables. It attaches its master secret key  $X_{gn}$  and the shared secret password  $X_{gui}$  to  $U_i$  verifiers ( $ID_i$  and  $PW_i$ ) during the registration phase. Accordingly, a malicious privileged user can't get any user sensitive information. Therefore, an attacker cannot impersonate the user. Furthermore, when the  $U_i$  initiates the authentication phase, the  $N_j$  forwards the hashed message to the GWN, whereby a privileged user cannot extract  $U_i$ 's password. The one-way property of the hash function prevents any attacker from getting any information. Consequently, the proposed scheme is resilient against both privileged Insider and Stolen-Verifier Attack.

**Other Type of Attacks.** *Many logged-in users with the same login-id attack, Password Change Attack and Replay attack:* Our proposed scheme uses a smartphone for a user's login or to Password Change. An attacker needs a legitimate smartphone to login or to change the password and also the user's fingerprint and password to successfully execute the login and change password phase. Timestamps are used in every message exchanged in login and authentication phase to prevent the replay attack. Therefore the proposed scheme is resilient against these attacks.

## 5.2 Performance Evaluation of the Proposed Protocol

**Computational Cost of the Proposed Protocol.** Computational cost varies from one scheme to another depending on the number of security features, number of attacks the scheme resists, and the type of cryptographic security primitives that the scheme uses. The proposed scheme uses the most lightweight cryptographic security primitives that are XOR and Hash; and thus provides a robust security against most of the well-known attacks. The security features comparison between our scheme and others authentication schemes is summarized in Table 2. In addition, the computational cost comparison of our scheme and others related schemes are summarized in Table 3.

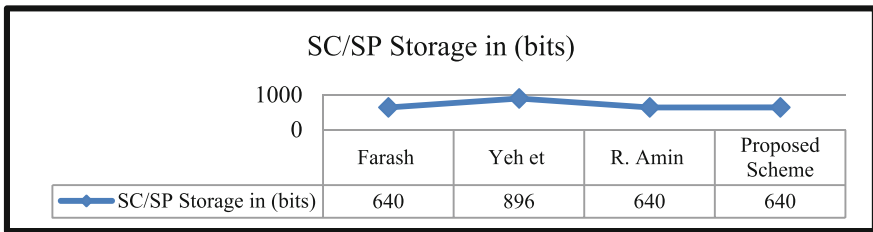
**Table 3.** Computational cost of the proposed protocol with other related protocols

Protocol	User	IoT sensor	Gateway	Total computational cost
Farash [19]	11 $T_h$	7 $T_h$	14 $T_h$	32 $T_h$
Yeh [30]	1 $T_h$ + 2 $T_{(d/e)}$	3 $T_h$ + 2 $T_{(d/e)}$	4 $T_h$ + 4 $T_{(d/e)}$	8 $T_h$ + 8 $T_{(d/e)}$
Amin [20]	13 $T_h$	5 $T_h$	16 $T_h$	34 $T_h$
Proposed scheme	<b>13 <math>T_h</math></b>	<b>7 <math>T_h</math></b>	<b>13 <math>T_h</math></b>	<b>33 <math>T_h</math></b>

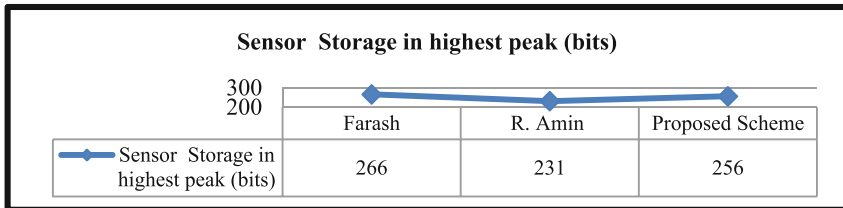
The proposed protocol uses a total number of 33 hashes. Although the protocol of Farash used 1 hash operation less than our protocol but we have solved the security drawbacks in Farash protocol as it fails to preserve user-anonymity, stolen-smartcard attacks, off-line password-guessing attack and user-impersonation attack. Our protocol also uses biometric for user login. Therefore for the extra security features that our protocol provides this difference can be neglected.

The author in [28] conducted an experiment to measure the energy cost on a sensor (i.e. CrossBow's MICA2) on an average message size of 24 bytes when hashed using SHA1 and for encryption/decryption using AES. The result was  $\approx 0.075$  J(Ws) and 0.241 J(Ws) for SHA1 and AES encryption/decryption respectively. Our scheme uses 7 hashes. Accordingly, the total energy cost consumed by the sensor is 0.525 J for each authentication cycle.

**Storage Cost of the Proposed Protocol.** Storage cost analysis is made for sensor and smartphone memory most of the protocols shown in Fig. 4 present the same storage cost for the smartphone memory. For sensor storage cost we have taken the measurements when the sensor has the maximum number of bits (moment of peak) it shows that the sensor in proposed protocol holds 256 bits as shown in Fig. 5 where its way far of typical sensor storage which is 128,000 bits.



**Fig. 4.** Smartphone storage cost of the proposed protocol and other related protocols



**Fig. 5.** Sensor storage cost of the proposed protocol and other related protocols

**Communication Cost of the Proposed Protocol.** In the proposed protocol four messages are exchanged between the  $U_i$ , the  $N_j$  and the GWN. In the first, third and fourth messages the packet size is 99 bytes and 98 bytes respectively. Their size is below the standard packet size (i.e. 127) and for that can be carried out without extra processing except for the second message that is sent from the  $N_j$  to the GWN as it carries both messages that come from the  $U_i$  and from the  $N_j$  as our protocol uses the direct approach where the user directly contacts the IoT device, not the gateway. The total number of bytes is 178 which can be handled by 6LoWPan (IPv6 over Low power Wireless Personal Area Networks) layer. The idea behind the design of 6LoWPan layer was for such situation where the packet size is more than 127 bytes of the regular



```

SUMMARY
SAFE

DETAILS
BOUNDED_NUMBER_OF_SESSIONS
TYPED_MODEL

PROTOCOL
/home/span/span/testsuite/results/autgus.if

GOAL
As Specified

BACKEND
CL-AtSe

STATISTICS

Analysed : 0 states
Reachable : 0 states
Translation: 0.37 seconds

```

Fig. 7. AVISPA output result of the proposed protocol

## 7 Conclusion

This paper proposed a remote biometric mutual authentication and key agreement protocol for the IoT environment. The user contacts the IoT node directly without contacting the gateway at first. It is best for a scenario where data has to be retrieved on-demand directly from the IoT node. We have conducted a deep security analysis for possible security attacks also we have implemented the protocol using AVISPA tool to make sure of its robustness and security. In addition, we have also done a performance evaluation of the protocol to prove its efficiency for the IoT environment.

The result shows that the proposed protocol resists to most known security attacks and lightweight in term of computation, memory, and communication costs which is suitable for the IoT environment.

## References

1. Bonino, D., et al.: ALMANAC: Internet of Things for smart cities. In: 2015 3rd International Conference on Future Internet of Things and Cloud (FiCloud), pp. 309–316, August 2015
2. Alkuhlani, A., et al.: Internet of Things (IOT) standards, protocols and security issues. <https://doi.org/10.17148/IJARCC.2015.411109>
3. Zhao, K., Ge, L.: A survey on the Internet of Things security. In: 2013 Ninth International Conference on Computational Intelligence and Security. IEEE (2013). <https://doi.org/10.1109/CIS.2013.145>
4. Miorandi, D., et al.: Internet of Things: vision, applications and research challenges. *Ad Hoc Netw.* **10**(7), 1497–1516 (2012)
5. Jing, Q., et al.: Security of the Internet of Things: perspectives and challenges. *Wirel. Netw.* **20**(8), 2481–2501 (2014)
6. Chatzigiannakis, I., et al.: True self-configuration for the IoT. In: 2012 3rd International Conference on the Internet of Things (IOT). IEEE (2012)
7. Sethi, P., Sarangi, S.R.: Internet of things: architectures, protocols, and applications. *J. Electr. Comput. Eng.* **2017** (2017). <https://doi.org/10.1155/2017/9324035>



8. Saied, Y.B., et al.: Lightweight collaborative key establishment scheme for the Internet of Things. *Comput. Netw.* **64**, 273–295 (2014)
9. Kalirai, J., Kumar, I.: Lightweight cryptography by simplification of hardware – a comparison study. In: *RFID Systems EE260*, Spring, 18 May 2015
10. Das, A.K., Sharma, P., Chatterjee, S., Sing, J.K.: A dynamic password-based user authentication scheme for hierarchical wireless sensor networks. *J. Netw. Comput. Appl.* **35** (52), 1646–1656 (2012)
11. Liu, J., Xiao, Y., Chen, C.P.: Authentication and access control in the Internet of Things. In: *2012 32nd International Conference on Distributed Computing Systems Workshops*. IEEE (2012). <https://doi.org/10.1109/ICDCSW.2012.23>
12. Turkanović, M., Hölbl, M.: An improved dynamic password-based user authentication scheme for hierarchical wireless sensor networks. *Elektronika Ir Elektrotehnika* **19**(6), 109–116 (2013)
13. Xue, K., Ma, C., Hong, P., Ding, R.: A temporal-credential-based mutual authentication and key agreement scheme for wireless sensor networks. *J. Netw. Comput. Appl.* **36**(1), 316–323 (2013)
14. Li, C.-T., Weng, C.-Y., Lee, C.-C.: An advanced temporal credential-based security scheme with mutual authentication and key agreement for wireless sensor networks. *Sensors* **13**(8), 9589–9603 (2013)
15. Brumen, B., Turkanović, M., Hölbl, M.: A novel user authentication and key agreement scheme for heterogeneous ad hoc wireless sensor networks, based on the Internet of Things notion. *Ad Hoc Netw.* **20**, 96–112 (2014)
16. Ndibanje, B., Lee, H.J., Lee, S.G.: Security analysis and improvements of authentication and access control in the Internet of Things. *Sensors* **14**(8), 14786–14805 (2014). <https://doi.org/10.3390/s140814786>
17. He, D., Kumar, N., Chilamkurti, N.: A secure temporal-credential-based mutual authentication and key agreement scheme with pseudo identity for wireless sensor networks. *Inf. Sci.* **321**, 263–277 (2015)
18. Amin, R., Biswas, G.P.: A secure light weight scheme for user authentication and key agreement in multi-gateway based wireless sensor networks. *Ad Hoc Netw.* (2015). <https://doi.org/10.1016/j.adhoc.2015.05.020>
19. Farash, M.S., Turkanović, M., Kumari, S., Hölbl, M.: An efficient user authentication and key agreement scheme for heterogeneous wireless sensor network tailored for the Internet of Things environment. *Ad Hoc Netw.* **36**, 152–176 (2016)
20. Amin, R., et al.: Design of an anonymity-preserving three-factor authenticated key exchange protocol for wireless sensor networks. *Comput. Netw.* (2016). <https://doi.org/10.1016/j.comnet.2016.01.006>
21. Arasteh, S., et al.: A new lightweight authentication and key agreement protocol for Internet of Things. In: *13th International ISC Conference on Information Security and Cryptology, ISCISC 2016, 7–8 September 2016, Shahid Beheshti University, Tehran, Iran* (2016)
22. Dhillon, P.K., Kalra, S.: A lightweight biometrics based remote user authentication scheme for IoT services. *J. Inf. Secur. Appl.* (2017). <https://doi.org/10.1016/j.jisa.2017.01.003>
23. Jiang, Q., et al: Lightweight three-factor authentication and key agreement protocol for internet-integrated wireless sensor networks. *IEEE Access.* <https://doi.org/10.1109/ACCESS.2017.2673239>
24. Chen, T.H., Shih, W.K.: A robust mutual authentication protocol for wireless sensor networks. *ETRI J.* **32**(5), 704–712 (2010)
25. Lumini, R., Nanni, L.: An improved BioHashing for human authentication. *Pattern Recognit.* **40**(3), 1057–1065 (2007)

26. Jin, A.T.B., Ling, D.N.C., Goh, A.: BioHashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognit.* **37**(11), 2245–2255 (2004)
27. Kocher, P., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener, M. (ed.) *CRYPTO 1999*. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999). [https://doi.org/10.1007/3-540-48405-1\\_25](https://doi.org/10.1007/3-540-48405-1_25)
28. Chang, C.-C., Nagel, D.J., Muftic, S.: Assessment of energy consumption in wireless sensor networks: a case study for security algorithms. In: 2007 IEEE International Conference on Mobile Adhoc and Sensor Systems, MASS 2007, pp. 1–6 (2007)
29. Armando, A., et al.: The AVISPA tool for the automated validation of internet security protocols and applications. In: Etessami, K., Rajamani, S.K. (eds.) *CAV 2005*. LNCS, vol. 3576, pp. 281–285. Springer, Heidelberg (2005). [https://doi.org/10.1007/11513988\\_27](https://doi.org/10.1007/11513988_27)
30. Yeh, H.-L., Chen, T.-H., Liu, P.-C., Kim, T.-H., Wei, H.-W.: A secured authentication protocol for wireless sensor networks using elliptic curves cryptography. *Sensors* **11**, 4767–4779 (2011)

# ECC Based Proxy Signature Scheme with Forward Security

Aparna Bannore<sup>1</sup> and Satish Devane<sup>2</sup>

<sup>1</sup> SIES Graduate School of Technology, University of Mumbai, Mumbai, India  
bannore.aparna@siesgst.ac.in

<sup>2</sup> Datta Meghe College of Engineering, University of Mumbai, Mumbai, India

**Abstract.** The Proxy signatures have fascinated many applications which entail delegating signing rights or power of decision making to parties. Applications demanding multilevel authority delegation can be implemented using forward secured proxy systems distinctly from traditional use of forward secrecy in terms of protecting the signed documents from leakage by key exposure attack. In pre-sent decade, focus shift from using traditional desktop environment to portable mobile devices as well secured IOT's using digital signature or proxy signature. Elliptic curve cryptography (ECC) based proxy signature algorithms with forward secrecy is the effective alternative to the traditional digital signature scheme in constrained resources. In this paper, our exclusive focus is on forward secure proxy signature scheme using ECC which deals with proxy key pair generation and forward secure the proxy secret key.

**Keywords:** Proxy signature · Forward secrecy · Elliptic curve cryptography

## 1 Introduction

Digital signatures are omnipresent in the era of Digital revolution. It is significant tool for identity authentication apart from data integrity and non-repudiation. In recent year's variant of digital signature called as proxy signature has attracted lot of attention where the original signer delegates its signing rights to proxy signer.

Proxy signatures allow an entity to delegate its signing capability to a designated person which can sign messages on behalf of the delegator. The original signer is called delegator and the entity to which signing rights are delegated is called proxy signer. This notion of proxy signer is very useful when some officer in an organization want to delegate his role to another officer without sharing the secret key associated with him. However the use of proxy signature scheme is not good solution when the delegated authorities are changing from time to time.

Applications likes e-commerce, m-commerce and e-governance make use of proxy signatures to sign on behalf of original entity. Mobile agent based technology is another usage area of proxy signature.

Similar to digital signature, the major security threat of key exposure by disgruntle owner remains the concern area for proxy signature. To protect the integrity and non-repudiation of signed documents from the key leakage, Forward secure signature schemes are more advantageous.

Forward secure signature (FSS) are designed to reduce the damage caused by key exposure by generating the time evolving secret key while keeping the same public key. Consider that the validity period of public key is  $T$  time unit then secret keys will go on changing for each  $t$  time interval. For example if validity period of public key belonging to some entity is 10 days then based on duration of interval say one day, there would be 10 different private keys generated. The notion of FSS is defined to tackle the key exposure and in turn signature forgery problem.

In e-governance service, proxy signature schemes can be used for role based authorization. In such case signing authority delegates his role to subordinate during his absence by sharing his role key and not the secret key. The subordinate officer using role key and his own identity key produces the proxy key and signs document by using proxy key and suitable signature scheme. Consider a situation where the designated officer is changing frequently then verifying system will face issues as the proxy key generated has identity associated with proxy signer. Role based authorization using proxy scheme is useful only when the delegated authority is not changing frequently. For such situation we propose the use of proxy key in a forward secure manner. In forward secure system proxy secret key will be generated in a time evolving manner and will be valid only for specified duration. New proxy secret key will be generated for next designated officer. Verification operation however can utilize the proxy public key generated for entire delegation period.

Multilevel proxy model allows one proxy signer to delegate signing rights to another which in turn can delegate his signing rights to third entity and so on. Mobile agent system in e-commerce can be implemented using proxy signature. Similarly for E-governance applications requiring authorization from various government agencies in a predefined manner will be helpful if implemented using forward secure proxy signature scheme.

With technology changes there is shift in authentication security requirements which should be made available on smart portable handheld devices. Complex applications like E-Governance applications where authentication security is of utmost importance require interaction with multiple government agencies are migrated to mobile devices. Gartner in its survey in 2016 for information security has identified the need of data security in the domain of Internet of Things. Signature algorithms based on elliptic curve cryptosystem (ECC) have evolved prioritizing this requirement.

Elliptic curve based signature algorithms promises authentication, integrity and non-repudiation security as with other cryptosystem but comparatively with much smaller key size. The feature like reduced computations and lesser power consumption potentials their use for applications working in constrained environment.

Signcryption schemes is another cryptographic primitive which performs the function of digital signature and public key encryption in order to provide confidentiality apart from authentication, integrity and non-repudiation security. In this paper our focus is on generating proxy signature in forward secure manner and not on signature generation and encryption.

Proxy signature algorithms based on ECC are explored by the researchers [8, 9] majorly targeting to enhancing the performance and security as compared to their counterparts using RSA or elgamal cryptosystem. The discrete logarithmic problem on the elliptic curve is considered as a hard problem which makes ECC based algorithms robust.

Present solution on ECC based forward secure proxy group signature scheme creates blinded identity for each of the proxy signer and is linked with proxy key pair [8]. This blinded identity protects other proxy signed documents if any attack on proxy key pair occurs.

Analyzing the current signature algorithms from perspectives like application requirements, performance and security requirements, we identified the need for cost efficient proxy signature with forward secrecy algorithm which will be appropriate for use in multilevel delegation in controlled environment. This paper propose a forward secure proxy signature scheme using ECC which will satisfy the security requirements of proxy signature scheme, generate proxy private evolving key suitable for multi-level delegation in controlled environment.

In Sect. 2 we introduce proxy signature scheme. Discussion on various proxy signature schemes is presented in Sect. 3. Section 4 discusses proposed proxy signature scheme. Security features and other work is discussed in Sect. 5 followed by conclusion in Sect. 6.

## 2 Proxy Signature

A proxy signature allows a designated person, called a proxy signer, to sign the message on behalf of the original signer. Proxy signatures are very useful tools when one needs to delegate signing capability to other party. Various other applications where proxy signature plays major role are e-cheques or digital documents to be signed by multiple users, mobile agents, e-cash, electronic commerce, distributed shared object systems etc.

Mambo et al. [1] presented proxy signatures in 1996. Their scheme allows an original signer to delegate his signing right to a proxy signer to sign the message on behalf of an original signer. Later, the verifier, which knows the public keys of original signer and a proxy signer, can check validity of a proxy signature issued by a proxy signer.

Proxy Signature schemes majorly differs from each other in three aspects namely degree of delegation, delegation capability and security features provided by implementation methodology. The signature schemes must satisfy the characteristics as specified in [4] like Verifiability, Unforgeability, and Distinguishability etc.

The application domain of proxy signature scheme is quiet expanding. In e-Governance services to improve efficiency, citizen's applications required to be processed in specified time frame and in transparent manner. In situation the authority is on leave or charge is transferred, the assigned work to that officer remains piling. In such case this role can be carried out by another designated officer. This role delegation can be made possible by using proxy scheme. With use of forward secure scheme, for each designated officer proxy private key will be generated and for the documents signed in that duration verifier can verify the signature with the single public key certificate.

### 3 Review of Some Proxy Signature Scheme

Proxy signatures were first proposed by Mambo et al. and later several different schemes were proposed by researcher using discrete logarithmic assumption.

Sunitha and Amberkar in [2] proposed a proxy signature scheme using elgamal cryptosystem which is useful in delegating the signing rights to the subordinates in controlled environment in financial transactions. In the proposed scheme verifier can verify the proxy public key by using the public key of original signer and that of proxy signer thus satisfying security property strong undeniability. This step in verification also protects any intruder to impersonate as proxy signer. Proxy private key generated by combining the received warrant and private key of proxy signer makes signature scheme satisfying strong identifiability property. From knowledge of proxy public key deciphering private key becomes discrete logarithmic problem therefore the scheme is secure against any kind of forgery attack. The proxy public key is not available publicly but is calculated by the verifier which essentially checks the authenticity of proxy signer thus protects the framing attack.

In 2012, Aboud and Yousef [12] proposed a variant of basic Mambo et al proxy signature scheme for partial delegation with warrant considering protected and unprotected scheme. The authors claimed computation time is reduced significantly in terms of modular inversion and hash function in warrant partial delegation scheme. The scheme designed is also secure against framing, impersonation and forgery by original signer attack.

Few proxy signature schemes are suggested based on integer factorization problem using RSA cryptosystem. Verma and Sharma [3] in 2013 proposed proxy signature and verification scheme based on RSA. The scheme satisfies security parameters required by proxy signature scheme and excels in terms of computation cost. In their paper comparison with other RSA based scheme namely LKK scheme and Shao's scheme is performed. For proxy key generation, signature generation and verification scheme proposed by [3] places less computation over-head as compared to other two schemes. However authors have not considered proxy revocation mechanism.

For FSS scheme specification of (a) Secret key updating algorithm, (b) Public key (c) Signing and verification algorithm is essential. In 2007 Amberkar et al. [6] identified that modifications are required in elgamal signature scheme to satisfy forward secure property and proposed the elgamal-like signature scheme. In proposed scheme secret key updation algorithm of Bellar-Miner is used but the public key generation is modified as by using secret key updation algorithm T number of times. In their proposed scheme signature generation constitutes message independent component, message dependent and secret key component along with time component. Elgamal-like FSS scheme is secure against message forgery and impersonation attack.

Authors Chen et al. [9] proposed proxy signature scheme using Bellar-Miner forward security scheme for multiple proxy signers. This scheme uses public key generation phase where common public key is used by all proxy signer. Corresponding to each proxy signer original signer has separate secret key. Proxy signer's uses seed secret key and apply key evolution algorithm. Authors mentioned that secret key generation by other proxy signers is not possible but is vulnerable to proxy signer's collusion attack.

In [5] authors Sunitha and Amberkar has modified their proxy signature scheme with multiple signers and made it forward secure with proxy revocation. In its scheme the original signer can delegate to multiple proxy signers in overlapping time duration based on number of proxy signers and equal division in the time duration. The novelty in the scheme is that the original signer sends delegation information at the beginning of the duration but the proxy signer can generate proxy signature only in its allocated duration therefore overcoming the intervention with proxy signer by sending messages.

In 2009 Qi and Chen [4] proposed proxy signature scheme with difficulty to solve ECDLP. It uses probabilistic encryption algorithm for protecting secrecy of proxy signer and distinguishability. Higher level of security is obtained with the use of one way trap door function. Even though the authors claim it to be proxy signature, focus is on group signature with verifier and less has been discussed about delegation of signing rights.

In the area of forward security various other alternatives of signature schemes are studied and the important one is signcryption schemes. In 2013, authors M Dutta, A K Singh and Ajay Kumar proposed the signcryption scheme using ECC [11] satisfying forward secrecy where the authentication of encrypted message and public verification of signature is performed. Authors have compared their proposed schemes with various existing schemes with respect to computation and communication cost. The scheme is designed to reduce the receivers computational cost.

Author Abdelfatah in [10] proposed a proxy signcryption scheme using ECC claiming to be computationally efficient than its counterpart schemes which secure majorly against key misuse and non-repudiation of signature delegation. The focus of research work is on proxy signature and encryption of the message and not to ensure forward secrecy. When compared with our proposed scheme the focus is on defining forward secure proxy signature scheme. Therefore the proxy private key generated is time evolving in nature.

### 3.1 Analysis of Existing Scheme

In depth study of various proxy signature schemes are carried out in previous section state that proxy signatures are important when leveraging the rights for signing the document to the subordinate Forward secrecy property further protects previous or future documents signed by authorized users in case of secret key exposure or Signature forgery. We emphasize the use of forward security in a novel way as to delegate the signing rights to more than one proxy signer in a linear manner. It important for the applications which require delegation of the authority and the delegated authority is changing from time to time. As specified in the introduction section forward secure proxy signature schemes will be useful in increasing the efficiency of many e-Governance services.

Further research reveals that sufficient work is carried on either forward secure proxy signature schemes based on RSA/Elgamal cryptosystem or proxy signature schemes based on ECC. But very less attention is paid on ECC based forward secure proxy signature scheme.

Digital signature defined on Elliptic curve cryptosystem has emerged as cost efficient and energy efficient alternative for existing signature schemes. Smart portable

devices technology has changed the norms of security and also has become challenge to provide it. For providing authentication and authorization security the use of ECC for smart hand held devices, pervasive systems like IOT has become transitive.

Need for robust authentication and authorization system using proxy signature scheme is required in various environment for portable devices and emerging application fields like IOT has motivated us to define new forward secure proxy signature scheme using ECC.

## 4 Proposed System

The need of forward secrecy is identified not only in terms of protecting from signature forgery for past or future documents but also in delegating the authority at multilevel with moving time. Applications using agent systems or e-Governance systems will require proxy delegation in a forward secure manner.

In India, E-governance applications provide citizens to submit their applications to government offices in terms of E-forms. These E-forms are processed by the officers and in many cases require interdepartmental approval in some specified sequence. These applications are processed and government officers perform digital signature along with approval/disapproval remark. Proxy signatures with forward security policy is useful in such application to provide multilevel authentication in a predefined manner. Proxy signatures are useful as they will allow speedy and transparent processing of the applications even in the absence of the officer. Forward security policy is important as can provide multilevel authorization of the application in terms of E-forms.

In our proposed method, original user sends signed warrant, which entrust proxy signer of original signer's authenticity. It further uses warrant along with its secret key for generation of proxy private key; which satisfies property of identifiability and undeniability. We propose to use this proxy secret key generated as seed key or initial key at time unit 0 for further evolution of time varying secret key in Forward secure proxy signature algorithm. Proxy public key is created by repeating the secret key updating algorithm for T unit of times which remains same for multilayer proxy signers. We have suggested use of this proxy key pair further for signing and verification operations.

With extensive use of mobile devices, security applications should be made available on portable devices including authentication and authorization. User identity and authentication algorithms containing DSS require complex operations to be performed which usually results in more power consumption we propose a new Forward secure proxy signature scheme for delegating the signing authority in a secure manner and with provable efficiency by using Elliptic curve system by combining the systems defined by [7, 8]. The proposed scheme is divided into two sections:

Section A: ECC based Proxy Signature scheme consists of four phases (1) System Initialization (2) Proxy key generation (3) Signing process (4) Signature verification. Section B: ECC based forward secure proxy signature scheme consisting of five phases (1) System Initialization (2) Proxy key generation (3) Forward secure proxy key evolution algorithm (4) Signing process (5) Signature verification.



As shown in Fig. 1. Scheme defined in section B uses the phases 1 and 2 described in section A. Proxy secret key generated in the scheme of section A is considered as input to the key evolution algorithm in section B.

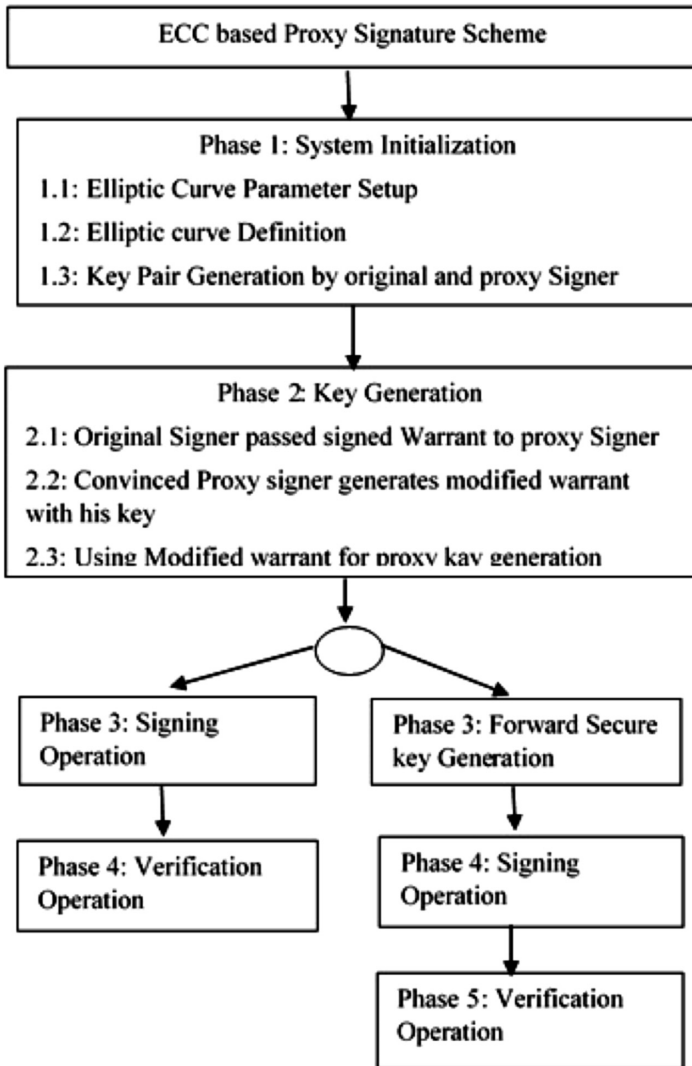


Fig. 1. Workflow of ECC based algorithm.

#### 4.1 ECC Based Proxy Signature Scheme

The application which requires delegation of signing rights for specific purpose for ex: software agents in reservation system signing on behalf of client require signing and verification operation after generation of proxy key pair.

- Phase 1: System Initialization: System parameters are set specifying Field with prime number, Curve specification, and base and generator point identification. Original signer initializes the scheme by generating secret key, corresponding public key and sends signed warrant message, originals signers' information to the proxy signer. Proxy signer on receiving checks the authenticity of received warrant and then modifies warrant to include proxy signer's details. The method for proxy authorization mentioned in [8] is modified to create a proxy key pair.
- Phase 2: proxy key generation: By using original signers secret information passed to proxy signer creates modified warrant information and this modified warrant is used as base for creating proxy secret key to be used by proxy signer. Proxy public key is obtained by using generator point on the curve.
- Phase 3: Signature algorithm:
- Phase 4: verification algorithm:

**Algorithm A: ECC based Proxy signature scheme**

System Initialization

A secure elliptic curves cryptosystem can be set up as the following.

Consider  $q$  as large prime number and  $F_q$  is a finite field; Elliptic curve in the finite field is defined as

$$E : y^2 = x^3 + ax + b \text{ where } (a, b) \in F_q$$

and

$$4a^3 + 27b^2 \text{ mod } q \neq 0$$

$P \in E(F_q)$  is a base point of the curve whose order is a large prime number such that  $\text{order}(P) = 1 \geq 160$ .

The generator point  $G$  is derived from base point  $P$

- (1) Original signer with  $(k_a, KA)$  as private, public key
- (2) Proxy signer with  $(k_b, KB)$  as private, public key
- (3)  $m$ : message,  $m_w$ : Warrant information;  $h()$ : one way secure hash function.

Step 1: Original Signer delegates signing rights by

- (1) Select random no  $u_0 = z_i^* u_0 \in Z_i^*$  and generate  $U = u_0 * G$
- (2) Computes:  $S_w = (k_a * h(m_w + U(x))) * G^{-1}$
- (3) Sends  $(S_w, KA, m_w, U)$  to proxy signer.

Step 2: Proxy signer authorizes the received quadruplet using public key of original signer and by computing  $h(m_w + U(x)) * KA + U$ . If matched with received  $S_w$  it guarantees that the proxy signing request is from authentic source otherwise rejects the further operation.

Step 3: Proxy key generation:  $(k_c, KC)$

$$\text{Calculate: } S'_w = S_w + k_b \text{ mod } l \quad (2)$$

$$k_c = S'_w \text{ mod } l \quad (3)$$

$$k_c = S'_w \text{ mod } l \quad (4)$$

$$KC = k_c * G \quad (5)$$

Step 4: Signing on the message m:

- (1) Select random integer  $v \in z_l^*$  which differs for each message
- (2) Generate  $V = v * G$  message independent portion of signature
- (3) Generate  $S = (v + h(m) * V(x) * k_c^{-1}) * KC$

In Eq. (5) message dependent portion of Signature i.e. S is calculated using  $k_c$  (Private Key)

- (4) Passes (S, V, U, m) as a signed document.

Step 5: Verifier verifies the signature by

$$\text{Calculate } S' = V + (h(m) * V(x)) * G \quad (6)$$

$$\text{Signature is valid if } S = S' \quad (7)$$

Correctness of the algorithm:

For Eq. (7)

$$S = ((v + h(m) * V(x) * k_c^{-1}) * KC$$

$$S = ((v + h(m) * V(x) * k_c^{-1}) * k_c * G$$

$$S = v * G + (h(m) * v(x)) * G$$

$$S = S' = R.H.S.$$

In step 3, separate proxy key pair is generated from key pairs of original as well as proxy signer.

## 4.2 ECC Based Forward Secure Proxy Signature Scheme

Applications which requires proxy signer to sign the documents in not fully trusted environment for longer duration or the proxy signer wish to delegate the signing rights further at next layer in the controlled manner the choice of forward secrecy in proxy secret key can be imparted. E-Governance services requiring government authorities to pass to signing rights to multiple subordinates or peer for different days during his leave duration, choice of FSS with proxy signature is most suitable.

As mentioned in earlier section of proposed system Phase 1 and 2 are same as that of ECC base proxy signature scheme.

Phase 3: Key evolution algorithm for proxy secret key: Forward secure secret key is generated by dividing the total time T in equal intervals. For ex if proxy signing rights are received for 10 days and he further wish to delegate the rights to individual for each day then 10 different forward secure keys will be generated. One way function utilizing the secret key of previous duration is used to generate secret key for next time slot. Initial seed key is considered as proxy secret key generated in previous phase. Forward secure proxy public key is obtained by applying algorithm T number of times.

Phase 4: Signature Algorithm: On receiving the proxy secret key for interval, proxy signer signs the message by using this secret key using algorithm [5, 7]. Signature components apart from message dependent and independent also have components which has private key for that duration.

Phase 5: Signature verification: On receiving (r, s, R, m) as a signature on the message verifier computes the components dependent on message and dependent on forward secure proxy public key and verifies the signed document

**Algorithm B: ECC based forward secure proxy signature scheme**

Step no 01 to 03 are same as in Algorithm A.

Step 4: Forward secure (FS) Key generation Procedure: Proxy secret key generated in previous algorithm is used as seed key  $k_{c_0} = k_c$  for generation of FS keys. Time evolving private key for each time interval  $t_i$  out of T time unit is calculated as

$$k_{c_i} = k_{c_{i-1}}^2 \text{ mod } l$$

(2) Public key for T time duration is calculated as  $Q = k_{c_0}^{2^T} * G$

Step 5: Signing operation: This operation is performed by proxy signer of that designated period. For each message m to be signed

(1) Select random integer  $v \in z_l^*$

$$(2) \text{ calculate } r = (v - k_{c_i}) * h(m)^{-1} \tag{8}$$

$$(3) \text{ calculate } s = v - r * k_{c_i}^{2^{T-i}} \tag{9}$$

$$(4) \text{ Computes } R = k_{c_i} * G \tag{10}$$

Passes (r, s, R, m) as a signed document

Step 6: Signature Verification:

(1) After receiving (r, s, R, m) receiver

$$\text{Calculate } X = (h(m) * r) * G + R \quad (11)$$

$$Y = s * G + r * Q \quad (12)$$

$$\text{Check if } X(x) \bmod l == Y(x) \bmod l \quad (13)$$

Verify the signature. Correctness of the algorithm:

From (11)

$$\begin{aligned} X &= (h(m) * r) * G + R \\ X &= \left( h(m) * (v - k_{c_i}) * h(m)^{-1} \right) * G + R \\ X &= v * G - k_{c_i} * G + k_{c_i} * G \\ X &= V(x, y) \end{aligned} \quad (14)$$

From (12)

$$\begin{aligned} Y &= s * G + r * Q \\ Y &= \left( v - r * k_{c_i}^{2^{T-i}} \right) * G + r * k_{c_0}^{2^T} * G \end{aligned} \quad (15)$$

Now

$$\begin{aligned} k_{c_i}^{2^{T-i}} &= (k_{c_{i-1}})^{2^{T-i+1}} \\ k_{c_i}^{2^{T-i}} &= (k_{c_{i-1}})^{2^{T-i+i}} \\ k_{c_i}^{2^{T-i}} &= (k_{c_0})^{2^T} \end{aligned} \quad (16)$$

Using (16) Eq. (15) can be rewritten as

$$\begin{aligned} Y &= \left( v - r * k_{c_0}^{2^T} \right) * G + r * k_{c_0}^{2^T} * G \\ Y &= v * G = V(x, y) \end{aligned} \quad (17)$$

Therefore Eq. (13) justifies verification

## 5 Security Analysis

In this paper we have presented security analysis in three ways viz security characteristics satisfied by the proposed scheme, comparison with other schemes in terms of computation cost and numerical example as a proof of concept.

The security properties satisfied by our forward secure proxy signature scheme are as follows:

- (a) Verifiability: Private proxy key is generated by using signed warrant by original signer. Therefore the receiver can be convinced that the rights of signing are delegated to proxy signer.
- (b) Distinguishability: Proxy signature private key is generated by using warrant signed by original signer and private key of proxy, so the signature generated will be easily distinguishable from that of proxy signer’s signature.
- (c) Unforgeability: Forging the proxy signature by intruder will be possible only if intruder could derive the private key from public key which is sheltered by ECDLP.
- (d) Resistance to Coalition Attack: If the original and other than designated proxy signer decide to secretly forge the signature, then the only way is to identify 2 parameters  $v$  (random variable per message basis) and  $k_{c_i}$  (secret key for the specified duration) from the information  $(r, s, R, m)$ , which is hard as per Elliptic curve Discrete logarithm problem (ECDLP).

**5.1 Performance Analysis**

As mentioned in the introduction section, there are very few schemes where both the aspects of forward secrecy and proxy signature using ECC are covered. Following notations are used for comparison of proposed scheme with the existing scheme given by [10]

- $T_{EC-M}$ : Total number of elliptic curve multiplication operation
- $T_{EC-A}$ : Total number of elliptic curve Addition operation.
- $T_{h()}$ : Total number of hashing function is used
- $T_{EXP}$ : Proxy private keys generated for each time interval

**Table 1.** Computation cost comparison with existing scheme

Forward secure proxy signature scheme based on ECC	$T_{EC-M}$	$T_{EC-A}$	$T_{h()}$	$T_{EXP}$
Zhou et al. [8]	9	4	04	–
Proposed scheme	10	3	3	4

**5.2 Numerical Example**

The proposed algorithm is implemented using Sage math 7.3, the sample execution obtained by considering very small values for the purpose of better understanding  $P = 23$  and Elliptic curve parameters are  $a = 1$  and  $b = 0$  and message  $m = 5$ . In Elliptic curve, private key is a random variable in the finite field and public key is a point on curve.

- (‘Original Signers key pair  $k_a, KA'$ , 1, (18: 13: 1))
- (‘proxy signer’s key pair  $k_b, KB'$ , 0, (0: 1: 0))
- (‘Proxy signature Secret and public key pair  $k_c, KC'$ , 3, (18: 13: 1))
- (‘Forward secure key generated for time unit 1 i.e.  $k_{c1}'$ , 1)
- (‘Signature on the message  $(r, s, R, m)$  in time unit 1’, 1, 1, (18: 13: 1), 5)

Receiver performs verification using forward secure public key  
 (\*Values of X and Y', (18: 10: 1), (18: 10: 1))  
 Forward secure signature verified by the receiver.

## 6 Conclusion

Proxy signature scheme is gaining importance in electronic transaction for providing authentication and authorization. Alternative schemes of proxy signature based on ECC or forward secure signature based on ECC exist but very less attention to the combination of them is paid by the researchers. Authorization applications requiring multi-level delegation of authority can be realized using property of forward secrecy. In this paper a forward secure proxy signature scheme based on ECC is presented with the aim to make it suitable for multilevel delegation and cost efficient. The signature scheme defined is resistant to coalition attack and antiforgery attack.

## References

1. Mambo, M., Usuda, K., Okamoto, E.: Proxy signature: delegation of the power to sign messages. *IEICE Trans. Fundam.* **E79-A**, 1338–1354 (1996). **2**(5), 99–110 (2016)
2. Sunitha, N., Amberkar, B.: Proxy signature schemes for controlled delegation. *J. Inf. Assur. Secur.* **3**(2), 159–174 (2008)
3. Verma, S., Sharma, B.: An efficient proxy signature scheme based on RSA cryptosystem. *Int. J. Adv. Sci. Technol.* **51**, 121–126 (2013)
4. Qi, C., Chen, X.: A new digital proxy signature scheme based on ECDLP. In: *International Conference on Computational Intelligence and Security 2009*, Beijing, China, pp. 473–477. IEEE press (2009)
5. Sunitha, N., Amberkar, B., Koulgi, P.: Controlled delegation in e-cheques using proxy signatures. In: *11th IEEE International Enterprise Distributed Object Computing Conference 2007*, MD, USA, pp. 414–419. IEEE (2007)
6. Amberker, B., Koulgi, P., Sunitha, N.: Forward security for an ElGamal-like signature scheme. In: *6th Annual Security Conference 2007*, Las Vegas, NV, pp. 20–29. IEEE (2007)
7. Bo, L., Yilin, L.: A new forward-secure digital signature scheme based on elliptic curve. In: *2nd International Conference on Industrial and Information Systems 2010*, vol. 2, pp. 152–155. IEEE (2010)
8. Zhou, X., Su, Y., Wei, P.: Further study on proxy authorization and its scheme. In: Zhou, J. (ed.) *Complex 2009*. LNICSSITE, vol. 5, pp. 1701–1718. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-02469-6\\_48](https://doi.org/10.1007/978-3-642-02469-6_48)
9. Chen, M., Kuo, M., Lai, C.: A new forward secure proxy signature scheme. In: *International Symposium proceedings on Communication 2005*, Taiwan (2005)
10. Abdelfatah, R.: A novel proxy signcryption scheme and its elliptic curve variant. *Int. J. Comput. Appl.* **165**(2), 36–43 (2017). FCS
11. Dutta, M., Singh, A.K.: An efficient signcryption scheme based on ECC with forward secrecy and encrypted message authentication. In: *3rd IEEE International Advance Computing Conference 2013*, pp. 399–403. IEEE (2013)
12. Aboud, S.J., Yousef, S.: Efficient undeniable threshold proxy signature scheme. In: *6th International Conference on Information Technology 2013*, pp. 150–153. IEEE (2013)

# Efficient and Robust Secure In-Network Aggregation in Wireless Sensor Networks

Radhakrishnan Maivizhi<sup>(✉)</sup> and Palanichamy Yogesh

Department of Information Science and Technology,  
College of Engineering, Anna University, Chennai, India  
maivizhil6@gmail.com, yogesh@annauniv.edu

**Abstract.** Data aggregation is a known and widely acknowledged approach to reduce energy consumption in Wireless Sensor Networks (WSNs). Preserving data confidentiality and integrity along with en-route aggregation is a great challenge. In this paper, we propose an Efficient and Robust Secure In-network Aggregation (ERSIA) protocol for additive aggregation function. This protocol employs privacy homomorphism to achieve data privacy and combines it with secret sharing to protect the integrity of the aggregated data. Security analysis shows that the proposed protocol is robust against active and passive attacks and securely computes the aggregation. The results of communication overhead analysis demonstrate that ERSIA outperforms other existing methods and increases the lifetime of the network while achieving end-to-end security.

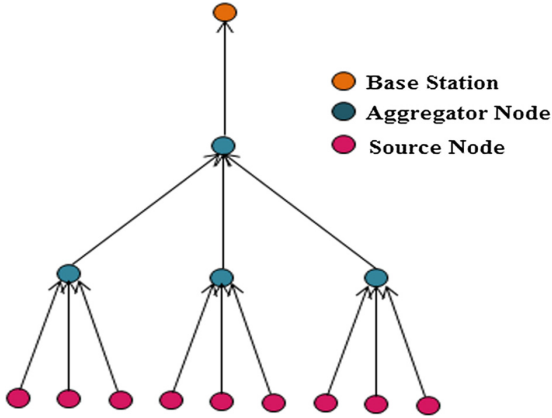
**Keywords:** Wireless sensor network · Confidentiality · Integrity  
Privacy homomorphism · Secret sharing

## 1 Introduction

Due to the developments in wireless networks, especially in mobile ad hoc networks, distributed sensing has become widespread and quite a lot of researches are undertaken in Wireless Sensor Networks (WSNs). WSNs have become a reality due to the advances that had happened in wireless communication and computing devices and they are used in a wide range of real world applications that face numerous challenges [1]. A wireless sensor node is a tiny device that has limited battery power, CPU cycles, bandwidth and storage. Among these resources, energy is one which directly affects the lifetime of WSNs. Since the energy required for communication is more than the energy required for computation, minimizing the communication cost is urgently required.

Removing the redundancy in the sensed data is an appropriate approach to reduce the resource consumption and redundancy removal shall be achieved through data aggregation. The goal of data aggregation is to combine data from the sensors and aggregate them using aggregation functions namely sum, average, count, min, max, standard deviation, variance, etc., and forwarding the result either to the upstream nodes or to the base station [2, 3]. Figure 1 shows an illustration of tree-based data aggregation scheme.





**Fig. 1.** Tree-based data aggregation

The main goal of in-network processing is to increase the longevity of a network by minimizing the resource utilization in the sensor nodes. However, the data aggregation protocols may affect major performance metrics such as latency, accuracy, fault-tolerance and security. Most of the protocols used to aggregate the data are vulnerable to a variety of attacks. For example, an attacker can attack both the confidentiality and integrity of aggregated data by compromising aggregator nodes. Therefore, Secure Data Aggregation (SDA) protocols [4–11] aim to perform data aggregation in a secured way. Confidentiality, integrity, authentication and freshness of data are the major requirements of a secured data aggregation scheme.

Secure data aggregation can be of two types: aggregation in which end to end security is preserved (end to end scheme) and aggregation in which security is ensured only between the neighbors (hop by hop scheme) [12]. In hop-by-hop secure data aggregation, every intermediate (or aggregator) node carries out encryption and decryption tasks. However, sensor data become malleable at these nodes. Hence it is mandatory to preserve the privacy of sensor readings at aggregator nodes. In end-to-end secure data aggregation, data is encrypted only once and can be decrypted only at the sink node. Privacy homomorphism [13], eliminates the need to decrypt the data at aggregator nodes and at the same time enables processing the ciphertext. Also, it reduces security vulnerabilities and additional computation overhead.

In this paper, we propose Efficient and Robust Secure In-network Aggregation (ERSIA) protocol for additive aggregation function in wireless sensor networks. ERSIA has been proposed on the basis of end-to-end secure data aggregation. The proposed approach utilizes privacy homomorphic encryption and secret sharing mechanism to provide confidentiality and integrity. Our approach is scalable, robust to various cryptographic attacks and incurs less communication overhead compared to existing approaches, thereby it leads to strong security guarantees.

The rest of the paper is organised as follows. In the next section, we provide the summary of related work on secure data aggregation. Section 3 formulates problem statement. Section 4 presents the Efficient and Robust Secure In-network Aggregation

(ERSIA) protocol. The security analysis of ERSIA is given in Sect. 5. In Sect. 6, we analyze the performance and discuss the results. Finally, concluding remarks are given in Sect. 7.

## 2 Related Work

A wireless sensor network consists of resource-constrained sensor devices. This mandates the development of new protocols to securely aggregate the sensed data. Sensors are not only unreliable, but can also be compromised by malicious adversaries. Thus the development of secure data aggregation protocols takes into account the new kinds of attacks that emerge with time. In this section, we discuss some of the major works done in the areas of preserving confidentiality and protecting integrity of the aggregated result.

Castelluccia et al. [4] developed an efficient and provably secure additive aggregation protocol (CMT cryptosystem). This scheme adopted one-time pad cipher for aggregation of ciphertext. Although the scheme is based on simple and cheap modular operations, it consumes more bandwidth as compared to other hop-by-hop secure data aggregation protocols. Moreover CMT does not protect data integrity.

Yang et al. [5] have proposed a Secure hop-by-hop Data Aggregation Protocol (SDAP). This protocol employs divide-and-conquer principle, which dynamically partitions the tree into numerous groups using a probabilistic approach and then generates group aggregates based on commitment based hop-by-hop aggregation. Although, the proposed protocol achieves confidentiality, integrity and source authentication, it incurs high energy utilization and transmission overhead.

Prakash et al. [6] have proposed Cluster-based Private Data Aggregation (CPDA) scheme. This data aggregation scheme preserves privacy for additive aggregation functions. It takes the advantages of both the clustering techniques and algebraic properties of polynomials for securely aggregating the data. The primary objective of CPDA is to bridge the gap between en-route data aggregation and data privacy. Less communication overhead is achieved by this scheme by compromising data integrity.

Nath et al. [7] have proposed a SECure Outsourced Aggregation via one-way chains (SECOA) framework. SECOA employs the unified use of one-way chains and supports different aggregate functions. The major advantage of this scheme is its ability to detect any malicious activity in aggregation without communicating with sensors. It is the state-of-the-art method for data integrity. The limitation of this framework is that it does not provide data confidentiality.

Poornima and Amberker [8] have proposed Secure End-to-End Data Aggregation (SEEDA) protocol. It is a hybrid approach which combines the prominent features of hop-by-hop and end-to-end aggregation schemes. This work is able to ensure data privacy in data aggregation. However data integrity is not within the scope of this work.

Liu et al. [9] have proposed High Energy-Efficient and Privacy-Preserving (HEEPP) secure data aggregation scheme. The proposed scheme modifies the slicing and assembling technology to preserve the privacy of aggregated data and developed a query mechanism to achieve accuracy while performing data aggregation. The advantage of this scheme is that it incurs low bandwidth consumption.

Alghamdi et al. [10] have proposed two approaches Sign-Share and Sham-Share. These schemes use secret sharing and signatures which allow the nodes designated as aggregators to aggregate the data. These methods have the advantages of resisting selective forwarding and modification attacks. The limitations are: the aggregators need more energy and the aggregators that are at distant locations from base station cannot communicate with the base station directly.

Many existing works focus either on confidentiality or on integrity. Only few works focus on both and they consume more resources from the network. In the proposed work we provide both data confidentiality and data integrity and at the same time we keep the communication overhead as small as possible.

### 3 Problem Definition

Hostile and unattended deployments of wireless sensor networks are prone to variety of attacks. An adversary may compromise a node and take control of the node or eavesdrop the wireless medium. In hop-by-hop encryption scheme, adversaries have the possibility of injecting false information into the data, as they are encrypted/decrypted at every intermediate node. Therefore, there is a need for encryption scheme for transmitting and aggregating data so that the adversaries cannot alter the data or reveal the confidentiality of the obtained information. In end-to-end encryption, a compromised node cannot extract or comprehend the information from the obtained ciphertext. The proposed scheme achieves end-to-end security and ensures that for an adversary it is much difficult to gain access to the data.

#### 3.1 Network Model

A WSN consists of a set of  $n$  sensors and employs a powerful and trusted base station (BS). To ensure secure communication, the base station should have enough memory and power. Since the sensor nodes are resource limited, energy available in the battery is exhausted at a faster rate when they communicate messages. We assume the network topology is tree-based and the nodes are stationary. Each sensor can be either a source,  $S$  or an aggregator,  $A$ .

An aggregation tree is constructed based on balanced ternary tree topology [11]. Let  $t$  be the range of possible sensor readings. We assume that, only the leaf node (source) generates the data and the intermediate (aggregator) node performs the aggregation. BS broadcasts an authenticated query with  $\mu$ TESLA [14]. Here, we focus on additive aggregation. For simplicity, we focus on SUM aggregation function. Before data transmission takes place, the sensed readings are encrypted and are decrypted and validated at the base station.

#### 3.2 Adversary Model

Wireless sensor networks pose several security challenges. The security challenges related to secure data aggregation are data confidentiality, data integrity, data

authentication and data freshness [15]. In this paper, we classify the adversaries into passive adversaries and active adversaries.

Passive adversaries eavesdrop the wireless channel for deducing a key or to collect sensitive information that affects the data privacy and data confidentiality. Without participating in the communication, the adversary can launch a series of attacks like ciphertext attacks, known plaintext attacks, etc., Encryption techniques helps to protect the network against passive adversaries.

Active adversaries are able to disturb the basic functioning of the network and degrade its performance [16] by altering the contents of the communication. False data injection is a popular attack carried out by the adversaries to affect the performance of the network. Active adversaries can launch malleability, replay attacks, node compromise attack and Denial-of-Service (DoS) attacks [17]. The proposed scheme guarantees end-to-end security against active adversary attacks.

### 3.3 Objectives

The primary objective of our proposed scheme is to achieve accurate, efficient, robust and secure data aggregation that satisfies confidentiality, integrity and authentication. Thus the proposed scheme provides the following security requirements:

<i>Data confidentiality:</i>	To ensure the secrecy of the transmitted data.
<i>Data Integrity:</i>	To ensure the originality of the received data without any alteration during transmission.
<i>Source authentication:</i>	To ensure the verification of the sender of the data.
<i>Efficiency:</i>	To reduce the additional bandwidth consumption considerably, to achieve added features in protecting the integrity of in-network aggregation scheme.
<i>Accuracy:</i>	To ensure the accuracy of the aggregated results as they play a major role in making critical decisions.
<i>Robustness:</i>	To ensure robustness against cryptographic attacks.

## 4 The Proposed Protocol

In this section, we discuss the building blocks employed in proposed protocol and present the ERSIA protocol.

### 4.1 Building Blocks

#### Privacy Homomorphism

Privacy Homomorphism (PH) is an encryption scheme with a homomorphic property for processing encrypted data. It was first introduced by Rivest et al. [13]. PH allows direct computations to be performed on ciphertext. The formal definition of privacy homomorphism is as follows:

**Definition**

Let  $E(\cdot)$  be the encryption function and  $D(\cdot)$  be the corresponding decryption function. Given an encryption of  $a$ ,  $E(a)$ , and an encryption of  $b$ ,  $E(b)$ , we can compute,

$$E(a) \oplus E(b) = E(a \otimes b)$$

The decryption process gives the same result

$$D(E(a \otimes b)) = a \oplus b$$

Cryptosystems supporting privacy homomorphism schemes use the same operator,  $\oplus$  or  $\otimes$ , or use different operators,  $\oplus$  and  $\otimes$  for plaintext and ciphertext. In this paper, we use  $\oplus$  operator for both plaintext and ciphertext. Symmetric key cryptosystems use same key for both encryption and its corresponding decryption but asymmetric key cryptosystems use different keys.

**Elliptic Curve ElGamal cryptosystem**

EC-ElGamal [18] cryptosystem is an elliptic curve based public key cryptosystem. It is defined on an elliptic curve over a finite field  $\mathbb{F}$ . Its key size is 160 bits and achieves the same level of security as 1024 bit key size of RSA achieves. The smaller key size improves energy utilization, bandwidth efficiency and storage capabilities of WSNs.

*Key Generation K:*

The secret key  $s_k$  is chosen randomly and the public key  $p_k$  is calculated as a function of secret key and a point  $P$  that lies in the chosen elliptic curve  $E(\mathbb{F}_p)$ . The field  $\mathbb{F}_p$  is decided by choosing a very large prime number  $p$ . The public key  $p_k$  is computed as follows.

$$p_k = s_k \cdot P \text{ on } E(\mathbb{F}_p)$$

*Encryption E:*

The plaintext  $M$  belongs to  $E(\mathbb{F}_p)$  and a random integer  $r$  is chosen. Then the two ciphertext  $C_1$  and  $C_2$  are computed as follows.

$$C_1 = r \cdot P \text{ and } C_2 = M + r \cdot p_k$$

In the above  $P$  is a chosen point that lies in  $E(\mathbb{F}_p)$  and  $p_k$  is the public key generated in the key generation phase.

*Ciphertext Aggregation A:*

The messages  $M_1$  and  $M_2$  are encrypted as

$$E(M_1) = \{C_{11} \cdot C_{12}\} \text{ and } E(M_2) = \{C_{21} \cdot C_{22}\}$$

Then the ciphertext are aggregated as

$$C_1 = C_{11} + C_{21} \text{ and } C_2 = C_{12} + C_{22}$$

The decryption of ciphertext  $C_1$  and  $C_2$  gives an aggregated plaintext  $M_1 + M_2$  on  $E(\mathbb{F}_p)$ .

*Decryption D:*

The ciphertext can be decrypted as

$$D(C) = C_2 - s_k \cdot C_1 = M$$

### Secret Sharing

Secret sharing [19] is a cryptographic technique for building secure authentication protocols. If  $s$  is a secret which is divided among  $n$  sensor nodes, then  $s$  can be recovered only when all  $n$  sensor nodes contribute. For this  $n-1$  independent random numbers  $ss_i$ ,  $1 \leq i \leq n-1$  are generated such that distributing one  $ss_i$  to each node while  $s - \sum_{i=1}^{n-1} ss_i$  is given to last sensor node. The value of  $ss_i$  is called a secret share. Thus the original secret  $s$  is equal to  $\sum_{i=1}^n ss_i$ . An adversary cannot compute the original secret  $s$  without knowing the secret shares of other nodes in the network. Since the share of an unauthorized node provides no useful information to breach the security, this secret sharing technique is highly secure from the perspective of information-theory.

## 4.2 Efficient and Robust Secure In-Network Aggregation Protocol

The proposed protocol performs secure en-route aggregation and ensures data privacy and protects integrity of the aggregated data. In addition, it ensures source authentication. To reduce the bandwidth consumption, the proposed scheme performs the costly decryption and integrity verification operations at the trusted and powerful base station. The symbols used in the protocol are depicted in Table 1. The proposed protocol consists of 5 phases: Setup phase, Initialization phase, Encryption phase, Aggregation phase and Verification phase.

### 1. Setup phase:

Using the EC-ElGamal cryptosystem, a key pair  $(p_k, s_k)$  is generated by the base station (sink node). Also, the base station generates random keys  $k_1, k_2, \dots, k_n$ . Finally, BS broadcasts the query to the system.

### 2. Initialization phase:

Every sensor  $S_i$  first generates its data value  $d_i$ . Then it calculates the secret share  $ss_i = \text{HM}_3(k_i)$ .  $\text{HM}_3(\cdot)$  is the HMAC PRF that uses SHA-3. Subsequently,  $S_i$  combines the data value  $d_i$  with security information  $ss_i$  to produce a message  $m_i$  as in Fig. 2.

### 3. Encryption phase:

This phase is executed when a sensor node  $i$  decides to send its sensed data. Each sensor (leaf) node  $S_i$ , encrypts  $m_i$  using the public key  $(p_k)$  of the base station to create a partial state record,  $\text{PSR}_i$ .

**Table 1.** Symbols used in ERSIA protocol

Symbol	Description
BS	Base station
S	Source (leaf node)
A	Aggregator (intermediate node)
n	Number of nodes in the network
$p_k$	Public key of the base station
$s_k$	Private key of the base station
PRF	Pseudo random function
$k_i$	Key known to BS and $S_i$
s	Secret to be verified at the base station
$ss_i$	Secret share generated by $S_i$
$d_i$	Data value generated by $S_i$
$m_i$	Plaintext of $S_i$ to be encrypted
x	Number of child nodes of intermediate node
E	Encryption function
D	Decryption function
$\oplus$	Addition operator
$PSR_i$	Partial state record generated by $S_i$
$HM_3(.)$	HMAC implemented with SHA-3

$$PSR_i = Ep_k(m_i) \quad 1 \leq i \leq n$$

$S_i$  sends  $PSR_i$  to its parent (aggregator) node. By encrypting the message, this phase ensures data confidentiality.

#### 4. Aggregation phase:

This procedure is triggered after the aggregator  $j$  gathers all PSR of its child nodes. This phase combines all PSRs into a single PSR as

$$PSR_j = \oplus_{i=1}^x PSR_i \quad 1 \leq x, j \leq n$$

Then the aggregator sends the result to BS.

#### 5. Verification phase:

Finally, BS receives a single PSR and using its private key, it decrypts the aggregated result.

$$m_f = D_{s_k}(PSR)$$

$m_f$  is the final decrypted message consisting of both the result and the secret. The first 4 bytes of  $m_f$  represent the result of SUM query and the remaining bytes represent the secret  $s = \sum_{i=1}^n ss_i$ . BS extracts the result and  $s$  separately. Next, it computes  $ss_i$  for each source  $S_i$  using  $HM_3(.)$  and calculates  $\sum_{i=1}^n ss_i$ .

$d_i$ (4)	$00\dots0$ ( $\log_2 n$ )	$ss_i$ ( $20^i$ )
--------------	------------------------------	----------------------

**Fig. 2.** Format of  $m_i$  in ERSIA: the size of each field in bytes is given in parenthesis. The reason for adding zeros in  $m_i$  is to avoid overflow caused by summation of  $n$  numbers. The additional bits required for adding  $n$  number is  $\log_2 n$  (up to 8 bytes). Hence,  $\log_2 n$  zeros are padded before  $ss_i$  in every  $m_i$ .

The base station validates the integrity of message by comparing  $\sum_{i=1}^n ss_i$  with the secret  $s$  extracted from  $m_f$ . If both are equal, it ensures integrity. By ensuring integrity, ERSIA guaranteed for accurate and secure data aggregation.

## 5 Security Analysis

In this section, we analyze the security strength of the proposed protocol. The security analysis covers robustness against some cryptographic attacks [20] namely ciphertext analysis, known plaintext attack, malleability, node capture attacks and impersonation.

### Ciphertext analysis

Ciphertext analysis interprets the gathered ciphertext to obtain information such as key, plaintext and statistical information. The proposed protocol uses asymmetric-key based EC-EIGamal cryptosystem for encrypting the messages. The probabilistic nature of this scheme assures robustness and security against ciphertext analysis since the ciphertext is produced in a highly random way.

### Known plaintext attack

In a known plaintext attack, an adversary tries to obtain plaintext for a ciphertext. For any asymmetric-key based cryptosystems, the availability of public key to all the nodes in the network enables anyone can generate plaintext-ciphertext pairs. Since our proposed approach uses EC-EIGamal, security is provided against this attack.

### Malleability

Cryptosystems supporting privacy homomorphism are inherently malleable. As EC-EIGamal cryptosystem is used in ERSIA, such an attack can be performed very well. Since the proposed protocol also uses secret sharing technique, integrity of the aggregated data is protected against malicious adversaries.

### Node capture attack

To encrypt the sensed readings, ERSIA employs EC-EIGamal cryptosystem and decryption is done only at the sink node. Even if an adversary compromises nodes, he/she cannot extract or comprehend the information from the obtained ciphertext. Hence, the privacy of the ciphertext at aggregator nodes is ensured when there exist node compromise attacks.



## Impersonation

An adversary could impersonate the base station and broadcast a false query to the sources. Since the actual aggregation process is not altered, the BS accepts the final result as correct one. However, the proposed protocol employs  $\mu$ TESLA protocol which ensures that each source verifies that the query indeed originated from BS. Hence ERSIA is secure and robust against querier impersonation.

## 6 Overhead Analysis

In this section, we compare the bandwidth consumption of ERSIA protocol with No-Aggregation (transmitting individual data packets), Hop-by-hop SDA [4] and End-to-end SDA [11]. For comparing the above schemes, we consider the network model discussed in Sect. 3.1. We calculate the number of bits transmitted for a balanced ternary tree with 7 levels.

**Table 2.** Comparison of bandwidth consumption of SDA schemes

Level	Nodes	All nodes responding				30% nodes not responding		
		No-aggregation	Hop-by-hop SDA	End-to-end SDA	ERSIA	Hop-by-hop SDA	End-to-end SDA	ERSIA
1	3	45927	73	75	475	72	3315	465
2	9	15309	71	75	475	70	1117	467
3	27	5103	69	75	475	69	422	468
4	81	1701	68	75	475	67	172	470
5	243	567	66	75	475	66	108	471
6	729	189	64	75	475	64	85	473
7	2187	63	62	75	475	61	52	475

### Communication overhead

We calculate the number of bits sent by a node as per the packet format of TinyOS [21]. In TinyOS, the size of packet header (hdr) is 56 bits and that of data payload is 232 bits. The communication overhead of ERSIA protocol and others are calculated in 2 different scenarios: (1) Number of bits sent when all nodes are responding (2) Number of bits sent when some percentage of nodes are not responding. Table 2 shows the bandwidth consumption of SDA schemes in terms of number of bits transmitted.

In No-Aggregation method, a node just needs  $\log_2(t)$  bits to encode its message. This is shown in Fig. 3. In this method, all leaf nodes encrypt their sensed data and forwarded the ciphertext to their corresponding parent node. After receiving the encrypted packets from their children, each intermediate node forward the packets transmitted by their children without performing any aggregation function. As a result the number of bits transmitted increases exponentially with levels. Although, this method ensures end-to-end privacy, it wastes bandwidth since aggregation does not take place. Moreover, the nodes nearer to the base station have the risk of reducing the

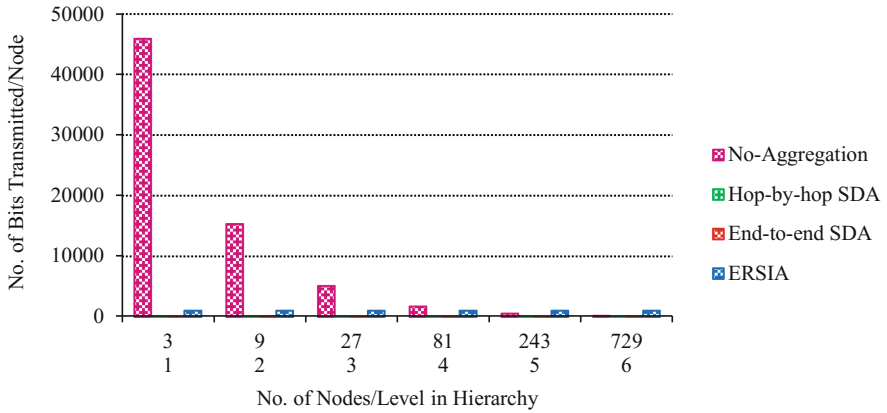


Fig. 3. Communication overhead

life of the nodes because of transmission of large number of bits. Such transmissions need more energy and the energy available in the battery is depleted at a faster rate. Also these nodes have the high probability of being attacked by the adversaries as they know all the aggregated data.

In hop-by-hop SDA method, the number of bits sent by a node depends on its level in the topology tree. As in No-Aggregation, leaf nodes transmit  $\log_2(t)$  bits whereas intermediate nodes receive aggregate data and thus send more number of bits. This method consumes more battery as the data is decrypted, aggregated and encrypted at each intermediate node but still the bandwidth consumption is less when compared to other schemes. This is shown in Fig. 4. Even though the communication overhead is less, there is an increase in the security infringements as the data is decrypted, processed, encrypted and forwarded at aggregator nodes. Especially, the node nearer to the sink would be hacked by the outside adversary as it knows all the aggregated data.

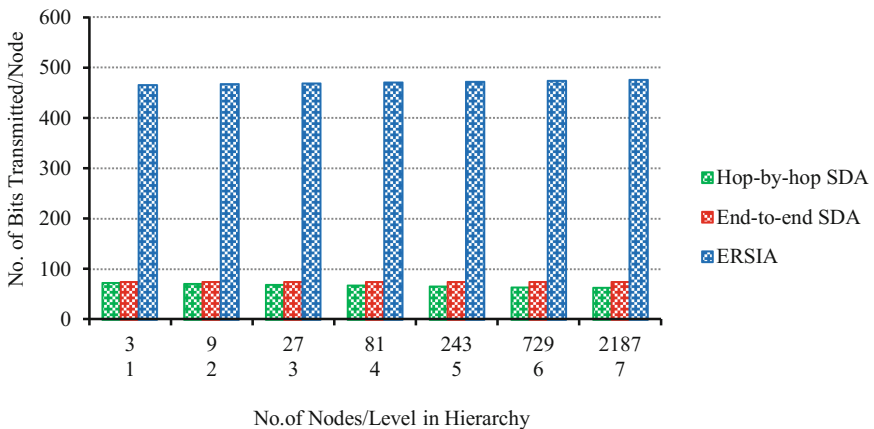


Fig. 4. Communication overhead (all nodes respond)

In end-to-end SDA, the modulus (M) size decides the maximum number of bits (i.e., maximum aggregate value) that can be transmitted by a node. For SUM aggregation function, each node transmits  $\log_2(t) + \log_2(n)$  bits which is  $\log_2(M)$  bits and remains constant. This method offers end-to-end privacy, but the added security features increases communication overhead than hop-by-hop method.

The bandwidth consumption of ERSIA is considerably high compared to other protocols. However, it offers much stronger security. The additional security features increases the bandwidth consumption in order to protect the integrity of the aggregated data. It ensures both end-to-end privacy and end-to-end integrity. ERSIA uses EC-EIGamal [18] cryptosystem for encrypting the sensor messages. It requires 160 bit key size for providing the security equivalent to RSA 1024 bit key size. Although, the messages expansion ratio of EC-EIGamal is 4-to-1, the ciphertext can be represented using fewer bits than asymmetric-key based cryptosystems. Furthermore, point compression techniques [22] reduce the ciphertext size. The total number of bits required for sending aggregated message payload is 363 bits. This results in 2 packets and therefore the total transmission cost is 475 bits ( $160 + 256 + 2 * 56$ ).

Figure 5 shows the communication overhead of the protocols when 30% nodes are not responding as in Castelluccia et al. [4, 11]. Since only the leaf nodes are sensors, we assume that the load of non-responding nodes is evenly distributed among all nodes. Therefore, the number of non-responding nodes that affects the communication cost is also taken as a parameter. The hop-by-hop method shows a very small increase in the number of bits sent by a node, whereas the performance of end-to-end method degrades significantly when there are non-responding nodes. This is because the ID's of non-responding nodes are appended to the aggregate. In end-to-end SDA, the number of non-responding nodes increases as we move up in the tree. However, ERSIA has constant overhead and ensures security objectives with reduced bandwidth consumption.

Besides communication cost, another factor that affects the performance of WSNs is the computation cost. Since WSNs consist of resource constrained sensor nodes, the

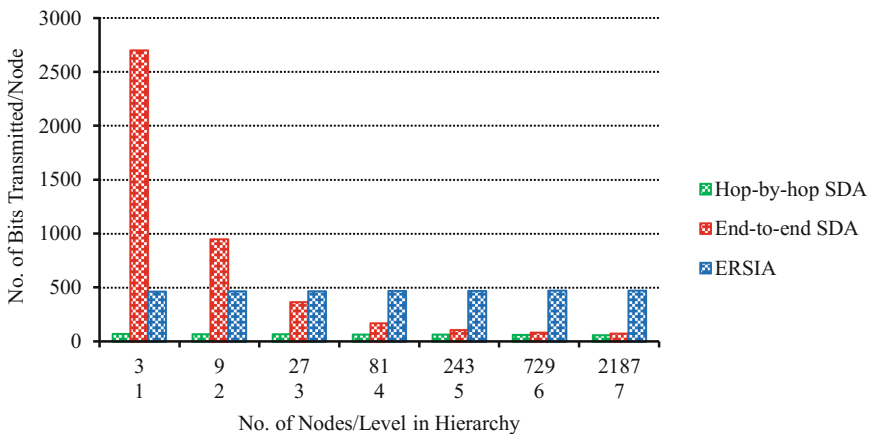


Fig. 5. Communication overhead (30% nodes not respond)

effective implementation of asymmetric key based cryptosystems is viable [23, 24]. As the energy needed for computation is negligible when compared to communication, the implementation of the proposed protocol is viable.

## 7 Conclusion

Performing efficient in-network aggregation while preserving data confidentiality and data integrity is a great challenge in wireless sensor networks. In this paper, we proposed an Efficient and Robust Secure In-network Aggregation (ERSIA) protocol for additive aggregation function. It combines privacy homomorphism and secret sharing to ensure data confidentiality, data integrity and source authentication. Security analysis reveals that the proposed protocol is secure and robust against ciphertext analysis, known plaintext attack, malleability and node capture attacks. Also, we compared the bandwidth consumption of the proposed protocol with prevailing secure in-network aggregation protocols. The results show that the proposed approach outperforms other existing techniques and provides end-to-end security thereby increasing the life time of the network. By changing the format of the message, the proposed protocol can be adopted to handle other sum based queries. A major limitation of the proposed protocol is that it cannot verify the integrity of aggregated data at intermediate node. Currently, we are extending our work to protecting integrity of aggregated data at intermediate nodes. For future work, we are interested in adopting these works for mobile sensor networks.

**Acknowledgements.** This research is supported by Visvesvaraya PhD Scheme for Electronics & IT, Ministry of Electronics and Information Technology, Government of India.

## References

1. Rehana, J.: Security of wireless sensor network. In: Seminar on Internetworking (2009)
2. Kafetzoglou, S., Papavassiliou, S.: Energy-efficient framework for data gathering in wireless sensor networks via the combination of sleeping MAC and data aggregation strategies. *Int. J. Sens. Netw.* **10**(1–2), 3–13 (2011)
3. Chen, X., Makki, K., Yen, K., Pissinou, N.: Sensor network security: a survey. *IEEE Commun. Surv. Tutor.* **11**(2), 52–73 (2009)
4. Castelluccia, C., Mykletun, E., Tsudik, G.: Efficient aggregation of encrypted data in wireless sensor networks. In: The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous), pp. 109–117. IEEE (2005)
5. Yang, Y., Wang, X., Zhu, S., Cao, G.: SDAP: a secure hop-by-hop data aggregation protocol for sensor networks. *ACM Trans. Inf. System Secur. (TISSEC)* **11**(4), 18 (2008)
6. Prakash, G.L., Thejaswini, M., Manjula, S.H., Venugopal, K.R., Patnaik, L.M.: Secure data aggregation using clusters in sensor networks. *World Acad. Sci. Eng. Technol.* **3**(3), 496–501 (2009)
7. Nath, S., Yu, H., Chan, H.: Secure outsourced aggregation via one-way chains. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, pp. 31–44. ACM (2009)

8. Poornima, A.S., Amberker, B.B.: SEEDA: secure end-to-end data aggregation in wireless sensor networks. In: Seventh International Conference on Wireless and Optical Communications Networks (WOCN), pp. 1–5. IEEE (2010)
9. Liu, C.X., Liu, Y., Zhang, Z.J., Cheng, Z.Y.: High energy-efficient and privacy-preserving secure data aggregation for wireless sensor networks. *Int. J. Commun Syst* **26**(3), 380–394 (2013)
10. Alghamdi, W.Y., Wu, H., Kanhere, S.S.: Reliable and secure end-to-end data aggregation using secret sharing in WSNs. In: Wireless Communications and Networking Conference, WCNC 2017, pp. 1–6. IEEE (2017)
11. Castelluccia, C., Chan, A.C., Mykletun, E., Tsudik, G.: Efficient and provably secure aggregation of encrypted data in wireless sensor networks. *ACM Trans. Sens. Netw. (TOSN)* **5**(3), 20 (2009)
12. Peter, S., Westhoff, D., Castelluccia, C.: A survey on the encryption of convergecast traffic with in-network processing. *IEEE Trans. Dependable Secure Comput.* **7**(1), 20–34 (2010)
13. Rivest, R.L., Adleman, L., Dertouzos, M.L.: On data banks and privacy homomorphisms. *Found. Secure Comput.* **4**(11), 169–180 (1978)
14. Perrig, A., Szewczyk, R., Wen, V., Culler, D., Tygar, J.D.: SPINS: security protocols for sensor networks, In: Proceedings of Mobile Computing and Networking (2001)
15. Papadopoulos, S., Kiayias, A., Papadias, D.: Exact in-network aggregation with integrity and confidentiality. *IEEE Trans. Knowl. Data Eng.* **24**(10), 1760–1773 (2012)
16. Othman, S.B., Bahattab, A.A., Trad, A., Youssef, H.: Confidentiality and integrity for data aggregation in WSN using homomorphic encryption. *Wirel. Pers. Commun.* **80**(2), 867–889 (2015)
17. Peter, S., Westhoff, D., Castelluccia, C.: A survey on the encryption of convergecast traffic with in-network processing. *IEEE Trans. Dependable Secure Comput.* **7**(1), 20–34 (2010)
18. Koblitz, N.: Elliptic curve cryptosystems. *Math. Comput.* **48**(177), 203–209 (1987)
19. Menezes, A.J., Van Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press, Boca Raton (1996)
20. Ozdemir, S., Xiao, Y.: Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. *Comput. Netw.* **55**(8), 1735–1746 (2011)
21. Karlof, C., Sastry, N., Wagner, D.: TinySec: a link layer security architecture for wireless sensor networks. In: Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, pp. 162–175. ACM (2004)
22. Hoffstein, J., Pipher, J.C., Silverman, J.H., Silverman, J.H.: An Introduction to Mathematical Cryptography, vol. 1. Springer, New York (2008). <https://doi.org/10.1007/978-0-387-77993-5>
23. Malan, D.J., Welsh, M., Smith, M.D.: A public-key infrastructure for key distribution in TinyOS based on elliptic curve cryptography. In: First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON 2004), pp. 71–80. IEEE (2004)
24. Gura, N., Patel, A., Wander, A., Eberle, H., Shantz, S.C.: Comparing elliptic curve cryptography and RSA on 8-bit CPUs. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 119–132. Springer, Heidelberg (2004). [https://doi.org/10.1007/978-3-540-28632-5\\_9](https://doi.org/10.1007/978-3-540-28632-5_9)

# Context-Aware Conditional Probabilistic Hyper-exponential Reputation Technique for Mitigating Byzantine Attacks

Geetha Achuthan<sup>1</sup>(✉), Sreenath Niladuri<sup>1</sup>, and A. G. Sareeka<sup>2</sup>

<sup>1</sup> Pondicherry Engineering College, Puducherry, India  
gachuthansareeka@gmail.com

<sup>2</sup> College of Engineering, Anna University, Chennai, India

**Abstract.** Data delivery in ad hoc networks requires co-operation as the predominant factor for ensuring pause-less communication between the nodes due to their self-organizing nature. This decentralized feature of ad hoc networks enables adversaries to compromise nodes through Byzantine attacks and induce them to act maliciously for degrading the performance of the network. These Byzantine nodes degrade the performance by selectively dropping the packets, modifying the packets or dropping the packets without forwarding for the sake of conserving their own energy. In this paper, we propose a Context-aware Conditional Probabilistic Hyper-exponential Reputation Scheme (CCPHRT) for improving the efficiency and effectiveness in detecting byzantine nodes. This CCPHRT mitigates byzantine nodes based on Hyper-exponential Conditional Survivability Factor (HCSF) as each mobile node's behavioural transition is hypo-exponentially distributed as the inter-transition time is independent and exponentially modeled. Extensive simulations are performed and the obtained results prove that CCPHRT is effective in improving the performance of the network than the compared techniques used for investigation. CCPHRT also facilitates a dynamic byzantine node isolation degree of 34% which is better than the baseline byzantine mitigation techniques used for study.

**Keywords:** Hyper-exponential · Conditional probability · Byzantine nodes  
Co-operation

## 1 Introduction

In MANET, every mobile node highly depends on their neighbouring nodes for their reliable data dissemination. In this category of network, the mobile nodes arbitrary movement makes the topology of network to be highly dynamic [1]. This dynamic topology of ad hoc network enforces the mobile nodes to interact directly or through intermediate nodes depending on the location pertaining to the communication radius [2] the extent of participation attributed by a node relies on the determination of reputation parameter that refers to its coordination towards reliable routing process [3–5]. However, some group of nodes do not participate in the routing activity are known as malicious nodes and such classes of mobile nodes drastically decreases the network performance [6].

The routing process in MANET must be effective enough in mitigating both internal and external attacks which arises due to non-cooperative and compromised nodes for confirming reliable data communication [7]. Researchers have contributed diversified techniques for mitigating attacks such as flooding attack, wormhole attack, rushing attack, selfish node etc., in MANET [8]. Some of the approaches are mainly based on cryptographic analysis, node trust factor based mechanism etc. One such cryptographic technique is that group key management technique which confirms the authorization against external attacks but fails to work properly against internal attacks like byzantine behaviour of mobile nodes. Byzantine nodes are conformed to be the most vulnerable since they crumble the network potential by introducing higher degree of re-transmissions by packet dropping.

The proposed CCPHRT focuses on mitigating byzantine mobile nodes with the aid of AODV protocol in which the data transmission between the source and the destination depends on the survivable parameter of the mobile nodes and distinct routes discovered in the network. CCPHRT determines Hyper-exponential Conditional Survivability Factor (HCSF) to estimate the extent of co-ordination maintained between the mobile nodes in the event of routing. CCPHRT is a distributed mitigation approach that aims in maintaining the balance in throughput degradation and packet drop.

The remaining part of the paper is organized as follows. Section 2 provides a brief about the related previous works including the compared byzantine mitigation techniques. Section 3 explains the proposed Context-Aware Conditional Probabilistic Hyper-Exponential Reputation Technique with its related algorithm. Section 4 details the performance evaluation of CCPHRT including performance metrics used, simulation environment, results and discussions and finally Sect. 5 concludes the paper.

## 2 Related Work

In this section, the major types of byzantine node detection techniques (DT) which analyzes the survivability of MANETs due to various kinds of byzantine attacks are presented. Out of it trust based detection techniques [9–14] uses incentives/credits to motivate the mobile nodes to effectively participate in routing process without involving in any malicious activity, punishment is given to nodes which involve in any misbehaviour. This kind of approach is not advisable to be used in MANET, where there is no possibility of having a central control authority to provide incentive/punishment.

Token based techniques rely upon distributed servers [15–18] for ensuring the security of ad hoc networks. The major disadvantage of this technique is if the server is attacked by a byzantine attacker, then the overall survivability and reliability of the network will be in risk.

Secure Routing Mechanisms utilize cryptographic algorithms [19–21]. Cryptographic algorithms are efficient in handling outsider attacks by ensuring authentication and authorization of participating mobile nodes in the network. Whereas they are not effective in handling insider attacks such as byzantine attacks which are performed by the authenticated mobile nodes of the network.

Acknowledgement-based mechanisms utilize a feedback mechanism [22, 23] which consumes lot of energy resource which is one of the major constraint in MANETs where all the participating nodes functions using their limited battery power.

Reputation-based Detection Techniques [24] are more efficient in handling the byzantine attacks. Past-history based reputation techniques [25–30] estimate the reputation of the nodes using first hand/second hand/centralized observations based on their past history. As the wireless ad hoc networks topology is highly dynamic in nature, there are high chances of false positive or false negative observations.

Bayesian conditional probabilistic reputation technique ARDBD [31] uses Bayes probability based factor for reputation computation, it uses first hand and centralized observation method of monitoring which induces overhead. Whereas NADBD [37] use improved Bayes probability factor for reputation computation, it uses first and second hand observation method of monitoring. The major drawback of it is it relies only on probe packets.

The mitigation techniques compared with proposed CCPHRT are Reputation Trust Factor-based Mitigation Scheme (RTFMS) [32], Dempster Shafer-Based Mitigation Scheme (DSBMS) [33] and Bayesian Probability-Based Mitigation Scheme (BPBMS) [34]. RTFMS is a Reliable Trust Factor-based Mitigation Technique that has been proposed for detecting byzantine nodes based on past performance of the nodes quantified using their packet forwarding efficiency. This RTFMS technique uses Gwet Kappa factor for estimating the performance of the mobile nodes under routing based on its contribution rendered to the other neighbouring nodes. The factors considered for detection in RTFS is only continuous in nature and hence it fails to consider discrete parameter into account during detection. The throughput and packet delivery ratio of RTFS is improved as it used multiple dynamic rules for detecting byzantine nodes. Further, Dempster Shafer-Based Mitigation Scheme (DSBMS) was proposed for combining evidences based on Dempster Shafer Theory for integrating evidences that could initiate predominant detection rate. The rules generated by DSBMS are dynamically optimized periodically depending on the number of packets needed to be forwarded by each interacting mobile nodes. In addition, Bayesian Probability-Based Mitigation Scheme (BPBMS) was proposed for detecting byzantine nodes based on Bayes theorem that helps in estimating the conditional probability of packet forwarding contributed by the mobile nodes. In BPBMS, the packet forwarding capability of mobile nodes are estimated only based on discrete parameter and ignores continuous parameters of influence that impacts on byzantine node detection. RTFMS, DSBMS and BPBMS possess the limitations of increased detection time and false positive rate since they fail to optimally combine maximum number of impactful parameters during byzantine node detection.

The specific contributions of CCPHRT are:

- To propose a mitigation mechanism that isolates byzantine nodes by estimating the degree of co-operation by using Hyper-exponential Conditional probabilistic Survivability Factor (HCSF).
- To develop a reputation scheme for mitigating byzantine for maintaining reliable network connectivity for enhancing resilience.
- To investigate and resolve the impacts produced by the byzantine nodes on the performance of the network and also to study the efficiency of the methodology.



### 3 Context-Aware Conditional Probabilistic Hyper-exponential Reputation Technique

Context-aware Conditional Probabilistic Hyper-exponential Reputation Scheme (CCPHRT) is proposed for facilitating the detection of byzantine nodes by computing Hyper-exponential Conditional Survivability Factor (HCSF). The computation of HCSF, a conditional probability-based trust parameter is carried out by integrating discrete and continuous time factors of routing. The obtained values are tabulated in Table 1.

**Table 1.** Identification of HCSF threshold of CCPHRT

Byzantine mitigation techniques	Detection range (0.36–0.40)	Detection range (0.30–0.35)
CCPHRT	24	30
RTFMS	22	25
DSBMS	19	21
BPBMS	16	19

In this scheme, an ad hoc network that comprises of ‘r’ routing path between the source and the destination is considered for implementing and investigating impacts of CCPHRT in byzantine behaviour detection. Each route ‘i’ of the network consist of different number of mobile nodes and the data delivery relies on the survivable factor of mobile nodes and its existing distinct routing paths. This lifetime of mobile node and routing path are considered to be exponentially distributed with parameter ‘ $\lambda_i$ ’ since it aids in estimating the lifetime of the specific route enabled for data dissemination. Here, ‘y’ and ‘x’ are considered as the lifetime of the path and the mobile node, respectively. The parameter ‘ $\lambda_i$ ’ pertains to the packet relying rate of each node in the distinct routing paths given by (1)

$$\lambda_i = \frac{P_{F-NEXT-HOP}}{P_{R-PREV-HOP}} \quad (1)$$

where ‘ $P_{F-NEXT-HOP}$ ’ and ‘ $P_{R-PREV-HOP}$ ’ refers to the total number of packets forwarded by a mobile node to its next-hop neighbour and total number of packets received from the predecessor hop node.

Then the probability density function of existence factor ( $f_{Y/X}(y/i)$ ) that highlights the possibility of each routing path ‘i’ to be trustworthy within its lifetime ‘y’ under high survivability of mobile nodes is given by Eq. (2)

$$f_{Y/X}(y/i) = \lambda_i e^{-\lambda_i y}, y > 0 \quad (2)$$

Further, the mobile node existing in a routing path is said to be survivable with probability  $P_{X(i)}$  at a particular instant of time ‘t’ by satisfying the constraint ( $\alpha_i$ ) which presents maximum threshold of survivability as presented in Eq. (3)

$$P_{X(i)} = \alpha_i \geq \lambda_i$$

$$\text{Where, } \sum_{i=1}^r \alpha_i = 1 \quad (3)$$

Furthermore, the integrated probability value  $f(i, y)$  which determines the survivable parameter of mobile nodes within the route stability 'y' is given by Eq. (4)

$$f(i, y) = f_{Y/X}(y/i)p_X(i) \quad (4)$$

The combined cumulative probability density function value ( $f_Y(y)$ ) related to the survivability of the network is determined using Eqs. (5), (6) and (7) as

$$f_Y(y) = \sum_{i=1}^r f(i, y) \quad (5)$$

$$= \sum_{i=1}^r \alpha_i f_{Y/X}(y/i) \quad (6)$$

$$= \sum_{i=1}^r \alpha_i \lambda_i e^{-\lambda_i y}, y > 0 \quad (7)$$

Finally, the quantified probability value determined using conditional probability-based resilience distribution of each routing path is expressed using Eqs. (8) or (9)

$$f_{Y/X}(y/i) = f_i(i, y) \quad (8)$$

Or

$$F_{Y/X}(y/i) = F_i(y) \quad (9)$$

The survivability factor (HCSF) for the participating mobile node under byzantine attack is computed using Eq. (9) based on Eq. (10)

$$HCSF = \sum_{i=1}^r \alpha_i F_i(y)$$

$$\text{Where, } f_i(y) = \sum_{i=1}^r \alpha_i F_i(y) \quad (10)$$

Based on the value of HCSF, the decision of isolating byzantine malicious nodes are performed rapidly during the routing activity. i.e., when the value of HCSF is less than 0.35, the byzantine nodes are isolated from the routing path [35]. HCSF also quantifies the conditional probabilistic factor that induces a mobile node to behave as a byzantine compromised node.

### 3.1 Algorithm – Computation of Hyper-exponential Conditional Survivability Factor

The formulated CCPHRT approach is implemented using the following algorithm.

**Notations:**

- N - Total number of nodes in the network
- GN - Group of nodes of the routing path
- SN - Source node in routing path
- DN - Destination node in routing path
- c - Each single session
- t - Number of sessions
- u - A node in GN
- $f_{Y/X}$  - Lifetime of mobile node participating in the network.
- $\lambda_i$  - Packet relaying rate of each mobile node.
- Y - Route stability
- $f_Y(y)$  - Resilience of the network
- k - Session
- $N_i$  - Mobile node
- i - Routing path
- x - Life time of each mobile node
- y - Life time of each routing path
- r - Total number of routing paths.

**Algorithm (Estimation of HCSF)**

1. Begin
2. For every mobile node ' $i$ ' in the network
3. For every session  $K = 1$  to  $k$  do
4. Compute conditional-probability based survivability of each node existing in the routing path at time 't' using
 
$$f_{Y/X}(y/i) = \lambda_i e^{-\lambda_i y}, y > 0$$
5. If ( $f_{Y/X}(y/i) > Y$ ) then
6. Determine probability density function of every node within the route stability 'y' using,
 
$$f(i, y) = f_{Y/X}(y/i) p_X(i)$$
6. If ( $f(i, y) \geq 0.4$ ) then
  7. Estimate survivability parameter of the network using
 
$$f_Y(y) = \sum_{i=1}^r \alpha_i \lambda_i e^{-\lambda_i y}, y > 0$$
  8. If ( $f_Y(y) > 0$ ) then,
  9. Estimate conditional survivability rate of each routing path (HCSF) as
 
$$HCSF = \sum_{i=1}^r \alpha_i f_i(y)$$
  10. If (HCSF (i)  $< 0.35$ ) then
  11.  $N_i$  is byzantine compromised
  12. Else
  13.  $N_i$  is cooperative
  14. End.

The quantified value of HCSF discriminates the behaviour of a particular node as either malicious compromised byzantine node or cooperative uncompromised node.

## 4 Performance Evaluation of CCPHRT

CCPHRT is comparatively analyzed with the existing Bayesian probabilistic mitigation mechanisms RTFMS, DSBMS and BPBMS. The performance investigation is also carried out based on the evaluation metrics Packet Delivery Ratio (PDR), throughput, control overhead, total overhead, delay and energy consumption under the influence of varying number of mobile nodes, number of byzantine attackers and number of source and destination pairs [36–38]. Performance metrics used, simulation set up used for comparative analysis, considered system and adversary models are detailed as follows.

### Performance metrics

- Packet Delivery Ratio: Ratio of the numbers of packets actually received by the destination to the total number of packets expected to be delivered at the destination.
- Throughput: Maximum number of data packets actually delivered in the destination node at a specified point of time.
- Control Overhead: Maximum bytes of packets that are essential for establishing end-to-end connectivity between the source and the destination nodes.
- Total Overhead: Ratio of the number of packets required for establishing end to end communication between the source and destination to the actual number of data packets received by the destination node.
- Energy consumption: Cumulative sum of energy utilized by the mobile nodes during the transmission, reception, idle and sleep states.

### Simulation Environment

Simulation experiments for CCPHRT, RTFMS, DSBMS and BPBMS are carried out using ns-2.32. The network contains of 100 mobile nodes distributed in a terrain size of  $1000 \times 1000$  meters. The packet size, channel capacity are 512 bytes, 2 Mbps respectively, constant bit rate (CBR) traffic model is used with 40 packets/sec. Further, each mobile node is made to contain 100 joules (J) of energy in the beginning and 10 joules (J) of energy is required for each time slot of communication. The MAC layer used for simulation is based on IEEE 802.11 distributed co-ordination function (DCF) with the two way ground channel propagation model that contains an interface queue at the MAC layer to hold up to 50 packets. Each mobile node transfers data packets varying from 4 packets to 40 packets with a maximum speed of 10 m/s.

In this model, we define the neighbor of each node  $N_i$  as a node that resides within the radio transmission range and the number of neighbors varies from time to time. Each node observes the abnormal behaviours that its neighbours conduct using first hand and second hand observations and updates the reputation details in its local storage. Both past and current behaviour of the nodes are taken into consideration for determining its byzantine nature. When a node needs to summarize its observation and

thereby form its local view of misbehaving nodes, it calculates the rate of abnormal behaviors over the overall behaviors it has observed for the node. For instance, if the packet drop behaviour of a node is to be observed, then packet drop rate (PDR) and node's residual energy (to make sure the node is byzantine and is not selfish node) is considered. Lifetime of mobile nodes and lifetime of routing paths are the factors that are considered for byzantine node detection and isolation.

We assume that the adversary aims to disrupt network operations by conducting a variety of misbehaviors such as packet dropping, and deliberate propagation of fake observations regarding the behaviour of other nodes. The adversary may alter the ratio of misbehaviour over time, and it can carry out the set of misbehaviors for any arbitrary length of time. The adversary is able to mix its misbehaviors at any ratio if it chooses to conduct multiple misbehaviors at the same time period.

The proposed and benchmarked byzantine detection schemes are implemented under the similar simulation set-up by utilizing probe packets to collect packet forwarding capability of each mobile node under routing. The probe packets are not sent with acknowledgment packets in order to reduce overhead in data routing. Sufficient numbers of control packets with re-transmissions were used for facilitating maximum rate of node discovery during packet forwarding. Selfish property of mobile nodes that contribute to byzantine behaviour of mobile nodes is also analyzed during the detection process.

## Results and Discussions

The performance investigation of CCPHRT is analyzed using four experiments, viz., (a) Performance analysis of CCPHRT by varying the number of mobile nodes (b) Performance analysis of CCPHRT by varying the number of byzantine nodes and (c) Performance analysis of CCPHRT by varying number of CBR connections.

### Identification of conditional probabilistic threshold

The following Table 1 portrays the HCSF threshold values identified during simulation by varying the detection range set for byzantine behaviour detection.

Initially, the performance of CCPHRT is studied by varying the number of mobile nodes of the network. In this experimental analysis, nearly 35% of the mobile nodes are considered as byzantine compromised. The throughput, total overhead and energy consumption obtained from experiment-1 are portrayed using Figs. 1, 2 and 3 respectively. Figure 1 transparently proves the throughput of the implemented CCPHRT approach. The decrease in performance of existing RTFMS, DSBMS and BPBMS approaches clearly specifies that they fail to handle the rate of data packets transfer with increase in the number of mobile nodes. But CCPHRT improves throughput of the network than the considered baseline mitigation mechanisms by integrating the impact of discrete and continuous changing behaviour of mobile nodes. CCPHRT improves PDR by 6%–8% over RTFMS, 10%–12% over DSBMS and 17%–19% over BPBMS. It is clear that CCPHRT improves the throughput by 5%–8% over RTFMS, 9%–11% over DSBMS and 17%–19% over BPBMS.

Figures 2 and 3 highlights the performance of CCPHRT based on total overhead and energy consumption. It is inferred that the total overhead and energy consumption of RTFMS, DSBMS and BPBMS increases with re-transmissions of packets that arises due to packet dropping behaviour of byzantine nodes that intentionally drops the

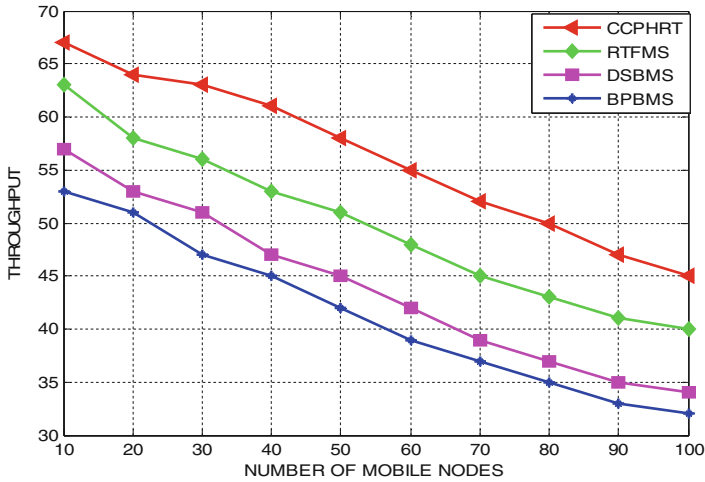


Fig. 1. Throughput vs. number of mobile nodes

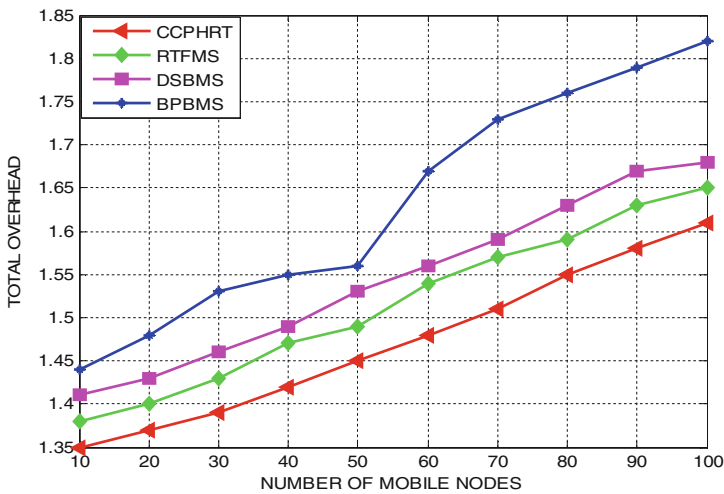


Fig. 2. Total overhead vs. number of mobile nodes

control packets and data packets from reaching the destination. But, as the byzantine nodes are eliminated from participating in routing, CCPHRT reduces the total overhead by 8%–11% over RTFMS, 13%–16% over DSBMS and 22%–25% over BPBMS. In general, the energy consumption proportionally increases with increase in the number of mobile nodes of the network as the energy consumed by each mobile node increases proportionally with respect to extra packets that gets introduced into the network. CCPHRT stabilizes the energy consumption by classifying mobile nodes based on their residual energy that induces improved detection and mitigation of byzantine nodes and

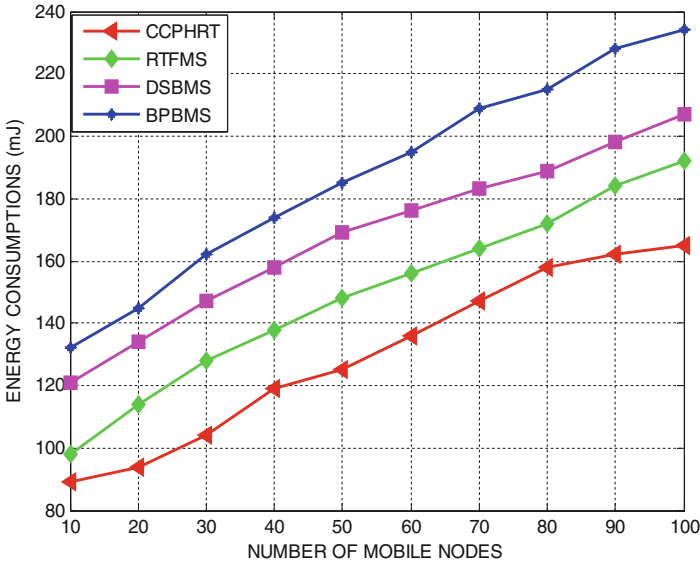


Fig. 3. Energy consumption vs. number of mobile nodes

thus it reduces the rate of energy consumption by 14%–16% over RTFMS, 18%–21% over DSBMS and 26%–28% over BPBMS.

Furthermore, the performance of CPHRT is compared with RTFMS, DSBMS and BPBMS based on different number of byzantine nodes using PDR, control overhead and packet latency. Figure 4 highlights the achieved PDR of CPHRT with increase in the number of byzantine nodes that forbids maximum number of packets in reaching

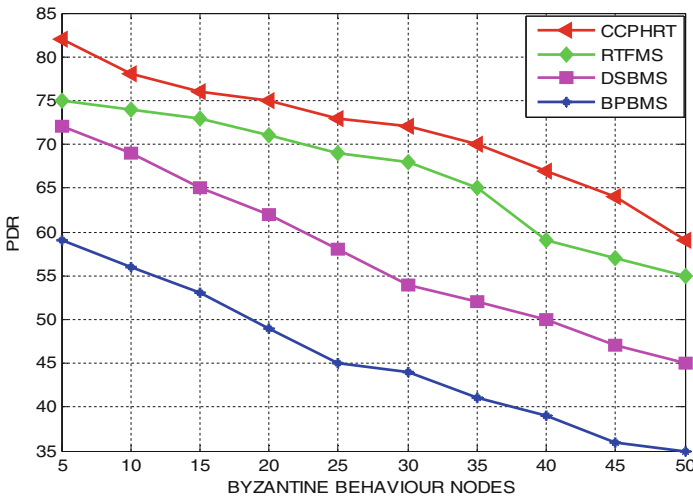
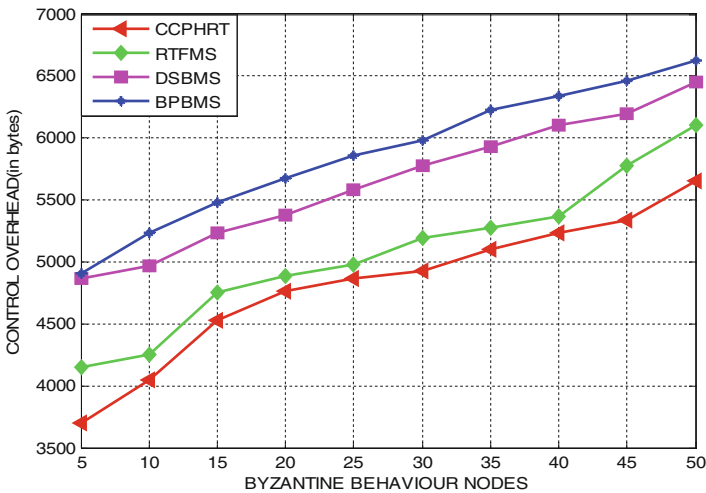


Fig. 4. PDR vs. number of byzantine nodes

the destination. CCPHRT improves PDR and throughput to a significant level by isolating byzantine nodes at a rapid rate of 16% superior to the baseline mitigation mechanisms considered for study.

CCPHRT improves the packet delivery ratio by 4%–6% over RTFMS, 9%–11% over DSBMS and 18%–20% over BPBMS. It also improves the throughput by 11%–13% over RTFMS, 15%–17% over DSBMS, 18%–20% over ARDBD and 22%–25% over BPBMS.

Figures 5 and 6 depict the performance of CCPHRT based on control overhead and packet latency. It is identified that increase in the number of byzantine nodes of the network induces the participating mobile nodes to re-transmit increased number of route request (RREQ) and route reply (RREP) packets for route discovery. CCPHRT balances the re-transmission of packets by establishing reliable routing paths that are capable in forwarding packets. It is inferred that CCPHRT reduces the control overhead by 12%–14% over RTFMS, 16%–18% over DSBMS and 25%–28% over BPBMS. It also reduces the packet latency by 9%–12% over RTFMS, 14%–17% over DSBMS and 22%–25% over BPBMS.



**Fig. 5.** Control overhead vs. number of byzantine nodes

Finally, the potential of CCPHRT is investigated with RTFMS, DSBMS and BPBMS by varying the number of CBR connections as presented in Figs. 7 and 8. Figure 7 confirms that the energy consumption of CCPHRT is better than RTFMS, DSBMS and BPBMS as it is advantageous in improving the detection rate even under the increase in CBR connections. The energy consumption of CCPHRT is reduced by 23%, 18% and 14% compared to RTFMS, DSBMS and BPBMS mitigation techniques.



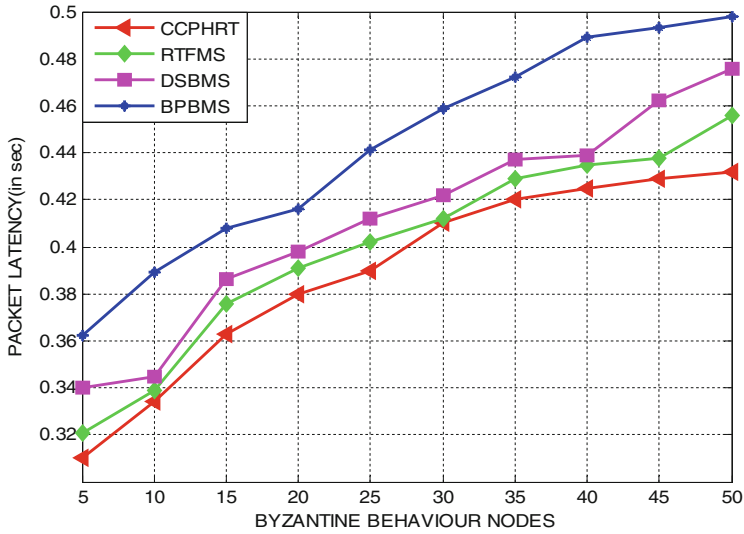


Fig. 6. Packet latency vs. number of byzantine nodes

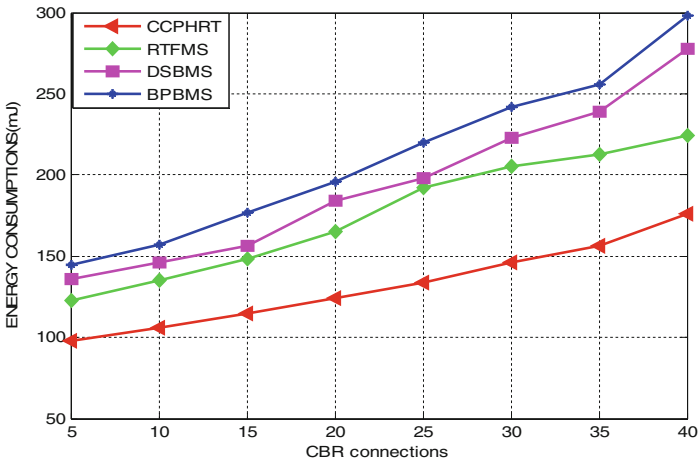


Fig. 7. Energy consumption vs. number of CBR connections

Figure 8 proves that routing overhead of CCPHRT is exceptionally reduced compared to RTFMS, DSBMS and BPBMS as it is potential in reducing the number of route discovery by facilitating minimized control packet re-transmission even when the number of CBR connections are monotonically varied. The routing overhead of CCPHRT is exceptionally reduced by 19%, 15% and 11% compared to RTFMS, DSBMS and BPBMS mitigation techniques.

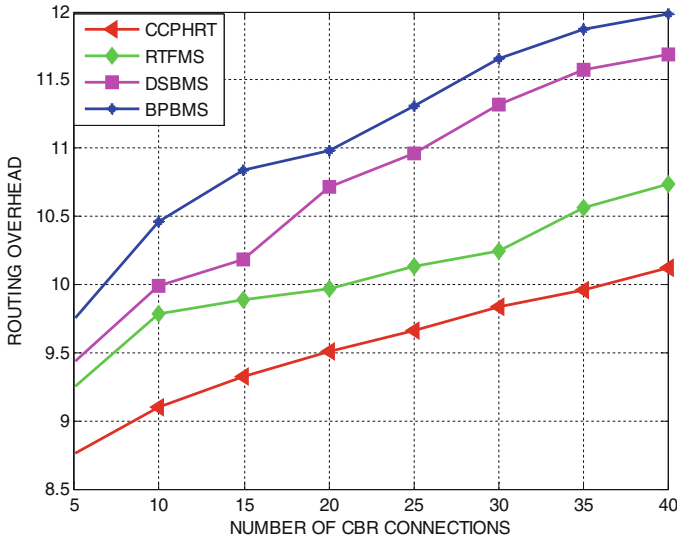


Fig. 8. Routing overhead vs. number of CBR connections

## 5 Conclusions

In this paper, CCPHRT was presented for mitigating byzantine nodes and ensuring the resilience of the network. The benefits of HCSF are investigated by integrating discrete and continuous behavioural parameters. The experimental results of CCPHRT outperform the existing probabilistic mitigation approaches in terms of throughput, control overhead, total overhead, packet delivery ratio, and energy consumption. CCPHRT is found to improve the byzantine node detection rate by 21% superior to the benchmark mitigation approaches considered for analysis. Further, CCPHRT frames a detection threshold point for realizing the severity of impact induced by the byzantine behaviour of mobile nodes. In addition, CCPHRT is also found to outperform the existing conditional probabilistic standard approaches in terms of packet latency and communication overhead.

## References

1. Xing, F., Wang, W.: On the survivability of wireless ad hoc networks with node misbehaviors and failures. *IEEE Trans. Dependable Secur. Comput.* **7**(3), 284–299 (2010)
2. Md. Akhtar, A.K., Sahoo, G.: Mathematical model for the detection of byzantine nodes in MANETs. *Int. J. Comput. Sci. Inform.* **1**(3), 25–28 (2008)
3. Buchegger, S., Boudec, J.Y.: Nodes bearing grudges: towards routing security, fairness and robustness in mobile ad-hoc network. Presented at Tenth Euromicro workshop on Parallel, Distributed and Network based Processing, Canary Islands, Spain (2002)
4. Marti, S., Giuli, T.J., Lai, K., Baker, M.: Mitigating routing misbehaviour in mobile ad hoc networks. *Mob. Comput. Netw.* **1**(1), 255–265 (2000)

5. Corradi, G., Janssen, J., Manca, R.: Numerical treatment of homogenous semi-Markov processes in transient case—a straightforward approach. *Methodol. Comput. Appl. Probab.* **6**, 233–246 (2004)
6. Sundarajan, T., Shanmugam, A.: Modeling the behavior of byzantine forwarding nodes to simulate cooperation in MANET. *Int. J.* **2**(2), 147–160 (2010)
7. Xing, F., Wang, W.: Modeling and analysis of connectivity in mobile ad hoc networks with misbehaving nodes. In: *Proceedings of IEEE International Conference on Communications*, vol. 4, no. 3, pp. 1879–1884 (2006)
8. Cárdenas, A.A., Radosavac, S., Baras, J.S.: Evaluation of detection algorithms for MAC layer misbehaviour: theory and experiments. *IEEE Trans. Netw.* **17**(2), 605–617 (2009)
9. Buttyan, L., Hubaux, J.P.: Self-organized public-key management for mobile ad hoc networks. *IEEE Trans. Mob. Comput.* **2**(1), 52–64 (2003)
10. Raghavan, B., Snoeren, A.C.: Priority forwarding in ad hoc networks with self-interested parties. In: *Proceedings of Workshop on Economics of Peer-to-Peer Systems*, vol. 2, no. 3, pp. 34–46 (2003)
11. Meka, H., Madria, S., Linderman, M.: Incentive based approach to byzantine nodes in mobile P2P networks. In: *Proceedings of IEEE 31st International Performance Computing and Communications Conference*, vol. 1, no. 6, pp. 352–359 (2012)
12. Huang, E., Crowcroft, J., Wassell, I.: Rethinking incentives for mobile ad hoc networks. In: *Proceedings of the ACM SIGCOMM Workshop on Practice and Theory of Incentives in Networked Systems*, vol. 2, no. 3, pp. 191–196 (2004)
13. Zhong, S., Chen, J., Yang, R.: Sprite: a simple, cheat-proof, credit-based system for mobile ad-hoc networks. In: *Proceedings of IEEE INFOCOM*, vol. 2, no. 5, pp. 1987–1997 (2012)
14. Demir, C., Comaniciu, C.: An auction based AODV protocol for mobile ad hoc networks with selfish nodes. In: *Proceedings of IEEE International Conference on Communications*, vol. 2, no. 7, pp. 3351–3356 (2007)
15. Zhou, H., Wu, J., Zhao, H., Tang, S., Chen, C., Chen, J.: Incentive-driven and freshness aware content dissemination in selfish opportunistic mobile networks. *IEEE Trans. Parallel Distrib. Syst.* **26**(9), 2493–2505 (2015)
16. Appenzeller, M.G., Baker, M.: User-friendly access control for public network ports. In: *Proceedings of IEEE INFOCOM*, vol. 1, no. 6, pp. 699–707 (1999)
17. Kathirvel, A., Srinivasan, R.: Single umpiring system for security of mobile ad hoc networks. *J. Adv. Wirel. Mob. Commun.* **2**(3), 46–54 (2009)
18. Sukumaran, S., Venkatesh, J.: Comparison of access control methods in mobile ad-hoc networks. In: *Proceedings of IEEE International Conference on Internet Multimedia Services Architecture and Applications*, vol. 4, no. 2, pp. 1–5 (2009)
19. Xu, Y., Scerri, P.: Token based resource sharing in heterogeneous multi-agent teams. In: Yang, J.-J., Yokoo, M., Ito, T., Jin, Z., Scerri, P. (eds.) *PRIMA 2009. LNCS (LNAI)*, vol. 5925, pp. 113–126. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-11161-7\\_8](https://doi.org/10.1007/978-3-642-11161-7_8)
20. Hallani, H., Shahrestani, S.: Improving the performance of wireless adhoc networks: accounting for the behaviour of selfish nodes. *Commun. IBIMA* **2**(4), 165–174 (2011)
21. Sanzgiri, K., Levine, B.N., Shields, C., Dahill, B., Royer, E.M.: A secure routing protocol for ad hoc networks. In: *Proceedings of 10th IEEE International Conference on Network Protocols*, vol. 3, no. 5, pp. 78–87 (2002)
22. Liu, K., Deng, J., Varshney, P., Balakrishnan, K.: An acknowledgment based approach for the detection of routing misbehavior in MANETs. *IEEE Trans. Mob. Comput.* **6**(5), 536–550 (2007)
23. Balakrishnan, V.C., Deng, J., Varshney, P.K.: TWOACK: preventing selfishness in mobile ad hoc networks. In: *Wireless Communications and Networking Conference*, vol. 4, no 1, pp. 2137–2142 (2005)

24. Michiardi, P., Molva, R.: CORE: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In: Proceedings of the IFIP TC6/TC11 6th Joint Working Conference on Communications and Multimedia Security: Advanced Communications and Multimedia Security, vol. 1, no. 1, pp. 107–121 (2001)
25. Refaei, M.T., Srivastava, V., Eltoweissy, M.: A reputation-based mechanism for isolating selfish nodes. In: Proceedings of 2nd Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services Ad Hoc Networks, vol. 2, no. 3, pp. 3–11 (2005)
26. Wang, Y., Giruka, V.C., Singhal, M.: Truthful multipath routing for ad hoc networks with selfish nodes. *J. Parallel Distrib. Comput.* **68**(6), 778–789 (2008)
27. Wang, F., Wang, F., Huang, B., Yang, L.T.: COSR: a reputation based secure route protocol in MANET. *EURASIP J. Wirel. Commun. Netw.* **20**(10), 1–11 (2010). Special issue on multimedia communications over next generation wireless networks archive
28. Tarag, F., Robert, A.: A node misbehaviour detection mechanism for mobile ad hoc networks. In: Proceedings of 7th Annual Post Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, vol. 1, no. 1, pp. 78–84 (2006)
29. Subramaniyan, S., Johnson, W., Subramaniyan, K.: A distributed framework for detecting selfish nodes in MANET using record- and trust-based detection (RTBD) technique. *EURASIP J. Wirel. Commun. Netw.* **4**(3), 31–39 (2014)
30. Hortelano, C., Calafate, T., Cano, J.C., Mecella, M.: Black-hole attacks in p2p mobile networks discovered through Bayesian filters. In: Proceedings of OTM Workshops, vol. 2, no. 6, pp. 543–552 (2010)
31. Chun, B.G., Chaudhuri, K., Wee, H., Barreno, M., Papadimitriou, C.H., Kubiawicz, J.: Selfish caching in distributed systems: a game-theoretic analysis. In: Proceedings of the 23th Annual ACM Symposium on Principles of Distributed Computing, vol. 7, no. 4, pp. 21–30 (2004)
32. Vallam, R.D., Antony Franklin, A., Siva Ramamurthy, C.: Modeling co-operative MAC layer misbehaviour in IEEE 802.11 ad hoc networks with heterogeneous loads. In: Proceedings of 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops, WIOPT, Berlin, Germany, vol. 1, no. 1, pp. 197–206 (2008)
33. Jaggi, N., Giri, V.R., Nambodiri, V.: Distributed reaction mechanisms to prevent byzantine misbehaviour in wireless ad hoc networks. In: Proceedings of IEEE Global Communications Conference, Houston, Texas, USA, vol. 1, no. 1, pp. 1–6 (2011)
34. Hernandez-Orallo, E., Serraty, M.D., Cano, J.-C., Calafate, T., Manzoni, P.: Improving byzantine node detection in MANETs using a collaborative watchdog. *IEEE Lett.* **16**(5), 642–645 (2012)
35. Fahad, T., Askwith, R.: A node misbehaviour detection mechanism for mobile ad hoc networks. In: *PGNet* (2006)
36. Buchegger, J., Boudec, J.Y.: Performance analysis of confidant protocol. In: Proceedings of 3rd International Symposium on Mobile Ad Hoc Networking and Computing, New York, USA, pp. 226–236 (2002)
37. Sathiyamoorthy, E., Iyenger, N.C., Narayana, S., Ramachandran, V.: Agent based trust management model based on weight value model for online auctions. *Int. J. Netw. Secur. Appl. (IJNSA)* **1**(3), 15–31 (2009)
38. Tang, J., Cheng, Y., Zhuang, W.: An analytical approach to real-time misbehaviour detection in IEEE 802.11 based wireless networks. In: Proceedings of 30th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, Shanghai, China, vol. 1, no. 1, pp. 1638–1646 (2011)

# **Social IoT**

# Illumination and Communication Using LED as Light Source in Underground Mines

B. Anitha Vijayalakshmi<sup>1</sup>(✉) and M. Nesa Sudha<sup>2</sup>

<sup>1</sup> Kings Engineering College, Chennai, Tamil Nadu, India  
anithaneil@yahoo.co.in

<sup>2</sup> Karunya University, Coimbatore, Tamil Nadu, India  
nesasudha@karunya.edu

**Abstract.** Lighting is a pre-requisite to work in an underground mine. In underground mining a much-regimented lighting system is mandatory. The presently used conventional lightning systems need lots of power and deals with major maintenance problems. As the crucial hazardous environment is prevailing in the underground mine, an enhanced communication technique is needed in mines to defeat the demerits in the present communication technology. In this paper, we propose an efficient way of delivering the disaster information inside the mine through light fidelity (Li-Fi) technology using LED as the light source which can offer both illumination and communication.

**Keywords:** Cap lamp · Li-Fi · VLC · LED

## 1 Introduction

In underground mines, communication is a crying need in terms of security and productivity [1, 2]. Communication inside the mine mostly deals with the transmission and reception of data from and to the miners and also among the miners. For smooth carrying out of mine works in such a dangerous environment, a reliable communication inside the mine is required. An instant message must be delivered for immediate rescue through a suitable and dedicated communication system from the underground mine functioning area to the outer surface area.

The major issues of underground mining are improper communication, locating miners positioning, and need of an efficient process for disaster relief. So far, Magneto phones are being used which are the elderly crack rays phone of the 20th century worked by DC batteries and AC signals [2]. Paging phones are party line wired phones for voice communication but with no tracking capability. Through the earth system (TTE) is an identified system to transmit a low-frequency signal to receivers that are incorporated into cap lamps [3].

Most of the existing systems are wire-based and they are unable to survive in disastrous situations and also dreadful in unapproachable places. The wired or wireless communication used in underground mine fails when faced with fires, roof fall, power failure, and explosion. Most of the communication schemes play a significant role in everyday utilization but fail during disasters. The wireless fidelity (Wi-Fi) based scheme might fail at the time of disaster due to unavailability of the appropriate frequency range

[4]. For smooth working in the underground mine, the infallible communication system is mandatory to ensure better safety in a hazardous environment.

Visible light communication (VLC) using Li-Fi technology is intrinsically safe as they are unaffected by interference from RF, EMF, and EMI. The features of VLC are non-licensed channels, larger bandwidth, and low power consumption. The Li-Fi source, even a single one which is few millimeters in size is capable of producing 2300 lumens of brilliant white light [5]. As visible light is used for communication in Li-Fi it's capable of providing high-speed internet at 10 Gbits/s. Li-Fi technology can be the novel trend in providing communication to and among the miners in the underground mine.

The latter part of the paper is itemized in the following manner. Section 2 explains the vital need of VLC, about the impediments of Wi-Fi in mines and the need of light fidelity with LED. Section 3 describes the importance of proper lighting, the conventional light sources used and its characteristics and about Cap lamps using the conventional light source in mines. The discussion about suggested development in an underground mine is dealt in Sect. 4. Section 5 provides the conclusion.

## 2 Vital Need of VLC

The information medium for both visible light and radio frequency communication is in the form of electromagnetic radiation [6]. Radio waves are capable of providing connectivity through often used substances. The inherent property of VLC, the waves in the visible region of the spectrum cannot penetrate through most surfaces that are there in our surroundings [7]. Wherever the VLC system is deployed, the information will be enclosed within the limited space of the exact location, which avoids the prospect of eavesdropping and eradicates interference. With the deployment of off-the-shelf illumination components, VLC can be realized as an intensity modulation and direct detection (IM/DD) scheme. The locations where RF communication does not suit can be employed with VLC. The goal of VLC is not to be competitive of RF; rather it complements the RF communication. The communication with low latency can be provided by VLC because of its high bandwidth and easier installation.

### 2.1 Impediments of Wi-Fi in Mine

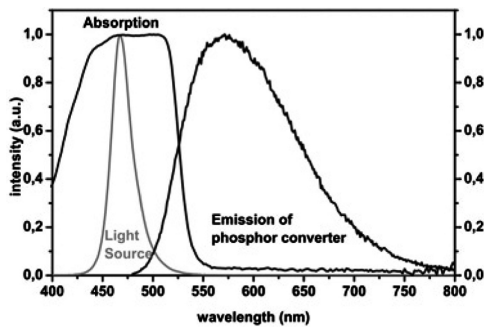
Inside the mine, the conventional switch design used by Wi-Fi network first cuts and then connects [3] i.e., the workstation initially breaks from the current access point (AP) and then commences the regular switching procedure and this could lead to a larger delay; that could be the problem of roaming when miners move in the wireless local area network (LAN). Then, the difficulty of stability, as the wireless network is simply troubled by other signals in the region around them due to the openness of the wireless channel and this ends up in instability in Wi-Fi signals and in turn it affects the quality of services.

## 2.2 Light Fidelity with LED

Li-Fi facilitates a wireless data transmission through LED. The escalating requirement in wireless data transmission, due to deficient in radio spectrum and concerns about unsafe electromagnetic contamination can be defeated through Li-Fi technology using the visible light spectrum. The LED flickers count as 1s and 0s, the input binary data. Data can be received using photodiode [8] within the area of light perception. Whenever LEDs are used, it not only offers light energy but is also realized as a wireless link at the same instance. It overcomes the issues of Wi-Fi in signal transmission like instability, larger delay and more.

The goal of Li-Fi Consortium is to foster the progress and allocation of optical wireless technologies such as communication, navigation, natural user interfaces and others [9]. There are around 14 billion light bulbs worldwide, they just need to be replaced with LED ones that transmit data, says Haas [9]. Light, in fact, has played an important role in our lives over millions and millions of years and does not have any major ill effect.

VLC using Li-Fi technology is economical than Wi-Fi as it uses light rather than radio-frequency signals. White light from LED can be generated by mixing specified quantities of Red, Blue and Green lights. Phosphor topped white LEDs radiate wide-band visible light spread over the entire visible spectrum as shown in Fig. 1.



**Fig. 1.** Radiation spectrum of phosphor white LED (Source: Phosphor Development Addresses Lower Efficacy of Warm-White LEDs)

## 3 Importance of Proper Lighting

Better lighting is needed in order to lessen accidents, health hazards and also to increase production. Miners commonly undergo the eye disease nystagmus, which gives the symptoms of irrepressible oscillation of the eyeball, headache, dizziness, and loss of night vision [10]. It's because of the low light working environment inside the mine. Underground miners are prone to hazards like spot fall of ground, unusual sliding and much more. The other hazard happenings inside the mine environment are hugging with dust, cramped spaces, low reflective surfaces and low visual contrasts. Yet, the major contributing factor for the accidents in an underground mine is poor lighting inside the mine. Lighting in mine is shown in Fig. 2.





Fig. 2. Lightning in mine

While designing lighting systems, the required features are that it must be essentially safe, harmless, ought to give intense brightness and intensity in horrible surroundings. The lighting system must give out less radiated heat to avoid any electrical hazards and also it must be resistant to shock, vibration, and there should not be any UV rays emission and it should be easy to install [11]. In an exclusive environment like underground mines, making high-quality lighting systems is the challenging task.

### 3.1 Conventional Light Sources and Its Characteristics

Incandescence means thermal emission. Heat is continually emitted from objects and when they heat up and get powerful enough, they attain the wavelength, starting with red and reach up to the spectrum. The wavelength/color of the light determines how much energy is being released. In an incandescent bulb, 90% of heat energy is released in the infrared spectrum which is just under the visible light and makes the lamp ineffective. The visible light produced by fluorescent lamp using fluorescence is more efficient than the light produced by incandescent lamps but flickers in fluorescent lamp cause irritation in eyes, eye strain, headaches, and migraines. The diffused light of the lamp is not good when used for the purpose of headlight or flashlight [12]. Some light sources when subjected to overheating results in firing. Some other sources of light are affected by the presence of other electronic devices and are subjected to interference from other sources. The comparison of characteristics between various light sources like the incandescent bulb, CFL, LED [12] is shown in Table 1.

Table 1. Comparison characteristics between incandescent, CFL and LED light source

S. No	Characteristics	Incandescent bulb	CFL	LED
1	Toughness	Not too tough	Not too tough	Very tough
2	Generation of heat	High	Medium	Too low
3	Hazard	Hazardous	Hazardous	Hazardless
4	Sensitivity to temperature	Moderate	High	Negligible
5	Mercury content	No	Yes	No
6	Carbon dioxide emission	68	15.8	6.8
7	Life time	1000 h	12000 h	100000 h

### 3.1.1 Cap Lamps Using Conventional Light Source in Mines

For every underground miner, the dependable and faithful appliance is cap lamp. The most common type of lighting used in mine projects is tungsten filament lamp. These lamps have limited life and efficacy [13]. The high operating temperature of the lamp causes blackening of the lamp and ends in failure. The improvisation in this family is tungsten halogen incandescent lamp which is filled with halogen gas. This avoids blackening of light bulb at high temperature. They are not recommended for lightning purpose due to low efficacy and limited life [13]. Fluorescent cap lamps are developed in the 1970s. As compared to an incandescent lamp, it offers better light output with a lesser amount of degradation over a 10-h shift [14]. Due to its size, weight, and high initial costs, fluorescent lamps are not used much in mines [14]. With Li-Fi using LED as a light source, a new class of intensity-based visible light communication technology can be employed. With energy efficiency, long lifetime and full spectrum availability, Li-Fi lighting applications using LED as a source in cap lamps will work better when measured up with conventional approaches.

## 4 Suggested Development in Underground Mine

In an underground mine, the recommended block is shown in Fig. 3 [3] to provide effective data transmission from the base station to receiving station. The system has a Li-Fi RF driver to sense the signal from the base station using the medium of light and the photo detector in the device captures the signal and passes the information to the monitoring section. '1' is transmitted if the LED is **on**; if it's **off** the transmitted output is '0'. LEDs can be toggled on and off very quickly, such that it offers pleasant opportunities for transmitting data. Li-Fi may not be able to replace conventional radios altogether, but it could enhance the development of wireless monitoring section in mines and make it easier to throw a wireless signal in underground mines [3].

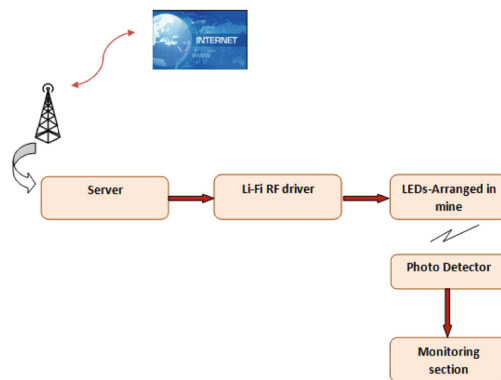
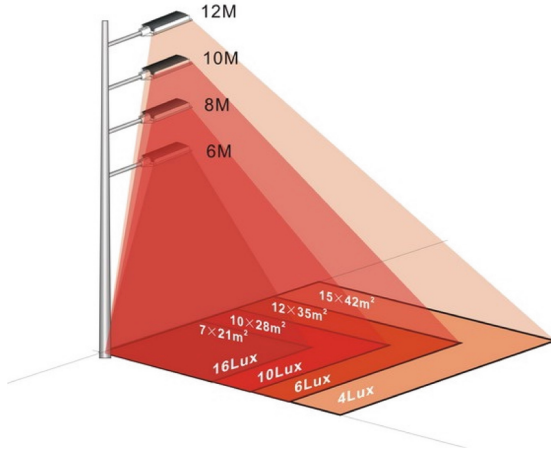


Fig. 3. Block diagram

Based on the height of the light source placed, the illumination level varies. The level of illumination based on the height of the LED luminaire placed [15] is shown in Fig. 4.



**Fig. 4.** Illumination based on height of the LED luminaire placed

The brightness of an illuminated surface is expressed as luminance. The light intensity released from the source has a cosine dependence on the angle of emission with respect to the normal surface [16]. The luminous intensity at angle  $\phi$  is given [16] by,

$$I(\phi) = I(0)\cos^{m_l}(\phi) \tag{1}$$

A horizontal illuminance  $E_{hor}$  at a point  $(x, y)$  [13] is given by,

$$E_{hor} = I(0)\cos^{m_l}(\phi) / D_d^2 \cos(\Phi) \tag{2}$$

where  $I(0)$  is the center illuminance intensity of the LED,  $\phi$  is the angle of irradiance,  $\Phi$  is the angle of incidence,  $D_d$  is the distance between the LED and the detector’s surface and  $m_l$  is the order of Lambertian emission [16] related to the LED’s semi-angle at half power  $\psi_{1/2}$ . Figure 5 [17] shows the illumination distribution results.

$$m_l = -\ln 2 / \ln \left( \cos \left( \psi_{1/2} \right) \right) \tag{3}$$

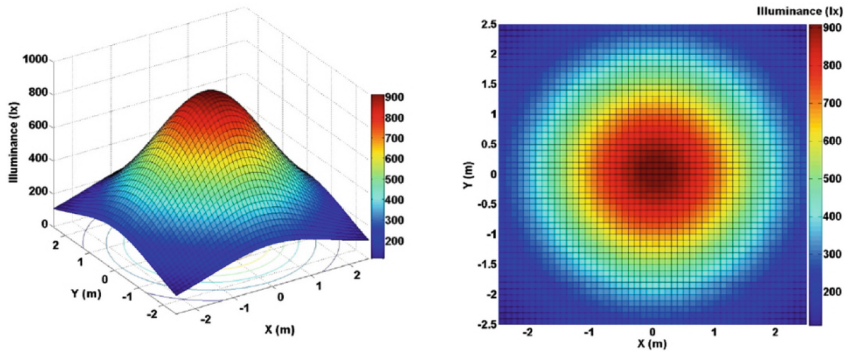


Fig. 5. Illumination distribution in the receiver using LED Source: (OAK CENTRAL)

## 5 Conclusion

An effectual wireless communication is required during disasters to ensure the location, in providing assistance and access to communication and much more. There is a dying need for more effective disaster warning and responsive system during calamity caused by man and nature. Visible light communication using Li-Fi technology will be the best solution in the future for better illumination and communication inside underground mines using LED luminaire for lighting system. LED cap lamps can be the talented competent to the conventional cap-lamps.

## References

1. Jayanthu, S., Jayadarshana, S.: Recent trends in communication systems for underground mines. In: National Seminar on Recent Trends in Mechanized Mining, pp. 27–28 (2014)
2. Patri, A., Nayak, A., Jayanthu, S.: Wireless communication systems for underground mines – a critical appraisal. *Int. J. Eng. Trends Technol.* **4**(7) (2013)
3. Vivek Priyan, R., Dinesh, S., Ilanthendral, J., Ramya, B.: Communication system for underground mines using Li-Fi 5G technology. In: IJLTEMAS, vol. III, no. IX (2014)
4. Kumar, A., Kumar, D., Singh, U.K., Gupta, P.S., Shankar, G.: Optimizing fibre optics for coal mine automation. *Int. J. Control Autom.* **4** (2011)
5. Sridharan, C., Srikanth, P., Thresphine, J.R.: Intelligence with Li-Fi technology. *Int. J. Comput. Eng. Sci.* (2014)
6. Fellers, T.J., Davidson, M.W.: The nature of electromagnetic radiation. National High Magnetic Field Laboratory
7. Manimegalai, M., Karthick, B.: Design and implementation of high data rate system for the diffuse indoor channel using VLC. *Int. J. Eng. Sci.* (2015)
8. Karthika, R., Balakrishnan, S.: Wireless communication using Li-Fi technology. *SSRG Int. J. Electron. Commun. Eng.* **2**(3) (2015)
9. Verma, P., Shekhar, J., Asthana, A.: Light-Fidelity (Li-Fi): transmission of data through light of future technology. *Int. J. Comput. Sci. Mob. Comput.* **4**(9), 113–124 (2015)
10. American Academy of Ophthalmology. <https://www.ao.org>. Accessed 1 Sept 2017
11. Wikipedia. <https://en.wikipedia.org/wiki/Radiation>. Accessed 26 Sept 2017

12. WikiHow. <http://www.wikihow.com/Diffuse-Light>. Accessed 25 June 2017
13. Lakshmiathy, N., Murthy, S.N., Aruna, M.: Problems encountered in the types of lighting systems generally used in surface mining projects a case study. *Int. J. Eng. Sci.* **3**(9), 61–72 (2014)
14. Sammarco, J.J., Carr, J.L.: Mine illumination: a historical and technological perspective
15. Mishra, H.: Study of application of LED lighting system in mines
16. Han, P.P., Sewaiwar, A., Chung, Y.-H.: 2 Gbit/s VLC scheme using time-frequency color-clustered MIMO based on BCYR LEDs. *J. Opt. Soc. Korea* **20**, 276–282 (2016)
17. Sewaiwar, A., Han, P.P., Chung, Y.-H.: 3-Gbit/s indoor visible light communications using optical diversity schemes. *IEEE Photonics J.* **7** (2015)

# Leveraging Social Networks for Smart Cities: A Case-Study in Mitigation of Air Pollution

Nagarathna Ravi<sup>(✉)</sup>, Manoranjani R., Vimala Rani P., Mercy Shalinie S.,  
and Karthick Seshadri

Department of Computer Science and Engineering,  
Thiagarajar College of Engineering, Madurai, Tamil Nadu, India  
rathnaravi2013@gmail.com, manoranjaniravichandran@gmail.com,  
vimalainfotechh@gmail.com, {shalinie,skcse}@tce.edu  
<http://www.tce.edu/>

**Abstract.** Air pollution is one of the pressing issues faced not only by human race but also by Earth's vegetation. There are numerous findings regarding the effects of air pollution, for instance air pollutants playing a notorious role in inducing asthma and in formation of acid rain. Apart from regular activities done by government like creating acts, setting standards and regular inspection in industries to provide people with pollution free environment, various mobile apps like Air Quality India, SAFAR Air have been developed and disseminated to create environmental awareness. But even then, air pollution persists due to gases from vehicles on roads, Choloro Fluro Carbon from home appliances, Carbon Monoxide from incomplete combustion of fuels or gas, particulate matters from cigarette smoking, cooking etc. Individuals are affected by these air pollutants which are predominant in the air they breathe in. It is necessary to educate people on how to ameliorate the effects of air pollutants and control air pollution. In this paper, we have presented a framework that utilizes data from social networks and supervised machine learning algorithm, which will facilitate in taking an initial step towards a smart environment in smart cities. A part of our proposed application was tested with a group of people. These experiments showed better accuracy in classifying people according to the place where they spend most of their time in a day, and also received good responses from the people.

**Keywords:** Social networks · Supervised learning · Health Environment · Air pollution · Smart cities

## 1 Introduction

The ultra fast pace of urbanization that includes conversion of green lands into urban sprawls and tremendous advancement in industrial activities have caused serious negative side effects on the clean atmosphere. The term air pollution can be defined as an increase in the concentration of green house gases and other

anthropogenic emissions of noxious chemicals [1,2]. Most of these toxins lead to long term adverse effects. The ChloroFluoroCarbons (CFCs) generated from air conditioners, aerosols propellants had no effect as far as they were in the troposphere. However over time when it moved to the stratosphere they had an adverse effect on the ozone [1]. The International Agency for Research on Cancer (IARC) has found that humans exposed to air pollutants that include carcinogens contracts to lung cancer [3,4]. Human nose and bronchioles lack the capacity to filter out ultra fine pollutants that leads to respiratory tract disorders [2]. A mix of nitrous oxide and PM10 produces lethal effects on people [4]. Exposure to air particulate matter that have a diameter lesser than  $2.5\mu\text{m}$  (PM2.5) have a correlation to cardiovascular mortality [5]. Pollution leads to neurodegenerative diseases such as Alzheimer's and Parkinson's diseases [2].

Earth's green cover also feels the ill brunt of air pollution through various ways. Ozone which acts as a protective shield from harmful ultraviolet rays of sun turns out to be an adverse air pollutant when it is at ground level and affects the yield of husbandry crops [6]. The Particulate Matter (PM) disturbs the normal physiological and biochemical status of plants [7]. Particulate air pollutant aerosols also play a significant role in rain scavenging [8].

Current era witnesses an increased usage of web among all ages. In particular most of the online time is spent on social networking sites. There are quite a lot of evidences in the usage of social networks in mass movement like the Arab Spring. Using this fact, we propose a novel application that uses social networks in educating people that will aid in moving towards a clean environment in smart cites. Merely posting facts about air pollution will not create a good result in achieving smart environment. It is necessary to create a widespread awareness among people and they also take smart steps to mitigate air pollution starting from their living area.

To address this problem we have propose the usage of a supervised learning approach in our model, as it is not possible to find the places where a user will spend most of their time in a day directly gathered from social networks.

## 2 Related Work

There have been various researches done in the sphere of air pollution, to confirm the ill effects of air pollutants and also identification of right measures to control air pollution.

Xie et al. [9] have done a study on the influence of the odd and even license plate model which was implemented in Beijing to control air pollution. They have concluded that the plate model has a good impact in controlling air pollution during an interim period, but gradually the good impact diminishes due to the increase in the number of private vehicles.

Ngo et al. [10] have made a study on how participation of people is crucial in controlling air pollution. They have found that when people were involved in measuring the PM2.5 levels in Nairobi and Kenya people expressed interest in knowing more information about air pollution.

Woodward and Levine [11] have done a study on the impact of air pollution on elder persons in US who are more vulnerable. They have proposed guidelines to minimize the elder people getting exposed to air pollutants by placing the older adult's care facilities at least 500 ft from the major roadways.

Adams and Kanaroglou [12] have proposed a neural network model to provide health risk information to people who live in places where air pollution monitoring systems are not available. The model utilizes various predictor variables like traffic in the area, meteorological conditions, characteristics of land usage and pollution statistics from nearby fixed monitoring stations.

Nowka et al. [13] made a survey through questionnaires to identify the extent of awareness of impact of air pollution on coronary heart disease among patients. The authors have concluded that patients lack the awareness of ill effects and there is a need for educating people about air pollution.

In India, Delhi Government has started to take measures to curb air pollution. They have proposed odd or even license rule, traffic police to use air mask, ban diesel cars above 2000 cc, hike in green cess, taxis should utilize CNG, prohibition on government agencies in procuring diesel vehicles, increase in public transport, fine on polluting agents, ban on burning waste and crop residues. We can see that Indian Government is also recently becoming very conscious of people's health getting affected due to air pollutants.

Apart from these research studies, government and other agencies have designed mobile apps to share air quality index through the levels of various pollutants. To aid in providing a smart environment to people and other living beings, and also safeguard individuals from ill impacts of air pollutants, we have developed an intelligent model that utilizes the k-nearest neighbor algorithm to educate people on ways to reduce air pollution.

### 3 Pollution Mitigation Framework

The proposed pollution mitigation framework is shown in Fig. 1. It uses data collected from the social networks and employs a three phased approach as detailed below:

#### 3.1 First Phase

**Step 1: Divide city into zone clusters** - Pollution control board of India regularly post updated statistics of major cities in their website. But a little more effort is required in aiding people in understanding that their participation is more required in mitigating air pollution and safeguarding themselves. For this, it is necessary to provide them air quality statistics of their area to attract people's attention [12]. As a first step, we propose the city be divided into many zones. We propose segmentation based on a constant and two variables namely density of people and air pollution levels in the zone in Algorithm 1. Arbitrarily fix a constant span value (we propose a value of 1 Km to have an effective coverage) to subdivide the whole range space of each city into clusters of equal



span. Most of the areas of city are non-residential or have relatively minimal air pollution levels. Such zones can be grouped into a new cluster. The terms *zone* and *cluster* are used interchangeably in the rest of this paper.

**Step 2: Collation of air quality statistics in each zone** - Installing air quality monitoring stations all over the zones is a cumbersome, time consuming and costly process. We propose the usage of Internet of Things (IoT) pollution sensors in mobile towers due to their widespread installation across the globe and on public transportation vehicles [14]. IoT pollution sensors are low cost and consume low power [15]. IoT nodes have a standard known as 6LoWPAN and RPL to perform node discovery, auto configuration of nodes and routing. These protocols facilitate collection of data and transmission to a central repository.

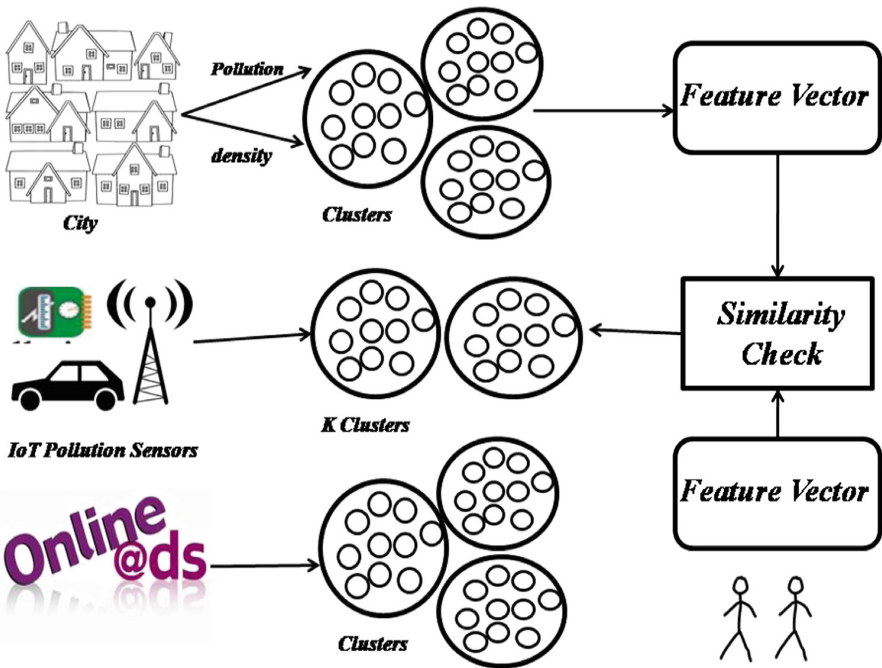


Fig. 1. Pollution mitigation framework

**Step 3: Pre-processing of social network’s user data** - Owing to an increase in the number of users, social network data is voluminous and needs to be pre-processed. The user id and name are obfuscated by a random unique id or name to protect anonymity and privacy. The present designation namely school student or college student or employee or own a business or others, school or college or work place or business site, address and e-mail ID are extracted from the social network profiles. The reason of extraction is that people generally spend most of their time in either their study place or workplace or their

residence. So the time and extent to which people are exposed to ill effects of air pollutants will be high in these places.

**Step 4: Model the zones and preprocessed data as feature vectors** - Each cluster and user should have a unique feature vector to aid the users getting mapped to the correct cluster. We propose the name of the roads and the various landmarks of the zone to be added in the feature vector of each cluster. The road and landmark near each user's lyceum or seminary or workplace or dwelling is added in the user's feature vector. To identify the elements of the feature vector, Google places API and Google roads API can be integrated with the application.

**Step 5: Mapping users to the zones** - To identify the cluster a user belongs to, there is a necessity for adding intelligence in our model. Supervised learning approach better fits our model. Training is done in step 1 and feature vectors act as training examples. This step just classifies the various training instances i.e. user to their respective zones. Instance based learning better suits the system as it is capable of working on training instances that are more complex. K-nearest neighbor (KNN) learning approach is employed, since the feature vector of a user comprises of the road and landmark closer to the user's place. There are ample chances of the feature vector matching two or more clusters as long roads may lie in two or more zones.

The Jaccard similarity coefficient is used as a distance measure for KNN to perform a similarity check. We find the ratio of the cardinality of common elements in the feature vector of each cluster and user, to the union of the elements. The jaccard similarity of the user with each cluster is added to a similarity vector. The user is assigned to the clusters which had top values in the similarity vector. Algorithm 2 illustrates the steps 3-5.

### 3.2 Second Phase

**Step 1: Dissemination of the air quality statistics** - Each government prescribes safe pollution levels. The air quality statistics are disseminated to the users based on certain conditions and the various routes from their residential address to work or study place which is illustrated in Algorithms 3 and 4. The condition includes checking if the air pollution level is close (20% less than the standard which has been arbitrarily set) or has exceeded the standards. The routes are calculated by integrating Google directions API. If the user's place or routes have polluted zones, then statistics are published to the user.

**Step 2: Recommendation of preventive and control measures** - Based on the air quality statistics of each zone the individuals are guided regarding the control and prevention of health problems due to air pollution. Based on which quantity of air pollutants is high in the zone, the users are educated how these pollutants will affect their health in short and long term and are provided appropriate guidance to improve their life sustenance.

### 3.3 Third Phase

Commercialization paves way to monetary benefit as well as provides users with extraneous information. We propose using pay per click on advertisements. The advertisers can purchase ad space only if they possess air pollution compliance certificate issued by standard environmental bodies. The advertisers can bid on the space where the preventive measure posted is related to their products. Companies who guarantee their products can be shipped anywhere can be posted to all the users who have been recommended to use the product or else only to the zone users where product can be shipped.

---

#### Algorithm 1. Subdividing city into zones

---

```

input : Span of zone(S), Area of city(A), Threshold density( $T_d$ ), Threshold
        pollution( $T_p$ )
output: clusters(C)

1  $I \leftarrow$  initial number of clusters ;
2  $C \leftarrow \emptyset$  ;
3  $I \leftarrow A \div S$  ;
4 Add the equally divided clusters to vector C;
5  $L \leftarrow$  zones whose density or pollution level is less than threshold ;
6  $L \leftarrow \emptyset$  ;
7 for  $j \leftarrow 1$  to  $I$  do
8    $D_j \leftarrow$  density of people in cluster  $j$  ;
9    $P_j \leftarrow$  air pollution in cluster  $j$  ;
10  if ( $(D_j < T_d)$  and  $(A_j < T_p)$ ) then /* Identifying less dense and
        polluted zones */
11     $L \leftarrow L \cup \{j\}$  ;
12  end
13 end
14 foreach  $k$  in  $L$  do
15    $C \leftarrow C \setminus \{k\}$  ;
16   Identify zones close to cluster  $k$ ;
17   Merge the nearby zones;
18    $Z \leftarrow$  Merged zone ;
19    $C \leftarrow C \cup Z$  ;
20 end
21 return  $C$  ;

```

---

**Algorithm 2.** k-nearest neighbour with Jaccard Index

---

**input:** Social Network user data(S), Clusters (C)

- 1 *Pre-process S to extract user's ID, name, present designation, school or college or work place or business site, address and email ID ;*
- 2 **foreach** *i in C do*
- 3      $landmark_i \leftarrow$  important places in the zone area;  
    /\* Find landmarks using the Google places API  
    integrated with the application \*/
- 4      $road_i \leftarrow$  name of roads in the zone area;  
    /\* Find the roads using the Google Roads API integrated  
    with the application \*/
- 5      $featurevector_i \leftarrow (landmark_i, road_i)$  ;
- 6 **end**
- 7 **foreach** *j in S do*
- 8      $R \leftarrow$  residential address ;
- 9      $L \leftarrow$  school or college or work place ;
- 10     $landmark_j \leftarrow$  important places nearby(R,L) ;  
    /\* Find landmarks using the Google places API  
    integrated with the application \*/
- 11     $road_j \leftarrow$  roads(R,L) ;  
    /\* Find the roads using the Google Roads API integrated  
    with the application \*/
- 12     $featurevector_j \leftarrow (landmark_j, road_j)$  ;
- 13    **foreach** *m in C do*
- 14         $Nr \leftarrow |F_j \cap F_m|$  ;  
       /\* F denotes feature vector \*/
- 15         $Dr \leftarrow |F_j \cup F_m|$  ;
- 16         $Jaccardindex \leftarrow Nr \div Dr$  ;  
       /\* Jaccard similarity measure to find which zone the  
       user belongs to \*/
- 17         $sim_j \leftarrow sim_j \cup Jaccardindex$  ;  
       /\* add the jaccard index to the similarity vector of  
       user j \*/
- 18    **end**
- 19    *Assign the user j to the k nearest neighbours (i.e.) k clusters that  
       have high similarity measure in vector sim\_j ;*
- 20 **end**

---

---

**Algorithm 3.** Dissemination of air quality statistics to people(1)

---

**input:** List A-air quality stats( $SO_2, NO_2, NO, CO, PM_{2.5}, PM_{10}$ ) of each zone, List S-safe levels of pollutants prescribed by any standard pollution control board, C-zone clusters

```

1 foreach  $k$  in  $C$  do
  | /* highly or moderately polluted zones */
2 | if  $((A > S) \text{ or } (A > S - (0.2 * S)))$  then
3 | | Post the air quality stats to the social network users in the zone
4 | | k;
5 | | Present preventive and mitigation measures of air pollution in
6 | | the zone;
7 | end
8 end

```

---



---

**Algorithm 4.** Dissemination of air quality statistics to people(2)

---

**input:** User's residence and work place/school/college (R), pollution standard levels (S), air quality stats of each zone

```

1  $I \leftarrow$  initial number of clusters ;
2 foreach  $k$  in User do
3 | if R data is available then
4 | | route  $\leftarrow$  direction ;
5 | | /* Find direction using the Google directions API
6 | | integrated with the application */
7 | |  $pol_{max} \leftarrow$  max(pollutants level across zones in route) ;
8 | | if  $((pol_{max} > S) \text{ or } (pol_{max} > S - (0.2 * S)))$  then
9 | | | /* Identifying highly or moderately polluted zones
10 | | | */
11 | | | Display if you are taking this route then you are breathing
12 | | | in air pollutants ;
13 | | end
14 | end
15 end

```

---

## 4 Experimental Setup and Evaluation

In order to test the efficiency of our proposed pollution mitigation framework we implemented a part of the framework using python.

### 4.1 Dataset Description

We tested our framework on a set of real time users, as feedback is necessary to ascertain that this model would be an effective means to take an initial step towards aiding individuals to help them learn about and safeguard their health

from air pollutants. A group of social network users were selected. The users were active Google social network users. A group of 700 users were selected arbitrarily from Chennai and Bangalore. The group comprised of a mix of high school students, college students (both undergraduate and postgraduate students) and other professionals of age ranging from 18 years to 54 years.

## 4.2 Pre-processing

Initially the features such as user ID, name, age, Gmail ID, residential address, school, college, period of study in school, college, profession and period of working was extracted from the dataset. Probe checks were made to find if the field either contained any value such as present or current, or the range included the current year. If either of the conditions was satisfied the school or college or the work area name was retained for the user and the rest of the data was filtered out. If not available, no filtering was done.

## 4.3 Air Quality Statistics

The pollution statistics were collated from central pollution control board of India (CPCB)'s website ([www.cpcb.nic.in](http://www.cpcb.nic.in)) for the cities Bangalore and Chennai. The CPCB has air quality monitoring stations in 5 places in Bangalore and 3 in Chennai. Based on this we divided Bangalore into 5 zones and Chennai into 3 zones. Table 1 provides a part of the air quality statistics for Bangalore city.

**Table 1.** Air quality stats

Zone cluster (Bangalore and Chennai)	Air quality	
	PM2.5	PM10
BTM layout	53.12	NA*
Peenya	108.4	NA*
BWSSB Kadabesanahalli	83.89	NA*
Saneguravahalli-KSPCB	NA*	17.41
City railway station-KSPCB	NA*	117.73
Alandur	112.51	NA*
IIT	95.96	NA*
Manali	252.82	NA*

\*Not Available

## 4.4 Computation of Feature Vectors

The Google places and roads API were integrated to model the users and clusters in the form of feature vectors by using the server key. The feature vector comprised of local businesses, points of interest, and the nearest road segments.

### 4.5 Jaccard Similarity

Naive implementation of Jaccard similarity is a computationally intensive task. So it is necessary to identify a probabilistic approach. Min Hash approach is ideal for data streams. The Min Hash independent permutations involves mapping the elements of the two vectors to distinct integers and determining minimal member with respect to a hash function that is the minimum of the values of  $h(x)$ , where  $h$  is the hash function and  $x$  is the set of distinct integers mapped.

$$J(U,C) = \Pr[\text{Min} \langle h(Y_1), h(Y_2), \dots \rangle = \text{Min} \langle h(Z_1), h(Z_2), \dots \rangle]$$

where,

$J$  denotes the Jaccard function

$U$  is the feature vector of user

$C$  is the feature vector of cluster

$\{Y_1, Y_2, \dots\}$  is the set of distinct integers mapped to the elements in user's feature vector

$\{Z_1, Z_2, \dots\}$  is the set of distinct integers mapped to the elements in user's feature vector

The plot of Fig. 2 clearly depicts the difference the time taken across various runs by the naive and minhash approach. In order to find the optimal number of permutation functions we executed various runs to identify the optimal point where the absolute error between min hash and jaccard index is minimum. From the plot in Fig. 3 it can be seen that the absolute error was minimum when the number of permutation function was 215.

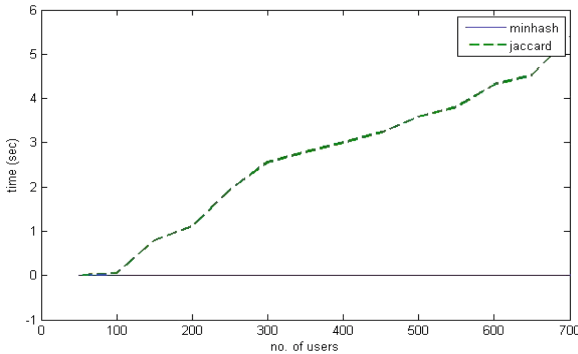
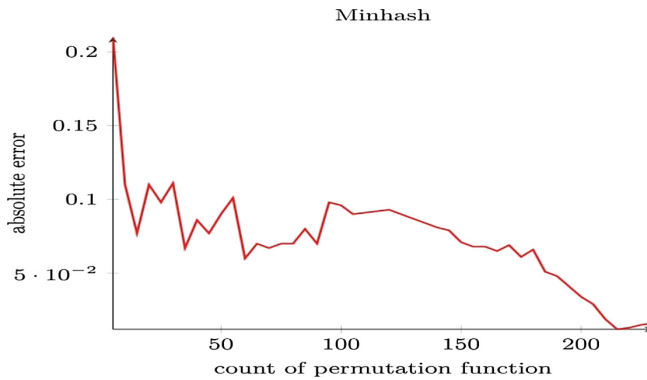


Fig. 2. Min hash vs jaccard

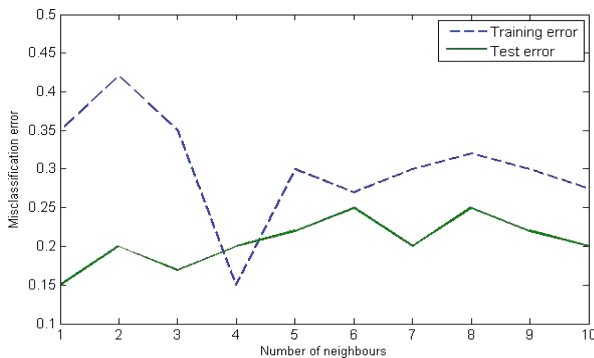
### 4.6 K Nearest Neighbor

The people were added to the k nearest neighbors based on the jaccard similarity coefficient vector. The accuracy of the classification algorithm is dependent on the reduced error rate. Choosing an optimal k value will lead to reduction in



**Fig. 3.** Optimal count of permutation functions to be used

error rates. When  $k$  value is larger, there is a good chance of minimization of variance due to noisy data. But it may introduce a bias, which will ignore smaller patterns that may have useful insights. The plot in Fig. 4 depicts that the optimal value of  $k$  is 4 for our test data. Overall we achieved an accuracy of nearly 87% in classifying people. The Fig. 5 depicts the results of the  $K$  nearest neighbor classifier of users in Bangalore alone through the confusion matrix.



**Fig. 4.** Optimal  $k$  value

#### 4.7 Dissemination of the Air Quality Stats

**Step 1:** The air quality statistics of each zone was e-mailed to the users in the zone along with preventive measures. To find if they were useful we sent a survey form through Google form. Tables 2, 3, 4, 5 and 6 show the responses received along with the corresponding questions. We received responses from 579 people out of the 700 people. This shows that nearly 83% of the people viewed our air quality related information. This survey has received a higher percentage of



		Predicted Label				
		BTM	Peenya(P)	BWSS(B)	SHALLI(S)	CRS
True Label	BTM	153	0	3	11	4
	P	0	103	9	4	1
	B	0	6	76	1	1
	S	0	2	0	15	0
	CRS	0	1	1	1	21

**Fig. 5.** Confusion matrix of users in Bangalore

positive response, which shows that our model is better to move towards smart and healthy city. We requested users to provide feedback and suggestions. We received positive feedback and few suggestions from the users.

Question 1: Did you find the air pollution related information useful?

**Table 2.** Question 1

Option	Response(%)
Yes	73
No	27

Question 2: Were you aware of the health related impacts due to air pollution before getting the statistics and other information from us?

**Table 3.** Question 2

Option	Response(%)
Yes	41
No	59

Question 3: What did you know about air pollution before? (Options-check box)

**Table 4.** Question 3

Option	Response(%)
Affects the health	52
Chemicals from industries	71
Vehicular fumes	63
Dust on roads	66
Other	44
No idea	36

Question 4: Have you taken any measure before, to safeguard yourself from ill impacts of air pollution?

**Table 5.** Question 4

Option	Response(%)
Yes	29
No	71

Question 5: Will you follow any of the suggestions provided by us to mitigate and safeguard from air pollution?

**Table 6.** Question 5

Option	Response(%)
Yes	76
No	11
May be	13

## 5 Conclusion and Future Work

The proposed pollution mitigation model can be a good start in achieving healthy smart city. We anticipate that this model when integrated on social networking platforms will accelerate the air pollution control at a better rate in urban areas of India. Owing to the limited availability of genuine air quality statistics, we were unable to quantify the efficiency of division of the city into zones. So we had to move into dividing the cities into zones as per the air quality monitoring

stations set by CPCB. Similarly, we were unable to check the effectiveness of the third phase of our model since our dataset is rather a small dataset when compared to huge population in India. However, our motive was to identify an initial step in leveraging widely used social networking sites as a platform to bring about smart and healthy city.

We suggest enhancing this model to include making a text analytics of the comments posted by the users regarding what measures they took to curtail the effects of air pollutants, and evolving a recommender system to disseminate these suggestions to users who also have similar pollution levels in their place.

## References

1. Zannetti, P., Al-Ajmi, D., Al-Rashied, S.: Ambient air pollution. The Arab School for Science and Technology and The EnviroCamp Institute, Fermont CA (2007)
2. Chen, R., Hu, B., Liu, Y., Xu, J., Yang, G., Xu, D., Chen, C.: Beyond PM2.5: the role of ultrafine particles on adverse health effects of air pollution. *BBA Gen. Subj.* **1860**(12), 2844–2855 (2016)
3. Hamra, G.B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J.M., Vineis, P., et al.: Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environ. Health Perspect.* **122**(9), 906–911 (2014)
4. Chen, X., et al.: Long-term exposure to urban air pollution and lung cancer mortality: a 12-year cohort study in Northern China. *Sci. Total Environ.* **571**, 855–861 (2016)
5. Cascio, W.: Proposed patho physiologic framework to explain some excess cardiovascular death associated with ambient air particle pollution: insights for public health translation. *Biochim. Biophys. Acta* **1860**(12), 2869–2879 (2016)
6. Matyssek, R., Clarke, N., Cudlin, P., Mikkelsen, T., Tuovinen, J.P., Wieser, G., Paoletti, E.: Climate Change Air Pollution and Global Challenges. Development in Environmental Science. Elsevier, Amsterdam (2013)
7. Rai, P.K.: Impacts of particulate matter pollution onplants: implications for environmental biomonitoring. *Ecotoxicol. Environ. Saf.* **129**, 120–136 (2016)
8. Elperin, T., Fominykh, A., Krasovitov, B.: Effect of raindrop size distribution on scavenging of aerosol particles from Gaussian air pollution plumes and puffs in turbulent atmosphere. *Process Saf. Environ. Prot.* **102**, 303–315 (2016)
9. Xie, X., et al.: Effect analysis of air pollution control in Beijing based on an odd-and-even license plate model. *J. Clean. Prod.* **142**(2), 936–945 (2016)
10. Ngo, N.S., et al.: Why participation matters for air quality studies: risk perceptions understandings of air pollution and mobilization in a poor neighborhood in Nairobi, Kenya. *Public Health* **142**, 177–185 (2015)
11. Woodward, N., Levine, M.: Minimizing air pollution exposure: a practical policy to protect vulnerable older adults from death and disability. *Environ. Sci. Policy* **56**, 49–55 (2015)
12. Adams, M.D., Kanaroglou, P.S.: Mapping real-time air pollution health risk for environmental management: combining mobile and stationary air pollution monitoring with neural network models. *J. Environ. Manag.* **168**, 133–141 (2016)
13. Nowka, M.R., Bard, R.L., Rubenfire, M., Jackson, E.A., Brook, R.D.: Patient awareness of the risks for heart disease posed by air pollution. *Environ. Sci. Policy* **53**, 379–384 (2011)

14. 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. In: HumTech 2015, pp. 480–484 (2015)
15. Krishna, G.G., Krishna, G., Bhalaji, N.: Analysis of routing protocol for low-power and lossy networks in IoT real time applications. In: ICRTCSE 2016, pp. 270–274 (2016)

# Smart Garbage Bin Systems – A Comprehensive Survey

Gulshan Soni<sup>(✉)</sup> and Selvaradjou Kandasamy

Department of Computer Science and Engineering,  
Pondicherry Engineering College, Puducherry 605014, India  
{gsoni, selvaraj}@pec.edu

**Abstract.** The neat and clean surrounding is the main driving force for any city to be called as “smart city”. Many modern cities are currently encumbered with various challenges such as smart transport system, smart grid, smart environment, and smart living. Now-a-days, proper waste management is the major concern for cities and urban areas. The traditional waste management approaches are not sophisticated enough to achieve a proficient and robust waste management. The smart Waste Management is on top priority in any smart city as it directly affects the lifestyle, healthcare and environment. This article deliberates a comprehensive survey of various proposed approaches for smart bin systems such as Smart Garbage Monitoring System, Wisely Waste Segregation System, and Smart Waste Collection System. In addition to this Survey, we propose a framework for smart Garbage Management System (GMS) that can be deployed in metro cities.

**Keywords:** Smart bin · Cloud computing · IoT · RFID · WSN · Sensors

## 1 Introduction

Integration of the two eminent technologies, Information and Communication Technology (ITC) and the Internet of Things (IoT) gave birth to the concept of smart city. In order to understand the concept of Smart Cities in depth, a suitable definition is needed. In the literature, various researchers proposed various definitions of smart city. The simplest definition of smart city is that “A city that monitors and integrates conditions of all of its critical infrastructure, including roads, bridges, tunnels, rail/subways, airports, seaports, communications, water, power, and even major building. The City can better optimize its resources & plan its preventive maintenance activities, and monitor security aspects while maximizing services for its citizens [1]”. Smart Cities can be identified along with six main dimensions [2], viz.

- (1) Smart Economy- Innovation and Competitiveness
- (2) Smart Mobility - Infrastructure and Transport
- (3) Smart Environment - Resources and Sustainability
- (4) Smart People - Creativity and Social Capital
- (5) Smart Living - Culture and Quality of Life
- (6) Smart Governance -Participation and Empowerment.

The major smart city applications [3] include Smart Street lighting, Smart Parking, Environmental monitoring, Information Beacons, Active Safety, Smart Journey Planning, Transport Sharing, Smart Bin Collection, Social & Health Care Cost Reduction, and Smart Social Housing.

We cannot imagine a smart city without having smart bin collection and monitoring system because this directly impacts the health of the citizens. Due to various reasons such as migration of people from villages to cities, rapid urbanization and modernization the population of urban areas is increasing rapidly worldwide. According to World Health Organization (WHO)'s Global Health Observatory (GHO) data, the population of urban areas in 2014 accounted for 54% of the total global population, up from 34% in 1960, and continues to grow, and it is estimated that by 2017, even in less-developed countries, a majority of people will be living within urban areas [4]. In urban areas, the Waste Management responsibility is typically carried out by the respective municipal corporations. The authorities of such corporations often instruct the citizen to deposit their household wastage at specific places in garbage bins. Due to the change in consumption habits, life-style of urban population and owing to the increase of population in urban areas, the volume of solid waste to be handled by such corporation is increasing enormously.

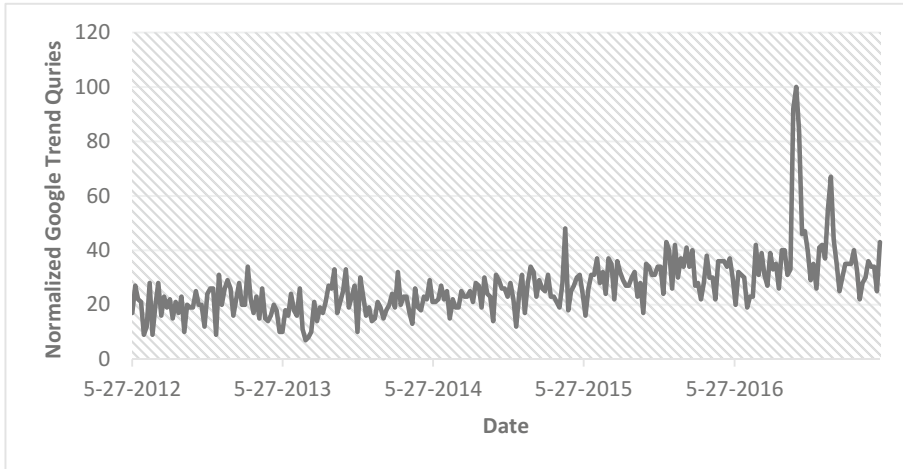
According to the World Bank's review report [5], the global Municipal Solid Waste (MSW) generation levels are expected to increase to approximately 2.2 billion tonnes per annum by the year 2025. Therefore, our traditional Waste Management system will be not able to handle such a huge volume of waste generated by cities; so "smart" bins are the need of time. If the municipal or corporation authorities are unable to manage garbage, then the garbage accumulation may become the root cause of illness and diseases such as diarrhea, dengue, etc. Degradation of garbage in open areas also causes bacterial and virus to grow, there by affecting the public health. Due to rapid urbanization, it is noticed that Waste Management became a crucial issue for municipal or corporation authorities as such bodies suffer from limited budgets and resource crunch. Traditional handling of garbage bins typically by human resources alone in large cities such as metros is no more a smart way of handling the Waste Management issues.

Many researchers and organizations conducted research on smart cities. Numerous interesting smart applications have already been implemented such as smart parking, smart transport, weather & environmental monitoring and many more. Among these applications, the smart Waste Management is considered to be a most important and yet challenging one.

We observed enough number of research articles and project implementations based on smart bin collection and monitoring system across the globe. The topic, smart bin gained popularity in last few years (Fig. 1), and the number of projects and research articles shows an increasing trend since 2012.

In this paper, we present a comprehensive survey of research papers in the area of smart monitoring of waste, smart bin collection and route optimization for garbage collection.

The rest of the paper is structured as follows. Section 2 outlines driving technologies of smart Waste Management system. Section 3 gives an overview of the existing and contemporary works in the area of Waste Management in Smart Cities, &



**Fig. 1.** Interest from Google research trends about smart bin.

Sect. 4 we propose a Framework for Smart Waste Management. Finally the Sect. 5 concludes with a brief summary of observations and for future scope.

## 2 Technology Support for Enabling Smart Waste Management Systems

Recent available technologies such as Pervasive Computing technologies, the Internet of Things (IoT), Cloud Computing, Big Data, and Wireless Sensor Networks (WSN) provides vast opportunities for researchers and developers to use these technologies the in area of smart bin collection system development. We observed a good number of research work in different aspect of smart garbage management such as smart monitoring of bin [13–27], selection of optimized routes for garbage collection [21–24], waste segregation [28]. Such systems typically use various recent technologies such as IoT [13–20], Integration of IoT and cloud computing [21–24], RFID [17–20], micro controller (Arduino Uno) [13–16], WSN [25–27], and Smart-M3 platform [30] to design and implement the smart bins.

The basic objective of Cloud Computing is to provide applications and services from data centers through the Internet to all over the world. The Cloud Computing provides services as the basis of pay-as-you-go model. The Cloud services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Infrastructure as a Service (IaaS) is a model that offers the Infrastructure. These cloud services may be useful for smart waste management system such as cloud storage used for storing real time sensor data, and CPU cycles used for processing of data.

The existing smart bins are equipped with various types of sensors that provides real time data for better management of resources used in waste management system. The Internet of Things refers to an idea of interconnection of uniquely identified physical objects (e.g. our daily usable items such as food, clothing, furniture, paper, etc.) via the Internet and with the help of standard protocols. These objects are typically connected through wireless technology such as Wi-Fi, Bluetooth, etc. On contrary to the objective of networking of computers in the conventional Internet; the IoT aims to provide network of any “things” (including PCs). IoTs have a huge number of potential applications, and these can be grouped into four domain, viz. Transportation and logistics, Healthcare, Smart environment (home, office, plant) and Personnel and social domain [6].

An object (garbage bin) can be attached with Radio-Frequency Identification (RFID) tag [7], which uses radio waves to read and capture information. RFID tags can store information and when compared with identification techniques such as bar code, RFID tag does not need direct line of sight for reading data. The RFID tags are widely used in smart bins for wireless data communication with the sink node.

The many smart bins implementations [13–16] used microcontroller board like Arduino Uno [8] (Uno a Italian word, which means one), Rasperi pi, LaunchPad, Nanode, Pinguino, STM32 Discovery, Teensy 2.0 etc., these boards uses limited power.

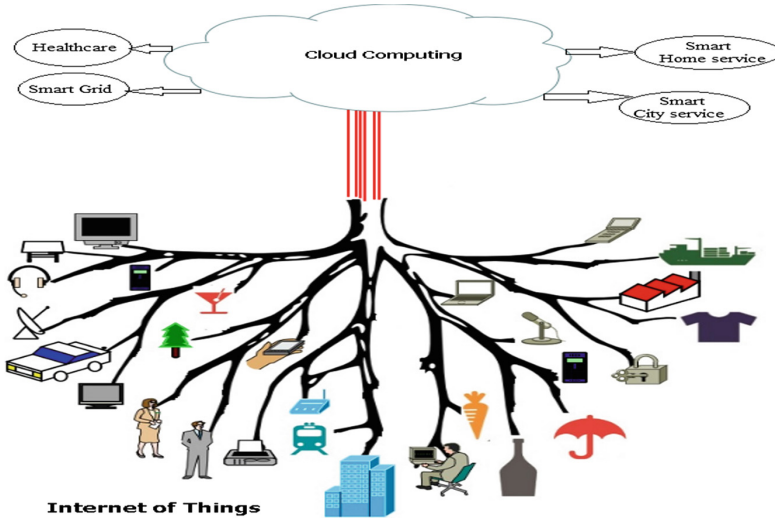
The number of interconnected devices (things) is increasing exponentially in last few years, and it is anticipated that the number will reach up to many trillions soon. These devices will generate a large volume of data. Indeed, IoTs will be one of the main sources of Big Data [9].

The IoTs and the Cloud computing technologies both have their own advantages and limitations. The characteristics of these two technologies complement each other when integrated in an application. As depicted in Fig. 2, the IoTs can be a benefit from the virtually unlimited capacities and resources of cloud to compensate its technological constraints (e.g. storage, processing, and communication) and Cloud can benefit from IoTs by extending its scope to deal with real-world things in a more distributed and dynamic manner, and for delivering new services in a large number of real scenarios [10]. The application of this integration are smart home, smart city, healthcare, smart grid etc. The good numbers of researchers [21–24] used this integration in smart waste management system.

A Wireless Sensor Network (WSN) [11] is a wireless network consisting of spatially distributed autonomous devices using sensors to monitor physical or environmental conditions (e.g. temperature, humidity, light, vibration, pressure, etc.) and designed to exchange their data through the network. The sensors connect the physical world with the digital world by monitoring and capturing the real-world conditions and converting these into a digital form. A wireless sensor node is capable of doing in-network analysis along with data collection with the help of on-board processing unit, chip radio and storage units. The WSN is used in various smart bins by many researchers [25–27].

Some of smart bin system is using Smart-M3 platform. Honkola et al. [12] proposed an open-source project called Smart-M3 platform that provides an environment in which different entities can share information and cooperate in a transparent way to the heterogeneities. The Smart-M3 as an information interoperability approach enables the devices to easily share and access local semantic information, while also allowing





**Fig. 2.** (M)any thing(s) generate data and the cloud provide services.

access to the locally relevant parts of the “giant global graph” to be available. The information is represented as semantic web, thus allowing easy exchange of global and local information.

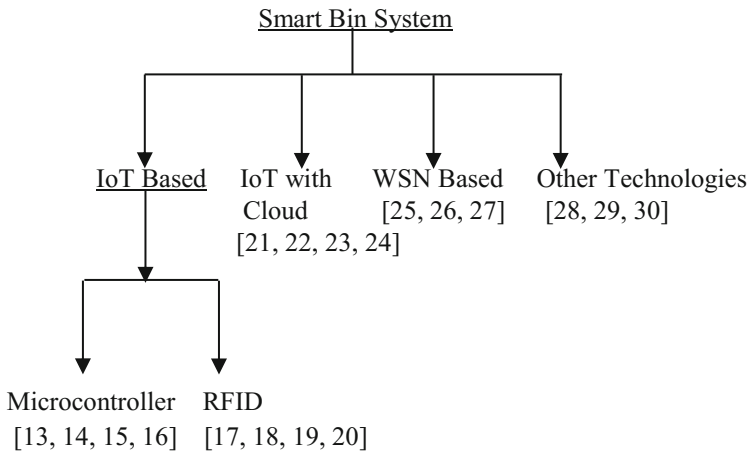
### 3 SWMS – A Comparative Study

Waste Management is a complex process that involves many steps: it includes generation, on-site handling and storage, collection, transfer, processing and disposal of solid wastes. The overall Waste Management involves three main types of entities:

- Users who generate waste,
- Waste collectors/city admin.
- Stakeholders.

As discussed in the earlier section, a huge number of research and projects have been done on various aspects of Waste Management. Such implemented projects typically use Arduino Uno micro controller based smart real time bin monitoring system [13–16], RFID technologies, Geographic Information Systems (GIS) and Geographic Positioning System (GPS) etc., Relatively all the works attempt to use cloud to access the stored sensor data and further analysis [21–24].

In general, we can classify smart bin system based on the applied technologies as follows:



### 3.1 IoT Based Approach

#### Microcontroller Based System

Yusof et al. [13], proposed an Arduino Uno micro controller based smart garbage monitoring system in order to measure the waste level in the garbage bin in real-time and when the garbage level is about to become full, then this system is designed to send an alert via SMS to the municipality so that the bin can be emptied, and garbage collected immediately. The system is equipped with ultrasonic sensor to measure the waste level, the GSM module to send the SMS, and an Arduino Uno which controls the system operation.

Adeyemo et al. [14], developed a smart city technology based architecture for refuse disposal management and implemented a proof of the concept prototype for the architecture. The system is again based on Arduino Uno micro controller board and in addition equipped with proximity sensor, refuse bin and a personal computer. The proximity sensors are attached on the five different position in a refuse bin and interfaced with the Arduino board with capture data set. To determine the appropriate classifier for realizing the pattern classification unit of the prototype, an experiment was performed using the acquired data set to train five different variants of the K-NN classifier.

Kumar et al. [15], developed a smart intelligent garbage alert system based on Arduino Uno interfaced with the ultrasonic sensor and if the garbage is filled, then this system sends the alert to the municipal web server. To perform the remote monitoring of the clearing process, an Android application is developed and linked to a web server to intimate the alerts from the system to the urban office. The notifications are sent by the Android application using Wi-Fi module.

Dugdhe et al. [16], proposed Waste Collection System architecture using the Internet of Things. The objective of this system is to schedule trucks by finding shortest path between the almost filled waste bins and bins, which have produced harmful gases

and gives an optimal route for garbage collection. The architecture consists of an embedded device with sensors and micro controller for sensing information of bins and sends to workstation. The system can also generate reports about waste gathering and fuel consumption for the logistics.

### **RFID Based System**

Chowdhury and Chowdhury [17], proposed a novel Automatic Waste Identity, Weight, and Stolen Bins Identification System (WIWSBIS), which is based on RFID and sensors. Using WIWSBIS, Waste Management service providers (e.g., municipalities, waste collectors) have a chance to track a waste identity (i.e., customer), weight missing/stolen bins quickly and accurately without any human intervention.

Issac and Akshai [18], introduced a system called SVASTHA (a Sanskrit word, which means “be healthy and hygienic”), which is based on RFID and GPS system that can effectively manage the Municipal Solid Waste. This embedded system has been developed to gather data from the RFID reader through Bluetooth’s connectivity and upload to a central server.

Wahab et al. [19], presented a smart recycle bin using RFID-based system integrating the web-based information system with the host server. This system works based on information in the smart card which automatically calculates the weight of waste and stores it into the card. This system assists the end user for waste classification and to know the status of smart bin.

RFID based Urban Solid Waste Collection system proposed by Karadimas et al. [20] uses ultrasonic sensors that provide ranging information, which is later translated to fill-level estimations.

## **3.2 IoT and Cloud Integrated Approach**

CloudSWAM, a cloud-based smart Waste Management mechanism [21] lets the sensor based garbage bins to notify its waste level status to the cloud. The stakeholders are able to access the desired data from the cloud and thereby perform route optimization for effective garbage collection.

With the integration of ultrasonic sensors and GPRS techniques, the micro controller based Dynamic Waste Management System [22], could measure the weight and volume of garbage and information sent to cloud server. Ant Colony Optimization (ACO) technique is used to find the shortest possible garbage collection route for each truck. The system is adaptable to dynamic changes, i.e. routes blocked during the waste collection process.

Another cloud computing based Smart Garbage Monitoring System [23] utilizes the concept of a network of smart garbage bins based on the Stack Based Front End approach of integrating WSN with the Cloud computing. To improve the efficiency of the garbage monitoring this approach uses Machine Learning techniques on the collected sensor data.

Use of surveillance systems as an assistive technology for high [24] quality of Service (QoS) in waste collection is proposed by Medvedev et al. Specifically, IoT components, viz. RFID tags, sensors, cameras, and actuators are incorporated into surveillance systems for efficient waste collection. This system consists of an advanced Decision Support System (DSS) for efficient waste collection and incorporates a model

for data sharing between truck drivers on real-time in order to perform waste collection and dynamic route optimization. Surveillance cameras are incorporated for capturing the problematic areas and provide evidence to the authorities concerned.

### 3.3 Wireless Sensor Network (WSN) Based Approach

Longhi et al. [25], proposed a Solid Waste Management Architecture based on sensor nodes and makes use of Data-Transfer Nodes. This WSN based system is used to control the filling of the bins, collecting data from embedded sensors. Remote monitoring on the web is facilitated as part of this system design. The system incorporates an integrated DSS for effective solid Waste Management by giving optimal solutions.

The Real-Time Solid Waste Bin Monitoring System Framework [26] is based on WSN and contains three levels: smart bin, gateway and control station that stores and analyzes the data for further use. The key objective of this framework is remote monitoring of solid waste bin in real time, via ZigBee-PRO and GPRS, to assist the solid Waste Management process. The waste collection route can be optimized by feeding the collected data into a decision support system.

The Smartbin system proposed by Folianto et al. [27] incorporates the concept of Wireless Mesh Network (WMN) and duty cycle techniques to reduce power consumption and to maximize operational time. The system is designed to collect data and to deliver the data through WMN. The Smartbin system has a three-tier architecture: outdoor nodes, analytics, and workstation. The Smartbin system was implemented and deployed on outdoor test bed.

### 3.4 Other Technologies

The main objective of GREENBIN [28] is segregation of waste at source so that the individual components of waste can be converted to useful electricity. The use of sensors like capacitive based moisture sensor, inductive based metal sensor, methane sensor and odour sensor helps to achieve this goal.

Thakker and Narayanamoorthi [29], introduced Smart Garbage Bin, which will alarm and inform the authorized person by SMS using GSM technology, when the garbage bin is about to fill. This system is designed to separate five types of plastic resins (which are not biodegradable) from garbage by using Near Infrared (NIR) spectroscopy and provides the details of all biodegradable substance that can be further used in bio gas plant.

Catania and Ventura [30], proposed an approach to smart waste collection based on Smart-M3 platform. The proposed approach can improve and optimize the handling of solid urban waste. The Smart-M3 platform helps solving the issue of interconnection among heterogeneous devices and data sharing involving a large amount of people. The real-time monitoring at the level of bin's fullness is made possible through sensors placed inside the containers. The data gathered is given for a decision system to determine the optimal number of waste vehicles or bins to distribute in the territory.

A comprehensive comparison of various waste management systems in terms of the sensors used, technologies applied and data transfer techniques used is presented in Table 1.



### 4 Proposed Framework for Smart Waste Management

To achieve a smart and efficient Waste Management System (WMS), we propose a framework that is capable for smart and smooth management of garbage in the city. Our proposed framework is illustrated in Fig. 3. If initially the garbage segregation is done, then it is easy to manage the garbage that lead to better WMS. The garbage segregation is performed by the citizen with the help of colour coded garbage bins. Basically, three colour coded garbage bins are green, blue and black respectively used for biodegradable waste, plastic & metal waste, and e-waste. The e-waste consists all electronic equipment, which are defective and obsolete like computers, TVs, cell phones, CD players, fax machines, printers etc. The various sensors that may be used in our proposed framework are listed in Table 2.

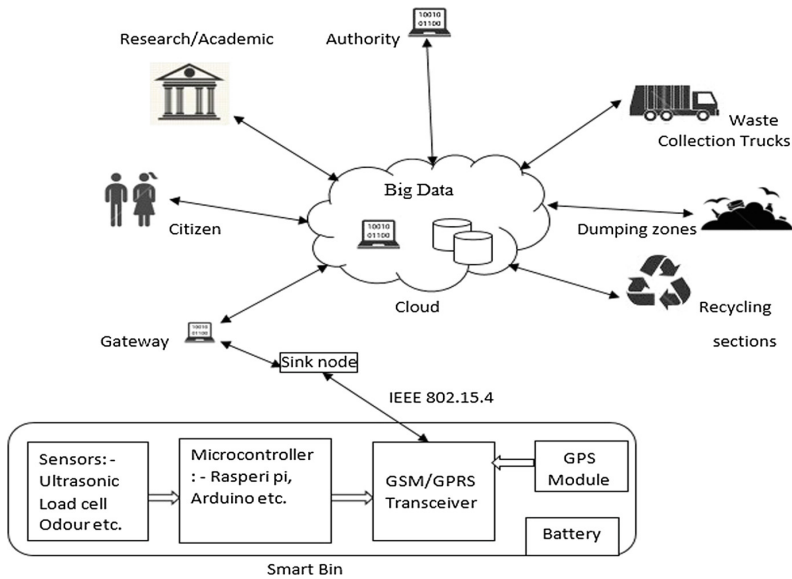


Fig. 3. The proposed framework for smart bin system.

Table 2. The different sensors used in colour coded garbage bins.

	Ultrasonic sensor	Load cell sensor	Proximity sensor	Gas equality sensor	Temp., humidity sensor	Metal sensor	Methane, odour sensor
Green bin	✓	✓	✓	✓	✓	✗	✓
Blue bin	✓	✓	✓	✗	✗	✓	✗
Black bin	✓	✓	✓	✓	✗	✗	✓

The green, blue and black bin may respectively equipped with various sensors that is showed in above Table 2. The sensors like methane, odour, temperature, and humidity is used in green bin because it contains biodegradable waste. The key objectives of these sensors to send alert to authorities, then they can send the garbage collecting vans as soon as possible. The ultrasonic sensor is used to identify the garbage level in garbage bin.

Big Data technologies and tools are used in the framework in order to handle the large volume information stored in the cloud.

This framework also contains GPS module so that their position is geographically identified that lead to make plan for optimal route for garbage collector vehicles across city. The microcontroller monitors operations of various sensors and controls voltage flows among sensors. The output of GPS and GSM/GPRS modules is controlled by microcontroller. The gateway receives data from smart bin with the help of IEEE 802.15.4 protocols such as ZigBee etc. and forward it to cloud storage. The cloud storage uses big data to analyze the huge amount of data and the result will be shared with different stakeholders such as city authority, research institute, dumping section, recycling section etc. This framework also provides a platform for citizen to give their feedback to authorities.

Following is the list of salient features of our proposed framework when compared with that of the existing works discussed in this paper:

- i. This framework is suggesting colour coded bin, which is equipped with various types of sensors. With the help of colour coded bin segregation of waste is done at collection point.
- ii. Our framework is capable to get real time data from various sensors and uses big data techniques & tools to get optimal result.
- iii. The GPS enabled smart bin in our proposed framework indeed helps arriving at optimal and efficient routes for garbage collection.

The following is a list of some, but not exhaustive open research problems applicable this framework:

- i. Developing energy efficient protocols at all layers of the network protocol stack such that the battery-powered embedded controllers mounted in the smart bins survive for a reasonable amount of life time.
- ii. Provisioning of security features in data collection that are susceptible to various attacks such as false-injection, black-hole and Denial of Service.
- iii. Handling the non-cooperative and novice citizens who do not follow the color codes while dumping the waste materials

## 5 Conclusion and Future Scope

The innovation in technologies via the Internet of things, cloud computing, big data, etc. provides opportunities for researchers to develop the smart bin system to realize “smart cities”. The importance of smart Waste Management cannot be ignored. In this paper, we discussed various enabling technologies and contemporary smart bin systems

with their advantages and limitations. This paper presented a framework that consist the smart bin and Garbage Management System. In future, this framework will be implemented and its performance studied through simulation and test beds.

## References

1. Hall, R.E., Bowerman, B., Braverman, J., Taylor, J., Todosow, H., Von Wimmersperg, U.: The vision of a smart city (No. BNL-67902; 04042). Brookhaven National Lab., Upton, NY, US (2000)
2. Research Institute for Housing, Urban and Mobility Studies (OTB): Smart cities ranking of European medium-sized cities. [http://smarcity-ranking.org/download/smart\\_cities\\_final\\_report.pdf](http://smarcity-ranking.org/download/smart_cities_final_report.pdf). Accessed 25 May 2017
3. Lewis, J.: The top 10 smart city applications of 2016. <http://www.telensa.com/2016/04/07/top-10-smart-city-applications-2016/>. Accessed 25 May 2017
4. Urban Population Growth. [http://www.who.int/gho/urban\\_health/situation\\_trends/urban\\_population\\_growth\\_text/en/](http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/). Accessed 25 May 2017
5. Hoornweg, D., Bhada-Tata, P.: What a waste: a global review of solid waste management. In: Urban Development Series Knowledge Papers, pp. 1–98. World Bank, Washington, DC, USA (2015)
6. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. *Comput. Netw.* **54**, 2787–2805 (2010). Elsevier
7. What is RFID? <http://www.epc-rfid.info/rfid>. Accessed 25 May 2017
8. Arduino UNO and Genuino UNO. <https://www.arduino.cc/en/main/arduinoBoardUno>. Accessed 25 May 2017
9. Dobre, C., Xhafa, F.: Intelligent services for big data science. *Future Gener. Comput. Syst.* **37**, 267–281 (2014). Elsevier
10. Botta, A., De Donato, W., Persico, V., Pescapé, A.: Integration of cloud computing and internet of things: a survey. *Future Gener. Comput. Syst.* **56**, 684–700 (2016). Elsevier
11. What Is a Wireless Sensor Network? <http://www.ni.com/white-paper/7142/en/>. Accessed 25 May 2017
12. Honkola, J., Laine, H., Brown, R., Tyrkkö, O.: Smart-M3 information sharing platform. In: IEEE Symposium in Computers and Communications, ISCC, Riccione, Italy, pp. 1041–1046. IEEE (2010)
13. Yusof, N.M., Jidin, A.Z., Rahim, M.I.: Smart garbage monitoring system for waste management. In: MATEC Web of Conferences Engineering Technology International Conference, vol. 97, p. 01098. EDP Sciences (2017)
14. Adeyemo, J.O., Olugbara, O.O., Adetiba, E.: Smart city technology based architecture for refuse disposal management. In: IST-Africa Week Conference, Durban, South Africa, pp. 1–8. IEEE (2016)
15. Kumar, N.S., Vuyalakshmi, B., Prarthana, R.J., Shankar, A.: IOT based smart garbage alert system using Arduino UNO. In: Region 10 Conference, TENCON, Singapore, pp. 1028–1034. IEEE (2016)
16. Dugdhei, S., Shelar, P., Jire, S., Apte, A.: Efficient waste collection system. In: International Conference on Internet of Things and Applications, IOTA, Pune, India, pp. 143–147. IEEE (2016)
17. Chowdhury, B., Chowdhury, M.U.: RFID-based real-time smart waste management system. In: Telecommunication Networks and Applications Conference, ATNAC, Christchurch, New Zealand, pp. 175–180. IEEE (2007)



18. Issac, R., Akshai, M.: SVASTHA: an effective solid waste management system for Thiruvalla Municipality in Android OS. In: Global Humanitarian Technology Conference: South Asia Satellite, GHTC-SAS, Trivandrum, India, pp. 254–259. IEEE (2013)
19. Wahab, M.H.A., Kadir, A.A., Tomari, M.R., Jabbar, M.H.: Smart recycle bin: a conceptual approach of smart waste management with integrated web based system. In: International Conference on IT Convergence and Security, ICITCS, Beijing, China, pp. 1–4. IEEE (2014)
20. Karadimas, D., Papalambrou, A., Gialelis, J., Koubias, S.: An integrated node for SmartCity applications based on active RFID tags; use case on waste-bins. In: 21st International Conference on Emerging Technologies and Factory Automation, ETFA, Berlin, Germany, pp. 1–7. IEEE (2016)
21. Aazam, M., St-Hilaire, M., Lung, C.H., Lambadaris, I.: Cloud-based smart waste management for smart cities. In: 21st International Workshop on Computer Aided Modelling and Design of Communication Links and Networks, CAMAD, Toronto, ON, Canada, pp. 188–193. IEEE (2016)
22. Sharmin, S., Al-Amin, S.T.: A Cloud-based dynamic waste management system for smart cities. In: Proceedings of the 7th Annual Symposium on Computing for Development, p. 20. ACM (2016)
23. Joshi, J., Reddy, J., Reddy, P., Agarwal, A., Agarwal, R., Bagga, A., Bhargava, A.: Cloud computing based smart garbage monitoring system. In: 3rd International Conference on Electronic Design, ICED, Phuket, Thailand, pp. 70–75. IEEE (2016)
24. Medvedev, A., Fedchenkov, P., Zaslavsky, A., Anagnostopoulos, T., Khoruzhnikov, S.: Waste management as an IoT-enabled service in smart cities. In: Balandin, S., Andreev, S., Koucheryavy, Y. (eds.) ruSMART 2015. LNCS, vol. 9247, pp. 104–115. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-23126-6\\_10](https://doi.org/10.1007/978-3-319-23126-6_10)
25. Longhi, S., Marzioni, D., Alidori, E., Di Buo, G., Prist, M., Grisostomi, M., Pirro, M.: Solid waste management architecture using wireless sensor network technology. In: 5th International Conference on New Technologies, Mobility and Security, NTMS, Istanbul, Turkey, pp. 1–5. IEEE (2012)
26. Al Mamun, M.A., Hannan, M.A., Hussain, A.: Real time solid waste bin monitoring system framework using wireless sensor network. In: International Conference on Electronics, Information and Communications, ICEIC, Kota Kinabalu, Malaysia, pp. 1–2. IEEE (2014)
27. Folianto, F., Low, Y.S., Yeow, W.L.: Smartbin: smart waste management system. In: Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP, Singapore, pp. 1–2. IEEE (2015)
28. Rajkamal, R., Anitha, V., Nayaki, P.G., Ramya, K., Kayalvizhi, E.: A novel approach for waste segregation at source level for effective generation of electricity—GREENBIN. In: International Conference on Science Engineering and Management Research, ICSEMR, Chennai, India, pp. 1–4. IEEE (2014)
29. Thakker, S., Narayanamoorthi, R.: Smart and wireless waste management. In: International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS, Coimbatore, India, pp. 1–4. IEEE (2015)
30. Catania, V., Ventura, D.: An approach for monitoring and smart planning of urban solid waste management using smart-M3 platform. In: Proceedings of 15th Conference of Open Innovations Association FRUCT, St. Petersburg, Russia, pp. 24–31. IEEE (2014)

# **Web of Things**

# Contextual Pattern Clustering for Ontology Based Activity Recognition in Smart Home

K. S. Gayathri<sup>1</sup>(✉), K. S. Easwarakumar<sup>2</sup>, and Susan Elias<sup>3</sup>

<sup>1</sup> Department of Computer Science and Engineering,  
Sri Venkateswara College of Engineering, Sriperumbudur, India  
gayasuku@svce.ac.in

<sup>2</sup> Department of Computer Science and Engineering,  
Anna University, Chennai, India

<sup>3</sup> School of Electronics Engineering, VIT University, Chennai, India

**Abstract.** Ambient Assisted Living (AAL) enabled in a smart home requires, the design of an activity recognition system. Generally, supervised machine learning strategies or knowledge engineering strategies are employed in the process of activity modeling. Supervised machine learning approaches incur overheads in annotating the dataset, while the knowledge modeling approaches incur overhead by being dependent on the domain expert for occupant specific knowledge. The proposed approach on the other hand, employs an unsupervised machine learning strategy to readily extract knowledge from unlabelled data and subsequently represents it as ontology activity model. The novelty in the proposed design is in the usage of Contextual Pattern Clustering (CPC) for activity modeling. The competence of the weighted Jaro Winkler similarity measure introduced in CPC lies in the utilization of contextual attributes for the composition of varied event patterns of an activity. Hierarchical strategy employed in CPC offers structured knowledge on activities and sub activities within a specific location. Additionally, the event organizer and habitual event generator subsystem introduced in the proposed framework derives knowledge related to event ordering and contextual description of an activity. The attained knowledge is later represented as a probabilistic ontology activity model to enable probabilistic reasoning over domain knowledge. An experimental analysis with a smart home dataset demonstrates the proficiency of the proposed unsupervised approach in activity modeling and recognition in comparison with that of the existing modeling strategies.

**Keywords:** Smart home · Activity recognition · Pattern clustering  
Probabilistic ontology

## 1 Introduction

Smart home designs aim at integrating intelligence into the surroundings to provide context aware services for better quality living [13]. Innovations in the field

of sensor technology, wireless communication, Internet of things, machine learning and ubiquitous computing have driven the design of smart homes to be a significant research area [2]. Ambient Assisted Living (AAL) offered by smart home assists the occupant autonomously in their habitual living impelling smart home to model various societal applications. Smart home mainly recognizes the ongoing activity of the occupant in order to identify abnormality or to automate events in the environment. This feature enables smart home to be deployed in variety of applications like hospitals, offices, classrooms, health care, elderly care and critical care [7, 13]. To enable AAL within smart home, it is necessary to perform activity modeling and recognition [2]. Activities of Daily Living (ADL) are generally considered for modeling and are the routine activities of the occupant, executed in complex patterns [3]. Activities in smart homes are mostly performed differently each time by the same occupant. The activity modeling strategy should thus proficiently capture this varied event patterns in the design so as to enhance the recognition accuracy. Furthermore, these event patterns are characterized with a set of sensors and preconditions required to execute an activity under various scenario [4].

Major issue in the design of activity recognition system lies in modeling uncertain data and contextual information [8]. The sensor data obtained during run time are imprecise in nature; hence uncertainty modeling should be enabled in the activity recognition framework [2, 3]. Simple activity in a smart home is a sequence of atomic event that is contextually described in terms of object used, location and time. The contextual description of events together with relative event ordering is crucial in the design of activity model [8]. Therefore the activity modeling and recognition system should enable context based reasoning within its framework [6]. The broad approaches for activity modeling are machine learning and knowledge engineering strategies [5], where machine learning strategy models from the dataset and knowledge engineering strategy models from the domain knowledge. Machine learning is further categorized as supervised and unsupervised approach based on the dataset considered for modeling [3]. The occupant specific knowledge are better modeled through machine learning strategy and general domain specific knowledge on routine activities are modeled effectively through knowledge engineering strategy. Furthermore temporal, probabilistic reasoning are presented well through machine learning strategies and context based reasoning are presented well through knowledge engineering strategies. Therefore, a hybrid data driven and knowledge driven approach is required for effective activity modeling.

The challenge in employing machine learning strategy in the smart home design appears in labeling the dataset. Voluminous sensor data is produced every moment in the environment creating an overhead for the annotator by labeling each single atomic event. Moreover, annotation is done manually and hence prone to errors. Therefore, an unsupervised machine learning strategy is preferred over supervised approach. To enable context based reasoning, ontology based modeling is considered. Ontology plays a significant role in modeling smart home applications [6]. Ontology is preferred over other knowledge representation

techniques because of its unified representation in modeling and semantically clear reasoning [6]. The customary practice in designing ontology is to gather knowledge from domain experts and later model the ontology structure [5]. But the restriction lies in obtaining occupant specific knowledge from the domain expert. Therefore, the proposed system brings in innovation by constructing the probabilistic ontology activity model from the knowledge derived from smart home dataset by executing an unsupervised machine learning algorithm.

The novelty in the proposed design of activity modeling and recognition framework lies in the:

1. strategy of unsupervised machine learning (Contextual Pattern Clustering) to extract activity patterns from sensor dataset
2. weighted contextual Jaro Winkler similarity measure introduced in clustering to evaluate the likeness of complex event patterns
3. approach of knowledge extraction to derive event ordering and dominant contextual definition of an activity
4. construction approach of probabilistic ontology activity model.

The remaining part of the paper is structured as follows: Sect. 2 reviews the related work on smart homes, presents various approaches for activity recognition and defines the motivation and scope of this paper. Section 3 describes the theoretical foundations of the proposed unsupervised approach of activity modeling and recognition framework. Section 4 describes the system prototype, experimental analysis and performance evaluation. Finally, Sect. 5 concludes the paper and outlines the future work.

## 2 Related Work

The strength of ambient assisted living offered by a smart home depends on the strategy of activity modeling and recognition. The literature review presented in this section details on the existing activity modeling systems and later justifies the need for proposed design. Depending on the kind of sensors employed in monitoring, activity recognition is categorized as vision based and sensor based technique [5]. Video cameras are employed in vision based approach to monitor the occupant. The collected video frames are analyzed through image processing techniques to recognize the ongoing activity of the occupant. Though, vision based activity recognition has several advantages in modeling smart environment it has its own setbacks while engaged in the design of smart home. First and foremost, privacy of occupant is a key concern in a home set up and thus camera is not a suitable means to monitor the occupant for recognition. Every frame of the captured video is to be analyzed for recognition hence, it is a complex process. Moreover, the vision based activity model built for one environment may not work well in another environment. Hence, reusable and scalable activity models could not be effectively developed through vision based technique. On the other hand, sensor based approaches monitors the occupant using varieties of miniature sensors which are either fixed on physical objects or worn by the occupant.

Primarily, sensor based technique defines an activity in terms of its object usage while additional body worn sensors are used in health care applications. The drawback of sensor based approach is the ease of use and battery life. Sensor based activity recognition is categorized as data driven and knowledge driven based on activity modeling approach [5]. Utilization of data mining, probabilistic and statistical methods is the core of data driven approach [1] and application of knowledge engineering and management techniques in activity modeling is the core of knowledge driven paradigm.

## 2.1 Data Driven Approach to Activity Recognition

Data driven approach are further categorized as generative and discriminative approach [1, 3, 5], where generative approaches uses probabilistic model to depict the input sensor data space and discriminative approaches represents activity as a classification model. Naive Bayes classifier (NB), Hidden Markov Model (HMM) and Dynamic Bayesian Networks (DBM) are the generative approaches used in the design of smart home. Whereas, Nearest Neighbor (NN), Support Vector Machines (SVM), Artificial Neural Network (ANN) and boosting techniques are the most commonly used discriminative approach [5]. Both generative and discriminative approaches are supervised machine learning strategy that requires labeled dataset for effective modeling. The general procedure to annotate the data is to manually record the dependency between atomic events and activity labels. Since, the occupant is continuously monitored in smart home; voluminous sensor dataset is produced. Labeling such sensor data creates an overload to the annotator and are prone to error. Therefore, unlabeled dataset is preferred for activity modeling in smart home and is enabled through unsupervised machine learning techniques.

Clustering is an unsupervised approach of activity recognition that groups related data together [1]. K-means clustering to detect abnormality in start time and duration of an activity is utilized in the existing system of [10]. This clustering approach organizes the objects into 'k' partitions based on the similarity measure. Efficiency of k-means algorithm depends upon the number of clusters, selection of cluster center, number of iterations involved in modeling. Dense based clustering is yet another approach in smart home that groups data based on the density. Since, the activity of the occupant is generally performed as sequence of events; the traditional data point clustering approach does not perform well in smart home. Hence clustering technique requires to group patterns rather than individual data points. The difficulty in pattern clustering is in defining the similarity measure to discover association between related complex activities. Various similarity measures like Euclidean distance, Manhattan distance and Levenshtein edit distance were used in pattern clustering, but the constraint is that, it is not capable to capture similarity between two event sequences that are composite in nature. Few existing systems on smart home have employed pattern clustering [14]. The issue with these systems is that it fails to model the varied event pattern behavior of an occupant. This limitation was effectively addressed in a EPAM system [9]. But this pattern clustering did not consider the

contextual attributes for measuring the similarity between event patterns which are very essential in describing an activity in a smart home. Additionally, event ordering among atomic events are also crucial to define activity, as activities are primarily described as a sequence of atomic events. This knowledge related to event ordering is not derived and modeled in this existing system. Moreover, the dominant contextual description of an activity is also not derived and represented. The ontology representation of activity model in this system does not perform uncertainty reasoning to recognize the ongoing activity of the occupant. Thus, the essential features of varied event pattern modeling, contextual modeling and uncertainty modeling need to be addressed.

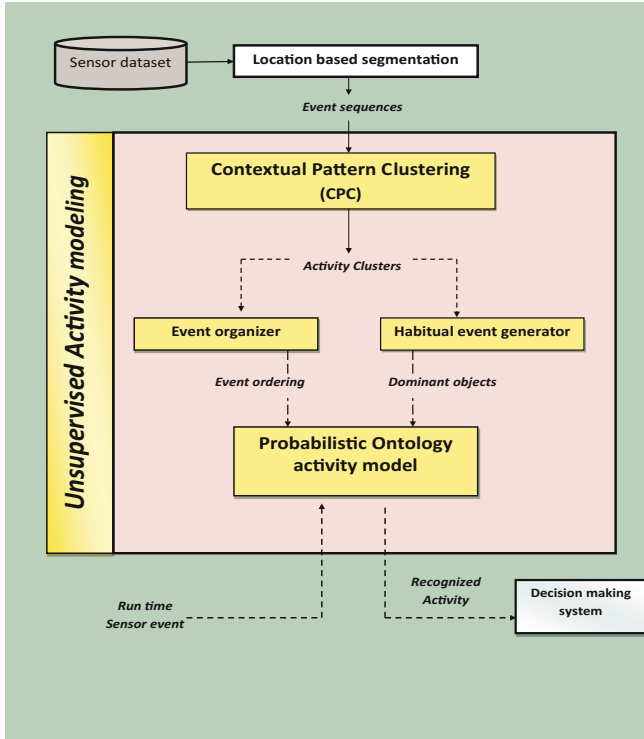
## 2.2 Knowledge Driven Approach to Activity Recognition

Knowledge driven approach constructs an activity recognition system as a reusable contextual model that represents and defines the knowledge related to the occupant behavior as relationship with objects, space, and time [5, 6]. Ontology activity models are preferred than first order logic because of the following reasons: Ontology is independent of algorithm choices thus facilitating portability, interoperability and reusability. Ontology based modeling is supported by a solid technological infrastructure. There are numerous technologies, tools and APIs that are offered to execute each task in ontology based approach, for example ontology editors for context and activity modeling, web ontology language for activity representation, semantic repository technologies for large scale semantic data management. The support of infrastructure gives ontology based approaches huge advantage in large scale adoption, application development and system prototyping.

Thus the proposed approach represents the activity model using ontology. However, imprecision could arise in ontology modeling due to inappropriate collection of occupant specific knowledge. So, the proposed system addresses this issue by extracting occupant specific knowledge automatically from the dataset using an unsupervised machine learning (pattern clustering) approach thereby integrating benefits of data driven approach and knowledge driven approach for activity modeling.

## 3 Proposed Unsupervised Activity Modeling Framework

The proposed unsupervised framework for probabilistic ontology activity modeling is presented in the Fig. 1. Sensor dataset intended for training the activity model is segmented into location based event sequences. A contextual pattern clustering module models activities from these event sequences and represents them using ontology. The hierarchical strategy introduced within the pattern clustering, groups activities with sub activities. This approach aids in deriving the inherent hierarchy for effective activity modeling. The clustered event patterns are further utilized to extract knowledge about event ordering and contextual description of each activity. The extracted knowledge is eventually



**Fig. 1.** Proposed unsupervised approach of activity modeling and recognition

represented as probabilistic ontology activity model. During runtime, the sensor events are presented to the modeled probabilistic ontology for activity recognition and the recognized activities are passed on to the decision making system for suitable actions.

### 3.1 Contextual Pattern Clustering (CPC)

The event sequences generated from sensor data through location segmentation are to be analyzed and grouped based on certain characteristics of the event patterns. Location based segmentation is preferred, as most of occupant activities are specific with respect to a spatial location. The Algorithm 1 describes the proposed contextual pattern clustering that follows an unsupervised approach in activity modeling. In a smart home, each event is associated with contextual entities such as object, location and time (part of the day). Thus, the pattern clustering should consider contextual attributes collectively rather than independently. For instance, when object in use is independently taken into consideration there could be ambiguity introduced by location and time of the event. Example ‘mug’ object located in kitchen is a varied activity than using ‘mug’ in bathroom location. Similarly using ‘mug’ in midnight is a varied activity than using ‘mug’



in morning in the same bathroom location. Thus the proposed pattern clustering considers all the contextual attributes together to measure the similarity of an event sequence and incorporated in Algorithm 1 (lines 6 to 8).

---

**Algorithm 1.** Context Pattern Clustering *CPC*


---

```

1 Function Contextual Pattern Clustering (EP)
   Input : Set of event patterns EP
   Output: set of activity clusters AC that includes a group of event
           patterns
2
3 begin
4   foreach event pattern  $e_i$  of EP do
5     foreach event pattern  $e_j$  of EP do
6        $JWO_{ij}$  = Calculate object based JW similarity between  $e_i$ 
           and  $e_j$ 
7        $JWL_{ij}$  = Calculate location based JW similarity between  $e_i$ 
           and  $e_j$ 
8        $JWT_{ij}$  = Calculate temporal based JW similarity between  $e_i$ 
           and  $e_j$ 
9        $WJW_{ij} = w_1 * JWO_{ij} + w_2 * JWL_{ij} + w_3 * JWT_{ij}$ 
           // where WJW represents weighted Jaro Winkler
           similarity
           // w represents the weights
10      end
11    end
12    set each event pattern  $e_i$  of EP as cluster pattern center
13    set clusnum = count of event patterns EP
14    repeat
15      merge contextual closest event patterns
16      update the WJW similarity measure for every event pattern
17      reduce clusnum by one
18    until clusnum == 1
19 end

```

---

$$d_j = \begin{cases} 0, & \text{if } m \text{ is } 0 \\ \frac{1}{3} \left( \frac{m}{|S_1|} + \frac{m}{|S_2|} + \frac{m-t}{m} \right) & \text{otherwise} \end{cases} \quad (1)$$

$$d_w = d_j + (l * p(1 - d_j)) \quad (2)$$

$$w_l = \frac{\max(|S_1|, |S_2|)}{2} - 1 \quad (3)$$

The Jaro Winkler (*JW*) similarity measure is given in Eq. 2. This similarity measure in comparison to other approaches (such as Levenshtein distance, Edit distance, Optimal String Alignment) provides an additional feature that allows

transposition of events within a permitted window length. The proposed approach extends the Jaro Winkler similarity to include other contextual entities for comparison. The Jaro distance measure utilized in  $JW$  is given in Eq. 1, outputs 1 for equal event patterns and 0 for complete dissimilar event patterns. The window limit within which the discontinuity is allowed is given in Eq. 3. In all the equations,  $S_1$  and  $S_2$  represent event patterns,  $l$  represents length of common prefix,  $p$  represents scaling factor,  $m$  represents number of matching events and  $t$  represents number of transpositions. The proposed weighted contextual Jaro Winkler similarity measure is given in line 9 of the Algorithm 1. If the similarity is measured only with location and temporal information it is difficult to learn the intrinsic behavior of an activity and thus the object becomes the significant contextual attribute in pattern clustering. Hence, weights are assigned to contextual attributes based on their importance in describing an activity. Thus a weighted Jaro Winkler similarity measure is proposed that compares event patterns based on the contextual attributes with weights. Among the contextual attributes, object is given a higher weight than location and time. The reason for assigning the least weight for time attribute because of its relatively insignificant contribution in defining the activity itself. Since the events patterns are specific to a location, hierarchical pattern clustering approach is preferred to find the activities and its sub activities specific to a location. The hierarchical clustering strategy considers each pattern as a separate group primarily and later groups these sequences based on the similarity measure as described in lines 12 to 18 of Algorithm 1. Thus, the knowledge related to activity hierarchy is obtained by executing contextual pattern clustering algorithm.

---

**Algorithm 2.** Event Organizer
 

---

```

1 Function Event organizer ( $AC$ )
  Input : set of activity clusters  $AC$  produced from  $CPC$ 
  Output: events ordering  $EO$  for each activity cluster in  $AC$ 
2
3 begin
4   foreach Cluster  $AC_i$  do
5      $EO_i = e_1$  in  $AC_i$  // Initialization of EO to the first event
       pattern in  $AC_i$ 
6     foreach event pattern  $e_j$  in Cluster  $AC_i$  do
7        $EO_i = \text{LCS}(EO_i, e_j)$  // LCS represents Longest Common
           Subsequence function
8     end
9   end
10 end

```

---

### 3.2 Event Organizer

Event organizer as presented helps in finding out the exact ordering of events among a set of event patterns. For example ‘using remote’ can happen only after ‘switching on the TV’. Such kind of total ordering can be obtained through this event organizer module and presented in Algorithm 2. Once the contextual pattern clustering is over, each activity group contains a set of event patterns. Meaningful facts are to be obtained from these event patterns in each group. This module focuses on discovering the common ordering patterns among event sequences that is, the precedence of one event over another event could be identified. The total event ordering is determined using ‘Longest Common Subsequence’ algorithm that identifies the longest set of event patterns that has the same ordering as shown in lines 6 to 8 of Algorithm 2. Thus the dependency between the objects used in performing an activity is identified. The captured ordering is incorporated into the ontology activity model to define the simple activity recognition system over event sequences in a unsupervised manner.

---

#### Algorithm 3. Habitual Event Generator

---

```

1 Function Habitual event generator ( $AC$ )
  Input : set of activity clusters  $AC$  produced from  $CPC$ 
  Output: Dominant Contextual Attributes  $DA$  for each activity cluster in  $AC$ 
  //  $DAO$ ,  $DAL$  and  $DAT$  represents Dominant Attribute for contextual
  // Object, Location and Time respectively
2 begin
3   foreach  $Cluster AC_i$  do
4     // Initialization of Dominant Attribute description for each
4     // activity cluster
4      $DAO_i =$  Object description,  $DAL_i =$  Location description,  $DAT_i =$ 
4     Temporal description of event pattern  $e_1$  in  $AC_i$ 
5     foreach  $event\ pattern\ e_j$  in  $ClusterAC_i$  do
6        $DAO_i =$  Object description of  $e_j \cap DAO_i$ 
7        $DAL_i =$  Location description of  $e_j \cap DAL_i$ 
8        $DAT_i =$  Temporal description of  $e_j \cap DAT_i$ 
9     end
10  end
11 end

```

---

### 3.3 Habitual Event Generator

It is also essential to identify the general objects used to perform an activity and the habitual event generator extracts it using Algorithm 3. This can be determined by identifying the common objects used in the various event patterns in a particular group as shown in line 6 of Algorithm 3. These common objects

can be considered to be dominant objects required by the particular activity. This could be extended over the location and time attribute so that the habitual location and part of the day can be identified in executing an activity as given in line 8 and 9 of Algorithm 3. The dependable objects needed to define an activity are thus obtained through the proposed approach automatically.

### 3.4 Probabilistic Ontology Activity Model

The extracted knowledge from contextual pattern clustering, event ordering generator, and habitual event generator are collectively utilized to define the ontology activity model. Terminology Box (TBox) of the ontology describes a structure for the activity model through concepts, relations and properties while, Assertion Box (ABox) of the ontology describes real world facts through labeled individuals. The ADLs form the concepts and are associated with roles (properties) in ontology that elucidate the characteristics of an activity [8]. The framework of the probabilistic ontology model presented in [8] is utilized in this work, where each activity is defined in terms of atomic events that are associated with objects, location and time (part of day) which is extended for simple activity modeling by annotating confidence in ontology based on event ordering.

## 4 Experimental Analysis

The proposed system is modeled and evaluated with a smart home dataset obtained from UCI machine learning repository [12]. This binary sensor dataset contextually describe (start time, end time, object, location, sensor triggered) each atomic event and associates it with an activity label. Since the proposed approach is an unsupervised technique, the activity labels are purely used to evaluate the process of contextual clustering and activity recognition, not for activity modeling. The contextual pattern clustering, event generator and habitual event generator subsystems of the proposed design are executed using ‘R’ language [16]. Protege [11], an open source tool for ontology modeling is utilized for the representation of probabilistic ontology activity model. For effective activity modeling, the experimentation is carried out as a threefold cross validation that considers 66% of sensor data as training dataset and the remaining 34% as test dataset. The process of activity modeling is evaluated in terms of accuracy and F-measure. Accuracy appraise the precision in grouping event sequences and F-measure evaluates the efficiency in activity recognition. The atomic events are segmented based on its location and as a result 253 location based event sequences were generated for the considered sensor dataset. Each of these event sequences are subsequently given to the pattern clustering for activity modeling. Weighted contextual Jaro Winkler similarity measure is used in pattern clustering, where weights are assigned to the contextual attributes based on its accuracy in clustering event patterns. After several rounds of experimentation, the object contextual attribute is associated with 70% weight, location with 20% weight and temporal entity with 10% weight.

## 4.1 Experimental Evaluation

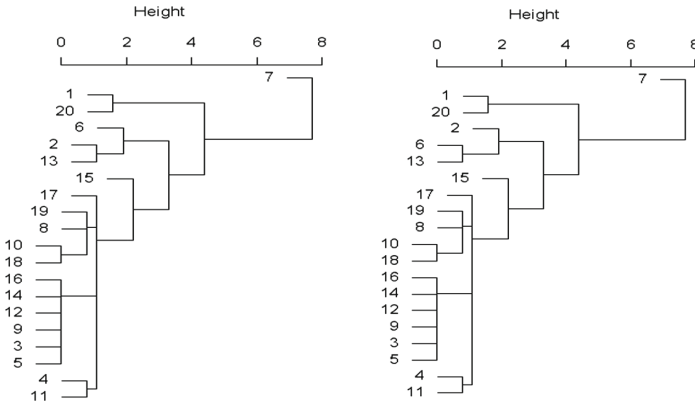
The efficiency of the pattern clustering approach lies in the similarity measure used for grouping and hence the first set of evaluation is done over various pattern similarity measures. Levenshtein distance (*LV*), Optimal String Alignment (*OSA*), Q-grams (*QG*) are the similarity measures considered for comparison. Each of these measures is evaluated in terms of its effectiveness in capturing similarity among varied event patterns of related activities. To demonstrate the efficiency of the proposed system, a sample of event sequences (obtained from dataset) is considered as shown in the Fig. 2. The visualization of hierarchical ordering among these samples of event sequences using contextual pattern clustering is presented in the Figs. 3, 4 and 5. It is evident from Fig. 4 that the event patterns are better grouped in the JW measure than that of other similarity measures. It is still more observed from Fig. 5 that the weighted contextual JW measure groups better than the object based JW similarity measure. The proposed similarity measure has enhanced mapping of event sequences to its corresponding activity as it proficiently captures inherent contextual characteristic of similar activities of varied event pattern through transposition features that allows disposition of events within a specific window bound. Though *LV*, *OSA* and *QG* are utilized in various string comparison applications, they fail in smart home modeling as the occupant performs an activity in a diverse manner every time.

EVENT SEQUENCE	ACTIVITY LABEL	NOTATION	OBJECT
[1,] "EKFH"	[1,] "Breakfast"	A	Basin
[2,] "CAL"	[2,] "Grooming"	B	Bed
[3,] "I"	[3,] "Spare_Time/TV"	C	Cabinet
[4,] "F"	[4,] "Snack"	D	Cooktop
[5,] "I"	[5,] "Spare_Time/TV"	E	Cupboard
[6,] "LA"	[6,] "Toileting"	F	Fridge
[7,] "EDHEFEF"	[7,] "Lunch"	G	Maindoor
[8,] "A"	[8,] "Grooming"	H	Microwave
[9,] "I"	[9,] "Spare_Time/TV"	I	Seat
[10,] "L"	[10,] "Toileting"	J	Shower
[11,] "E"	[11,] "Snack"	K	Toaster
[12,] "I"	[12,] "Spare_Time/TV"	L	Toilet
[13,] "AL"	[13,] "Toileting"		
[14,] "I"	[14,] "Spare_Time/TV"		
[15,] "GG"	[15,] "Leaving"		
[16,] "I"	[16,] "Spare_Time/TV"		
[17,] "B"	[17,] "Sleeping"		
[18,] "L"	[18,] "Toileting"		
[19,] "J"	[19,] "Showering"		
[20,] "EFHK"	[20,] "Breakfast"		

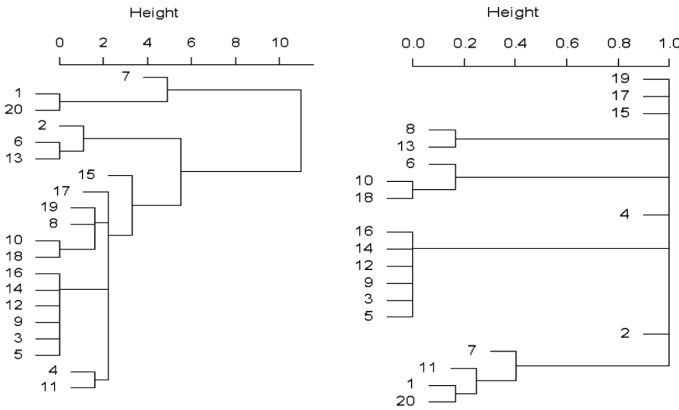
Fig. 2. Sample event sequence from sensor dataset

The Table 1 compares the accuracy of various similarity measures. It is observed that Jaro Winkler measure combined with contextual pattern clustering offer better accuracy in comparison with other measures. The reason for

high accuracy is the ability of *JW* to estimate the similarity in varied event patterns onto an activity using a transposition feature that allows disposition of events. Moreover, the weighted contextual similarity measure compares the events collectively with object, location and time. The weights assigned in similarity measure were good enough to learn the inherent characteristics of an activity in terms of its contextual attributes.

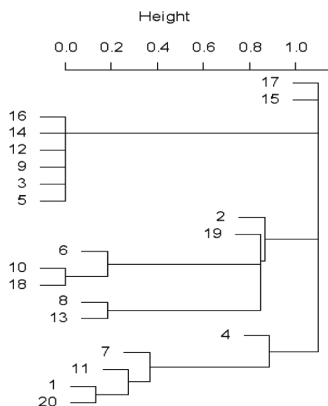


**Fig. 3.** Hierarchy of event patterns using weighted contextual LV and OSA



**Fig. 4.** Hierarchy of event patterns using weighted contextual QG and JW (only with object)

The hierarchical organization of patterns, total ordering among events and contextual definition of activities are eventually modeled for activity recognition. Probabilistic ontology approach is preferred for activity modeling to enable probability reasoning over domain ontology. The extracted knowledge on hierarchical organization and contextual description of activities are utilized to construct



**Fig. 5.** Hierarchy of event patterns using weighted contextual JW

**Table 1.** Accuracy of various similarity measure in contextual pattern clustering

	Contextual pattern clustering			
	LV	OSA	QG	JW
Breakfast	0.65	0.56	0.68	1
Grooming	0.64	0.54	0.62	0.94
Leaving	0.67	0.56	0.65	1
Lunch	0.61	0.55	0.62	0.93
Showering	0.65	0.58	0.67	1
Sleeping	0.67	0.59	0.68	1
Snack	0.64	0.57	0.65	0.98
SpareTimeTV	0.65	0.53	0.66	1
Toileting	0.55	0.57	0.71	0.95

the *TBox* of the ontology and knowledge derived on event ordering is utilized to enable probability reasoning within ontology. Annotations in ontology are generally performed with domain knowledge, but the proposed system annotates with the knowledge derived from pattern clustering. High confidence is assigned to the derived event ordering so as to enhance the recognition accuracy. The probabilistic ontology activity model is evaluated and compared with other activity modeling techniques. Hidden Markov Model (HMM) is preferred in data driven approach and domain ontology modeling is preferred in knowledge driven approach for comparison. Thus, experiments are extended by modeling HMM and ontology activity model with the preferred smart home dataset as presented in [6,15] respectively. From the Table 2, it is observed that the F-measures of the existing systems are low when compared to the proposed system. The reason for low F-measures is that contextual reasoning is not performed in HMM and probabilistic reasoning, occupant specific modeling is not appropriately done in

domain ontology. F-measures are high for the proposed system for the reason that it integrates unsupervised machine learning, context based reasoning and probabilistic reasoning within a single framework for effective activity modeling.

**Table 2.** Comparison between existing and proposed activity recognition system

	Activity model (F-measure)		
	HMM	Ontology	Proposed ontology
Breakfast	0.81	0.72	0.91
Grooming	0.88	0.78	0.94
Leaving	0.85	0.74	0.95
Lunch	0.83	0.73	0.91
Showering	0.87	0.77	0.95
Sleeping	0.89	0.78	0.97
Snack	0.81	0.75	0.93
SpareTimeTV	0.89	0.77	0.95
Toileting	0.89	0.79	0.93

## 5 Conclusion

Smart home offers ambient assisted living through modeling activity recognition system. The unsupervised approach of proposed activity modeling overcomes the dependency on the annotators and domain experts for activity modeling. The proposed pattern clustering ably groups the event patterns contextually. Weighted contextual Jaro Winkler similarity measure introduced in the design collectively clusters varied event patterns to related activities. The hierarchical approach of pattern clustering derives knowledge related to activity organization. Furthermore, the event organizer and habitual event generator extracts knowledge related to event ordering and contextual description of an activity. The obtained knowledge were utilized in the construction of probabilistic ontology activity model. The experimental study showed the proficiency of the proposed unsupervised pattern clustering in the design of activity modeling and recognition in comparison with existing systems. The future work would be to extend the pattern clustering approach to dynamically redefine the activity model during run time.

## References

1. Atallah, L., Yang, G.-Z.: The use of pervasive sensing for behaviour profiling - a survey. *Pervasive Mob. Comput.* **5**(5), 447–464 (2009)



2. Augusto, J.C., Nakashima, H., Aghajan, H.: Ambient intelligence and smart environments: a state of the art. In: Nakashima, H., Aghajan, H., Augusto, J.C. (eds.) *Handbook of Ambient Intelligence and Smart Environments*, pp. 3–31. Springer, Boston (2010). [https://doi.org/10.1007/978-0-387-93808-0\\_1](https://doi.org/10.1007/978-0-387-93808-0_1)
3. Aztiria, A., Izaguirre, A., Augusto, J.C.: Learning patterns in ambient intelligence environments: a survey. *Artif. Intell. Rev.* **34**(1), 35–51 (2010)
4. De Carolis, B., Ferilli, S., Redavid, D.: Incremental learning of daily routines as workflows in a smart home environment. *ACM Trans. Interact. Intell. Syst.* **4**(4), 20:1–20:23 (2015)
5. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Zhiwen, Y.: Sensor-based activity recognition. *IEEE Trans. Syst. Man Cybern. Part C* **42**(6), 790–808 (2012)
6. Chen, L., Nugent, C.D., Wang, H.: A knowledge-driven approach to activity recognition in smart homes. *IEEE Trans. Knowl. Data Eng.* **24**(6), 961–974 (2012)
7. de la Concepción, M.Á.Á., Morillo, L.M.S., García, J.A.Á., González-Abril, L.: Mobile activity recognition and fall detection system for elderly people using Ameva algorithm. *J. Pervasive Mob. Comput.* **34**, 3–13 (2017)
8. Gayathri, K.S., Easwarakumar, K.S., Elias, S.: Probabilistic ontology based activity recognition in smart homes using markov logic network. *Knowl.-Based Syst.* **121**(Supplement C), 173–184 (2017)
9. Gayathri, K.S., Elias, S., Shivashankar, S.: An ontology and pattern clustering approach for activity recognition in smart environments. In: Pant, M., Deep, K., Nagar, A., Bansal, J.C. (eds.) *Proceedings of the Third International Conference on Soft Computing for Problem Solving. AISC*, vol. 258, pp. 833–843. Springer, New Delhi (2014). [https://doi.org/10.1007/978-81-322-1771-8\\_72](https://doi.org/10.1007/978-81-322-1771-8_72)
10. Lotfi, A., Langensiepen, C.S., Mahmoud, S.M., Akhlaghinia, M.J.: Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour. *J. Ambient Intell. Humaniz. Comput.* **3**(3), 205–218 (2012)
11. Noy, N.F., Crubézy, M., Ferguson, R.W., Knublauch, H., Tu, S.W., Vendetti, J., Musen, M.A., et al.: Protege-2000: an open-source ontology-development and knowledge-acquisition environment. In: *Proceedings of the Annual Symposium of the American Medical Informatics Association (AMIA-2000)* (2003)
12. Ordóñez, F.J., de Toledo, P., Sanchis, A.: Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. *Sensors* **13**, 5460–5477 (2013)
13. Peek, S.T.M., Aarts, S., Wouters, E.J.M.: Can smart home technology deliver on the promise of independent living? A critical reflection based on the perspectives of older adults. In: van Hoof, J., Demiris, G., Wouters, E.J.M. (eds.) *Handbook of Smart Homes, Health Care and Well-Being*, pp. 203–214. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-01583-5\\_41](https://doi.org/10.1007/978-3-319-01583-5_41)
14. Rashidi, P., Cook, D.J., Holder, L.B., Schmitter-Edgecombe, M.: Discovering activities to recognize and track in a smart environment. *IEEE Trans. Knowl. Data Eng.* **23**(4), 527–539 (2011)
15. Singla, G., Cook, D.J., Schmitter-Edgecombe, M.: Recognizing independent and joint activities among multiple residents in smart environments. *J. Ambient Intell. Humaniz. Comput.* **1**(1), 57–63 (2010)
16. RC Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2013, 2014)

# E-FPROMETHEE: An Entropy Based Fuzzy Multi Criteria Decision Making Service Ranking Approach for Cloud Service Selection

B. Akshya Kaveri, O. Gireesha, Nivethitha Somu,  
M. R. Gauthama Raman, and V. S. Shankar Sriram (✉)

School of Computing, Centre for Information Super Highway (CISH),  
SASTRA University, Thanjavur 613401, Tamil Nadu, India  
akshyakaveri@gmail.com, {gireesha,nivethitha,  
gauthamaraman\_mr}@sastra.ac.in, sriram@it.sastra.edu

**Abstract.** The immense popularity and rapid adoption of cloud computing has led to the emergence of various cloud service providers, offering functionally-equivalent services. This scenario complicates the identification of an appropriate and trustworthy cloud service provider with respect to the unique Quality of Service (QoS) requirement of the users. Trust based service selection approaches prove to be a prominent solution for the cloud service selection problem since trust evaluation of the cloud services exploits the intrinsic relations between the QoS attributes. This paper presents Shannon entropy based Fuzzy PROMETHEE service ranking approach for the identification of trustworthy cloud service providers. A case study using real world QoS data from Cloud Harmony demonstrates the effectiveness and robustness of the proposed approach in terms various quality metrics (trustworthiness, untrustworthiness, uncertainty) and sensitivity analysis.

**Keywords:** Cloud service ranking · Entropy · Fuzzy logic  
Multi Criteria Decision Making · Trustworthy service providers

## 1 Introduction

Cloud computing – an internet and business paradigm uses virtualization technologies to provide ‘Anything as a Service’ (XaaS) in a ‘Pay-As-You-Use’ fashion from the state-of-the-art data centers [1–3]. The immense popularity and rapid adoption of cloud computing have led to the proliferation of numerous cloud service providers offering functionally-equivalent cloud services at different performance and price levels [4]. This scenario provides a healthy competition among the cloud service providers, however it poses a significant challenge to the users with respect to the identification of trustworthy cloud service providers [5]. In simpler terms, the existence of functionally-similar cloud services and the diverse set of user requirements complicates the process of service selection.

Over the past few decades, trust based service selection models and approaches have been gaining more attention by the research communities since trustworthiness is a measure of compliance between the service provision and the user requirements [6].

In general, trustworthiness is evaluated based on objective (QoS monitoring) and subjective (user feedbacks) assessments [7]. However, most of the existing approaches assess trust based on the QoS monitoring values, since the user feedbacks reflects the biased values [8]. The dynamic nature of cloud environment makes the QoS experienced by the user different from the QoS claimed by the service provider. This difference in the QoS values depends on various factors like user device type, network congestion, context, etc. [9, 10]. For example, in real-world WS-DREAM QoS dataset#2, the response time and throughput experienced by the two user varies over 64 time intervals for the same service [11]. The uncertainty and vagueness of the QoS values affects the accuracy of the cloud service selection model by providing imprecise service ranking.

The state-of-the-trust based service selection models employs various intelligent and statistical techniques to evaluate the trustworthiness of the cloud service providers. Out of these, Multi Criteria Decision Making (MCDM) approaches such as Analytic Hierarchy Process (AHP), Technique for Order Performance by Similarity to Ideal Solution (TOPSIS), Vlsekriterijumska Optimizacija I Kompromisno Resenje (VIKOR), Preference Ranking Organization METHod for Enrichment Evaluations (PROMETHEE), etc. were found to be most appropriate for cloud service selection problem since it reveals the intrinsic relations among the multiple criteria and the decision [6, 12, 13]. A broad survey on the existing MCDM approaches reveals the prevalence of fuzzy PROMETHEE in addressing uncertainty related issues in cloud service selection [14]. However, fuzzy PROMETHEE computes the importance/weights of each criteria based on the inconsistent QoS data, which in turn leads to imprecise service ranking. Hence, this paper presents E-FPROMETHEE, a novel service ranking approach based on Shannon entropy, Fuzzy logic and PROMETHEE to overcome the above said challenges in the process of obtaining accurate service ranking. E-FPROMETHEE integrates Triangular Fuzzy Numbers (TFNs) due to its computational efficiency and simplicity [15]. A case study using Cloud Harmony real world QoS reports demonstrate the effectiveness and robustness of E-FPROMETHEE over Fuzzy based MCDM approaches in terms of trustworthiness, untrustworthiness, and uncertainty and sensitive analysis [16]. Further, the performance of E-FPROMETHEE was compared with the existing MCDM methods like AHP [17] and improved TOPSIS [18].

The rest of the paper is organized as follows: Section 2 presents a broad literature survey on the state-of-the-art MCDM approaches for cloud service ranking. Section 3 describes the working of E-FPROMETHEE. Section 4 presents a case study to analyze the performance of the proposed ranking approach. Section 5 concludes the paper.

## 2 Related Works

According to Jagpreet Sidhu and Sarbjeeth Singh, “MCDM is a sub discipline of operations research that explicitly evaluates multiple conflicting criteria and decision making; MCDM is concerned with structuring and solving decision and planning problems involving multiple criteria” [14]. In general, MCDM based approaches were extensively used to solve several research problems in a wide range of applications such as energy planning [19], water resources management [20], supplier selection [21], risk management [22] and so forth due to the involving complex, vague, and uncertain

nature of the data. Further, this instigated several researchers and industry professionals towards the exploration and integration of MCDM approaches with statistical and intelligent techniques. In recent years, MCDM approaches has gained its attraction in the cloud service selection which involves the interdependent relations among the multiple attributes for the identification of appropriate cloud service providers via ranking. Table 1 presents few recent MCDM based service selection approaches. A common and significant challenge posed by the various approaches given in Table 1 is with respect to imprecise service ranking due to the incomplete and uncertain nature of the QoS values that forms the base for decision making.

**Table 1.** Related works

Authors	Technique proposed	Dataset	Validation
Ma et al. [10]	ELECTRE	WS-DREAM dataset	Difference degree and prediction accuracy
Alabool and Mahmood [23]	Modified VIKOR + Fuzzy set	Case study-5 CSPs, 5 DMs and 15 criteria	–
Rai and Kumar [24]	TOPSIS + VIKOR	Case study-10 CSPs and 3 criteria	Average method and instance method
Kumar et al. [25]	AHP + Fuzzy TOPSIS	Case study- 6 CSPs and 10 criteria	Sensitivity analysis
Sun et al. [26]	Fuzzy AHP + Fuzzy TOPSIS	Synthetic dataset	Threshold, recall, precision, fallout, F-measure
Zhuo et al. [27]	PROMETHEE and Block Nested Loop (BNL)		Effectiveness and efficiency
Singh and Sidhu [14, 18, 28]	AHP + Improved TOPSIS	Cloud Armour	Trustworthiness, untrustworthiness, and uncertainty
	AHP, TOPSIS, PROMETHEE	Cloud Harmony	
Radulescu and Radulescu [29]	Extended TOPSIS (E – TOPSIS)	Case study-10 CSPs and 3 criteria	–
Liu et al. [30]	Statistical variance (SV), improved TOPSIS, SAW, and Delphi–AHP	Case study-4 CSPs, DMs, and criteria	–
Sun et al. [31]	Fuzzy AHP, Fuzzy TOPSIS	Simulation	Precision, recall, fallout and F-measure
Shetty and D’Mello [32]	REMBRANDT approach	Case study-4 CSPs and 6 QoS attributes	Cost, availability, response time, security, and scalability

### 3 Proposed Methodology

Preference Ranking Organization METHod for Enrichment of Evaluations (PROMETHEE), an outranking approach is advantageous over the state-of-the-art MCDM approaches since it address the uncertainty related issues [14]. However, it suffers from rank reversal problem since it assigns weights to criteria based on the uncertain QoS data. In order to overcome the above-said challenges, this work presents E-FPROMETHEE, a novel service ranking approach based on Shannon entropy for weight computation, triangular fuzzy rule to handle vague judgments of the decision makers in fuzzy linguistic terms and defuzzify them to remove uncertainty in the data, and PROMETHEE to derive accurate service ranking. The step-by-step procedure of E-FPROMETHEE is detailed as follows,

**Step-1.** Determine the alternatives, linguistic variables (criteria), linguistic terms, and experts. Construct an evaluation matrix of order  $q \times r$ , where  $q$  represents the alternatives (cloud service providers) and  $r$  represents the criteria (QoS parameters). Each criteria can be represented in the form of a linguistic variable, in turn each linguistic variable is expressed as a linguistic term ( $VL, L, M, H, VH, E$ ). Further, each linguistic terms are mapped to the fuzzy intervals using triangular fuzzy number properties (Table 2). For example, the linguistic terms ( $VL, L, M, H, VH, E$ ) with their corresponding TFNs are shown in Table 3.

**Table 2.** Triangular fuzzy number properties

Operation	Meaning
Scalar summation ( $A^- \oplus B^-$ )	$(a_l, a_m, a_u) \oplus (b_l, b_m, b_u) = (a_l \oplus b_l, a_m \oplus b_m, a_u \oplus b_u)$
Scalar subtraction ( $A^- \ominus B^-$ )	$(a_l, a_m, a_u) \ominus (b_l, b_m, b_u) = (a_l \ominus b_u, a_m \ominus b_m, a_u \ominus b_l)$
Scalar multiplication ( $A^- \otimes B^-$ )	$(a_l, a_m, a_u) \otimes (b_l, b_m, b_u) = (a_l \otimes b_l, a_m \otimes b_m, a_u \otimes b_u)$
Scalar division ( $A^- \div B^-$ )	$(a_l, a_m, a_u) \div (b_l, b_m, b_u) = (\frac{a_l}{b_u}, \frac{a_m}{b_m}, \frac{a_u}{b_l})$
Scalar multiplicative inverse ( $\bar{A}^{-1}$ )	$(a_l, a_m, a_u)^{-1} = (1/a_l, 1/a_m, 1/a_u)$

**Table 3.** Linguistic terms – Triangular fuzzy numbers

Linguistic terms	TFNs scale
Very Low (VL)	(0.0, 0.0, 0.2)
Low (L)	(0.0, 0.2, 0.4)
Medium (M)	(0.2, 0.4, 0.6)
High	(0.4, 0.6, 0.8)
Very High	(0.6, 0.8, 1.0)
Excellent	(0.8, 1.0, 1.0)

**Step-2.** Construct Fuzzy Evaluation Matrix. Construct a fuzzy evaluation matrix  $F_{ij}$  ( $i = 1, 2, \dots, q; j = 1, 2, \dots, r$ ) with  $q$  alternatives (CSPs) and  $r$  criterias (QoS parameters) based on the linguistic variables and terms as follows,

$$F_{ij} = \begin{matrix} \text{CSP}_1 \\ \text{CSP}_2 \\ \text{CSP}_3 \\ \vdots \\ \text{CSP}_{q-1} \\ \text{CSP}_q \end{matrix} \begin{bmatrix} C_{p_1} & C_{p_2} & C_{p_3} & \dots & C_{p_{r-1}} & C_{p_r} \\ C_{p_1(\alpha_1, \beta_1, \gamma_1)} & C_{p_1(\alpha_2, \beta_2, \gamma_2)} & C_{p_1(\alpha_3, \beta_3, \gamma_3)} & \dots & C_{p_1(\alpha_{(r-1)}, \beta_{(r-1)}, \gamma_{(r-1)})} & C_{p_1(\alpha_r, \beta_r, \gamma_r)} \\ C_{p_2(\alpha_1, \beta_1, \gamma_1)} & C_{p_2(\alpha_2, \beta_2, \gamma_2)} & C_{p_2(\alpha_3, \beta_3, \gamma_3)} & \dots & C_{p_2(\alpha_{(r-1)}, \beta_{(r-1)}, \gamma_{(r-1)})} & C_{p_2(\alpha_r, \beta_r, \gamma_r)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{p_{(q-1)}(\alpha_1, \beta_1, \gamma_1)} & C_{p_{(q-1)}(\alpha_2, \beta_2, \gamma_2)} & C_{p_{(q-1)}(\alpha_3, \beta_3, \gamma_3)} & \dots & C_{p_{(q-1)}(\alpha_{(r-1)}, \beta_{(r-1)}, \gamma_{(r-1)})} & C_{p_{(q-1)}(\alpha_r, \beta_r, \gamma_r)} \\ C_{p_q(\alpha_1, \beta_1, \gamma_1)} & C_{p_q(\alpha_2, \beta_2, \gamma_2)} & C_{p_q(\alpha_3, \beta_3, \gamma_3)} & \dots & C_{p_q(\alpha_{(r-1)}, \beta_{(r-1)}, \gamma_{(r-1)})} & C_{p_q(\alpha_r, \beta_r, \gamma_r)} \end{bmatrix}$$

In the above fuzzy decision matrix ( $F_{ij}$ ),  $CSP_1, CSP_2, \dots, CSP_q$  are the  $q$  cloud service providers and  $C_{p_1}, C_{p_2}, C_{p_3}, \dots, C_{p_r}$  are the compliance levels of  $r$  parameters.  $C_{p_{i,j}}$  denotes the compliance level of  $j$  th parameter of  $i$  th cloud service provider.

**Step-3.** Normalize the fuzzy evaluation matrix. In general, normalization controls the dominance of high valued parameters, thereby provides dimensionless units to represent heterogeneous data. Max-min and min – max normalization method (for benefit and cost criteria) is used to normalize the fuzzy decision matrix to enhance the accuracy of E-FPROMETHEE. The resultant fuzzy normalized matrix ( $FN_{ij}$ ) is represented using Eqs. (1) and (2).

$$FN_{ij} = \frac{F_{ij}(\alpha, \beta, \gamma) - \text{Min}_{j(\alpha, \beta, \gamma)}}{\text{Max}_{j(\alpha, \beta, \gamma)} - \text{Min}_{j(\alpha, \beta, \gamma)}}, i = 1, 2, \dots, q; j = 1, 2, \dots, r \tag{1}$$

$$FN_{ij} = \frac{\text{Max}_{j(\alpha, \beta, \gamma)} - F_{ij}(\alpha, \beta, \gamma)}{\text{Max}_{j(\alpha, \beta, \gamma)} - \text{Min}_{j(\alpha, \beta, \gamma)}}, i = 1, 2, \dots, q; j = 1, 2, \dots, r \tag{2}$$

Eq. (1) - benefit criteria and Eq. (2) - cost criteria.

**Step-4.** Determine the importance (weights) of each criteria. The importance of each criteria is computed based on Shannon entropy from the fuzzy normalized matrix ( $FN_{ij}$ ) matrix. The weights of each criteria is represented as fuzzy numbers using TFN (Eqs. (3)–(6)).

$$E_j = -k \sum_{j=1}^r FN_{ij}(\alpha, \beta, \gamma) \ln FN_{ij}(\alpha, \beta, \gamma) \tag{3}$$

where,  $k = (\ln(q))^{-1}$

$$d_j = 1 - E_j(\alpha, \beta, \gamma); \quad j = 1, 2, \dots, r \tag{4}$$

where,  $d_j$  is the divergence. The divergence value is directly proportional to the importance of the criteria  $q$ .

$$W_j = \frac{d_j}{\sum_j d_j}; \quad j = 1, 2, \dots, r \tag{5}$$

where,  $W_j$  is the weight of the criterion.

$$W_j = [W_1, W_2, W_3, \dots, W_r] \quad W_r = (\alpha^W, \beta^W, \gamma^W) \quad j = 1, 2, \dots, r \tag{6}$$

**Step-5.** Compute the fuzzy preference function. Design a generalized preference function  $p_j(q, r)$  for each criteria  $\hat{c}_j$ . Let  $C_{p_j}(m)$  be the value of a criteria  $r$  for  $CSP_m$ . The preference function  $p_j(q, r)$  is expressed as  $d_j(CSP_m, CSP_n)$  i.e. the difference between the alternative  $CSP_m$  and  $CSP_n$ .

$$p_j(CSP_m, CSP_n) = F(d_j((CSP_m, CSP_n))) \tag{7}$$

$$d_j(CSP_m, CSP_n) = C_{p_j}(CSP_m) - C_{p_j}(CSP_n) \tag{8}$$

$$p_j(CSP_m, CSP_n) = \begin{cases} 0, & DM_{qj} \leq DM_{rj} \\ 1, & DM_{qj} > DM_{rj} \end{cases} \quad j = 1, 2, \dots, r \tag{9}$$

Eq. (8) can be expressed as fuzzy number using TFN as in Eq. (10).

$$d_j(CSP_{m(\alpha,\beta,\gamma)}, CSP_{n(\alpha,\beta,\gamma)}) = C_{p_j}(CSP_{m(\alpha,\beta,\gamma)}) - C_{p_j}(CSP_{n(\alpha,\beta,\gamma)}) \tag{10}$$

**Step-6.** Aggregate the values. Aggregate the fuzzy weights of each criteria using Eq. (11), fuzzy normalization matrix using Eq. (12) and fuzzy preference function among pairs of alternatives using Eq. (13).

$$W_j = (\alpha^W, \beta^W, \gamma^W) = \frac{1}{3}(\alpha^W + \beta^W + \gamma^W) \tag{11}$$

$$FN_{ij} = (\alpha_{N_{ij}}, \beta_{N_{ij}}, \gamma_{N_{ij}}) = \frac{1}{3}(\alpha_{N_{ij}} + \beta_{N_{ij}} + \gamma_{N_{ij}}); \quad i = 1, 2, \dots, q; \quad j = 1, 2, \dots, r \tag{12}$$

$$p_j(CSP_m, CSP_n) = (\alpha_{p_j}(CSP_m, CSP_n), \beta_{p_j}(CSP_m, CSP_n), \gamma_{p_j}(CSP_m, CSP_n)) = \frac{1}{3}(\alpha_{p_j}(CSP_m, CSP_n) + \beta_{p_j}(CSP_m, CSP_n) + \gamma_{p_j}(CSP_m, CSP_n)) \tag{13}$$

**Step-7.** Calculate global index values. Define the aggregated weight preference function as in Eq. (14).

$$\prod(CSP_m, CSP_n) = \sum_{j=1}^r W_j * p_j(CSP_m, CSP_n) \tag{14}$$

**Step-8.** Compute the leaving, entering and net flows. For each possible decision  $m$ , calculate the leaving flow  $\emptyset^+(CSP_m)$ , entering flow  $\emptyset^-(CSP_m)$  and net outranking flow  $\emptyset(CSP_m)$  using Eqs. (15)–(17).

$$\emptyset^+(CSP_m) = \frac{1}{r-1} \sum_{j=1}^r W_j * p_j(CSP_m, CSP_n) \tag{15}$$

$$\emptyset^-(CSP_m) = \frac{1}{r-1} \sum_{j=1}^r W_j * p_j(CSP_m, CSP_n) \tag{16}$$

$$\emptyset(CSP_m) = \emptyset^+(CSP_m) - \emptyset^-(CSP_m) \tag{17}$$

**Step-9.** Compute the Trustworthiness, Untrustworthiness, and Uncertainty. Compute the trustworthiness ( $T_i$ ), untrustworthiness ( $UT_i$ ), and uncertainty ( $U_i$ ) using Eqs. (18)–(20) for  $CSP_i$  from the variance, leaving flow and entering flow using Eqs. (15)–(17) respectively.

$$T_i = \frac{\emptyset^+(CSP_m)}{\emptyset^+(CSP_m) + \emptyset^-(CSP_m) + V_i} \tag{18}$$

$$UT_i = \frac{\emptyset^-(CSP_m)}{\emptyset^+(CSP_m) + \emptyset^-(CSP_m) + V_i} \tag{19}$$

$$U_i = \frac{V_i}{\emptyset^+(CSP_m) + \emptyset^-(CSP_m) + V_i} \tag{20}$$

where, variance  $V_i$  obtained from the normalization matrix.

**Step-10.** Rank the alternatives. Rank the CSPs (alternatives) with respect to trustworthiness ( $T_i$ )

$T_i$ ,  $UT_i$ , and  $V_i$  are used to derive the trustworthiness from the compliance values and can be employed by any of the cloud entities to compute the trustworthiness from its point of view.



## 4 Experimental Results – Case Study

This section presents a case study to determine the effectiveness and robustness of E-FPROMETHEE. The case study consists of six CSPs (alternatives) with ten compliance parameters, a sample dataset extracted from the real world QoS dataset, CloudHarmony Reports - Cloud Database Servers (CDS) with three instances (Small (S), Medium (M), Large (L)). This case study includes six CDSs (alternatives), namely Amazon Web Services (AWS), DigitalOcean, Google, Rackspace, SoftLayer and Microsoft Azure and ten compliance parameters (criteria), namely Cost on demand ( $C_1$ ), Network Latency ( $C_2$ ), SequentialDiskRead/Write Performance Consistency ( $C_3$ ), Random Disk Read/Write Performance Consistency ( $C_4$ ), CPU integer performance ( $C_5$ ), CPU floating point performance ( $C_6$ ), Memory Performance On Scale ( $C_7$ ), Memory Performance On Triad ( $C_8$ ), Sequential Read/Write Disk Performance ( $C_9$ ), Random Read/Write Disk Performance ( $C_{10}$ ) [16]. Initially, a fuzzy evaluation matrix ( $F_{ij}$ ) was generated by the experts or DMs for six CDSs and ten criteria using the linguistic terms given in Table 4.

The fuzzy evaluation matrix ( $F_{ij}$ ) was normalized using min-max normalization technique for the construction of fuzzy normalization matrix ( $FN_{ij}$ ) (Eqs. (1) and (2)). Further, aggregate the fuzzy normalization matrix ( $FN_{ij}$ ) using Eq. (12). Tables 5 and 6 presents the fuzzy normalization matrix ( $FN_{ij}$ ) and its aggregation.

The importance or the weights ( $W_j$ ) for each criteria was determined using Shannon entropy (Eqs. (3)–(6)). The weights of the criteria were aggregated using (Eq. 11). Based on the aggregated weights computed using Eq. 11, the ten compliance parameters (criteria) were ranked as  $C_1 \cong C_5 \cong C_6 \cong C_7 \cong C_8 > C_3 > C_4 \cong C_{10} > C_2 \cong C_9$ . From the above ranking, the most important criteria for the considered dataset are  $C_1, C_5, C_6, C_7$  and  $C_8$  (Table 7).

The preference degree among the CDSs (alternatives) were computed using Eqs. (7) and (10). If the value of the difference,  $d \leq 0$ , then the preference degree value is *zero*. Otherwise, it is *one* (Table 8).

Further, the global index values, which is a product of aggregate weight and preference function values were calculated using Eq. (14) (Table 9).

Finally, the flow values, i.e. leaving flow, entering flow, and net flow of alternatives were calculated using Eq. (15)–(17). The flow measures can be defined as “Leaving flow: the preference of an alternative over all other alternatives; Entering flow: the preference of all other alternatives over an alternative; Net flow: the difference of leaving and entering flow” [33, 34]. The leaving flow (Eq. 15), entering flow (Eq. 16), and net flow (Eq. 17) denotes the relative strength, relative weakness and their differences of an alternative respectively. The net flow is directly proportional to the ranking order i.e. the high net flow value denotes the best alternative. Table 10 presents the flow values of six alternatives.

Finally, the trustworthiness ( $T_i$ ), untrustworthiness ( $UT_i$ ) and uncertainty ( $U_i$ ) of the CSPs were computed using Eq. (17)–(20) respectively.

**Table 4.** Triangular fuzzy evaluation matrix

Criteria (L, M, H)	CSP (alternatives)					
	Amazon EC2	Digital ocean	Google	Microsoft Azure	Rackspace	Softlayer
$C_1$	L (0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	M (0.2, 0.4, 0.6)	L (0.0, 0.2, 0.4)
$C_2$	VL (0.0, 0.0, 0.2)	M (0.2, 0.4, 0.6)	H (0.4, 0.6, 0.8)	M (0.2, 0.4, 0.6)	M (0.2, 0.4, 0.6)	VL (0.0, 0.0, 0.2)
$C_3$	L (0.0, 0.2, 0.4)	VL (0.0, 0.0, 0.2)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	M (0.2, 0.4, 0.6)	L (0.0, 0.2, 0.4)
$C_4$	L (0.0, 0.2, 0.4)	VL (0.0, 0.0, 0.2)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	M (0.2, 0.4, 0.6)	M (0.2, 0.4, 0.6)
$C_5$	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	VL (0.0, 0.0, 0.2)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)
$C_6$	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	VL (0.0, 0.0, 0.2)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)
$C_7$	H (0.4, 0.6, 0.8)	H (0.4, 0.6, 0.8)	H (0.4, 0.6, 0.8)	M (0.2, 0.4, 0.6)	H (0.4, 0.6, 0.8)	H (0.4, 0.6, 0.8)
$C_8$	H (0.4, 0.6, 0.8)	H (0.4, 0.6, 0.8)	H (0.4, 0.6, 0.8)	M (0.2, 0.4, 0.6)	H (0.4, 0.6, 0.8)	H (0.4, 0.6, 0.8)
$C_9$	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	VL (0.0, 0.0, 0.2)	L (0.0, 0.2, 0.4)	H (0.4, 0.6, 0.8)	M (0.2, 0.4, 0.6)
$C_{10}$	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	VL (0.0, 0.0, 0.2)	L (0.0, 0.2, 0.4)	L (0.0, 0.2, 0.4)	M (0.2, 0.4, 0.6)

**Table 5.** Fuzzy normalization matrix

Criteria (L, M, H)	CSP (alternatives)					
	Amazon EC2	Digital ocean	Google	Microsoft Azure	Rackspace	Softlayer
$C_1$	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(0, 0, 0)	(1, 1, 1)
$C_2$	(1, 1, 1)	(0.5, 0.33, 0.3)	(0, 0, 0)	(0.5, 0.67, 1)	(0.5, 0.33, 0.33)	(1, 1, 1)
$C_3$	(1, 0.5, 0.5)	(1, 1, 1)	(1, 0.5, 0.5)	(0, 0.5, 1)	(0, 0, 0)	(1, 0.5, 0.5)
$C_4$	(1, 0.5, 0.5)	(1, 1, 1)	(1, 0.5, 0.5)	(0, 0.5, 1)	(0, 0, 0)	(0, 0, 0)
$C_5$	(0, 1, 1)	(0, 1, 1)	(0, 1, 1)	(0, 0, 0)	(0, 1, 1)	(0, 1, 1)
$C_6$	(0, 1, 1)	(0, 1, 1)	(0, 1, 1)	(0, 0, 0)	(0, 1, 1)	(0, 1, 1)
$C_7$	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(0, 0, 0)	(1, 1, 1)	(1, 1, 1)
$C_8$	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(0, 0, 0)	(1, 1, 1)	(1, 1, 1)
$C_9$	(0, 0.33, 0.33)	(0, 0.33, 0.33)	(0, 0, 0)	(0, 0.33, 0.33)	(1, 1, 1)	(0.5, 0.67, 0.67)
$C_{10}$	(0, 0.5, 0.5)	(0, 0.5, 0.5)	(0, 0, 0)	(0, 0.5, 0.5)	(0, 0.5, 0.5)	(1, 1, 1)

**Table 6.** Aggregation of normalization data

Criteria (L, M, H)	CSP (alternatives)					
	Amazon EC2	Digital ocean	Google	Microsoft Azure	Rackspace	Softlayer
$C_1$	1	1	1	1	0	1
$C_2$	1	0.39	0	0.72	0.39	1
$C_3$	0.67	1	0.67	0.50	0	0.67
$C_4$	2	3	2	1.50	0	0
$C_5$	0.67	0.67	0.67	0	0.67	0.67
$C_6$	0.67	0.67	0.67	0	0.67	0.67
$C_7$	1	1	1	0	1	1
$C_8$	1	1	1	0	1	1
$C_9$	0.22	0.22	0	0.22	1	0.61
$C_{10}$	0.33	0.33	0	0.33	0.33	1
Variance	0.24	0.62	0.38	0.26	0.18	0.10

**Table 7.** Computation of weights for each criteria

Criteria	$E_j$	Divergence	Sum (divergence)	Weights	Aggregation of weights
$C_1$	(0, 0, 0)	(1, 1, 1)	(9.23, 6.55, 7.09)	(0.14, 0.15, 0.11)	0.1341
$C_2$	(0.58, 0.56, 0.41)	(0.42, 0.44, 0.59)		(0.06, 0.07, 0.06)	0.0654
$C_3$	(0, 0.78, 0.58)	(1, 0.23, 0.42)		(0.14, 0.12, 0.05)	0.0674
$C_4$	(0, 0.58, 0.39)	(1, 0.42, 0.61)		(0.14, 0.06, 0.07)	0.0863
$C_5$	(0, 0, 0)	(1, 1, 1)		(0.14, 0.15, 0.11)	0.1341
$C_6$	(0, 0, 0)	(1, 1, 1)		(0.14, 0.15, 0.11)	0.1341
$C_7$	(0, 0, 0)	(1, 1, 1)		(0.14, 0.15, 0.11)	0.1341
$C_8$	(0, 0, 0)	(1, 1, 1)		(0.14, 0.15, 0.11)	0.1341
$C_9$	(0.19, 0.76, 0.76)	(0.81, 0.24, 0.24)		(0.11, 0.04, 0.03)	0.0523
$C_{10}$	(0, 0.77, 0.77)	(1, 0.23, 0.23)		(0.14, 0.12, 0.02)	0.0583

Table 11 and Fig. 1 provides the service ranking based on trustworthiness, untrustworthiness and uncertainty. The six CDSs were ranked based on their respective measure of trustworthiness, untrustworthiness and uncertainty as follows,

**Table 8.** Preference degree

CSP (Alternatives)	Amazon EC2	Digital Ocean	Google	Microsoft Azure	Rackspace	Softlayer
Amazon EC2	0	0.2	0.2	0.4	0	0
Digital Ocean	0.1	0	0.2	0.4	0	0.1
Google	0.1	0.3	0	0.5	0.1	0.1
Microsoft Azure	0.1	0.2	0.2	0	0	0.1
Rackspace	0.5	0.4	0.5	0.8	0	0.4
Softlayer	0.3	0.4	0.3	0.7	0.1	0

**Table 9.** Preference function

CSP (Alternatives)	Amazon EC2	Digital Ocean	Google	Microsoft Azure	Rackspace	Softlayer
Amazon EC2	0	0.20	0.20	0.40	0	0
Digital Ocean	0.10	0	0.20	0.40	0	0.10
Google	0.10	0.30	0	0.50	0.10	0.10
Microsoft Azure	0.10	0.20	0.20	0	0	0.10
Rackspace	0.50	0.40	0.50	0.80	0	0.40
Softlayer	0.30	0.40	0.30	0.70	0.10	0

**Table 10.** Metrics - Leaving Flow, Entering Flow and Net Flow

CSP (Alternatives)	Metrics			Ranking
	Leaving Flow	Entering Flow	Net Flow	
Amazon EC2	0.16	0.22	-0.06	4
Digital Ocean	0.16	0.22	-0.06	5
Google	0.22	0.20	0.02	3
Microsoft Azure	0.12	0.18	-0.06	6
Rackspace	<b>0.52</b>	<b>0.19</b>	<b>0.32</b>	<b>1</b>
Softlayer	0.36	0.12	0.23	2

Trustworthiness ( $T_i$ ): *Softlayer* > *Rackspace* > *Google* > *Amazon EC2* > *Microsoft Azure* > *Digital Ocean*

Untrustworthiness ( $UT_i$ ): *AmazonEC2* > *Microsoft Azure* > *Google* > *Digital ocean* > *Rackspace* > *Softlayer*

Uncertainty ( $U_i$ ): *Digital Ocean* > *Google* > *Microsoft Azure* > *Amazon EC2* > *Rackspace* > *Softlayer*

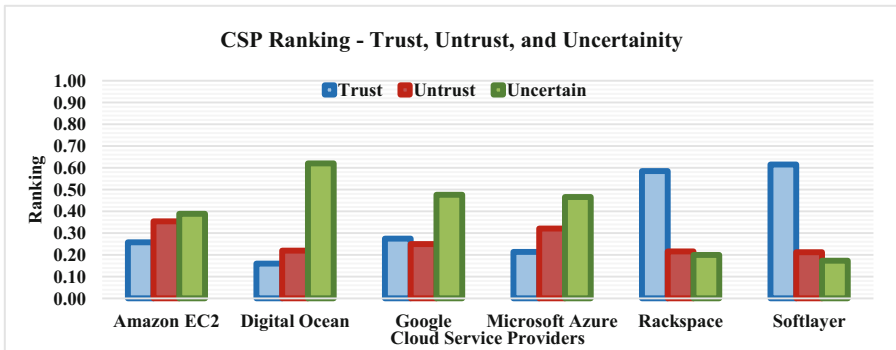


Fig. 1. Cloud service ranking based on trustworthiness, untrustworthiness and uncertainty

Table 11. CSP Ranking based on trustworthiness, untrustworthiness and uncertainty

CSP (Alternatives)	Trust	Ranking	Untrust	Ranking	Uncertain	Ranking
Amazon EC2	0.2574	4	<b>0.3539</b>	<b>1</b>	0.3887	4
Digital Ocean	0.1599	6	0.2197	4	<b>0.6205</b>	<b>1</b>
Google	0.2744	3	0.2495	3	0.4760	2
Microsoft Azure	0.2137	5	0.3206	2	0.4656	3
Rackspace	0.5847	2	0.2159	5	0.1994	5
Softlayer	<b>0.6148</b>	<b>1</b>	0.2118	6	0.1734	6

The efficiency of E-FPROMETHEE was demonstrated by comparing the performance of the proposed service ranking approach with the existing MCDM approaches like AHP [17] and improved TOPSIS [18]. From Table 12 and Fig. 2, it is clear that the service ranking of E-FPROMETHEE has some degree of similarity with the ranking provided by AHP and improved TOPSIS.

Table 12. Service Ranking – AHP, Improved TOPSIS, and E-FPROMETHEE

MCDM Approach	Ranking
AHP	Amazon EC2 > Softlayer > Rackspace > Google > Microsoft Azure > Digital Ocean
Improved TOPSIS	Softlayer > Amazon EC2 > Digital Ocean > Rackspace > Microsoft Azure > Google
<b>E-FPROMETHEE</b>	Softlayer > Rackspace > Google > Amazon EC2 > Microsoft Azure > Digital Ocean

Further, sensitivity analysis was performed to show the robustness of E-FPROMETHEE with the change in the criteria weights under similar assumptions [25]. Sensitivity analysis creates different scenario or new condition with the change in the weights that has a significant impact on the priority of the alternatives. In simpler terms, if the change in the criteria weights affects the ranking order, then it is said to be sensitive, else robust. Sensitivity analysis was carried out by interchanging the criteria

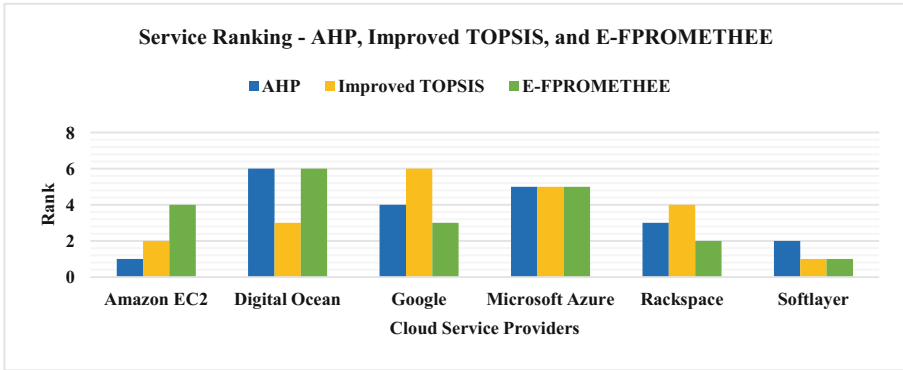


Fig. 2. Service Ranking – AHP, Improved TOPSIS, and E-FPROMETHEE

weights with the others. We have conducted 21 experiments to determine the impact of weights in the selection of suitable service provider. The service ranking obtained based on the highest score in the sensitivity experiments is *Softlayer* > *Rackspace* > *Google* > *AmazonEC2* > *MicrosoftAzure* > *DigitalOcean* (Fig. 3).

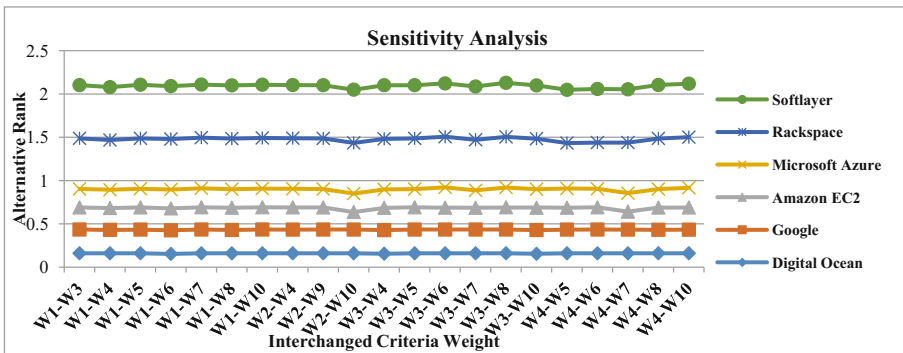


Fig. 3. Sensitivity analysis

## 5 Conclusion

Cloud service selection, a multi criteria decision making problem which exploits the intrinsic relations among the multiple criteria or QoS attributes. This work put forth E-FPROMETHEE, a novel service ranking approach based on Shannon entropy (weight assignment), Triangular fuzzy rule – Fuzzy logic (uncertainty), and PROMETHEE – MCDM (ranking) to address the uncertainty related issues in the state-of-the-art service selection approaches. The validity and effectiveness of E-FPROMETHEE over the existing MCDM based service selection approaches were demonstrated using real world QoS reports from Cloud Harmony (6 CSPs and 10 criteria) in terms of trustworthiness, untrustworthiness, uncertainty, and sensitivity analysis.

## References

1. Chiregi, M., Navimipour, N.: A new method for trust and reputation evaluation in the cloud environments using the recommendations of opinion leaders' entities and removing the effect of troll. *Comput. Hum. Behav.* **60**, 280–292 (2016)
2. Mell, P., Grance, T.: The NIST definition of cloud computing (2011)
3. Nivethitha, S., Kannan, K., Sriram, V.S.S.: A computational model for ranking cloud service providers using hypergraph based techniques. *Future Gener. Comput. Syst.* **68**, 14–30 (2016)
4. Somu, N., Kirthivasan, K., Sriram, V.S.S.: A rough set-based hypergraph trust measure parameter selection technique for cloud service selection. *J. Supercomput.* **73**(10), 4535–4559 (2017)
5. Thampi, S., Bhargava, B., Atrej, P.: *Managing Trust in Cyberspace*. CRC Press, Boca Raton (2013)
6. Qu, L.: *Credible service selection in cloud environments* (2016)
7. Tang, M., Dai, X., Liu, J., Chen, J.: Towards a trust evaluation middleware for cloud service selection. *Future Gener. Comput. Syst.* **74**, 302–312 (2016)
8. Mao, C., Lin, R., Xu, C., He, Q.: Towards a trust prediction framework for cloud services based on PSO-driven neural network. *IEEE Access* **5**, 2187–2199 (2017)
9. Ma, H., Zhu, H., Hu, Z., Tang, W., Dong, P.: Multi-valued collaborative QoS prediction for cloud service via time series analysis. *Future Gener. Comput. Syst.* **68**, 275–288 (2017)
10. Ma, H., Hu, Z., Li, K., Zhang, H.: Toward trustworthy cloud service selection: a time-aware approach using interval neutrosophic set. *J. Parallel Distrib. Comput.* **96**, 75–94 (2016)
11. Al-Masri, E., Mahmoud, Q.H.: QoS-based discovery and ranking of web services. In: *Proceedings - International Conference on Computer Communications and Networks, ICCCN*, pp. 529–534 (2007)
12. Costa, P.M.A.C.: *Evaluating cloud services using multicriteria decision analysis* (2013)
13. Sun, L., Dong, H., Hussain, F.K., Hussain, O.K., Chang, E.: Cloud service selection: state-of-the-art and future research directions. *J. Netw. Comput. Appl.* **45**, 134–150 (2014)
14. Sidhu, J., Singh, S.: Design and comparative analysis of MCDM-based multi-dimensional trust evaluation schemes for determining trustworthiness of cloud service providers. *J. Grid Comput.* **15**, 197–218 (2017)
15. Efe, B.: An integrated fuzzy multi criteria group decision making approach for ERP system selection. *Appl. Soft Comput.* **38**, 106–117 (2016)
16. Cloud harmony cloud reports: <http://static.lindsberget.se/state-of-the-cloud-compute-0714.pdf> (2014)
17. Garg, S.K., Versteeg, S., Buyya, R.: A framework for ranking of cloud computing services. *Future Gener. Comput. Syst.* **29**, 1012–1023 (2013)
18. Sidhu, J., Singh, S.: Improved TOPSIS method based trust evaluation framework for determining trustworthiness of cloud service providers. *J. Grid Comput.* **15**(1), 81–105 (2017)
19. Ervural, B.C., Zaim, S., Demirel, O.F., Aydin, Z., Delen, D.: An ANP and fuzzy TOPSIS-based SWOT analysis for Turkey's energy planning. *Renew. Sustain. Energy Rev. Part 1* **82**, 1538–1550 (2018)
20. Lee, K.C., Tsai, W.H., Yang, C.H., Lin, Y.Z.: An MCDM approach for selecting green aviation fleet program management strategies under multi-resource limitations. *J. Air Transp. Manag.* 1–10, July 2017
21. Yazdani, M., Chatterjee, P., Zavadskas, E.K., Zolfani, S.H.: Integrated QFD-MCDM framework for green supplier selection. *J. Clean. Prod.* **142**, 3728–3740 (2017)

22. Büyüközkan, G., Karabulut, Y., Arsenyan, J.: RFID service provider selection: an integrated fuzzy MCDM approach. *Measurement* **112**, 88–98 (2017)
23. Alabool, H.M., Mahmood, A.K.: Trust-based service selection in public cloud computing using fuzzy modified VIKOR method. *J. Basic Appl. Sci.* **7**, 211–220 (2013)
24. Rai, D., Kumar, P.: Instance based multi criteria decision model for cloud service selection using TOPSIS and VIKOR. *Int. J. Comput Eng. Technol.* **7**, 78–87 (2016)
25. Kumar, R.R., Mishra, S., Kumar, C.: Prioritizing the solution of cloud service selection using integrated MCDM methods under Fuzzy environment. *J. Supercomput.* **73**(11), 4652–4682 (2017)
26. Sun, L., Dong, H., Hussain, F.K., Hussain, O.K., Ma, J., Zhang, Y.: A hybrid fuzzy framework for cloud service selection. In: *IEEE International Conference on Web Service*, pp. 313–320 (2014)
27. Zhao, Z., Jiang, Y., Zhao, X.: SLA\_oriented service selection in cloud environment : a PROMETHEE\_based Approach. In: *2015 4th International Conference on Computer Science Network Technology*, pp. 872–875 (2015)
28. Singh, S., Sidhu, J.: Compliance-based multi-dimensional trust evaluation system for determining trustworthiness of cloud service providers. *Future Gener. Comput. Syst.* **67**, 109–132 (2017)
29. Rădulescu, C.Z., Rădulescu, I.C.: An extended TOPSIS approach for ranking cloud service providers. *Stud. Inform. Control* **26**, 183–192 (2017)
30. Liu, S., Chan, F.T.S.S., Ran, W.: Decision making for the selection of cloud vendor: an improved approach under group decision-making with integrated weights and objective/subjective attributes. *Expert Syst. Appl.* **55**, 37–47 (2016)
31. Sun, L., Ma, J., Zhang, Y., Dong, H., Hussain, F.K.: Cloud-FuSeR: fuzzy ontology and MCDM based cloud service selection. *Future Gener. Comput. Syst.* **57**, 42–55 (2016)
32. Shetty, J., D’Mello, D.A.: Quality of service driven cloud service ranking and selection algorithm using REMBRANDT approach. In: *Proceedings of 2015 International Conference on Smart Technology Management for Computing, Communication, Controls, Energy and Materials ICSTM 2015*, pp. 126–132 (2015)
33. Dwivedi, A.N.: *Handbook of Research on Information Technology Management and Clinical Data Administration in Healthcare*. IGI Global, Hershey (2009)
34. Malczewski, J., Rinner, C.: *Multicriteria Decision Analysis in Geographic Information Science*. Springer, Heidelberg (2015)



# Workflow Scheduling Using IOT Enabled Reputation of Service Providers in the Cloud

K. Kanagaraj<sup>1</sup>(✉) and S. Swamynathan<sup>2</sup>

<sup>1</sup> Department of MCA, MEPCO Schlenk Engineering College, Sivakasi, India  
kanagaraj@mepcoeng.ac.in

<sup>2</sup> Department. of IST, CEG, Anna University, Chennai, India  
swamyns@annauniv.edu

**Abstract.** Cloud computing has become the standard and popular resource provisioning mechanism, in which hardware and software resources are provided in pay per use model. Workflow execution in the cloud is a major research area with lots of open and interesting problems. The key challenges in workflow execution in the cloud are, finding a suitable service provider and effective workflow scheduling. Due to the availability of large number of service providers, selecting a reputed service provider is very essential. Selecting a service provider with low reputation may lead to several problems like incorrect resource allocation, over charging and slippage of deadline, and selecting a highly reputed service provider may cost more. Hence this paper propose a reputation based workflow scheduling strategy that calculates the reputation of a cloud service provider (CSP) based on user rating and the actual performance of Virtual Machines (VMs) obtained using IOT enabled devices. Finally a scheduler is used to schedule the workflow with the service provider having the user's preferred level of reputation and also recommend a suitable scheduling algorithm for executing the workflow.

**Keywords:** Workflow · IOT · User rating · Reputation · Scheduling  
Dynamic · Deadline · Cost · Optimization

## 1 Introduction

Cloud and IoT are considered as complementary technologies. They provide support for the improvement of distributed, heterogeneous and complex applications that demand large scale storage space, huge volume of data, powerful computing facilities, greater flexibility and availability, and interoperable network and communication links. These systems contain a collection of smart devices of which are interconnected and controlled through services with the help of cloud infrastructure. Thus the convergence of cloud with the IoT has the potential to provide new levels of services in various sectors including businesses, education, science and research. Computing paradigms have a great influence in the way how business is carried in the society. Computing supports the day to day activities of most individuals such as sending email, using a credit card for booking an air ticket, etc. Computing has evolved from centralized computing in which computing is done at a central location to distributed computing

where computation is performed with a group of computing systems which may locate in and around the world. Apart from these two, other computing paradigms like parallel computing, which solve a single problem with the help of two or more processes, grid computing, that handle the distributed information and tasks in a group of networked computers with the help of a single main computer and utility computing, which is intended to provide services to the customer, based on their need and the service providing systems that give the resources as per the demand, were also popularly used.

Cloud computing is the state-of-the-art resource provisioning mechanism with several features like elasticity, virtualization and pay per use. Cloud computing, maximize the utilization of computing, storage and networking resources with the help of virtualization and helps users to access them with the help of internet. Due to the user friendly nature of the cloud, many researchers use it for their research and development activities. Workflow scheduling in the cloud is a major research area with lots of unaddressed open research problems.

Any application that consists of a series of tasks can be modelled as a workflow that can be represented using Directed Acyclic Graph (DAG). Workflow is a convenient model to represent most of the scientific applications that have high computational and data flow requirements [5]. Like other computational problems, workflows can also be executed in a cloud environment and can gain great benefits like dynamic resource provisioning, pay per use, scalability, elasticity, etc. As there are numerous cloud service offerings from plenty of vendors, an important challenge for customers is to determine the suitable cloud service provider to execute their workflow. Therefore, it is not adequate to just discover multiple cloud service providers but it is also mandatory to evaluate them and determine the most suitable cloud service provider [17]. As it is difficult to request the large number of cloud users to rate service providers against a large set of multifaceted criteria, it can be defined as the collective opinion of a community towards that entity based on several major aspects of performance [16].

Cloud Service Measurement Index Consortium (CSMIC) [17] has identified measurement indexes that are combined in the form of Service Measurement Index (SMI) that is very important for the evaluation of cloud services. These measurement indexes can be used by customers to compare different cloud services such as accountability, agility, cost, performance, assurance, security, privacy and usability. The fuzzy logic based trust management system [18] uses a fuzzy logic controller to calculate the reputation of the CSPs. Ontology based ranking technique [19] proposes a ranking algorithm for evaluating web services. The QoS-based ranking algorithm adopts Analytic Hierarchy Process (AHP), a multiple criteria decision making technique, as an underlying mechanism for developing a flexible and dynamic ranking algorithm. It can be adopted to rate cloud services only after incorporating it with some minor changes suitable for evaluating CSPs.

Most of the existing models depends on user ranking and social media to evaluate the reputation of the cloud service provider. However the actual reputation can be obtained only if the actual performance of the Virtual Machines is known. Hence this paper propose a reputation calculation process that uses the user rating and the actual performance of the VMs obtained using IOT enabled devices. User ratings represent several parameters including Quality of Service (QOS), availability, security, response time, accessibility and price. IoT enabled devices helps to understand the Throughput and power requirements of the VMs. Using the above details, the proposed model

classifies the cloud service providers into three categories such as High (**H**), Medium (**M**) or Low (**L**), which give some freedom for the users to select a suitable service provider depending on their need. Once a suitable CSP is identified, the next step is to schedule the workflow with them. As scheduling of scientific applications is an NP-hard problem [6, 7], it demands optimized scheduling policies to save cost and time. For many years heuristic algorithms are widely used for scheduling workflows, but they don't fit well for scheduling workflows in the cloud. Due to this, many researchers have modified the algorithms used for scheduling workflows in Grids and clusters and adopted them in the cloud.

Dynamic Resource provisioning is the approach in which the user has the freedom to select a suitable resource provider based on their requirements and just pay only for the period of usage. However, it does not provide any information about the reputation of the service provider. Another important issue in dynamic resource provisioning is the scheduling of workflow only in the provisioned resources. There are a plenty of work done on workflow scheduling. However they do not contain information about the reputation of the service provider and does not allow the user to select the scheduling strategy.

This paper presents an efficient environment that helps the user to pick up a service provider as well as a scheduling strategy as per their need. The algorithm has two phases. In the first phase, the algorithm calculates the reputation of a service provider using user rating and IOT enabled real VM performance data. User rating is a widely used approach to recommend items in the forms of ratings for items in a given domain and exploiting similarities in rating behavior amongst several users in determining how to recommend an item. The Internet of things is the network of physical devices and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data. Reputations are assigned to service providers such as High, Medium or Low. In the second phase the algorithm helps the user to select a scheduling strategy such as Deadline Constrained, Cost Constrained or Deadline and Cost Constrained. Based on the user's need, a suitable resource provider and a suitable scheduling strategy can be decided for executing the workflow.

## 2 Literature Review

Abrisami et al. [1] have proposed the ICPCP algorithm for scheduling workflows on cloud by modifying the PCP algorithm they used for scheduling workflow on Grids. The modified version of the Particle Swam Optimization (PSO) algorithm proposed by Rodriguez and Buyya [2] is useful in scheduling workflows in heterogeneous cloud environment. The dynamic dataflow introduced by Kumbhare et al. [3]. shows that flexibility can be introduced in application composition which in turn will improve the overall performance of the workflow. The dynamic load balancing algorithm developed by Zomaya and Teh [4] uses the threshold values, information exchange policies, and IPC to perform load balancing. The Heterogeneous Earliest Finish Time algorithm developed by H.Topcuoughlu was useful to schedule workflows effectively. Sadjadi et al. [8] and Pietri et al. [9] presented a performance prediction model to estimate the workflow execution time of scientific workflows considering the structure and system dependent characteristics of the workflow. The truthful dynamic workflow scheduling

algorithm [10] proposed by Fard et al. used a pricing model and truthful mechanism for dynamic scheduling of a single task in commercial multicloud environment. They optimize the two important factors of the workflow i.e. makespan and monetary cost. The control structure reduction algorithm (CSR) proposed by Li et al. [11] uses the float distribution value, the total time float is assigned to every task depending on the critical tasks in the workflow. Huang and Nicol [12] proposed a reasonable strategy to identify the reputation of the service provider. They evaluate the trust of a user based on competency, goodwill, and domain-specific expectation. Almost all the resource provisioning techniques are based on optimizing either deadline or cost. Zhang et al. [13] used k-means clustering algorithm that divide the workload into tasks of similar characteristics to minimize the overall energy consumption. The combinatorial auction based dynamic resource provisioning algorithm proposed by Zaman et al. [14] is designed to meet the changing needs of the user requirements.

There are a good number of contributions from researchers in identifying the reputation of the cloud service providers [15–19]. The trust mechanisms for cloud computing recommended by Huang and Nicol [16] evaluate the trust worthiness of a cloud service provider based on five categories such as reputation, SLA verification, transparency mechanisms (self-assessment and information revealing), trust as a service, and formal accreditation, audit, and standards. The work is a detailed study based on several factors involved in deciding the CSPs trust. However using this approach will be complex and require more details before selecting a CSP which is suitable only for learned users and difficult for end users.

Garg et al. [17] propose a framework for comparing and ranking cloud services called as Service Measurement Index (SMI), considering several factors like Accountability, Agility, Cost, Performance, Assurance, Security, privacy and Usability, called as Key Performance Indicators (KPI). The KPIs are divided into two types such as quantitative and qualitative. The framework contains three levels to take a decision. At first the customer requests are received by the SMI Cloud broker and the service providers with the required QOS is selected at the second level called the Monitoring and the last level contains the features promised by the cloud providers. Supriya et al. [18] proposed a mathematical model using Fuzzy Logic to estimate the trust values of cloud service providers. It is a two stage process in which the first stage is the implementation with the help of Mamdani Fuzzy Inference System and the second stage is the implementation using Sugeno FIS which calculates the trust rating for cloud service providers. The QoS-based ranking algorithm proposed by Tran et al. [19], for ranking web services, uses Analytic Hierarchy Process (AHP) by considering several parameters. But adopting it directly for cloud services will not yield the expected outcome. Resource provisioning is widely studied by many researchers.

The benefits of integrating IOT and Cloud is extensively studied by Stergiou et al. [20]. The network dependent routing algorithm that identify the energy economic path to the data centre, called GreeDi was proposed by Baker et al. [21]. The algorithm was also evaluated using the real Italian physical network topology. The drawbacks and weakness of the different service provides in multicloud environment was presented by Aldawsari et al. [22]. Baker et al. [23] propose a method to create cloud-dependent service compositions from the requirements metadata. The service requirements are converted into Intention Workflow Model (IWM) using the situation calculus. The

structure aware resource estimation, an novel model for effective resource provisioning and scheduling of workflows depending on the arrangement and the dependency of tasks in it is proposed by Kanagaraj et al. [24]. Further they also reduce the data transfer time between the tasks in a workflow for the benefit of data intensive workflows. Lent [25] proposed a methodology to predict the exact power and performance of VMs by aggregating the power and performance parameters of CPUs, hard disk drives, memory and other networking devices.

### 3 Reputation Based Workflow Scheduler

Due to the tremendous growth of cloud service providers, as well as consumers, there is a pressing need for the users to identify a suitable service provider for executing their tasks. The global metrics used to rate the user is the reputation of the service provider. Reputation is the collective opinion of a service provider by the users. User ratings represent several parameters, including Quality of Service (QOS), Availability, Security, response time, accessibility and price. While many different techniques have been evolved in the past few years, the interest in this area still remains high due to rising demand on sensible applications in the form of workflows requiring high level of importance and security. Hence this paper try to combine three entities such as workflow, scheduling and reputation together. However the explanation is restricted only for workflow scheduling and reputation calculation.

#### 3.1 Reputation Evaluator

Every user will rate the service provider in a 1 point scale (0-1) where 1 represents high reputation and 0 represents low reputation. The power and throughput of the VMs are obtained using IOT enabled devices. Total Reputation Score (TRS) is the sum of scores from all the users added with the values of power and throughput. The TRS is normalized by dividing TRS by the number of users and the number of IOT parameters called as the normalized score. Once Normalized score is obtained, the classification process starts. The classifier uses a simple comparison logic. As most of the existing classifiers separates the service providers as trusted or not, the users have to depend only on the trusted service providers and have to pay high rates fixed by them. On the other hand some service providers will be classified as untrusted due to marginal difference in the scores.

In order provider a fair classification among the service providers, the classification algorithm proposed in this work classifies the service provider into three categories. A service provider will be assigned High reputation (**H**), if the normalized score is greater than **0.7**. If the normalized score is between **0.4 and 0.7** then the reputation is Medium (**M**) and Low reputation (**L**) will be assigned for those whose scores are below **0.4**. This classification greatly helps the users to select the service providers with a suitable reputation level for executing their workflows. Table 1 represents the scores given by three users U1, U2, U3, and the actual VM performance indicators power and threshold of five different service providers SP1, SP2, SP3, SP4 and SP5. The reputation score can be represented by a square matrix  $\mathbf{M}(\mathbf{t}) = [\mathbf{R}_{ij}(\mathbf{t})]$ .

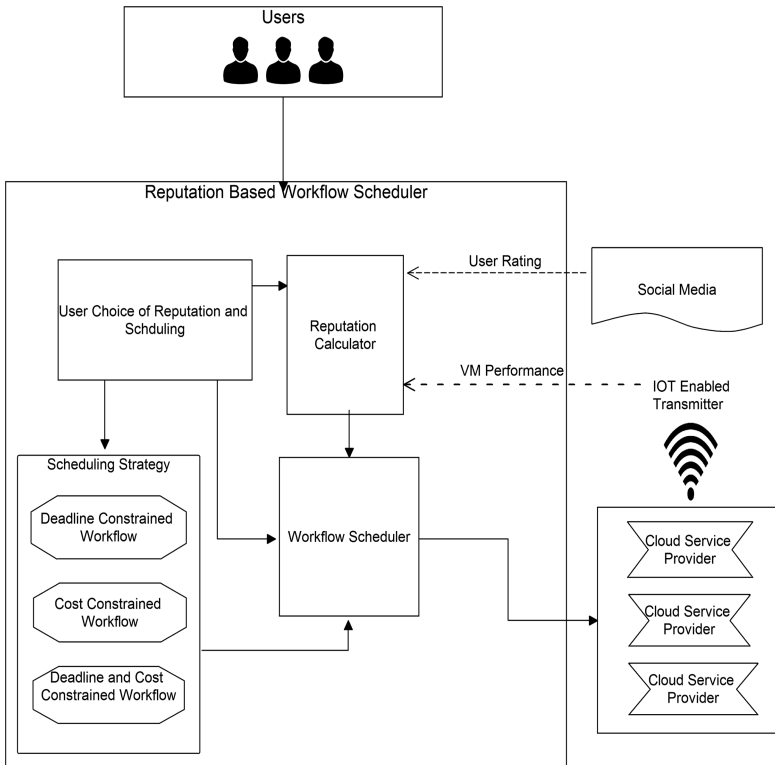
**Table 1.** Ratings of five service providers by 3 users and 2 IOT devices.

	User1	User2	User3	Power	Throughput	
SP1	0.9	0.9	0.8	0.8	0.6	$4.0/5=.8$ $2.2/5=.44$ $3.5/5=.7$ $1.0/5=.2$ $2.5/5=.5$
SP2	0.4	0.5	0.3	0.2	0.8	
SP3	0.6	0.7	0.8	0.7	0.7	
SP4	0.2	0.1	0.3	0.2	0.2	
SP5	0.4	0.7	0.4	0.6	0.4	

Where  $R_{ij}(t)$  is the local score issued by user  $j$  about service provider  $i$ , at time  $t$ . All scores are in range of 0–1. As the service providers are evaluated with respect to time, it depicts their performance at any instance and updates the ratings dynamically. Though we are considering only three users and two IOT parameters, we represent the number of users as  $n$  (Fig 1).

For each Service Provider  $i$ ,

$$Total\ Reputation\ Score\ (TRS_i) = \sum_{j=1}^n R_{ij}, \quad i = 1, 2, 3, 4, 5 \tag{1}$$



**Fig. 1.** Architecture of Reputation Based Workflow Scheduler

where,

$R_{ij}$  is the rating provided by user  $j$  for Service Provider  $i$ .

The total reputation score for the five different service providers are 4.0, 2.2, 3.5, 1.0 and 2.5, calculated using (1).

The TRS represents the total score obtained by each service provider, however the number of users who rate the service provider cannot be same for all. So we have to normalize the score based on the number of users and IOT parameters involved in rating a particular service provider. The score can be normalized by dividing the TRS of a service provider by the number of users involved in rating, as per Eq. 2.

$$\begin{aligned}
 & \text{For each Service Provider } i, \\
 & \text{Normalized Score (NRS}_i) = \frac{1}{n} * \text{TRS}_i, i = 1, 2, 3, 4, 5 \tag{2}
 \end{aligned}$$

The Normalized Score obtained by the service providers  $SP1, SP2, SP3, SP4$  and  $SP5$  are 0.8, 0.44, 0.7, 0.2 and 0.5 respectively.

Once the normalized scores are obtained, the classification algorithm will classify the service providers as High Reputation (H) or Medium Reputation (M) or Low Reputation (L), based on the normalized score obtained by them (Table 2).

$$\begin{aligned}
 & \text{Reputation of each} \\
 & \text{Service Provider } R(SP_i) = \begin{cases} \text{Low,} & NRS_i < 4 \\ \text{Medium,} & 4 \leq NRS_i \leq 7 \\ \text{High,} & NRS_i > 7 \end{cases}
 \end{aligned}$$

After selecting a service provider with a particular reputation, the next step is to perform resource provisioning. The reputation of the service provider has a direct impact in workflow execution cost. When the reputation increases the workflow cost also increases. Considering the same five service providers in our earlier discussion, if the user want a service provider with high reputation then the workflow can be executed using SP1. If the user opts for medium reputation then there are three service providers SP2, SP3 and SP5 and if the users want to go with low reputation then the choice is SP4.

**Table 2.** Service providers classified using reputation score

High reputation	Medium reputation	Low reputation
SP1	SP2, SP3, SP5	SP4

### 3.2 Workflow Scheduler

A set of related tasks that need to be executed in an order is called as a workflow. To automate any application, defining it in the form of a workflow and executing with suitable resources is a simple solution. The usual notation to represent a workflow is in the form of Directed Acyclic Graph (DAG), with nodes representing the set of tasks and edges denoting the dependencies. A Sample workflow is shown in Fig. 2.

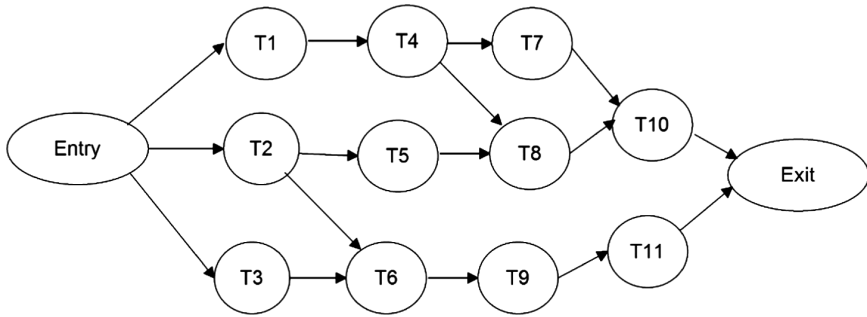


Fig. 2. Sample Workflow

Cloud computing is a mechanism in which the resources are assumed to be unlimited and there are number of service providers lending these resources. Cloud provides the users with a plenty of choices to select the best resource provider and a resource with specific configuration, which best suits their requirements. Dynamic resource provisioning is the mechanism in which the user selects resources on the spot. Based on the type of the application, the user may want to execute the workflow within a particular time (Deadline) or within a price (Cost) or both time and price (Deadline and Cost). Hence the user has the following three possible choices.

- Choice 1: Deadline Constrained (DC) Scheduling: Execute the workflow within the minimum possible time.
- Choice 2: Cost Constrained (CC) Scheduling: Execute the workflow within the minimum possible price.
- Choice 3: Deadline and Cost Constrained (DCC) Scheduling: Execute the workflow within the optimal time and price.

Based on the user's preference, several interesting analysis need to be done here.

- Case 1:** If the user prefer to execute the workflow with a service provider having high reputation then the execution cost will be comparatively high. Since the user pays high cost there is no constraint on cost. The scheduling strategy suitable for this user is DC scheduling.
- Case 2:** If the user prefer to execute the workflow with a service provider having low reputation then the execution cost will be comparatively low. Since the user pays low cost, the constraint is on the cost. The scheduling strategy suitable for this user is CC scheduling.
- Case 3:** If the user prefer to execute the workflow with a service provider having Medium reputation then the execution cost and time will be important. The scheduling strategy suitable for this user is DCC scheduling. Though we recommend the scheduling strategy, the final decision is with the user.

The sample workflow show in Fig. 2 has 11 tasks with a single entry and a single exit node. The arrows connecting the tasks represents the data dependency between the tasks, if any node has a preceding arrow then it shows that the task can be processed



only if all its predecessors are completed. The best method to estimate the runtime of individual tasks and the runtime of the workflow is the Critical Path Method (CPM), which is a mathematical model for scheduling tasks.

The critical path method is used to identify tasks that can be executed in parallel, the shortest time in which a workflow can be completed, required resources, the relationship between the tasks, time required to complete individual task and the overall time to complete the workflow. The critical path method also helps to identify the Earliest Start Time (EST), Earliest Finish Time (EFT), Latest Start Time (LST) and Latest Finish Time (LFT) for all the activities in the workflow. The Deadline of the workflow will be the Earliest Finish Time of the Last task in the workflow. Though the workflow is same the execution time and cost varies according to the scheduling algorithm used.

### 3.2.1 Deadline Constrained (DC) Scheduling

Scheduling tasks in a workflow such that every task is executed at the earliest possible time is called as Deadline Constrained Scheduling. In a real world scenario there are huge number of service providers available to execute a task at different time and cost. The responsibility of the DC scheduler is to find the service provider who can execute the task at the earliest. This need to be carried out for all the tasks and suggest a set of service providers for executing the workflow.

$$DC(t_i) = \text{Min}(ET(t_{is})), \text{ for all, } t_i \in T, \text{ and } s \in S, i = 1, 2, 3, 4, 5 \quad (3)$$

where,

T is the set of all tasks in the workflow

S is the set of all service providers

$ET(t_{is})$  is the execution time of task  $t_i$  with service provider  $s$ .

$DC(t_i)$  is the execution time of task  $t_i$  with DC scheduling.

### 3.2.2 Cost Constrained Scheduling

In Cost Constraint workflow scheduling, the workflow need to be executed as cheap as possible. The simple solution for this is to select a service provider for each task with the cheapest execution price as per Eq. 2.

$$CC(t_i) = \text{Min}(EC(t_{is})), \text{ for all, } t_i \in T \text{ and } s \in S, i = 1, 2, 3, 4, 5 \quad (4)$$

where,

T is the set of all tasks in the workflow

S is the set of all service providers

$EC(t_{is})$  is the execution cost of task  $t_i$  with service provider  $s$ .

$CC(t_i)$  is the execution cost of task  $t_i$  with CC scheduling

### 3.2.3 Deadline and Cost Constrained Scheduling

This scheduling provides an optimal resource provisioning technique that optimizes both cost and time to execute the workflow. The service provider here is selected based

on the minimum value obtained when multiplying the cost and time taken by them to execute a task. The service provider can be selected using the Eq. 3.

$$DCC(t_i) = \text{Min} (ET(t_{is}) * EC(t_{is})), t \in T, r \in S, i = 1, 2, 3, 4, 5 \quad (5)$$

where,

T is the set of all tasks in the workflow

S is the set of all service providers

ET( $t_{is}$ ) is the execution time of task  $t_i$  with service provider  $s$

EC( $t_{is}$ ) is the execution cost of task  $t_i$  with service provider  $s$ .

DCC( $t_i$ ) is the execution time for task  $t_i$  with DCC scheduling.

The overall functions of the Reputation Based Workflow Scheduler is represented in Algorithm 1. The algorithm takes three inputs such as workflow, the set of available service providers and the reputation of all the service providers. Based on these input, the algorithm schedules the workflow

---

Algorithm: 1 Reputation Based Workflow Scheduler

---

RBWS (Workflow (DAG), Service Providers (SP))

Start

Input: User Choice of Reputation (UR) & User Choice of Scheduling (UC)

if ( $UC \leftarrow$  DC Scheduling)

  for  $i \leftarrow$  1 to set of all tasks ( $t \in T$ )

    for  $s \leftarrow$  1 to all service providers ( $s \in S$ ), and  $R(Sp_i) = UR$

$DC(t_i) = \text{Min}(ET(t_{is}))$

    end for

  end for

else if ( $UC \leftarrow$  CC Scheduling)

  for  $i \leftarrow$  1 to set of all tasks ( $t \in T$ )

    for  $s \leftarrow$  1 to all service providers ( $s \in S$ ), and  $R(Sp_i) = UR$

$CC(t_i) = \text{Min}(EC(t_{is}))$

    end for

  end for

else // DCC scheduling

  for  $i \leftarrow$  1 to set of all tasks ( $t \in T$ )

    for  $s \leftarrow$  1 to all service providers ( $s \in S$ ), and  $R(Sp_i) = UR$

$DCC(t_i) = \text{Min}(ET(t_{is}) * EC(t_{is}))$

    end for

  end for

end

---

## 4 Implementation

The proposed logic is implemented in workflowsim, a popular tool for scheduling workflows in the cloud. The workflow shown in Fig. 2 is considered for implementation. Prior to implementation some preprocessing is done to populate the values for Table 3, which is used as the resource pool for this implementation. The table contains the reputation of all the service providers and the time and cost quoted by the service provider to execute each task in the workflow.

**Table 3.** Execution time, cost and reputation of service providers

Task	SP1 reputation: H		SP2 reputation: M		SP3 reputation: M		SP4 reputation: L		SP5 reputation: M	
	Time	Cost	Time	Cost	Time	Cost	Time	Cost	Time	Cost
	T1	2	4	2	5	13	2	2	2	4
T2	4	6	4	7	5	6	3	5	6	6
T3	5	10	8	7	8	9	4	7	6	6
T4	2	3	1	3	3	2	2	2	2	2
T5	6	7	4	8	6	5	6	4	5	7
T6	2	5	2	2	2	3	2	2	3	3
T7	6	7	5	5	6	5	6	4	5	4
T8	4	14	6	4	5	10	4	8	7	3
T9	5	6	14	1	5	3	5	3	5	2
T10	3	14	5	8	3	12	3	11	6	7
T11	3	12	14	4	8	9	8	5	10	6

For simplicity, the service providers with medium reputation (M) are only considered for implementation. In the sample resource pool shown in Table 3, the service providers SP2, SP3 and SP5 are having medium reputation. Hence for the workflow scheduler they are the only set of available service providers.

### 4.1 Deadline Constrained Scheduling

In deadline constrained scheduling the workflow need to be executed at the earliest possible time. Hence service providers are selected using Eq 1.

For Task  $t_1$ :  $DC(t_1) = \text{Min}(ET(t_{1s})), s \in \{SP2, SP3, SP5\}$

i.e.  $DC(t_1) = \text{Min}(2, 13, 4) = 2$ .

$DC(t_2) = \text{Min}(4, 5, 6) = 4$ .

Similarly the execution time for all tasks in the workflow can be obtained. Then by applying the critical path algorithm, the actual execution time of individual tasks and the overall execution time of the workflow based on the task dependencies is calculated and the result obtained is shown in Table 4.

According to Critical Path method, the execution time of the workflow is the Earliest Finish time of the last node. The execution cost of the workflow is obtained by adding the actual execution cost of each task in the workflow, represented in the last column in Table 4. Hence, the total time for executing the sample workflow using deadline constrained scheduling is **21** min and the total cost is **USD 68**.

**Table 4.** Execution time and cost for executing workflow using DC scheduling

Task	Durtion (in minutes)	Earliest Start Time (EST)	Earliest Finish Time (EFT)	Latest Start Time (LST)	Latest Finish Time (LFT)	Service provider	Cost in USD
T1	2	0	2	10	12	SP2	5
T2	4	0	4	2	6	SP2	7
T3	6	0	6	0	6	SP5	6
T4	1	2	3	12	13	SP2	3
T5	4	4	8	9	13	SP2	8
T6	2	6	8	6	8	SP2	2
T7	5	3	8	13	18	SP5	4
T8	5	8	13	13	18	SP3	10
T9	5	8	13	8	13	SP5	2
T10	3	13	16	18	21	SP3	12
T11	8	13	21	13	21	SP3	9

## 4.2 Cost Constrained Scheduling

In cost constrained scheduling the workflow need to be executed at the minimum possible cost. Hence select service providers using Eq. 2.

$$\text{For Task } t_1: CC(t_1) = \text{Min}(EC(t_{1s})), s \in \{SP2, SP3, SP5\}$$

$$\text{i.e. } CC(t_1) = \text{Min}(2, 3, 5) = 2$$

$$CC(t_2) = \text{Min}(6, 6, 7) = 6.$$

Similarly the execution time for all tasks in the workflow can be obtained. Then by applying the critical path algorithm, the actual execution time of individual tasks and the overall execution time of the workflow based on the task dependencies is calculated and the result obtained is shown in Table 5.

According to Critical Path method, the execution time of the workflow is the Earliest Finish time of the last node. The execution cost of the workflow is obtained by adding the actual execution cost of each task in the workflow, represented in the last column in Table 4. Hence, the total time for executing the workflow using cost constrained workflow scheduling is **36** min and the total cost for execution is **42** USD.

## 4.3 Deadline and Cost Constrained Scheduling

In deadline and cost constrained scheduling the workflow need to be executed at the optimal cost and time. Hence select service providers using Eq. 3.

**Table 5.** Execution time and cost for executing workflow using CC scheduling

Task	Duration (in minutes)	Earliest Start Time (EST)	Earliest Finish Time (EFT)	Latest Start Time (LST)	Latest Finish Time (LFT)	Service provider	Cost in USD
T1	13	0	13	8	21	SP3	2
T2	6	0	6	0	6	SP 5	6
T3	6	0	6	0	6	SP 5	6
T4	2	13	15	21	23	SP 5	2
T5	6	6	12	17	23	SP 3	5
T6	2	6	8	6	8	SP 2	2
T7	5	15	20	25	30	SP 5	4
T8	7	15	22	23	30	SP 5	3
T9	14	8	22	8	22	SP 2	1
T10	6	22	28	30	36	SP 5	7
T11	14	22	36	22	36	SP 2	4

For Task  $t_j$ :  $DCC(t_j) = \text{Min} (ET(t_{1s}) * EC(t_{1s}))$ ,  $s c \{SP2, SP3, SP5\}$

i.e.  $DCC(t_1) = \text{min} (2*5, 13 *2, 4*3) = \text{Min} (2, 3, 5) = 2$

$DCC(t_2) = \text{min} (4*7, 5 *6, 6*6) = \text{Min} (6, 6, 7) = 6$

Similarly the optimal value for all tasks in the workflow can be obtained. Then by applying the critical path algorithm, the actual execution time of individual tasks and the overall execution time of the workflow based on the task dependencies is calculated and the result obtained is shown in Table 6.

**Table 6.** Execution time and cost for executing workflow in DCC scheduling

Task	Duration (in minutes)	Earliest Start Time (EST)	Earliest Finish Time (EFT)	Latest Start Time (LST)	Latest Finish Time (LFT)	Service provider	Cost in USD
T1	2	0	2	14	16	SP2	5
T2	4	0	4	2	6	SP2	7
T3	6	0	6	0	6	SP5	6
T4	1	2	3	16	17	SP2	3
T5	6	4	10	11	17	SP3	5
T6	2	6	8	6	8	SP2	2
T7	5	3	8	19	24	SP5	4
T8	7	10	17	17	24	SP5	3
T9	5	8	13	8	13	SP5	2
T10	3	17	20	24	27	SP3	12
T11	14	13	27	13	27	SP2	4

According to Critical Path method, the execution time of the workflow is the Earliest Finish time of the last node. The execution cost of the workflow is obtained by

adding the actual execution cost of each task in the workflow, represented in the last column in Table 6. Hence, the total time for executing the workflow using Deadline and Cost constrained workflow scheduling is 27 min and the total cost for execution is 53 USD.

## 5 Comparison

The performance of the three scheduling algorithms are analyzed in this section. The sample workflow in Fig. 2 and service providers with medium (M) reputation are considered for the analysis and the results shows that the DCC algorithm out performs the other scheduling algorithms (Table 7).

**Table 7.** Comparison of performance of the scheduling algorithms

Scheduling algorithm	Time in minutes	Cost in USD
Deadline Constrained (DC)	21	68
Cost Constrained (CC)	36	42
Deadline and Cost Constrained (DCC)	27	53

## 6 Conclusion and Future Enhancements

Due to availability of large number of service providers, trusting a service provider is crucial in the resource provisioning phase of a workflow execution. To calculate the exact reputation of the service provider, user rating and IOT enabled performance indicators are used. On the other hand effective workflow scheduling algorithm is required to satisfy the users. Hence, in this paper, an effective scheduler that gives the users, the freedom to select a service provider having the desired level of reputation and the scheduling algorithm that optimized the time, cost or both, is proposed. The implementation is done in workflowsim. The experimental output shows that the algorithm that optimizes both deadline and cost is efficient when compared to the other two categories. Hence it is advisable to select a service provider with the desired reputation and adopt the scheduling algorithm that optimizes both deadline and cost. In future this work can be extended by including clustering technique to further optimize the cost and time for workflow execution.

## References

1. Abrishami, S., Epema, D.H.J.: Deadline constrained workflow scheduling algorithms for infrastructure as a service clouds. *J. Future Gener. Comput. Syst.* **29**, 158–169 (2013). Elsevier
2. Rodriguez, M.A., Buyya, R.: Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds. *IEEE Trans. Cloud Comput.* **2**(2), 223–235 (2014)

3. Kumbhare, A.G., Simmhan, Y., Frincu, M., Prasanna, V.K.: Reactive resource provisioning heuristics for dynamic dataflows on cloud infrastructure. *IEEE Trans. Cloud Comput.* **3**(2), 105–118 (2015)
4. Zomaya, A.Y., Teh, Y.H.: Observations on using genetic algorithms for dynamic load balancing. *IEEE Trans. Parallel Distrib. Syst.* **12**(9), 899–911 (2001)
5. Lin, C., Shiyong, L., Fei, X., Chebotko, A., Pai, D., Lai, Z., Fotouhi, F., Hua, J.: A reference architecture for scientific workflow management systems and the VIEW SOA solution. *IEEE T. Serv. Comput.* **2**(1), 79–92 (2009)
6. Topcuoglu, H., Hariri, S., Wu, M.: Performance-effective and low-complexity task scheduling for heterogeneous computing. *IEEE Trans. Parallel Distrib. Syst.* **13**, 260–274 (2000)
7. Garey, M.R., Johnson, D.S.: *Computer and Intractability: A Guide to the Theory of NP-Completeness*. Free-man, San Francisco (1979)
8. Sadjadi, S.M., Shimizu, S., Figueroa, J., Collazo-Mojica, X.J., et al.: A modeling approach for estimating execution time of long-running scientific applications. *IEEE* (2008)
9. Pietri, I., Juve, G., Deelman, E., Sakellariou, R.: A performance model to estimate execution time of scientific workflows on the cloud, New Orleans, Louisiana, USA (2014)
10. Fard, H.M., Prodan, R., Fahringer, T.: A truthful dynamic workflow scheduling mechanism for commercial multicloud environments. *IEEE Trans. Parallel Distrib. Syst.* **24**(6), 1203–1212 (2013). <https://doi.org/10.1109/TPDS.2012.257>
11. Li, H., Liu, H., Li, J.: Workflow scheduling algorithm based on control structure reduction in cloud environment. In: *IEEE International Conference on Systems, Man, and Cybernetics* (2014)
12. Huang, J., Nicol, D.M.: Trust mechanisms for cloud computing. *J. Cloud Comput.: Adv. Syst. App.* **2**, 9 (2013). Springer
13. Zhang, Q., Zhani, M.F., Boutaba, R., Hellerstein, J.L.: Dynamic heterogeneity aware resource provisioning in the cloud. *IEEE Trans. Cloud Comput.* **2**(1), 14–28 (2014)
14. Zaman, S., Grosu, D.: A combinatorial auction based mechanism for dynamic VM provisioning and allocation in clouds. *IEEE Trans. Cloud Comput.* **1**(2), 129–141 (2013)
15. Abrishami, S., Naghibzadeh: Cost-driven scheduling of grid workflows using partial critical paths. *IEEE Trans. Parallel Distrib. Syst.* **23**(8), 1400–1414 (2012)
16. Huang, J., Nicol, D.M.: Trust mechanisms for cloud computing. *J. Cloud Comput.: Adv. Syst. Appl.* **2**, 9 (2013)
17. Garg, S.K., Versteeg, S., Buyya, R.: SMICloud: a framework for comparing and ranking cloud services. In: *Fourth IEEE International Conference on Utility and Cloud Computing* (2011)
18. Supriya, M., Venkataramana, L.J., Sangeeta, K., Patra, G.K.: Estimating trust value for cloud service providers using fuzzy logic. *Int. J. Comput. App.* **48**(19), 28–34 (2012). ISSN 0975-8887
19. Tran, V.X., Tsuji, H., Masuda, R.: A new QoS ontology and its QoS-based ranking algorithm for web services. *J. Simul. Model. Pract. Theory* **17**, 1378–1398 (2009). Elsevier
20. Stergiou, C., Psannis, K.E., Kim, B.-G., Gupta, B.: Secure integration of IoT and cloud computing. *J. Future Gener. Comput. Syst.* **78**(3), 964–975 (2018, in Press)
21. Baker, T., Al-Dawsari, B., Tawfik, H., Reid, D., Ngoko, Y.: GreeDi: an energy efficient routing algorithm for big data on cloud. *Ad Hoc Netw.* **35**(1), 83–96 (2015). <https://doi.org/10.1016/j.adhoc.2015.06.008>
22. Aldawsari, B., Baker, T., England, D.: Towards a holistic multi-cloud brokerage system: taxonomy, survey and future directions. In: *IEEE IUCC* (2015). <https://doi.org/10.1109/cit/iucc/dasc/picom>

23. Baker, T., Rana, Omer F., Calinescu, R., Tolosana-Calasanz, R., Bañares, J.Á.: Towards Autonomic Cloud Services Engineering via Intention Workflow Model. In: Altmann, J., Vanmechelen, K., Rana, Omer F. (eds.) GECON 2013. LNCS, vol. 8193, pp. 212–227. Springer, Cham (2013). [https://doi.org/10.1007/978-3-319-02414-1\\_16](https://doi.org/10.1007/978-3-319-02414-1_16)
24. Kanagaraj, K., Swamynathan, S.: Structure aware resource estimation for effective scheduling and execution of data intensive workflows in cloud. *Future Gener. Comput. Syst.* **79**, 878–891 (2017). <https://doi.org/10.1016/j.future.2017.09.001>
25. Lent, R.: Evaluating the performance and power consumption of systems with virtual machines. In: International Conference on Cloud Computing Technology and Science (2011). <https://doi.org/10.1109/CloudCom.2011.120>



# SCICS: A Soft Computing Based Intelligent Communication System in VANET

Mamata Rath<sup>1</sup>(✉) and Binod Kumar Pattanayak<sup>2</sup>

<sup>1</sup> Department of Information Technology, C. V. Raman College of Engineering,  
Bhubaneswar, Odisha, India

mamata.rath200@gmail.com

<sup>2</sup> Department of Computer Science and Engineering, SOA University,  
Bhubaneswar, Odisha, India

binodpattanayak@soauniversity.ac.in

**Abstract.** Delicate registering procedures are tolerant of imprecision, expected on guess, concentrates on vulnerability and in light of incomplete truth. Current real issue relating to congested activity in smart city is inescapably uncertain and in this manner plan of brilliant movement control framework is a challenging issue. Because of expanding rate of vehicles at movement focuses in keen urban communities, it makes startling deferrals amid travel, chance of mishaps are increasing, superfluous fuel utilization is an issue and unhygienic condition because of contamination additionally debases the wellbeing state of general individuals in an ordinary city situation. To stay away from such issues many keen urban communities are right now actualizing enhanced activity control frameworks that work on the guideline of movement robotization with avoidance of the above indicated issues. The fundamental test lies in utilization of ongoing examination performed with on-line movement data and effectively applying it to some activity stream. In this exploratory article, an upgraded activity administration framework called SCICS (Soft Computing based Intelligent Communication System) has been proposed utilizing swarm knowledge as a delicate registering procedure with astute correspondence between smart vehicles and movement of tower focuses. It utilizes an enhanced course redirection system with executed rationale in nano robots. Under a Vehicular Ad-hoc Network (VANET) situation, the correspondence between wise sensors happen through nano robots cooperatively. Reenactment results done utilizing Ns2 test system indicates empowering results in terms of better execution to control issue.

**Keywords:** VANET · Smart city · Traffic management · Smarm intelligence

## 1 Introduction

Basic characteristics of designing an improved traffic control system includes connecting traffic signals and traffic control centres with GIS enabled digital road map of the town using intelligent computational power of data analytics as a key module. In this context, the basic challenge lies in usage of real time analytics on on-line traffic information and correctly applying it to some basic traffic flow and data analytics tools

takes data from the Traffic Management System and using GIS mapping under real time support they provide useful information to the drivers in the vehicles and help reducing the traffic congestion. Additionally, basic tourist information such as visiting places, parking area and distance are also projected in real time basis on large digital screens installed at city centres entrance points to guide the drivers towards their destination. This helps to save fuel and finally to save a lot of time spent in searching various visiting places. The smart living style in metro cities [8] is also fulfilled as the environment becomes pollution free and more hygienic. Soft computing techniques are tolerant of imprecision, uncertainty, approximation and partial truth. As the human mind can assess the probability of some event in chances, similarly soft computing methodologies also use some intelligent based techniques to assess real time problem with analytical models. In the proposed approach an Intelligent Swarm Smart Controller (ISSC) module embedded in nano robot has been designed to function during decision making in a smart traffic control system to divert vehicles in other direction at some stage of heavy traffic jam at traffic points. Depending on the traffic density a congestion level is set by the proposed algorithm and accordingly vehicles are re-directed towards less congested routes of other neighbour traffic points. Swarm intelligence provides an intelligent approach to optimization problems in distributed manner in a smart city [9]. Its logic is embedded in nano robots to collaboratively perform distributed tasks intelligently. The idea used here is that multiple number of simple nano robots which act as micro-controllers at various points of traffic way can do complex decision making logic in a collaborative and distributed way in a group to avoid congestion. In swarm intelligence, the collaborative behaviour of social insects, such as the honey-bee's dance, the wasp's building of nest, the construction of termite mound or the ants making a path with a colony are measured as strange aspect of biology. Out of inspiration of this concept, an improved methodology has been used in the current research work.

## 2 Related Work

An extensive literature study has been carried out in this section to invest various similar traffic control systems and their methodology. In [1] an improved process has been used in soft computing to control the robotic movement carefully in network. Research in [2] proposes establishing an intelligent transportation system with a network security mechanism in an Internet of Vehicles (IoV) environment, with emphasis on integrating it with traffic signal control to aid emergency vehicles more promptly arriving towards its destination. In [3] soft computing methodologies have been analyzed and assessed with appropriate case studies. Automation of Intersection Control (AIC) points mainly highlights on collision avoidance and effective traffic control. In [4] an optimized AIC technique has been used to improve traffic performance. To avoid collision, a scheduled rule has been formulated that determines priority of vehicles in the traffic point considering the travel time of vehicles. Performance evaluation of the projected algorithm has been simulated and the superior results were projected demonstrating the usefulness of the approach. We have considered this novel approach of AIC to compare the results of our proposed approach for route diversion.

In [5], a SDN-enabled connectivity-aware geographical routing protocol of VANETs for urban environment has been proposed that routes the traffic during congestion in a controlled manner.

### 3 The Proposed Approach

The proposed system has different modules such as video control system, Traffic Control System, Supervisory computer control system and peripheral devices. The Traffic Control System manages and controls the heavy traffic during congestion. It uses the video monitoring system at every traffic tower to identify excess traffic flow through video camera and when the amount of vehicles in particular route increases a pre-calculated threshold value, it executes the proposed congestion control mechanism and prevents any further vehicle to enter in that path. Traffic Control system includes the controlling of signal communication between traffic tower, sending and receiving the appropriate control frames, transmission of control frames to ISSC nano robot agents and correct control of route diversion. Similarly, supervisory computer control system is located at the control room and continuous traffic scenario is monitored through it by the human co-coordinators who are basically the transport officials. Peripheral devices include the sensors deployed at control points to provide input to the ISSC nano robots based on which decision can be taken. Similarly the response of any real time applications are triggered through actuator devices. In the discussed approach, the SCICS controller refers to the main controller that communicates with other controlling points through nano robots by sending control signals. Unique features of the proposed system are as follows.

In a smart city, there are multiple traffic points and all the routes are connected to multiple traffic points. So, it is possible to divert the path of two, three or four wheelers to some other directions when congestion takes place in a particular traffic point. The proposed approach keeps track of traffic congestion of all the traffic points and diverts the vehicles as per the congestion density [10] and as per the vehicle driver's response. It performs fair load balancing among all the available traffic points by evenly distributing vehicles in all the branch routes of every traffic instead of getting a particular traffic overloaded. It checks the priority of the vehicle in case of emergency and allows to pass such vehicles with higher priority. In case of any symptom of accident ahead, it alarms the vehicles in queue and re-directs them. Security mechanism is in-built in vehicle configuration. In case of any attack or crime signal, the alarm sensor triggers the alarm which produces alarm so that it can be detected and precaution can be taken at the nearest traffic point by the human operators. Figure 1 shows the block diagram of the proposed approach.

The SCICS controller at the central point of the traffic has been set up that observes traffic density at a route in traffic and according to the increasing density of traffic in a route, it diverts the other vehicles towards alternate routes of other traffic points of the city. This anticipate the ant colony optimization concept where ants wait and watch for the deposited information in a particular path and then continue to go ahead in the same path as long as the information deposition is the maximum at one point. So our novel

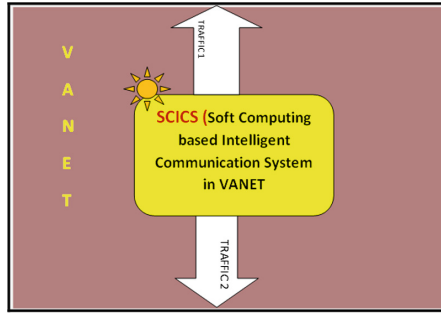


Fig. 1. Block diagram of the proposed approach

approach gets continuous status information from other neighbor traffics and to store the continuous frames an optimized data mining approach has been used [7]. When it senses congestion problem in the local traffic point, it diverts the route of the vehicle in other appropriate direction as per the embedded logic in nano robot of SCICS. Swarm intelligence helps us to find solution of distributed optimization problems in such a way that centralized control of global model is not required in Vehicular Ad-hoc Network. Communication of information frame takes place between different traffic towers.

It is assumed that the traffic control towers are deployed at every traffic point and they are wirelessly connected with each other in a VANET. Communication between towers involve transmission of control signals with frames which are periodically sent by one traffic tower to all its neighbor towers in a similar way as the beacon frames are relayed in VANETs. There control frames carry bits of information indicating congestion status of a specific route under a specific tower. Figure 2 shows the flowchart of the SCICS operation. The SCICS controller has been connected with two sensor points from one kilometer distance of the central traffic centre. If there is a congestion of vehicles sensed within the area of one kilometer of the traffic post by the SCICS, then it

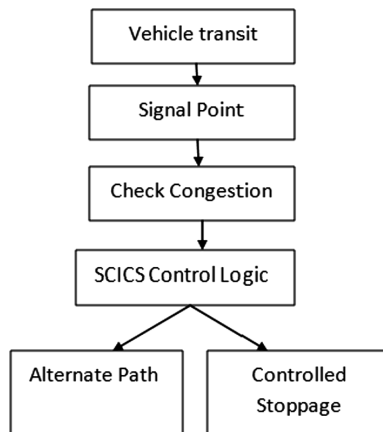


Fig. 2. Flowchart of SCICS operation

triggers one sensor to choose an option from the vehicle driver either to divert the route or stop and wait. Depending on the response from the vehicle,, if divert is selected, then the route of the incoming vehicle is diverted to another route. Determination of an alternate route is carried out by SCICS using status of the received frames from other traffic tower. Then the route is diverted by sending divert signal with the proper alternate route no. This route diversion is achieved by collaboration among swarm robots as per the signal from SCICS controller.

#### 4 Simulation Setup and Results

The proposed traffic model has been simulated using ns2 [6] with the following parameters as given in Table 1. In this simulation work, we have used the V2I approach in which vehicles interact and response with the road side sensor units through network routing devices at the road side. To simulate a realistic VANET

**Table 1.** Network parameters used in simulation

Parameter	Value
Channel type	Wireless channel
Propagation	Radio-propagation model
Network interface	Wireless phy
MAC type	MAC/802_11
Interface queue type	Drop tail
Antenna model	Antenna/omniantenna
Max packet size in ifq	100
Simulation tool	Ns2

scenario, ns2 has been integrated with SUMO and MOVE tool.

The simulation has been carried out under different traffic scenario with variable number of vehicles and the results are obtained by extracting output values from trace file created after simulation. The simulation results of current system has been compared with AIC approach [4] and the results are depicted with graphs. Figure 3 shows the comparative delay analysis of our proposed traffic model with three scenario.In SCICS traffic the delay is minimum as there is no time spent for wait and the minor delay occurred is due to the time for deciding the option to be selected to re-route the vehicle direction in alternate way.

Figure 4 shows the analysis of processing time in the said three scenario. The x-axis indicates the range of vehicles and y-axis indicates the average processing time. It can be seen that as the traffic congestion increases, the average processing time also increases in all the approaches. But due to swarm intelligence and better collaboration among nano robots in SCICS, the processing time is comparatively very less than other approaches.

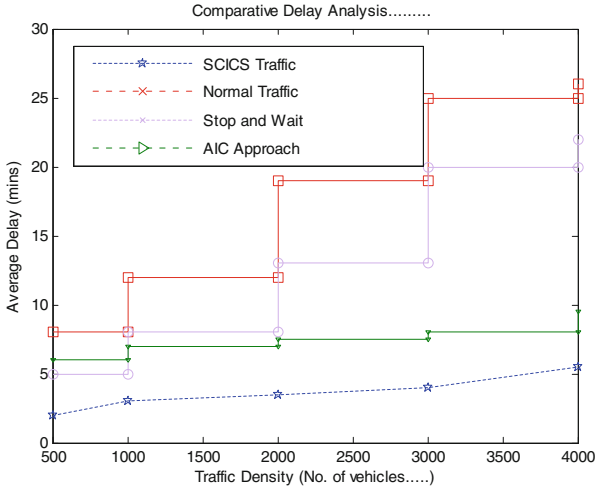


Fig. 3. Comparison of delay

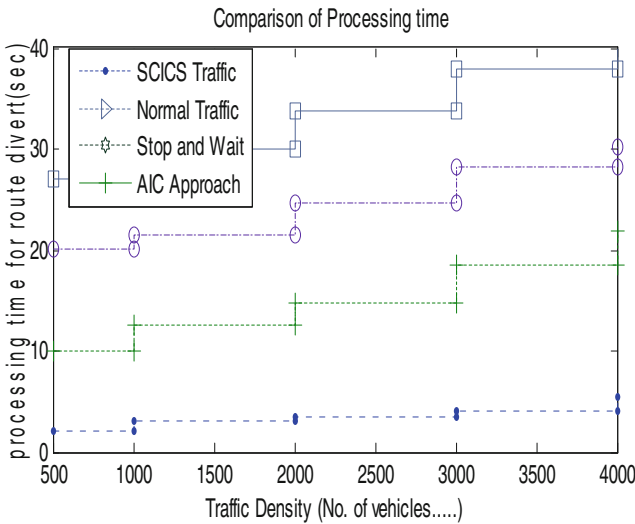


Fig. 4. Comparison of processing time

## 5 Conclusion





With the advancement of emerging technology, industrial and educational development there are more opportunity of employment and better scope of education as well as research in developing cities. The life style of people in metro cities with large volume of population is equally affected by various application and service systems. Therefore currently most of the cities are in the process of transforming into smart cities by

adopting automated systems in all possible sectors. With an objective of developing a new transport system for vehicles in a smart city, this article proposed a modern traffic control system using soft computing techniques for implementing improved route diversion mechanism with an integrated approach of solving general traffic related issues in a high volume traffic gateway. For the overall benefit of the traffic system, nano robots are used which work collaboratively as agents in the proposed swarm intelligence based approach. Simulation results show that it has an improved rate of congestion control in traffic points as it uses advanced technology of automating vehicles, big data analytics and swarm intelligence.

## References

1. Huang, H.C.: Fusion of modified bat algorithm soft computing and dynamic model hard computing to online self-adaptive fuzzy control of autonomous mobile robots. *IEEE Trans. Ind. Inform.* **12**(3), 972–979 (2016)
2. Wu, H.-T., Horng, G.-J.: Establishing an intelligent transportation system with a network security mechanism in an internet of vehicle environment. *IEEE Access*, 2169–3536 (2017). IEEE. Translations and content mining. <https://doi.org/10.1109/ACCESS.2017.2752420>
3. Shamshirband, S., et al.: Soft-computing methodologies for precipitation estimation: a case study. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **8**(3), 1353–1358 (2015)
4. Dai, P., Liu, K., Zhuge, Q., Sha, E.H.M., Lee, V.C.S., Son, S.H.: A convex optimization based autonomous intersection control strategy in vehicular cyber-physical systems. In: 2016 International IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress, Toulouse, pp. 203–210 (2016)
5. Venkatramana, D.K.N., Srikantaiah, S.B., Moodabidri, J.: SCGRP: SDN-enabled connectivity-aware geographical routing protocol of VANETs for urban environment. *IET Netw.* **6**(5), 102–111 (2017)
6. Bhalerao, P.O.: Communication system design and simulation for future micro grids in NS2. In: 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, pp. 688–690 (2016)
7. Siddique, K., Akhtar, Z., Yoon, E.J., Jeong, Y.S., Dasgupta, D., Kim, Y.: Apache Hama: an emerging bulk synchronous parallel computing framework for big data applications. *IEEE Access* **4**, 8879–8887 (2016)
8. Singh, D., Vishnu, C., Mohan, C.K.: Visual big data analytics for traffic monitoring in smart city. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, pp. 886–891 (2016)
9. Alshawish, R.A., Alfagih, S.A.M., Musbah, M.S.: Big data applications in smart cities. In: 2016 International Conference on Engineering & MIS (ICEMIS), Agadir, pp. 1–7 (2016)
10. Ianuale, N., Schiavon, D., Capobianco, E.: Smart cities, big data, and communities: reasoning from the viewpoint of attractors. *IEEE Access* **4**, 41–47 (2016)

# Classification and Recommendation of Competitive Programming Problems Using CNN

S. Sudha<sup>(✉)</sup> , A. Arun Kumar , M. Muthu Nagappan ,  
and R. Suresh 

Department of Computer Science and Engineering,  
College of Engineering Guindy, Chennai, India  
sudhaksmanian@gmail.com, arunkumaras10@gmail.com,  
muthugajal410@gmail.com, sureshravib728@gmail.com

**Abstract.** Artificial Neural Networks were first made as a component of extensive research attempts around man-made brainpower. They have discovered most of their usage in applications that are hard to express with conventional computer algorithms utilizing rule based programming. In competitive programming, the online judge presents a set of logical or mathematical problems to competitors. The contenders are required to develop computer programs to solve them. At present, the problems are labeled by users and are dubious. There is no reliable framework to recommend analogous problems. Our proposed system comprises of building a Convolution Neural Network (CNN) to perceive programming techniques utilized in the C++ program solutions. In our experiment, the considered domains are segment tree, binary search, dynamic programming and graph. The end goal of our system is to determine the approach required to solve the problem. Problems are tagged in view of the programming approach found in the solutions that are acknowledged by the online judge. The system prescribes to undertake challenges that belong to the same domain and can be tackled with similar approaches. Solving similar problems will improve the programmer's proficiency in that particular domain.

**Keywords:** CNN · Machine learning · Competitive programming  
Neural networks · Classification · Recommendation  
Dropouts · Pattern recognition

## 1 Introduction

Aspiring computer programmers improve their efficiency and skill by solving programming problems. Competitive programming is a mind sport where the contestants are given a set of programming problems and they have to build computer programs to solve them. Programs are submitted to an online judge, which checks whether they have delivered the right answer and if its execution time surpasses the given time confinement. Members are scored by how rapidly they present a right and quick program. It's a way to concentrate on the fundamentals and fine tune one's knowledge in algorithms, mathematics and data structures and reinforce the learning by applying it to different problems.



Recent studies show that Convolutional Neural Networks are effective in pattern recognition tasks and nonlinear classifications. Their ability can be controlled by fluctuating the dimensions and strides. They make a firm prediction about the nature of the program. Subsequently, when contrasted amidst standard fully connected neural systems with also estimated layers, CNNs have many less associations and parameters. Thus, they are simpler to prepare and train. CNN has fewer learning parameters than fully connected neural networks, as the parameters are shared. Number of small independent logical blocks constitutes the algorithm of the solution. Source code within a region of the program is more likely to be related than that are in different regions. CNN exploits this spatial invariance much better than the standard fully connected neural networks. A simple sorting logic may appear anywhere in the solution. Because the weights are shared in CNN, weights learnt to identify the sort logic in one part of the program can be used to recognize the same, in other parts of the program as well. Therefore, we decided to proceed with Convolutional Neural Networks.

The extent of our system and training dataset made over-fitting a huge issue. Therefore, we adopted regularization penalties and dropout techniques to improve the model performance.

The remainder of the paper is structured as follows. Sections 1.1 and 1.2 describe the need for classification and recommendation system and our objectives respectively. Section 2 portrays the related pattern recognition work using CNN. Section 3 illustrates our system design. Performance of the system is evaluated in Sect. 4. Conclusion and future works are presented in Sects. 5 and 6 respectively.

## 1.1 Need for Classification and Recommendation

Novice programmers do not have sufficient clarity in choosing the problems. Conventionally, the problems are categorized into easy, medium and hard. Since the difficulty is related to the knowledge of the programmer in different programming approaches, this kind of classification becomes ineffective. Hence, a classifier that classifies problems into different domains based upon the logic behind the problem is essential. Solving problems without any order in terms of class or domain of programming results in slackened learning. In contrast to this approach, learning is enhanced when solving multiple problems of a particular class before proceeding to another class. Repeatedly solving problems that require a specific technique helps the developer in acing that approach.

## 1.2 Objectives

- Identify the algorithm/data-structure pattern in a computer program
- Classify a problem based upon the programming technique/data-structure pattern discovered in solutions that are accepted by the online judge
- To provide recommendations to problems that belongs to the same domain category and can be solved with a similar approach.

## 2 Related Work

Deep learning architectures based on Convolutional Neural Networks (CNN) are very successful in image recognition tasks. Bezak [1] proposed a model capable of recognizing the right object in the photographs of various historical buildings in the town Trnava. Their approach based on deep learning enables automated extraction of high dimensional sets of image features and this relatively shows better accuracy and prediction rate. Simard et al. [2] proposed a set of concrete best practices that document analysis researchers can use to get good results with neural networks. They expanded the dataset by adding a new form of elastically distorted data. Ciresan et al. [3] proposed a method to improve recognition rates using committees of neural networks. They produced a group of classifiers whose errors on various parts of the training set differ as much as possible. The architecture proposed by Krizhevsky et al. [4] reduced over-fitting on image data by artificially enlarging the dataset using label preserved transformations (Data Augmentation) and by adding dropouts proposed by Srivastava et al. [5].

## 3 Proposed System

The proposed system consists of a CNN model as shown in Fig. 1 which is trained to learn the pattern of each class of problems. The system is split into four modules which are explained in the following subsections. Figure 2 shows the block diagram of the proposed system and Fig. 3 depicts the system flow.

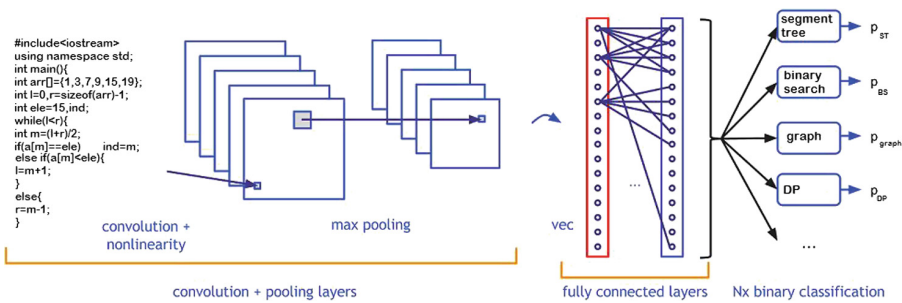


Fig. 1. CNN model

### 3.1 Data Scraping

The dataset is prepared from the C++ solutions submitted by competitive programmers around the world for the problems hosted by the Codeforces Online Judge [6]. This dataset contains solutions of the problems which belong to the four classes namely Segment Tree, Binary Search, Dynamic Programming and Graph. There are almost 6000 data points for each class. The dataset is scrapped from the Online Judge using DOM parser. Data is chosen manually with care to minimize noise.

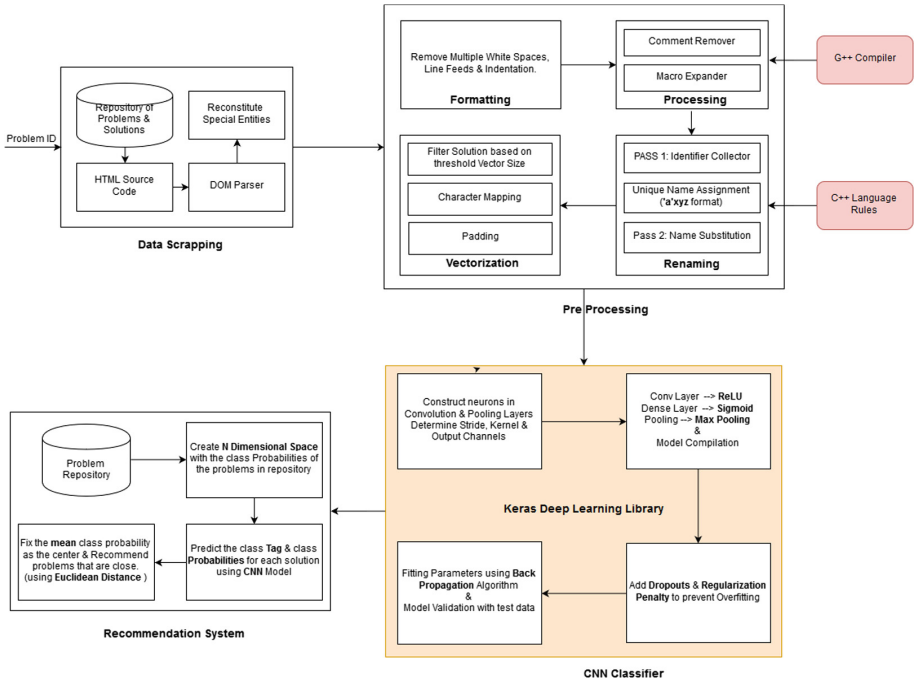


Fig. 2. Architecture diagram

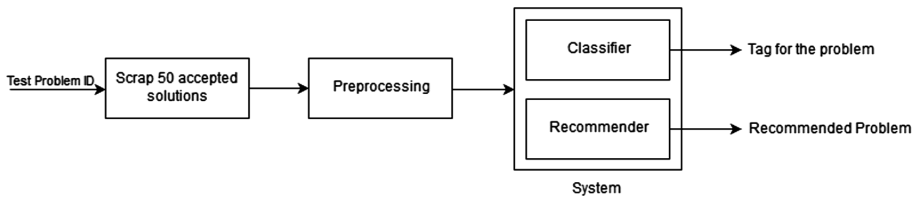


Fig. 3. System flow

### 3.2 Pre-processing

As the first step, some rudimentary pre-processing steps like whitespace and comments removal are carried out. Then the macros are expanded. The identifiers in the source code are renamed in a standard arrangement like 'aXYZ' where XYZ is a unique integer for each identifier. This step is done to enhance the code uniformity. As every solution has to be converted to a vector of same size, a threshold value is set for the vector size and the solutions whose length is greater than the threshold are discarded. Each character in the solution is then mapped onto the range [0–95] and zeroes are padded if necessary.

### 3.3 Training the CNN Model

The dataset is split into test and train data in the ratio 80:20. A CNN model is built with 2 convolution layers followed by 2 pooling layers. First kernel is of size 50 by 1, moving with 1 by 1 stride and has 16 output channels. Second kernel is of size 25 by 1, moving with 1 by 1 stride and has 32 output channels. Both the pooling layers are max-pool layers with window size 4 by 1 and moving with 4 by 1 stride. Starting with 800 by 1 vector, after first pooling layer we have 16200 by 1 vectors, after second pooling layer we have 3250 by 1 vectors. Then we have a fully connected layer with 256 neurons followed by a softmax layer with  $n$  outputs where  $n$  represents the number of classes. The activation function for both the convolutional layers and the first fully connected layer is RELU (Rectified Linear Unit). Overfitting could be a serious issue in neural networks. Srivastava et al. [5] proposed a strategy known as dropouts to avoid overfitting. Dropout prevents overfitting and provides a way of roughly combining many different neural network architectures efficiently. The term dropout refers to dropping out units (hidden and visible) in a neural network. By dropping a unit out, we mean temporarily removing it from the network, together with all its incoming and outgoing connections. The decision of which units to drop is random. Each unit is retained with a probability  $p$ , where  $p$  is a hyper parameter. As a measure to avoid overfitting, a dropout of probability 0.25 is added after each pooling layer and a dropout of probability 0.5 is added after the first fully connected layer. The model is trained for 20 epochs.

### 3.4 Classification and Recommendation

The basic assumption made is that there are atleast 50 accepted solutions for any problem in an online judge. To classify a problem, almost 50 accepted solutions that are acknowledged by the online judge are scraped. The pre-processing steps as explained in the above section are applied to all these solutions. Once the preprocessing step is done, we have the vector representation of each of the solution. The vectors are then passed to the trained CNN model and it yields a set of predictions that represent the probabilities of the solution being in each class. As any prediction system cannot be completely reliable, some solutions may be classified as one class and few solutions may be classified as some other class. The class which is predicted the most number of times will only be considered and the problem will be labeled by that class. An  $N$ -dimensional space, as shown in Fig. 4, is assumed for the purpose of recommending similar problems where  $N$  is the number of classes. Each axis represents a class and the values along the axis represent the probability of a problem being that class. The problem in consideration is also assumed to be plotted in that  $N$ -dimensional space. All problems in the online judge are also assumed to be in the problem space. Then the Euclidean distance between the problem in consideration and all the other problems is calculated. If two problems are closer to each other in the  $N$ -dimensional space, it intuitively means that those two problems are having similar approach and it is extremely possible for them to be in the same class. Hence those problems with the least distance to the problem in consideration are recommended to the user. As an example, we can assume the famous Rod cutting dynamic programming problem and

we want to tag it. As we have already trained the CNN model with many dynamic programming solutions, the system will be able to identify that pattern. Now we scrap 50 accepted solutions for that problem and pre-process them. Then the vector for each solution is passed to the trained CNN model and it gives a vector of probabilities for each solution. As we expected, most of the solution will be recognized as dynamic programming. So the problem will also be classified as dynamic programming. The other dynamic programming problems will be closer to the rod cutting problem. One such problem with the least Euclidean distance to the rod cutting problem will be recommended.

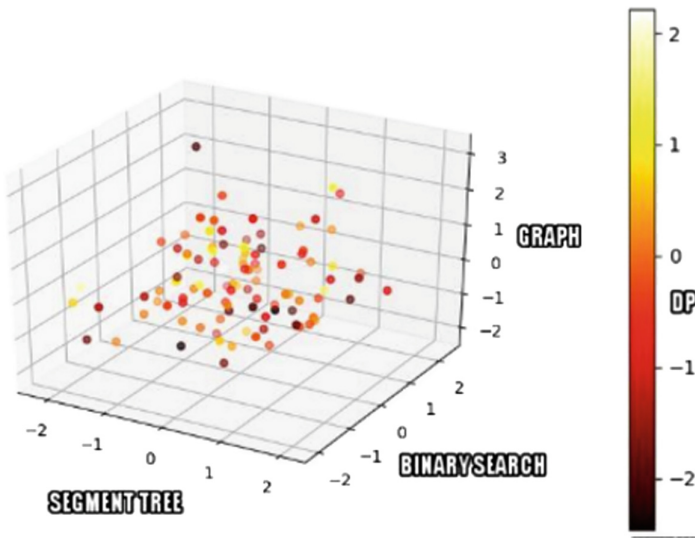


Fig. 4. N-dimensional space

## 4 Results

### 4.1 Evaluation Criteria

For evaluating the model, source codes (100 in number - 25 in each class) were chosen such that it does not coincide with the dataset picked for training and tested against the model tagging system. The recommendation system is evaluated by judging the relevancy of the problems recommended for each of the 40 problems (10 problems of each class) chosen as the test dataset. The relevancy is curated by checking if the recommended problem has the same tag as the given problem. Besides Codeforces, the test dataset is extracted from online programming portals, for example, geeksforgeeks. The following metrics are used to evaluate the model.

**Confusion Matrix.** A confusion matrix is a table that is used to outline the efficiency of a classification model on a set of test data for which the true values are known. It

contains information about actual and predicted classifications performed by a classification system. Each column of the matrix depicts an instance in a predicted class while each row depicts an instance in an actual class.

Figure 5 depicts the visualization of performance of the tagging system.

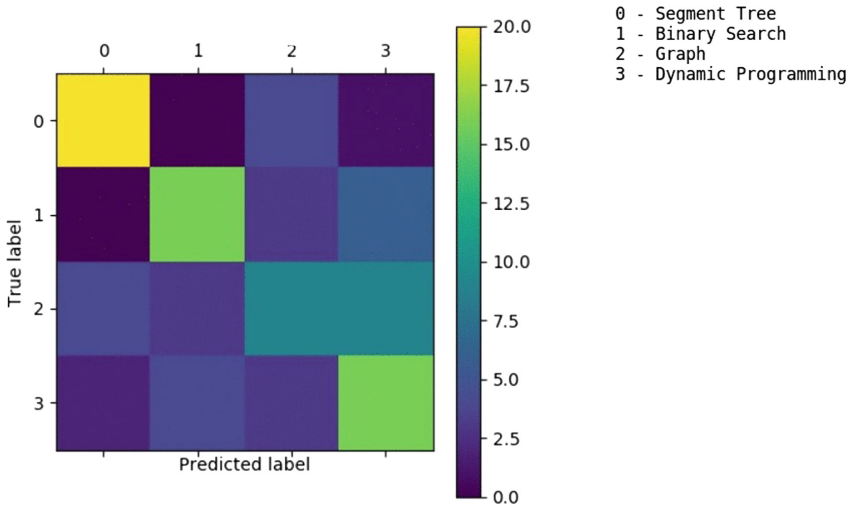


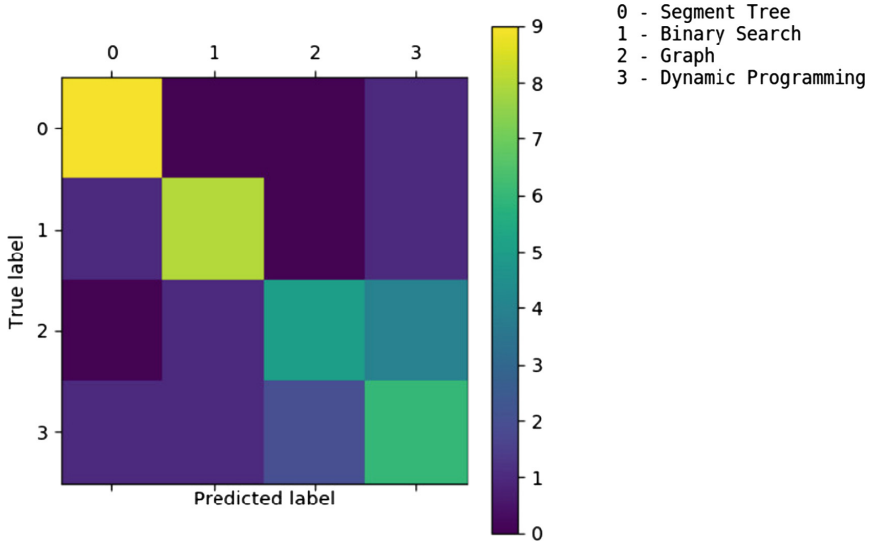
Fig. 5. Confusion matrix for the tagging system

- Out of 25 segment tree problems, 20 were predicted correctly.
- Out of 25 binary search problems, 16 were predicted correctly.
- Out of 25 graph problems, 9 were predicted as DP and 9 were predicted as graph because of the similarity in pattern between graph and DP and noise in the dataset.
- Out of 25 DP problems, 16 were predicted correctly.

Figure 6 depicts the visualization of performance of the recommendation system.

- For 10 segment tree problems, out of the 10 problems recommended, 9 of them were grouped correctly under segment tree.
- For 10 binary search problems, out of the 10 problems recommended, 8 of them were grouped correctly under binary search.
- For 10 graph problems, out of the 10 problems recommended, 5 of them were grouped correctly under graph. Because of the similarity of code pattern between Dynamic Programming and Graph, the recommended problems for graph are distributed between those two classes.
- For 10 dynamic programming problems, out of the 10 problems recommended, 6 of them were grouped correctly under dynamic programming.

**Recall.** Recall measures the proportion of positives that are correctly identified as such. It can be thought of as a measure of a classifier’s completeness.



**Fig. 6.** Confusion matrix for the recommendation system

Recall rate is given by Eq. 1.

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

**Precision.** Precision is the fraction of relevant instances among the retrieved instances. It can be thought of as a measure of a classifier's exactness.

Precision is given by Eq. 2.

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

**Accuracy.** Accuracy is the number of correct predictions made out of the total number of predictions made.

Accuracy is given by Eq. 3.

$$Accuracy = \frac{TP + TN}{Totalobservations} \quad (3)$$

The average accuracy of tagging system is calculated to be 80.5

The average accuracy of recommendation system is calculated to be 85

- True Positive (TP): number of correctly recognized observations for any class C.
- False Positive (FP): number of observations that were incorrectly assigned to any class C.

- True Negative (TN): number of correctly recognized observations that do not belong to any class C.
- False Negative (FN): number of observations that were not recognized as belonging to any class C.

Figure 7 depicts the performance metrics of tagging system and Fig. 8 depicts the performance metrics of recommendation system.

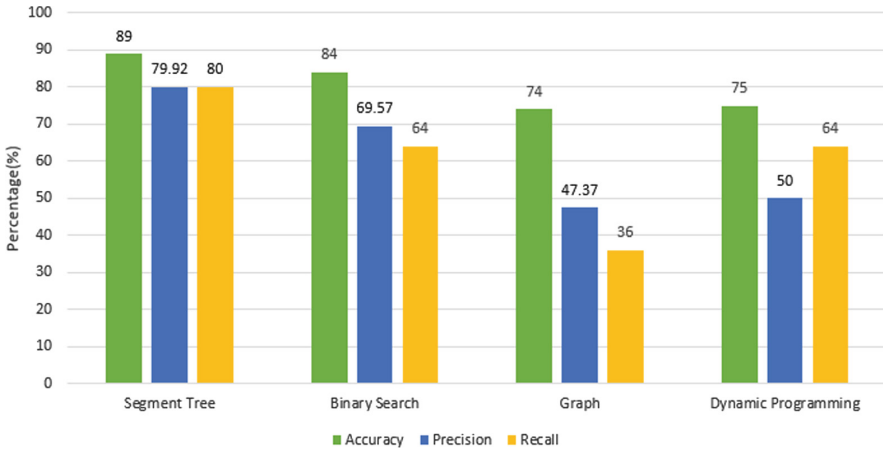


Fig. 7. Performance metrics of tagging system

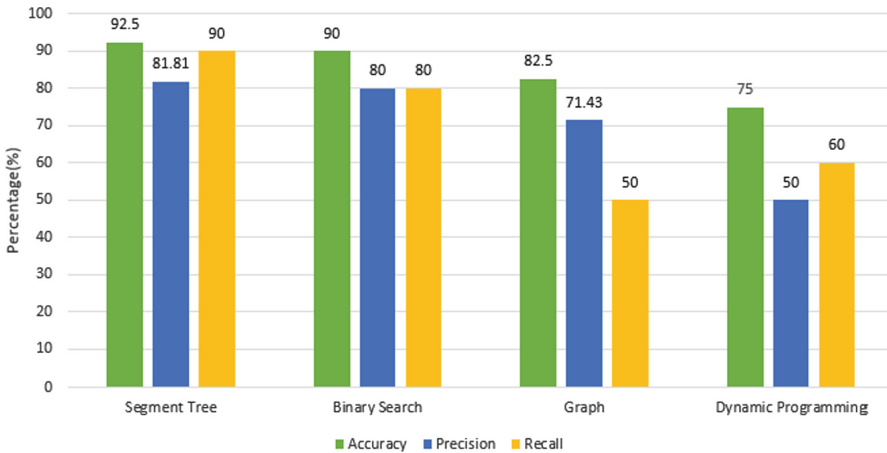


Fig. 8. Performance metrics of recommendation system



## 5 Conclusion

Programming is variable but concepts are constants. Programming varies from language to language but algorithms and data structures form the base of programming. So, it is very important for every programmer to have a profound knowledge of these concepts. This paper mainly aims to provide better Competitive Programming platforms for beginners to improve their logical and problem solving skills. Convolution Neural Network which is mostly utilized for the characterization of images is tweaked to distinguish designs in a C++ source code. The problems are classified based on the pattern perceived among them. The system also suggests problems that require similar approach to solve. These two components of the system will help in building an intelligent competitive programming stage. These features help the programmers to develop knowledge about algorithms and data structures and when to use them. This results in improvement of programming skills and helps to develop efficient code.

## 6 Future Works

The research can be further extended by making the CNN model language independent if the vectorization is done from intermediate code or object code of the source program. The way memory is read or written for each class can also be used to improve the efficiency of the system. The model can be further extended to recognize other algorithm patterns. The system can be enhanced to decide further deeper labels, for example, bottom up or top down Dynamic Programming, instead of simply dynamic programming. Hints can also be provided at every stage of coding otherwise the programmer has to look at the entire solution which retards her/his thinking ability. These hints can be logical errors which cannot be determined by compilers and low level implementation tricks like stack, queue, map or hashing.

## References

1. Bezak, P.: Building recognition system based on deep learning. In: Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), pp. 1–5, Lodz, Poland (2016). <https://doi.org/10.1109/ICAIPR.2016.7585230>
2. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition, vol. 3, pp. 958–962 (2003)
3. Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Convolutional neural network committees for handwritten character classification. In: IEEE International Conference on Document Analysis and Recognition, pp. 1135–1139 (2011). <https://doi.org/10.1109/ICDAR.2011.229>

4. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of Conference on Neural Information Processing System, pp. 1097–1105 (2012)
5. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
6. Codeforces. <http://codeforces.com>

# A Generic Context-Aware Service Discovery Architecture for IoT Services

S. Sasirekha<sup>1</sup>(✉), S. Swamynathan<sup>2</sup>, and S. Keerthana<sup>1</sup>

<sup>1</sup> Department of Information Technology, SSN College of Engineering, Chennai, India  
sasirekhas@ssn.edu.in

<sup>2</sup> Department of Information Science and Technology,  
Anna University, Chennai, India  
swamyns@annauniv.edu

**Abstract.** Internet of Things (IoT) is the networks of physical objects or things that are connected to the Internet via the embedded devices to provide appropriate services to the user. These services stimulate the physical object to communicate and interact with the global environments efficiently. In the current state-of-art, with the increase in the number of services provided by the physical objects, the discovery of desirable services is a key challenge. Hence, in this work, a generic architecture for context-aware service discovery among the existing IoT services has been proposed. The context aware service discovery architecture investigates and understands the context based on the submitted users request and provides the users with the relevant information and services. In order to performs a comparison between the properties of the user request and the registered instance, this discovery architecture has been developed using the Web Ontology Language (OWL) based context-aware computing and a service oriented approach. To justify the performance of the proposed context aware solution, test scenarios and user survey has been carried out on a weather forecasting scenario to determine the relevancy retrieval ratio.

**Keywords:** Internet of Things · Ontology · Semantics · Context-aware

## 1 Introduction

The Internet of Things (IoT) is the global networks of the physical objects with a high level of connectivity to the Internet and built intelligence into the object, devices and the system [1]. This allows the physical object to be sensed, communicated and interacts with the surrounding environment. It improves the efficiency, accuracy and economic benefit. It is connected individually identifiable through the embedded devices inwards the existing Internet infrastructure to offer advanced connectivity of devices, systems, and services. It has emerged as a global Internet-based information architecture facilitating the exchange of goods and services [2]. The IoT connects the various technologies to enable new

applications by connecting physical objects, each other using knowledge decision-making [3]. IoT goes beyond M2M (machine-to-machine) communications and covers a variety of protocols, domains, and applications. In order to use the services and data generated in the IoT effectively, search and discovery mechanisms are crucial. These mechanisms should support locating resources and services related to an entity of interest in the physical world. Traditional web service discovery approaches are not suitable for service discovery in IoT, due to the differences between real world services and traditional web services.

A sensor as a service gives functional aspects of the sensor as services by hiding technical details of the sensors from the user. It helps to create, managing, discovering and delivering sensor functionalities and capabilities as services as well as to build the IoT infrastructure and services. The service has been experienced with similar challenges on service level, and service has already developed a number of standardized solutions to address such challenges. Therefore, the key idea is to address the sensor by lifting one step up and transferring them into service-world challenges so that one can exploit all the existing service-world standards to cope with sensor-world challenges.

Service discovery is the action of finding a service provider for a requested service. When the location of the demanded service is retrieved, the user may further access and use it. The service discovery mechanism is to create a highly dynamic infrastructure wherever users would be able to obtain explicit services of interest, and service suppliers providing those services would be able to announce and advertise their capabilities to the network [4]. Furthermore, service discovery minimizes human intervention and allows the network to be self-healing by automatic detection of services, which have become unavailable. Once services have been discovered, devices in the network could remotely control each other by adhering to some standard of communication. The efficient service discovery for the IoT remains a challenge. IoT environments are highly dynamic such as physical mobility, low power and it involves a massive amount of heterogeneous nodes focused on different applications. These features raise various issues for an effective and efficient discovery (e.g., availability, scalability, interoperability), which needs a high degree of automation (e.g., self-configuring, self-managing, self-optimizing).

The context-aware computing [5] defines context as information is used to characterize the things of an entity. An entity can be an individual, place, or object that has thought of relevant to the interaction between a user and an application as well as the user and application themselves. Context awareness, on the other hand, is defined as a property of a system that uses context to provide relevant information and/or service to the user, where relevance depends on the user's task. Consequently, context-aware service discovery is defined because of the ability to make use of context data to get the most relevant services for the user. Context-aware systems are capable of changing their operations to the current context without change user interference and its increasing serviceability and effectiveness of carrying out environmental context into account.

A traditional service discovery system contains two main areas are network architecture and service descriptions. Designing an effective and an efficient discovery mechanism faces many new challenges and IoT devices in the physical world provide requirements such as Real-world services. Normally, data of real world objects and events are available globally and in vast amounts this not suitable for service discovery in IoT, because of the search and discovery mechanisms are important.

Context awareness plays an important role in the discovery framework to enable services according to the current situation with minimal human intervention ontology to encode the context information and match queries with services to select the most appropriate services.

Ontology provides a common understanding of the context, and it can help the discovery service to infer relationships between entities and context. This challenge requires the discovery mechanism to be scalable with respect to a large number of objects. Moreover, due to the mobility of physical entities and devices and other dynamic changes, the relations between IoT Services, IoT devices, and the physical environments may also change over time. Hence, in this work a generic architecture for context-aware service discovery among the existing IoT services has been proposed. The context aware service discovery architecture investigates and understands the context based on the submitted users request and provides the users with the relevant information and services. In order to it performs a comparison between the properties of the user request, the registered instance, this discovery architecture has been developed using the Web Ontology Language (OWL) based context-aware computing, and a service-oriented approach [6].

In the remainder of the paper, the related work is presented in Sect. 2, In Sect. 3, the proposed system architecture is briefly illustrated. Then, in Sect. 4, implementation and results are demonstrated. Finally, in Sect. 5, it is concluded with remarks.

## 2 Related Works

There are many research works carried out related to context aware computing in IoT domain. To justify the objective of the proposed work, some of the most relevant works about the context aware service discovery are discussed as follows. Li et al. [7] discusses a context-aware semantics- based service discovery mechanism the service discovery is a challenging task for the Internet of things due to the large number of services provided by the physical things so they proposed a LOCA (LOcation-preserving Context-Aware discovery framework) that can discover services based on the context requirement efficiently. The discovery framework was built based on the distributed peer-to-peer (P2P) architecture. LOCA is entirely scalable and robust, it improves the integrity and security of the discovery model, and they proposed an ontological model to provide context information.

Similarly, Cubo et al. [8] has defined a Context - aware sensor search, selection and ranking model for IoT, connecting large number of sensors are connected and generated the huge amount of data. Whereas the increasing number of sensors in the external environment. It is difficult to find the sensor and its overlapped with other sensors and it obtains some redundant functionality so, they proposed the search and sensor selection based on the user request and the priorities and the multimodal space techniques is used to rank and index sensors. CASSARAM recognize the characteristics of sensors for searches like reliability and accuracy, etc.

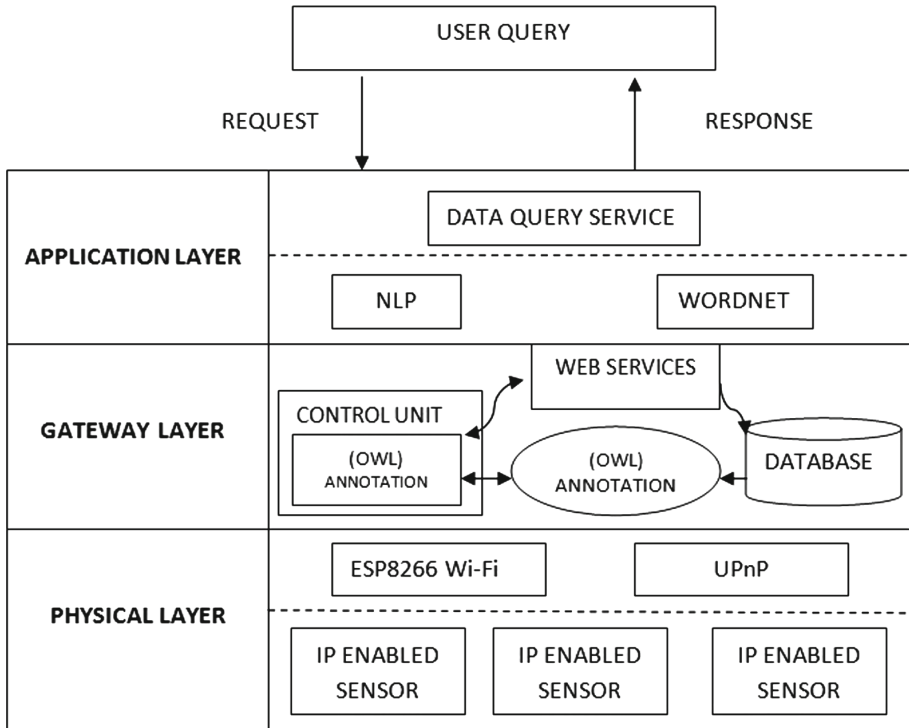
Butt et al. [9] has proposed a Trendy: an adaptive and context-aware Service Discovery Protocol for the IoT. This protocol employs CoAP primarily based restful web services that change application-layer integration of constrained domains and the Internet. Trendys resource directory provides service discovery with a context aware service, selection using user- and network-based context. A trendy adaptive timer based on demand manages the trade-off between status, maintenance, load and dependability. It permits the service hosts to share their load with the resource directory and decreases the service invocation delay. They introduced a context-based grouping technique wherever the resource directory divides the network at the application layer, by making location-based groups. This grouping of nodes localizes the control overhead and provides the base for service composition, localized aggregation and process of information. They proposed the trendys techniques decrease the control overhead, energy consumption and service invocation delay.

Likewise, Xiao et al. [10] discusses a context modeling approach which might dynamically handle various context types and values. They proposed an approach to dynamically derive a context model from ontologies and suggest services using context. Given a group of available context types and values, our approach will dynamically find their relations and construct a context relation model. By discovering the semantic relations among context values, this approach can establish a users wants hiding within the context values and generate looking criteria for service discovery.

It is evident from the above works, that there is need for an optimal service discovery mechanism to retrieve the relevant service in IoT. Hence, in this work a generic architecture for context-aware service discovery among the existing IoT services has been proposed.

### 3 Context Aware Service Discovery Architecture

The system architecture contains four layers such as Application layer, Gateway layer, Communication layer and Physical layer as shown in Fig. 1. The application layer is responsible for receiving a request and sending the response to external applications using HTTP (Hypertext transfer protocol). It receives the request and processes the request submitted by the user. After the request is processed, the application layer redirects the action to the gateway layer. The gateway layer act as a handler for the service provider and it is responsible for



**Fig. 1.** System architecture for service discovery.

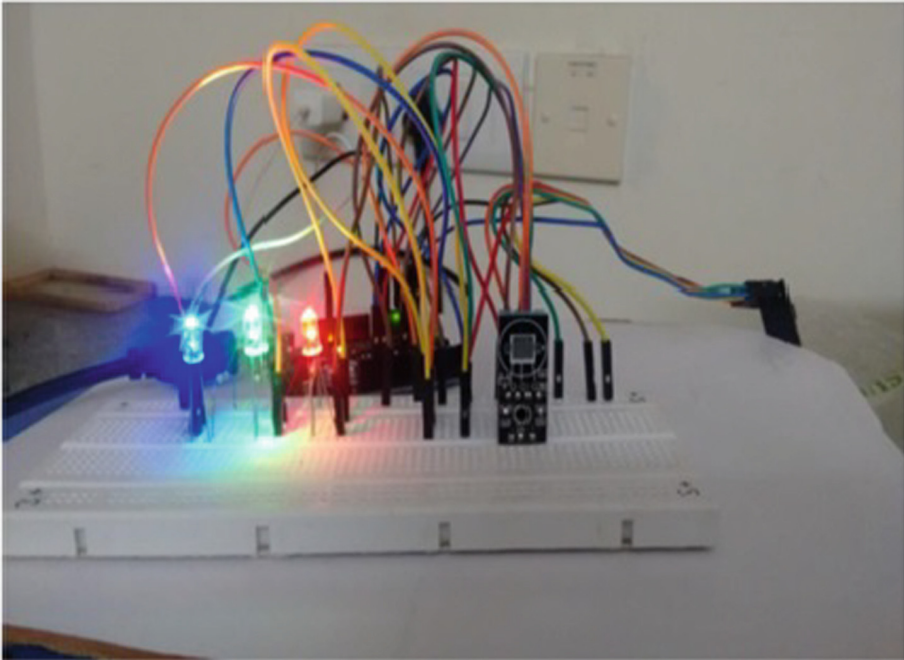
discovering the relevant web services using the suitable OWL file and based on the URL mapped with the available services. The communication layer interconnects the physical layer using Ethernet, Wi-Fi (ESP8266) and UPnP. The Physical layer has one or more IP-based sensors for receiving data from the communication layer and sends the response back to the gateway layer, using communication layer again.

The user from the application layer requests for sending request, it is processed by the proposed context aware IoT architecture. The service composition module of the architecture handles the service request. This, in turn, invokes the desired REST services through the BPEL process definition. It executes the set of all possible sequences of valid tasks to deliver the expected result to the application user. In order to handle the BPEL composition semantically, the composition module, in turn, accesses the service discovery and the knowledge base, which has entities information about IoT devices as merged ontologies and annotated semantic web, services information. It is difficult to find the most optimal workflow sequence path that can satisfy the input request for an IoT based application as it dynamically connects billions of things in the network. Therefore, in this work, considering the scalability and dynamic nature of the IoT, the composition workflow is bounded with the annotated semantic web

service, which is available in the knowledge base module of the middleware. This ontology provides the meta-data information of the IoT services that can provide the highest probability of seeking the appropriate service solution to generate the approximate composition workflow. The composition process, in parallel also searches for the service discovery module to invoke the appropriate service based on the ontological workflow solution. For this purpose, initially the IoT services are registered in the service registry with meta-data information such as the purpose of sensing, deployment location and physical characteristics. The lookup process of the IoT services for its approximate solution is bounded with the merged ontology available in the knowledge base.

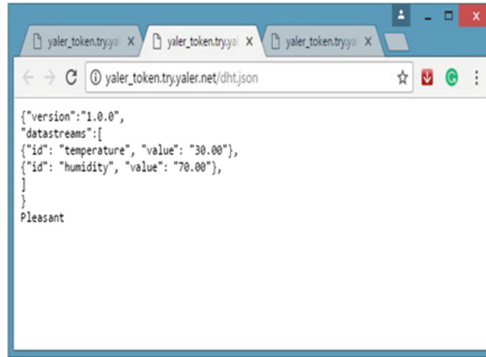
## 4 Implementation and Results

The implementation setup was build for a weather forecasting scenario. The design of physical layer contains sensors such as DHT11 Sensors and LM35 sensors. To measure both the moisture (humidity) and air temperature a DHT11 sensor is used. Then to measure the temperature in degree centigrade ( $^{\circ}\text{C}$ ) with decent accuracy a LM35 sensor is used. These sensors are finally integrated with the microcontroller board such as Arduino UNO and Arduino mega2560 board [11]. After building the setup environment and ready to monitor the deployed environment, the microcontroller starts to processes the sensors, retrieves the data,



**Fig. 2.** Physical layer setup.





**Fig. 3.** Sample JSON data retrieved with semantic annotation.

and passes them to the gateway layer through the Wi-Fi module connected to it as shown in Fig. 2. The similar prototype setup is deployed in different location in order to monitor the weather status of that particular region. For instance, the set up was deployed in the various location points of our college campus.

The Arduino connects both hardware and software with the ESP8266 Wi-Fi module using the UPnP protocol. The UPnP is a set of protocol that is able to communicate with the networked devices within the network. The Ethernet shield allows an Arduino device to connect to the Internet using the Ethernet cable. It is used to enable the Arduino to send and receive the data to the web server over Ethernet. The ESP8266 is a compatible and a self-contained Wi-Fi module. It is used to connect the Wi-Fi networks and interact with the external world over the TCP/IP stack.

The Gateway layer is connected link between the physical layer and the communication layer. It is responsible for collecting data from the sensor and discovering the web service. In this work, to collect the data from different formats first, a yaler.net relay service is used to access the ESP8266 from the web. Next, the data or URL are collected using a web server. Finally, using a NEO6MV2 GPS module, geo-location is received using a GPS receiver and then the OWL file is fetched along with the service URL of the relevant service based on the user request. The Application layer is the top most layer of IoT, which are responsible for providing services to the users. This layer includes the protocols and interfaces to identify the devices and communicate with each other. It contains the HTTP Protocols that are the lightweight protocol to access the data on WWW (World Wide Web). It utilizes the TCP connection to sending the client request and receiving the server response to the web browser. Here, the application layer displays the sensor information based on the context aware system.

Figure 3 shows the sample data retrieved with semantic annotation using the above described setup environment in the JSON format. The data is semantically annotated using the semantic modeling techniques. The data is predefined to be generated in the JSON format as shown in the Fig. 4.

```
// assemble the json output
jsonOut += "{\n";
jsonOut += pin;
jsonOut += "\":\n";
jsonOut += outValue;
jsonOut += "\n}";
// return value with wildcarded Cross-origin policy
client.println("HTTP/1.1 200 OK");
client.println("Content-Type: text/html");
client.println("Access-Control-Allow-Origin: *");
client.println();
client.println(jsonOut);
```

Fig. 4. Code snippet to predefine data generation in JSON format.

The screenshot shows a web form titled "NODE REGISTRATION". It has four main input sections: "Node Name" (text input with "Node2"), "Enter the URL to register the service:" (text input with "http://yaler\_token.try.yaler.net/lm35.json"), "Enter the place:" (dropdown menu with "Seminar Hall" selected), and "Enter the service offered:" (dropdown menu with "Temperature" selected). A "Register" button is at the bottom.

Fig. 5. Node data registration process.

The yaler web server, which acts as gateway server in this setup environment, receives the sensed data and populates the database for later analysis on the environmental condition. Next, for providing additional knowledge to the ontology about the service provided and the environmental condition a node registration service is included. The node registration process is used to add the semantic information of the sensed data as shown in Fig. 5.

Finally, to include the knowledge information in the OWL file [12]. Therefore, the OWL file is created using the protégé tool [13] as shown in the Fig. 6. The ontology source is defined to provide a knowledge structure for the weather forecast for a particular region [14].

For the users to experience the Context aware service discovery in the IoT environment, a user interface is designed to submit the user request. This portal acts as a service requester for the context aware service discovery architecture. The request submitted using this portal is process and the relevant service are returned as shown in Fig. 7.

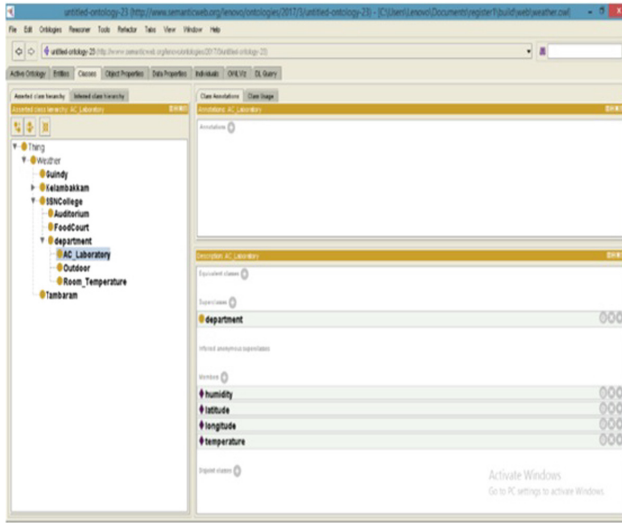


Fig. 6. Ontology structure of weather forecasting system.



Fig. 7. User interface for submitting user request.

The evaluation on the proposed architecture was carried out to justify the performance of the system. It was tested for the relevancy of the service discovered in a particular context. It is inferred from the results that accuracy of the relevancy has improved extensively when compared with the analysis carried out on the same example without semantic service. Table 1 shows the relevancy retrieval ratio estimated for different test cases executed on a weather forecasting scenario. It is observed from the precision and the recall ratio that the proposed system provides an optimal solution for discovering relevant context aware services.

**Table 1.** Relevancy retrieval ratio

Discovered services	Relevant services	Retrieved relevant services not retrieved	Irrelevant services retrieved	Precision (%)	Recall (%)
17	10	5	7	58.82	66.67
14	9	4	5	64.29	69.23
16	10	7	6	62.50	58.82
10	7	6	3	70.00	53.85
6	4	3	2	66.67	57.14
13	8	4	5	61.54	66.67
9	4	4	5	44.44	50.00

## 5 Conclusions

In this paper a context-aware service discovery architecture was proposed for retrieving useful and the most appropriate services for requesting user. The proposed solution was tested on a weather forecasting scenario. The architecture was developed with the service in semantic environment OWL to Semantically Annotate Web-Services. It is inferred from the results that accuracy of the relevancy has improved extensively when compared with the analysis carried out on the same example without semantic service. It is observed from the precision and the recall ratio that the proposed system provides an optimal solution for discovering relevant context aware services. The development of generic context aware service discovery architecture can serve as a platform to create context-aware and personalized IoT-based services and applications.

## References

1. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. *Future Gener. Comput. Syst.* **29**(7), 1645–1660 (2013)
2. Chen, S., Xu, H., Liu, D., Hu, B., Wang, H.: A vision of IoT: applications, challenges, and opportunities with china perspective. *IEEE Internet Things J.* **1**(4), 349–359 (2014)
3. Cirani, S., Davoli, L., Ferrari, G., Léone, R., Medagliani, P., Picone, M., Veltri, L.: A scalable and self-configuring architecture for service discovery in the internet of things. *IEEE Internet Things J.* **1**, 508–521 (2014)
4. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M.: Internet of things: a survey on enabling technologies, protocols, and applications. *IEEE Commun. Surv. Tutor.* **17**(4), 2347–2376 (2015)
5. Talal, B.K., Rachid, M.: Service discovery - a survey and comparison. *CoRR* abs/1308.2912 (2013)
6. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L.A.: OWL web ontology language reference. Technical report, W3C, February 2004. <http://www.w3.org/TR/owl-ref/>

7. Suraci, V., Mignanti, S., Aiuto, A.: Context-aware semantic service discovery. IEEE Conference Publication
8. Li, J., Zaman, N., Li, H.: A decentralized locality-preserving context-aware service discovery framework for internet of things. In: 2015 IEEE International Conference on Services Computing, pp. 317–323, June 2015
9. Cubo, J., Canal, C., Pimentel, E.: Context-aware service discovery and adaptation based on semantic matchmaking. In: 2010 Fifth International Conference on Internet and Web Applications and Services, pp. 554–561, May 2010
10. Butt, T.A., Phillips, I., Guan, L., Oikonomou, G.: Adaptive and context-aware service discovery for the internet of things. In: Balandin, S., Andreev, S., Koucheryavy, Y. (eds.) NEW2AN/ruSMART -2013. LNCS, vol. 8121, pp. 36–47. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-40316-3\\_4](https://doi.org/10.1007/978-3-642-40316-3_4)
11. Banzi, M.: Arduino board and pin configuration (2005). <http://arduino.cc>
12. Xiao, H., Zou, Y., Ng, J., Nigul, L.: An approach for context-aware service discovery and recommendation. In: 2010 IEEE International Conference on Web Services, pp. 163–170, July 2010
13. Protege: The protégé ontology editor and knowledge acquisition system (1999). <http://protege.stanford.edu>
14. Yang, S., Xu, Y., He, Q.: Ontology based service discovery method for internet of things. In: 2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing, pp. 43–47, October 2011

# **IoT Services and Applications**

# Understanding How Adversarial Noise Affects Single Image Classification

Amit Adate, Rishabh Saxena<sup>(✉)</sup>, and Don.S

VIT University, Vellore 632014, Tamil Nadu, India  
{adateamit.sanjay2014,rishabh.saxena2014,don.sasikumar}@vit.ac.in

**Abstract.** In recent trends, computer vision applications have seen massive implementation of supervised learning with convolutional neural networks. In this paper, we have analyzed image classifiers and their classification accuracy. Also, we have measured their robustness upon introduction to various noise layers. Furthermore, we have implemented a generative adversarial network for the generator task of adversarial noise generation and the discriminator task of single image classification on the handwritten digits database. Our experiments are yielding progressive results and we have performed conditional and quantifiable evaluation of the generated samples.

**Keywords:** Convolutional neural network  
Generative adversarial network · Adversarial noise · Image classification  
Handwritten digits

## 1 Introduction

In recent times, Generative models have made quantum leaps and have become a dominant heuristic in the field of data and image generation. With the advent of Generative Adversarial Networks by Goodfellow [1], they showed how Adversarial Learning could be used to train two networks by competing them with each other. In this paper, we outline an adversarial learning approach to generate an attacking scheme for image classifiers. It is generally noted that adversarial approach performs better at generating images. With several previous models already capable of generating images which are identical to the dataset they were trained from [2], we have created a DCGAN based model that can introduce the aforementioned noise. A few existing models which have been tested and proven to be efficient are: Deep Convolutional Generative Adversarial Networks [3] which includes a deconvolutional neural network as a generator that generates images from a given set of data and a convolutional neural network discriminator that distinguishes the image from the generator whether it was from the original dataset or if it was generated by the DeconvNet generator. Radford et al. [3] applied various techniques and niche methods to optimize the training and output of their networks which includes using the famous All-Convolutional Neural Network [4] that replaces the max pooling layer with another convolutional layer

with 2 strides that learns its own spatial downsampling and batch normalization to improve training on the dataset; Wasserstein GAN [5] which uses the Earth Mover’s distance [6] as the loss function for comparing the probability distribution between the generated image and the original image to evaluate the work needed to transform one distribution from one histogram to another; and Bayesian GAN [7] which uses the Bayesian loss function for calculating the probability density of the generated and original image and maximized that as a loss function.

All these architectures stand as testimony to the adversarial method of training but none of them can be applied to generate random noise for interference with an image to produce a misclassification attack on their vanilla structure. This is because these networks take their samples from a known database and introduce random noise to produce an image closer to the original image. But for image noise generation, we need a random noise image and change the noise in that image in tandem with the original image to produce an original image classification. The classification must be different from the classification the image would have taken had there been no introduction of the noise. This is introduced by Goodfellow et al. [8] with a simple methodology called the Fast Gradient Sign Method where the noise generator and trained with the loss function of the image classifier as follows

$$x^* = x + \epsilon \text{sign}(\nabla_x J(\Theta, x, y))$$

where  $x^*$  is the image with the noise layer,  $x$  is the original image,  $\epsilon$  is the magnitude of the perturbation,  $y$  is the final label and  $\Theta$  is the noise parameter.  $J(\Theta, x, y)$  is the loss function.

Another viable method for such noise generation is the Jacobian-based Saliency Map attack proposed by Papernot et al. [9] that uses input image  $x$  for a model  $F$  with classification  $j$  and a target classification  $t$  where we increase the probability of classification for  $t$  and reduce all classifications that are  $j \neq t$  by the following equation:

$$S(X, t)[i] = \begin{cases} 0 & \text{if } \frac{\partial F_t(X)}{\partial X_i} < 0 \text{ or } \sum_{j \neq t} \frac{\partial F_j(X)}{\partial X_i} > 0 \\ \left( \frac{\partial F_t(X)}{\partial X_t} \mid \sum_{j \neq t} \frac{\partial F_j(X)}{\partial X_i} \right) & \text{otherwise} \end{cases}$$

This paper looks at the problem of generating an noise layer which can be added to an image to fool a discriminator or a classifier into misclassifying the data. It looks at the use of this technique for attacks on classifiers in general. The remainder of the paper is organized as follows: Sect. 2 looks at the background that led to our work and provides a gist of DNNs and Noise introduction, Sect. 3 explains the experimental setup of the architecture, Sect. 4 represents the results of our experiments and the last section provides with the conclusion of our work.

## 2 Background

Deep Neural Networks have been around for some time and have shown exceptional results in almost all fields of science and technology, with applications



ranging from predictions to generation of data. Neural Networks, which have been reasonably the most successful algorithms for data generation, can also be used for generating noise and images in general. This section looks at the previous work done in this field.

In their work, Dodge and Karam [10] take a look at how distortion and perturbations in an image affect the classification paradigm of a Deep Neural Networks and a brief outlook on how adversarial samples affect the performance of the DNN. They performed training on the ImageNet 2012 dataset with four fairly common networks namely Caffe Reference, VGG-CNN-S [11], VGG16 [11] and GoogLeNet [12]. They used two image compression techniques mainly JPEG Compression, JPEG2000 compression; added noise; and used two transformations on the images that were blurring and contrasting the input images at the time of experimentation.

Their results showed an interesting observation: it was seen that the four network’s performance decreased significantly as noise, blur and contrast levels were increased, with the exception of GoogLeNet and VGG16 which had deeper network structures. But for compression techniques like JPEG and JPEG2000, the networks were more robust to increase in distortion levels and the addition of noise to the image had significantly more chance of creating a performance related misclassification than the two compression techniques. Finally, they concluded that the VGG16 network performed best under given distortions and image stresses.

There are various noise introduction methods that can introduce adversarial noise which is a specifically designed noisy image which can be layered on top of the original image to produce a targeted misclassification of the original image.

Kurakin et al. [13] proposed an iterative method for the aforementioned Fast Gradient Sign Method where they apply the noise to the image  $x$  iteratively and test the image for misclassification with the following equation:

$$x_i^* = clip_{x,\epsilon}(x_{i-1}^* + \epsilon sign(\nabla_{x_{i-1}^*} J(\Theta, x_{i-1}^*, y)))$$

The authors clip the values of  $x_i$  so that they are  $\pm\epsilon$  of the original image  $x$ . This creates an image where the noise isn’t extremely dominant and allows for a more human believable object.

The Carlini and Wagner method [14] of creating adversarial samples uses not one but three simultaneous attacks namely the  $L_0$  attack, the  $L_2$  attack and the  $L_\infty$  attack which finds an unrestricted perturbation using the previously defined perturbation:

$$\delta_i^* = \frac{1}{2}(tanh(\omega_i + 1)) - x_i$$

And then optimizes  $\delta$  over  $\omega$  with the following equation:

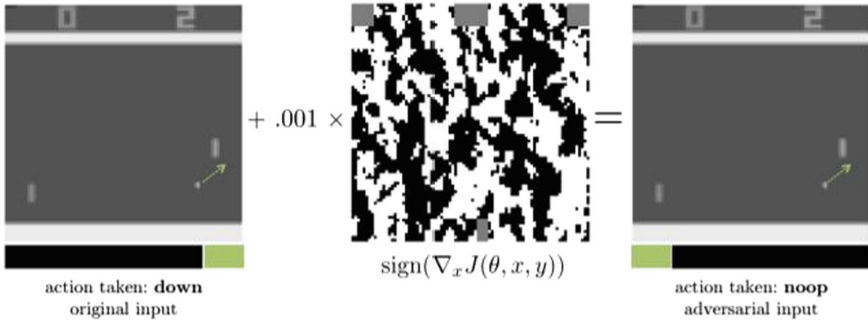
$$min_\omega = ||\frac{1}{2}(tanh(\omega) + 1) - x||_2^2 + cf(\frac{1}{2}tanh(\omega) + 1)$$

And finally takes the resultant  $\delta^*$  and transforms it into a restricted perturbation. The gradient is calculated by the following equation:

$$f(x) = max(max\{Z(x)_i : i \neq t\} - Z(x)_t, -\kappa)$$

Work by Huang et al. [15] shows that adversarial examples can also affect Deep Reinforcement Learning algorithms that like those that were used in Atari vision systems and the famous Chinese game Go. Their paper gives us an insight into how noise distortions can affect the end-to-end policies of the image to action sequence in an algorithm. They look at two methods, White-Box method where the adversary does not have access to the training environment and only has the final output of the given data policy and it learns from those actions to produce a noise level that can produce targeted action mismatch in the output by learning the different noise to action mechanisms.

The other approach used what the black-box technique where the adversary has access to the training environment and the hyperparameters, and can use the weight update policy as well that allows it to create a more accurate noise perturbation in the input image and cause end-to-end policy changes. They used FGSM method with  $L_0$ ,  $L_2$  and  $L_{infinity}$  normalization for the generation of adversarial examples which can be seen in Fig. 1.



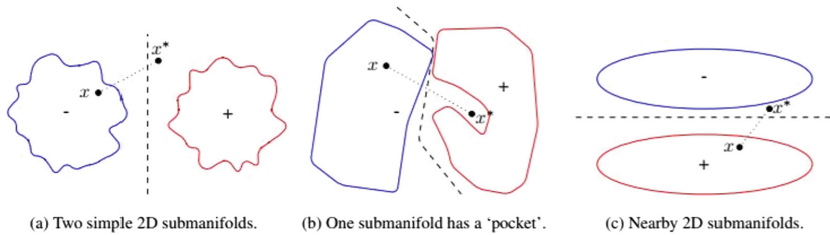
**Fig. 1.** This adversarial example is computed using the fast gradient sign method (FGSM) [10] with an  $L_{\infty}$ -norm constraint on the adversarial perturbation; the adversarial input is equivalent to the original input when converted to 8-bit image encodings, but is still able to harm performance [15].

In conclusion to their work, we see that in many cases an  $L_{\infty}$ -norm FGSM adversary with  $\delta = 0.001$  decreases the network’s performance by 50% or more; when converted to 8-bit image encodings, these adversarial inputs are indistinguishable from the original inputs.

Feinman et al. [16] explains the performance change by reconfiguring the images in terms of high-dimensional data and assuming that in that this high-dimensional data lies a lower dimensional manifold, which can be carefully traversed to change the underlying image label, thereby creating distortions. This perturbation point can lie in three regions of the manifold, as seen in Fig. 2, where  $x$  is the original image,  $C_x$  is the class, which can be:

1.  $X^*$  lies far away from the submanifold of  $C_x$ .
2.  $X^*$  lies near the submanifold but not on it and it is far from the classification boundary of  $C_x$  and  $C_{x^*}$ .
3.  $X^*$  lies near the submanifold but not on it and is near the boundary of the classification for  $C_x$  and  $C_{x^*}$ .

This work essentially looks at the last case where the adversarial sample is such that it lies near the submanifold and the classification boundary which creates a misclassification.



**Fig. 2.** Feinman et al. [16] explains the submanifold concept via diagrammatic representation of the higher dimensional plane.

### 3 Experiment

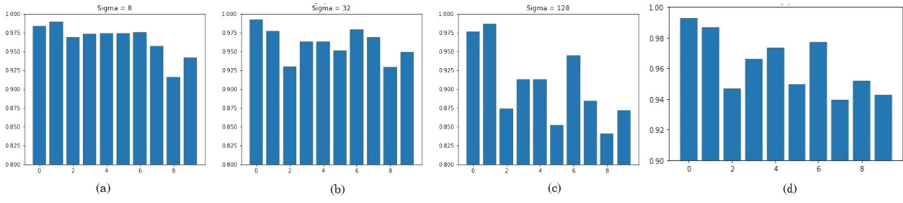
The experimental setup uses a basic convolutional Neural Network based on 2 convolutional layers and 2 fully connected layers [17]. It takes an input image size of  $28 \times 28$  from the MNIST hand written image database. This database consists of 55,000 training set images, 5,000 validation set examples and 10,000 test set examples. The architecture was trained on a GTX 1060 GPU. The activation function used was the softmax function at the fully connected layer. The convolution layers involved normal padding and a stride of 3.

The first loss function was optimized using Adam optimizer [18]. This optimizer is used to train the classification function of the network that allows the CNN to classify hand written digits with an accuracy of 89.01%.

The second optimizer tries to optimize the additive loss of the L2 loss function and the softmax function with the Adam optimizer to produce a noisy image that is used to misclassify the image of 2 as 6 (Fig. 3).

We also include the dropout function [19] as previous work on the same has shown that the dropout function improves the training time of the network as well as helps avoid mode collapse [20] during training.

We take the delta parameter as the threshold for the noise levels in the adversarial sample that can be be adjusted for variable outcome of the data. The delta parameter keeps the noise in the image under the maximum pixel value to avoid overflow and above the minimum pixel value to avoid underflow of the noise layer during training. After training on one batch, the noise layer atop the original image is clipped so that additionally, the values remain under the normalized 0 to 1 range for the entire image.



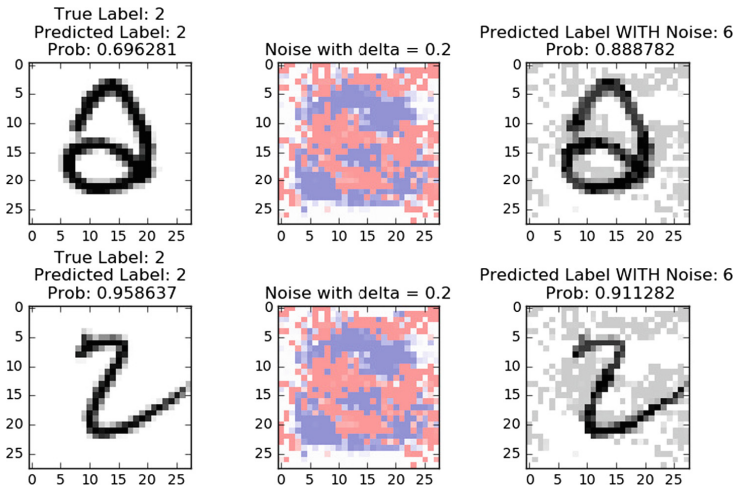
**Fig. 3.** Adam optimizer function with parameter: (a) For  $\Sigma = 8$  (b) For  $\Sigma = 32$  (c) For  $\Sigma = 128$  (d) The average accuracy vs digit class for the final classification metric using adversarial samples along with the original images

### 4 Results

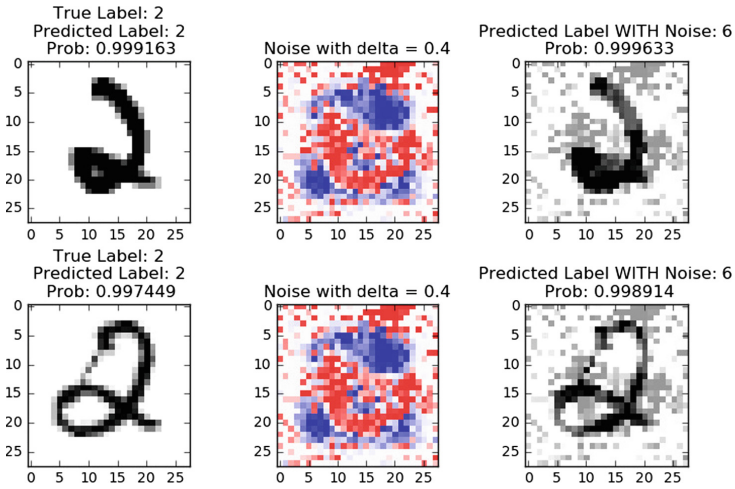
With the implementation, it can be seen that an adversarial sample can cause misclassification when adversarially trained on a vanilla Convolutional Neural Network alongside the usual training. This is a case of whitebox adversarial training when the adversary has access to the entire training environment and can use the training procedure at a more rudimentary level to extra information about the classification process.

Figures 4, 5 and 6 provide with the results of our work, showing the noise layer over the image layer of the MNIST digit of 2.

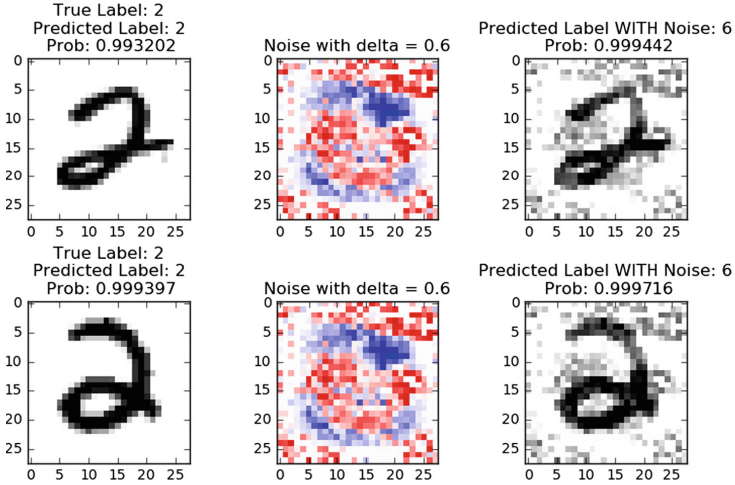
We cycled the delta values through 0.2, 0.4 and 0.6 with variable noisy images produced with each value. When the noise limit (delta) is small (i.e. 0.2) the noise values will be brighter hence more pixels in the original image will become brighter. Consequently, the effect of the noise will not be major. Simply, the CNN may be able to still recognize the image of 2 as 2 with high level of accuracy.



**Fig. 4.** (From left to right) original MNIST digit 2, the noise layer using  $\delta = 0.2$ , superposition of noise layer and the original MNIST digit image with the misclassification



**Fig. 5.** (From left to right) original MNIST digit 2, the noise layer using  $\delta = 0.4$ , superimposition of noise layer and the original MNIST digit image with the misclassification



**Fig. 6.** (From left to right) original MNIST digit 2, the noise layer using  $\delta = 0.6$ , superimposition of noise layer and the original MNIST digit image with the misclassification

On the other hand, when the noise limit (delta) is high (i.e. 0.8) the noise values will be darker hence more pixels in the original image will become darker. Consequently, the effect of the noise will be major. Simply, the CNN will not be able to recognize the image of 2 as 2 anymore, it will always see it as 6.

To view our experiments and images generated:

<https://github.com/amitadate/gan-noise>

## 5 Conclusion

In this work, we present a model that produces a misclassification for a specific digit from 2 to 6 that is a direct result of adding adversarial noise to the image. The limitations of this paper are that the current model provides a noise layer for a single digit misclassification rather a generalized adversarial noise sample for the entire domain of digits. There is further scope in a application of this model for multi-image classification.

## References

1. Goodfellow, I.J.: NIPS 2016 tutorial: generative adversarial networks. CoRR, abs/1701.00160 (2017)
2. Arjovsky, M., Bottou, L.: Towards principled methods for training generative adversarial networks. ArXiv e-prints, January 2017
3. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR, abs/1511.06434 (2015)
4. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.A.: Striving for simplicity: the all convolutional net. CoRR, abs/1412.6806 (2014)
5. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. ArXiv e-prints, January 2017
6. Hou, L., Yu, C.P., Samaras, D.: Squared earth mover's distance-based loss for training deep neural networks. CoRR, abs/1611.05916 (2016)
7. Saatchi, Y., Wilson, A.G.: Bayesian GAN. ArXiv e-prints, May 2017
8. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. ArXiv e-prints, December 2014
9. Papernot, N., McDaniel, P.D., Jha, S., Fredrikson, M., Celik, Z.B., Swami, A.: The limitations of deep learning in adversarial settings. CoRR, abs/1511.07528 (2015)
10. Dodge, S.F., Karam, L.J.: Understanding how image quality affects deep neural networks. CoRR, abs/1604.04004 (2016)
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)
12. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. CoRR, abs/1409.4842 (2014)
13. Kurakin, A., Goodfellow, I.J., Bengio, S.: Adversarial examples in the physical world. CoRR, abs/1607.02533 (2016)
14. Carlini, N., Wagner, D.A.: Towards evaluating the robustness of neural networks. CoRR, abs/1608.04644 (2016)
15. Huang, S.H., Papernot, N., Goodfellow, I.J., Duan, Y., Abbeel, P.: Adversarial attacks on neural network policies. CoRR, abs/1702.02284 (2017)
16. Feinman, R., Curtin, R.R., Shintre, S., Gardner, A.B.: Detecting adversarial samples from artifacts. ArXiv e-prints, March 2017
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

18. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR, abs/1412.6980 (2014)
19. Metz, L., Poole, B., Pfau, D., Sohl-Dickstein, J.: Unrolled generative adversarial networks. CoRR, abs/1611.02163 (2016)
20. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors. CoRR, abs/1207.0580 (2012)

# Top- $k$ Category Search for an IP Address-Product Network

Ramalingeswara Rao Thottempudi<sup>1</sup>(✉), Pabitra Mitra<sup>2</sup>, and Goswami Adrijit<sup>1</sup>

<sup>1</sup> Department of Mathematics, Indian Institute of Technology Kharagpur,  
Kharagpur 721302, India

[trrao@iitkgp.ac.in](mailto:trrao@iitkgp.ac.in), [goswami@maths.iitkgp.ernet.in](mailto:goswami@maths.iitkgp.ernet.in)

<sup>2</sup> Department of Computer Science and Engineering,  
Indian Institute of Technology Kharagpur, Kharagpur 721302, India  
[pabitra@cse.iitkgp.ernet.in](mailto:pabitra@cse.iitkgp.ernet.in)

**Abstract.** Due to the vast number of online business transactions on World Wide Web, mining and analyzing relevant data from the web log data for the users navigational behavior is a challenging task. Finding similar objects and mining top- $k$  objects has a great significance in web recommender systems and social networks. In this paper, we define similar behavior of users about different categories and some propositions in the context of structural similar behavior of nodes in a network. We present an efficient algorithm for top- $k$  categories based on early associates notion (NATBEAN) that mines top- $k$  categories with most similar IP addresses in a descending order. NATBEAN is useful to forecast similar visiting behavior of the users through IP addresses for different categories in the structural context of a bipartite network. This leads to find popular products and less influenceable products in a network of web log data. Initially, we run both Naive approach and NATBEAN for finding top- $k$  categories on a clickstream dataset whose attributes are IP addresses and product categories, then we run our algorithm on three other datasets and compare running times of both the algorithms.

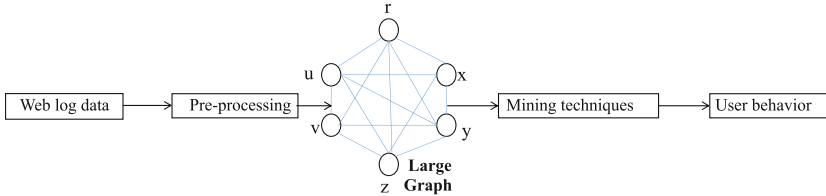
**Keywords:** Structural equivalence · Similar users · Bipartite network  
Top- $k$  category

## 1 Introduction

Due to rapid development of social networks, the analysis of network data for finding user accessing behavior has a great significance in social media mining and data mining. Finding similar behavior, abnormal behavior and collective behavior are three types of applications in network data analysis. The notion of similarity is an essential and broadly used concept in many applications of different areas like classification, clustering, social network analysis, citation analysis, etc. Transforming the pre-processed clickstream data, social network data into large graphs (see Fig. 1) and then analyzing the data using data/graph mining techniques to interpret user's navigational behavior for predicting the user's



interests is an emerging and challenging area in Graph based data mining and Recommender Systems [1]. Finding similar behavior of users about distinct categories and retrieving top- $k$  categories from large networks are two fundamental problems in graph based data mining. The structural properties of different nodes in a network provide more information about the interactions between several nodes. Common neighbors, jaccard similarity, preferential attachment and Adamic/Adar are the most popular local structural similarity measures to find the similarity between two nodes [2].



**Fig. 1.** Web graph mining process for user behavior

The structural properties of various nodes in a network produce intuitive information about interactions among various nodes. Identifying the similar behavior of nodes through structural properties provides useful information about node behavior in various applications of data mining. In this paper, we aim to find the top- $k$  categories on an IP address-Product network based on structural similar behavior of nodes with respect to object-to-object relationship. We apply the concept of node-to-node relationship between IP addresses and hence we find top- $k$  categories through that IP addresses. The problem of top- $k$  category search with respect to node-to-node relationship for an IP address-Product network is new in the existing literature and play a crucial role in the analysis of social networks.

### **Our Contribution:**

In this paper, we aim to find top- $k$  product categories with maximum number of IP addresses in a descending order. We define similar behavior of users about product categories with respect to structural context and present some propositions which represent the similar behavior of users about single and more categories. Moreover, we propose an efficient algorithm for finding top- $k$  categories with maximum number of similar IP addresses in a descending order known as NATBEAN using early associates approach. The mining efficiency of NATBEAN is better than the baseline in terms of running time. We conduct various experiments on four distinct datasets to show the efficiency of our proposed algorithm with the naive approach. To the best of our knowledge we present a novel approach to describe the similar access behavior for top- $k$  categories through IP addresses based on object-to-object relationship for an IP address-Product network.

## 2 Related Work

Lorrain and White [3] extended the notion of structural equivalence to understand the interrelations among relations within concrete social groups. Lin [4] presented several intuitions of similarities between ordinal values, strings, feature vectors, words and semantic similarity in a generalized anatomy. A measure of structural similarity to find the similarity between two objects with respect to their neighbors is known as SimRank [5]. A wide range of similarity measures are available in the literature and can be broadly divided into two categories: (1) content related similarity measures and (2) structural similarity measures. The structural similarity measures find the similarity between object-to-object relationships (web pages, articles, movies, etc.) with respect to the links or edges in a graph [5], whereas the content related similarity measures consider each entity as a set of items [6]. Fagin et al. [7] discovered two algorithms Fagin’s algorithm (FA) and Threshold algorithm (TA) for finding top- $k$  objects with highest overall grades. Deshpande and Karypis [8] discovered item based top N recommendation algorithms by defining similarity between items. Holme and Huss [9] presented a method for estimating the role of nodes in networks based on role similarity measures of nodes. Leicht et al. [10] presented a structural similarity measure in the sense that “two nodes are similar if their next neighbors in the network are themselves similar”, and this measure can be viewed as a weighted sum of all paths of different lengths between the nodes in the network. Sun et al. [11] proposed algorithms on bipartite graphs to find similar nodes (neighborhood formation) and abnormal nodes (anomaly detection). Liben-Nowell and Kleinberg [2] discussed link-based similarity measures that works on node neighborhood methods in the link prediction problems of large graphs. Rossi et al. [12] presented local node and global node metrics which are useful in finding the node predictions and link predictions in a network. Cai et al. [13] constructed user behavior networks for web traffic in a bipartite graph representation of server and client nodes to analyze the communities of clients. Zweig and Kaufmann [14] proposed a new way to evaluate one-mode projection of bipartite networks by introducing general interestingness measures. Xu et al. [15] represented internet backbone links using one-mode projection and presented a novel approach to find the similar social-behavior of internet end-hosts. Xu et al. [16] used bipartite networks in one mode projection to discover similar social behavior among distinct end hosts in the same network. Jakalan et al. [17] discovered a novel method using bipartite graph representation of IP addresses of inside a network with the connectivity of outside network to find the clusters of IP addresses with similar behavior. Taheri et al. [18] proposed HellRank measure to detect most behavioral users in bipartite social networks by avoiding one mode projections. However, one-mode projection of bipartite graphs are less informative and loses information about some nodes. From the literature, we observe that representing an IP address-Product network in one-mode projection causes loss of information. Hence we use node neighborhood approach to find top- $k$  categories with their corresponding IP addresses using early associates technique.

### 3 Problem Definition

Similarity is a principal and widely used key concept. The similarity between two nodes in a connected network can be computed either based on the network similarity or content similarity. The notion of similarity in any network, measures the role of a node that is concerned to its structural properties. Hence, nodes with similar roles can be identified using several similarity measures [10]. Node similarity is defined by how similar their interactions are. Depending on the information available of a network, similar behavior of two nodes in a network can be find by measuring their structural equivalence. The definitions with respect to the notion of similarity of nodes in a graph are given below.

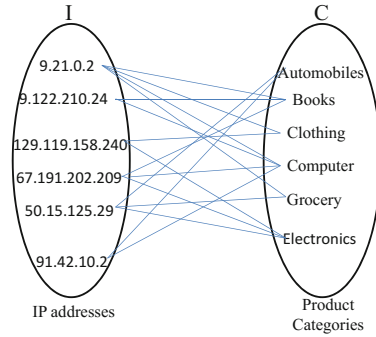


Fig. 2. IP address-Product network

Let  $G = (I \cup C, E)$  be an undirected graph, where  $V = I \cup C$ , set of IP addresses and categories (see Fig. 2). For each category  $C_j$  ( $1 \leq j \leq n$ ) find the maximum number IP addresses. i.e., The category shared by majority of IP addresses is  $C_j$ . For a given positive integer  $k$ , find top- $k$  categories  $C_j$  ( $j = 1, 2, \dots, k$ ) with most similar IP addresses (users) in a descending order. Let  $\{(IP_i, C_i), 1 \leq i \leq n\}$  be the set of IP addresses (users) and categories. Top  $k$ -Category =  $\{(C_j, IP_i), 1 \leq j \leq k \text{ and } IP_1 \geq IP_2 \geq \dots \geq IP_k \geq IP_i \geq \dots, \geq IP_n\}$  mines most similar IP addresses/users for each category. i.e., a particular user access a category through IP address  $IP_1$  first, next through  $IP_2$  and so on. Similar IP addresses we mean, the IP addresses that share a particular category. In otherwords, the IP addresses through which users access the same category. In the proposed algorithms we assume that only one user can access the products through a particular IP address.

**Definition 1.** *Two nodes are said to be structurally equivalent in a graph, if they share the same neighboring nodes [3, 19]. Structurally equivalent nodes have same degree, clustering coefficient, centrality, belong to the same cliques. The most familiar measure for finding the structural similarity between two nodes in a network is common neighbors. The nodes having more common neighbors that share, are more similar.*

**Definition 2.** Two nodes  $u_i$  and  $u_j$ ,  $i \neq j$  and  $(u_i, u_j) \notin E$  are said to be similar if their neighbors are also similar [20], i.e., the nodes  $u_i$  and  $u_j$  have same neighbors, i.e.,  $\Gamma(u_i) = \Gamma(u_j)$ .  $\Gamma(u_i)$  represents the set of all neighbors of  $u_i$ . The measure of a node similarity can be denoted by

$$\sigma(u_i, u_j) = |\Gamma(u_i) \cap \Gamma(u_j)|.$$

Two users  $u_i$  and  $u_j$  are having similar accessing behavior about a product category  $\{c_i\}$ , if  $\Gamma(u_i) = \Gamma(u_j) = \{c_i\}$ .

**Definition 3.** A set of users  $X = \{u_1, u_2, \dots, u_m\} \subseteq U, m \leq n$  is said to have similar access behavior about a category  $c_j \in C$ , a set of product categories,  $j = 1, 2, \dots, n$ , if  $\bigcap_{i=1}^m \Gamma(u_i) = c_j$ . Every user in  $X$  shares the same neighbor  $c_j \in C$ .

**Definition 4.** A subgraph  $B = \langle U, V \rangle$  is a Biclique of a Graph  $G$ , containing two non-empty disjoint sets of nodes  $U$  and  $V$  such that for any two nodes  $u \in U$  and  $v \in V$ ,  $\exists$  an edge  $(u, v) \in E$  [21].

Let  $U$  and  $V$  be the two disjoint sets of a Bipartite Graph  $G$ . If two users  $\{u_i, u_j\}$  have same common neighbors  $\{c_i, c_j\}$  and also the common neighbors of  $\{c_i, c_j\}$  are  $\{u_i, u_j\}$ , then the sets  $\{u_i, u_j\}$  and  $\{c_i, c_j\}$  form a biclique. i.e., a biclique of order 2 represents the similar behavior of two users about two products.

**Proposition 1.** Let  $G = (U \cup C, E)$  be a graph, where  $C = \{c_1, c_2, \dots, c_n\}$  be the set of product categories accessed by the set of users  $U = \{u_1, u_2, \dots, u_n\}$ . Let  $u_i, u_j \in U$  ( $i \neq j$ ) and if  $u_i$  and  $u_j$  are similar about a particular category  $c_i \in C$  then  $\sigma(u_i, u_j) = c_i$ .

*Proof.* Given  $u_i, u_j \in U$  are similar about a node  $c_i \in C$ . Then both  $u_i$  and  $u_j$  access the same neighbor  $c_i \in C$ , i.e., Both  $u_i$  and  $u_j$  share the same category  $c_i$ . Therefore,  $\sigma(u_i, u_j) = c_i$ .

**Proposition 2.** Let  $G = (U \cup C, E)$  be a graph. Let  $u_i, u_j \in U$ ,  $i \neq j$ , be two distinct users. If  $u_i$  and  $u_j$  are similar about a set of nodes  $X \in C$ , then  $\sigma(u_i, u_j) = X$ .

*Proof.* Let  $u_i, u_j \in U, i \neq j$  be two distinct users in the set  $U$ . Let  $X \subseteq C$  be a non-empty set of categories in  $C$ .

Suppose  $u_i, u_j$  are similar about the non-empty set  $X \subseteq C$ , a set of categories. Then  $\Gamma(u_i) = \{c_i/c_i \in C \text{ and } \forall c_i \in X, (u_i, c_i) \in E\}$ . Also  $\Gamma(u_j) = \{c_i/c_i \in C \text{ and } \forall c_i \in X, (u_j, c_i) \in E\}$ . That is, the users  $u_i$  and  $u_j \in U$  shares each and every category  $c_i \in X$ , i.e.,  $\Gamma(u_i) = \bigcup_{i=1}^m c_i, c_i \in X, m \leq n$  and  $\Gamma(u_j) = \bigcup_{j=1}^m c_i, c_i \in X, m \leq n$ . Therefore,  $\sigma(u_i, u_j) = \bigcup_{i=1}^m c_i = X$ . Hence,  $u_i$  and  $u_j$  are similar about a set of categories  $X \in C$ . More generally, Two sets  $X \subseteq U$  and  $Y \subseteq C$  are said to be similar  $\Leftrightarrow \exists u_i \in X$  and  $\exists c_i \in Y$  such that  $\Gamma(u_i) = Y$  and  $\Gamma(c_i) = X, i = 1, 2, \dots, n$ .

**Proposition 3.** *Let  $u_i, u_j \in U$  and  $c_i, c_j \in C, i \neq j$ . Suppose  $\sigma(u_i, u_j) = \{c_i, c_j\}$ . Also  $\sigma\{c_i, c_j\} = \{u_i, u_j\}$ . Then the sets  $\{u_i, u_j\}$  and  $\{c_i, c_j\}$  form a complete Biclique of order two.*

*Proof.* Let  $u_i, u_j \in U$  be the two nodes. Since  $u_i$  and  $u_j$  have common neighbors  $\{c_i, c_j\}$ , we have  $\sigma(u_i, u_j) = |\Gamma(u_i) \cap \Gamma(u_j)|$ , i.e.,  $\Gamma(u_i) = \{c_i, c_j\}$  and  $\Gamma(u_j) = \{c_i, c_j\}$ . Also,  $\text{degree}(u_i) = \text{degree}(u_j) \geq 2$ . Therefore, the nodes  $u_i$  and  $u_j$  are similar about the two nodes  $\{c_i, c_j\} \in C$ . Also,  $\sigma(c_i, c_j) = |\Gamma(c_i) \cap \Gamma(c_j)|$ . And  $c_i$  and  $c_j$  are similar about  $\{u_i, u_j\} \in U$  and  $\text{degree}(c_i) = \text{degree}(c_j) \geq 2$ . Hence, the two sets  $\{u_i, u_j\}$  and  $\{c_i, c_j\}$  form a Biclique.

More generally if  $\sigma(u_1, u_2, \dots, u_k) = \{c_1, c_2, \dots, c_l\}$ , where  $k \leq m$  and  $l \leq n$  and  $\sigma(c_1, c_2, \dots, c_l) = \{u_1, u_2, \dots, u_k\}$ , then the sets  $\{u_1, u_2, \dots, u_k\}$  and  $\{c_1, c_2, \dots, c_l\}$  form a bi-clique of order  $j = \text{minimum}\{k, l\}$ .

One important and interesting property is that there is a one-one correspondence between similar behavior and bicliques. Mining complete bicliques of order 2 give the similar behavior of two users about two categories. Mining a maximal biclique represents the similar behavior of  $n$  users about  $m$  product categories in a bipartite graph of users and categories.

## 4 Proposed Algorithms

---

### Algorithm 1. Naive Approach

---

```

input :  $G = (I \cup C, k)$ 
output: Top- $k$  Categories with descending order of similar IP
         addresses/users
1 String  $S = \phi$ ;
2 Find total product categories from the dataset and remove the duplicates;
3 Find total IP addresses from the data set and remove the duplicates;
4 For each category  $C_i$  find  $\Gamma(C_i)$ ;
5 for ( $\text{int } i = 1; i \leq n; i++$ ) do
6   | for each category  $C_i$  find  $\Gamma(C_i)$ ;
7   |  $S = S \cup C_i \cup \Gamma(C_i)$ ; // Forms a list of IP addresses that
   |   share the corresponding category  $C_i$ 
8 end
9 for ( $i = 1; i \leq n; i++$ ) do
10  | find  $C_j$  for which  $|\Gamma(C_j)|$  is maximum;
11 end
    | // Find the category that is shared by most of the IP
    |   addresses
12 Find top- $k$  categories  $C_j$  with the corresponding IP addresses;

```

---

Algorithm 1 finds the top- $k$  categories with most similar IP addresses in decreasing order for a single source problem. That is for a source set of IP addresses  $IP_i$  find the total product categories  $C_i, i = 1, 2, \dots, n$ . The naive approach of Algorithm 1 finds similar access behavior of users through distinct IP addresses for distinct product categories and retrieve top- $k$  categories among them. In this case similar IP addresses, we mean that the IP addresses that share the same product category or the IP addresses through which users accessed a particular product category. The Algorithm 1 initially finds product categories that exist in

the dataset and removes the duplicates. Then step 3 finds the total IP addresses and removes duplicates. In step 4, for each category  $C_j, j = 1, 2, \dots, n$  the naive approach finds the corresponding IP addresses. Step 7 forms a list of categories and their respective IP addresses (users) based on Proposition 2. Step 10 finds the total number of IP addresses for each category and then top- $k$  retrieves  $k$  categories with maximum number IP addresses in step 12. However, the Naive approach of finding top- $k$  categories with maximum number of IP addresses is a straight forward approach and it takes  $O(n^2)$  time. Moreover, removing redundant data externally and high computational cost are the disadvantages of Naive approach.

We now propose an efficient algorithm to find top- $k$  categories with their corresponding similar IP addresses in descending order known as NATBEAN. This algorithm works on the principle of early associates that the users who visit a particular product will be associated to that the respective category on first come first serve basis. In this algorithm, we find the top  $k$  categories in the context of structural similar behavior of users through IP addresses about product categories. We assume that only one user can access through a particular IP address. An advantage of NATBEAN is it eliminates redundant categories and users automatically with out any external effort. This algorithm also finds total number of categories and IP addresses that exist in the dataset.

For the NATBEAN algorithm we assume that a single user can access only through a particular IP address. Initially the procedure `Category_Identifier` finds distinct categories by eliminating redundant elements while traversing along the categories of the dataset. It assigns a unique number for each category. Then each category is replaced by corresponding IP address in `User-Location`. Then all the IP addresses associate to the particular respective category in `Early-associates`. Finally the procedure `Top-k` retrieve  $k$  categories with maximum number of IP addresses in a descending order.

### Example:

The run of NATBEAN algorithm for the given example is organized as follows.

- Input data:  $IP_4, C_2; IP_3, C_1; IP_1, C_5; IP_1, C_1; IP_3, C_3; IP_1, C_4; IP_3, C_2; IP_1 C_3; IP_4, C_1; IP_2, C_4; IP_2, C_1; IP_3, C_5; IP_3, C_4; IP_2, C_3; IP_5, C_5; IP_4 C_6;$
- The Procedure **Category\_Identifier** assigns a unique number to each category. It returns the output  $C_2, 1; C_1, 2; C_5, 3; C_3, 4; C_4, 5; C_6, 6;$
- The Procedure **Store-Count**(String b, String e[ ]) takes the input for **b** from the `Category_Identifier` and stores into locations of array **e**. In the array **e**,  $C_2$  is stored in location 1,  $C_1$  is stored in location 2,  $C_5, C_3, C_4$  and  $C_6$  are stored in locations 3, 4, 5 and 6 respectively, i.e.,  $e[1] = C_2, e[2] = C_1, e[3] = C_5, e[4] = C_3, e[5] = C_4$  and  $e[6] = C_6$ . The procedure **Store-Count** returns the total number of locations which is 6 for this example.
- The procedure **User-Location**(String b, String a) takes the input for **b** from the output of **Category\_Identifier** and **a** is the original dataset. It takes the input for **b** =  $\{C_2, 1; C_1, 2; C_5, 3; C_3, 4; C_4, 5; C_6, 6;\}$  and **a** as in input step

then returns the output  $u_4, 1; u_3, 2; IP_1, 3; IP_1, 2; IP_3, 4; IP_1, 5; IP_3, 1; IP_1, 4; IP_4, 2; IP_2, 5; IP_2, 2; IP_3, 3; IP_3, 5; IP_2, 4; IP_5, 3; IP_4, 6;$

– **Early-Associates:**

---

**Algorithm 2. A Novel Algorithm for Top- $k$  category Based on Early Associates Notion (NATBEAN)**

---

```

input :  $G = (I \cup C, k)$ 
output: Top- $k$  Categories with descending order of similar IP addresses
1 int  $k, n$ ; //  $n$  is the input size,  $k \leq n$ 
2 String[]  $m_1 = \text{new String}[n]$ ;
3 String[]  $m_2 = \text{new String}[n]$ ;
4 String  $a, b, c$ ;
5 Read the dataset into  $a$  which consists of both IP addresses and categories
6  $a = rdstring()$ ; // Read the whole dataset of IP addresses and
   categories  $(IP_i, C_j) \ i = 1, 2, \dots, n, \ j = 1, 2, \dots, m.$ 
7  $b = \text{Category\_Identifier}(a)$ ; // find distinct categories  $C_j$  and
   assigns a unique number to  $C_j$ 
8  $n = \text{Store} - \text{Count}(b, m_2)$ ; // stores the distinct categories in
   their respective locations of  $m_2[ ]$ 
9  $c = \text{User} - \text{Location}(b, a)$ ;
10  $\text{Early} - \text{Associates}(c, m_1, n)$ ; // early IP addresses will be
   associated to their corresponding locations  $m_1[ ]$ .
   //  $m_2[i]$  gives categories and their corresponding IP
   addresses based on early associates approach
11 for ( $i = 1; i \leq n; i++$ ) do
12   |    $m_2[i] = m_2[i] + m_1[i]$ ; // Categories with their
       |   corresponding similar IP addresses
13 end
14 Read  $k$  value ;
15 Top- $k(m_2, n, k)$ ; // Top- $k$  categories with descending order of
   similar IP addresses/users

```

---

**Procedure(Category\_Identifier)**

---

```

input :  $G = (I \cup C, k)$ 
output: No.of distinct categories
1 int in, count=0; String  $S = \phi$ ;
   // Identify distinct categories and assign a unique number to
   it
2 for ( $i = 1; i \leq n; i++$ ) do
3   |   find the category  $C_j$  in  $(IP_i, C_i), j=1, 2, \dots, n.$ 
4   |   if ( $S.indexOf(C_j) == -1$ ) then
5   |   |   count++;
6   |   |    $S = S \cup (C_j, \text{count})$ ;

```

---

**Procedure Store-Count**

---

```

input :  $G = (\text{String } b, \text{String } e[ ])$ 
output: finds the total number of distinct categories
   //  $b$  is the output of Category_Identifier
1 int c=0;
2 while ( $\text{Given input } b \text{ is not empty}$ ) do
3   |   for each  $(C_i, i) \in b$ 
4   |   |    $e[i] = C_i$ 
5   |   |    $c++$ ; // count number of distinct products

```

---

---

```

Procedure User-Location
input:  $G = (\text{String } b, \text{String } a)$ 
1 The input for  $b$  is the Category_Identifier and  $a$  is the original dataset.
output: find the IP addresses and locations
2 while (Given input string a is not empty) do
3   for each  $(IP_i, C_j)$  in  $a$ 
4     find the IP addresses  $IP_i$  and corresponding category  $C_i$ 
5     find the location of each category in string  $b$ . // find the
        index of  $C_j$  in category identifier
6     Associate the corresponding location of  $C_j$  to IP address
         $IP_i$ .

```

---

```

Procedure Early-Associates
input:  $(\text{String } c, \text{String } m[ ], \text{int } n)$ 
1 The String  $c$  takes the input from User-Location. The input for  $n$  is the
  output of Store-Count procedure.
output: IP addresses will be associated to the respective locations in
  array  $m$ .
2 int  $i$ ;
3 for  $(i = 1; i \leq n; i++)$ 
4    $m[i] = \phi$ ;
5   while (Given input string c is not empty) do
6     for each substring  $(IP_i, \text{location})$  in  $c$ 
7       find the user  $IP_i$  and corresponding location  $i$ ;
8        $m[\text{location}] = m[\text{location}] + "" + IP_i$ ; // Different IP
        addresses associate to their corresponding
        locations of array  $m$ 

```

---

```

Procedure Top- $k$ 
input:  $(\text{String } a, \text{int } n, \text{int } k)$ . The output of User-Location is the input  $a$ 
  for Top- $k$ 
output: Categories with maximum number of IP addresses appear first
  // find the maximum number of IP addresses and retrieve top- $k$ 
  categories
1 for  $(i = 1; i \leq k; i++)$  do
2   int  $p$ ;
   // find location of maximum string with most IP
   addresses
3    $p = \text{loc.max.string}(a, n)$ 
4    $\text{print}(a[p])$ ;
5    $a[p] = \phi$ ;

```

---

**Table 1.** Output of Early-associates

$m_1[ ]$	Early-Associates
$m_1[1]$	$IP_4, IP_3$
$m_1[2]$	$IP_3, IP_1, IP_4, IP_2$
$m_1[3]$	$IP_1, IP_3, IP_5$
$m_1[4]$	$IP_3, IP_1, IP_2$
$m_1[5]$	$IP_1, IP_2, IP_3$
$m_1[6]$	$IP_4$

The above Table 1 shows that all the similar IP addresses associated to the corresponding categories based on the principle that early user who visited  $C_2$



through an IP address will be associated first, then the next user, and so on. For example for category  $C_2$ , IP address  $IP_4$  will be associated first and then  $IP_3$ . The stepwise output of NATBEAN for the above example is given Table 4.

**Table 2.** Output of  $m_2[i]$  values

$m_2[i]$	$m_2[i]$	$m_1[i]$
$m_2[1]$	$C_2$	$IP_4, IP_3$
$m_2[2]$	$C_1$	$IP_3, IP_1, IP_4, IP_2$
$m_2[3]$	$C_5$	$IP_1, IP_3, IP_5$
$m_2[4]$	$C_3$	$IP_3, IP_1, IP_2$
$m_2[5]$	$C_4$	$IP_1, IP_2, IP_3$
$m_2[6]$	$C_6$	$IP_4$

The procedure Top- $k$  retrieves the top  $k$  product categories in such a way that the category with maximum number of IP addresses/users first and remaining categories in a descending order. Table 2 shows the  $m_2[i]$  values with all categories and their corresponding IP addresses  $k=6$ . Also the procedure Top- $k$  arranges categories from the obtained  $m_2[i]$  values in descending order of number of IP addresses which are shown in Table 3. From this result, we can find the category that has maximum and minimum number of number of IP addresses. Hence based on number of IP addresses the most popular and less influential products can be identified. The major steps involved in NATBEAN process are shown in Table 4. Table 4 illustrates that there are only two cases in Category\_Identifier, (i) when the category is distinct and (ii) when the category is repeated. When the category is distinct, a unique number is assigned for each category. For example for first input ( $IP_4, C_2$ ), the unique number for  $C_2$  is 1 and the category  $C_2$  will be stored in  $m_2[1]$ . Then User-Location assigns the corresponding IP address  $IP_4$  to location 1. And in Early-Associates  $IP_4$  will be stored in  $m_1[1]$ . Similarly for 7th input ( $IP_3, C_2$ ) from Table 4, for the category  $C_2$ , the corresponding IP address  $IP_3$  will be stored in location 1 of  $m_1[1]$ . i.e.,  $m_1[1] = IP_4, IP_3$ . In

**Table 3.** Output of Top- $k$  categories procedure

$m_2[i]$	$m_2[i]$	$m_1[i]$
$m_2[2]$	$C_1$	$IP_3, IP_1, IP_4, IP_2$
$m_2[3]$	$C_5$	$IP_1, IP_3, IP_5$
$m_2[4]$	$C_3$	$IP_3, IP_1, IP_2$
$m_2[5]$	$C_4$	$IP_1, IP_2, IP_3$
$m_2[1]$	$C_2$	$IP_4, IP_3$
$m_2[6]$	$C_6$	$IP_4$

this case the user visits category  $C_2$  first through  $IP_4$  and then through  $IP_3$ . Similarly if the category is new and then a new number will be assigned for it and the respective IP address will be stored in new location of array  $m_1[i]$ . Table 4 shows the systematic steps of various procedures in NATBEAN for different input data pairs. NATBEAN is useful in taking the decisions to improve the product quality on a particular website by having the most and least accessing behavior for different product categories through distinct IP addresses.

**Table 4.** Overall process of NATBEAN

S.no	$(IP_i, C_j)$	Category_identifier	$m_2[i]$	User-location	Early-associates $m_1[i]$
1	$(IP_4, C_2)$	$C_2(1)$	$m_2[1] = C_2$	$IP_4(1)$	$m_1[1] = IP_4$
2	$(IP_3, C_1)$	$C_1(2), C_2(1)$	$m_2[2] = C_1$	$IP_4(1), IP_3(2)$	$m_1[2] = IP_3$
3	$(IP_1, C_5)$	$C_5(3), C_1(2), C_2(1)$	$m_2[3] = C_5$	$IP_4(1), IP_3(2), IP_1(3)$	$m_1[3] = IP_1$
4	$(IP_1, C_1)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2)$	$m_1[2] = IP_3, IP_1$
5	$(IP_3, C_3)$	$C_3(4), C_5(3), C_1(2), C_2(1)$	$m_2[4] = C_3$	$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4)$	$m_1[4] = IP_3$
6	$(IP_1, C_4)$	$C_4(5), C_3(4), C_5(3), C_1(2), C_2(1)$	$m_2[5] = C_4$	$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5)$	$m_1[5] = IP_1$
7	$(IP_3, C_2)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1)$	$m_1[1] = IP_4, IP_3$
8	$(IP_1, C_3)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4)$	$m_1[4] = IP_3, IP_1$
9	$(IP_4, C_1)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2)$	$m_1[2] = IP_3, IP_1, IP_4$
10	$(IP_2, C_4)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2), IP_2(5)$	$m_1[5] = IP_1, IP_2$
11	$(IP_2, C_1)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2), IP_2(5), IP_2(2)$	$m_1[2] = IP_3, IP_1, IP_4, IP_2$
12	$(IP_3, C_5)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2), IP_2(5), IP_2(2), IP_3(3)$	$m_1[3] = IP_1, IP_3$
13	$(IP_3, C_4)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2), IP_2(5), IP_2(2), IP_3(3), IP_3(5)$	$m_1[5] = IP_1, IP_2, IP_3$
14	$(IP_2, C_3)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2), IP_2(5), IP_2(2), IP_3(3), IP_3(5), IP_2(4)$	$m_1[4] = IP_3, IP_1, IP_2$
15	$(IP_5, C_5)$			$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2), IP_2(5), IP_2(2), IP_3(3), IP_3(5), IP_2(4), IP_5(3)$	$m_1[3] = IP_1, IP_3, IP_5$
16	$(IP_4, C_6)$	$C_6(6), C_4(5), C_3(4), C_5(3), C_1(2), C_2(1)$	$m_2[6] = C_6$	$IP_4(1), IP_3(2), IP_1(3), IP_1(2), IP_3(4), IP_1(5), IP_3(1), IP_1(4), IP_4(2), IP_2(5), IP_2(2), IP_3(3), IP_3(5), IP_2(4), IP_5(3), IP_4(6)$	$m_1[6] = IP_4$

## 5 Experiment and Results

In this section, we conduct several experiments to show the efficiency of proposed algorithm in terms of running time.

**Data Sets:** We tested both Naive approach and NATBEAN algorithm on four different networks, namely Web-Google, ca-GrQc, Global Superstore<sup>1</sup>, and Clickstream<sup>2</sup>. Out of these datasets Global Superstore and Clickstream are web log datasets. We extracted the Clickstream web log dataset using Hortonworks data platform. In Web-Google and ca-GrQc datasets the attributes are numeric. The attributes in GlobalSuperstore and Clickstream datasets are (user-id, product category) and (IP address, product category) respectively. The description of the four datasets are given in Table 5. We implemented both Naive approach and NATBEAN algorithm in java programming. The experiments were performed on a computer with an Intel Xeon 5670 (2.93GHz) processor and 8 GB in RAM, running Fedora 14.04 distribution of the 64 bit GNU Linux operating system.

**Table 5.** Dataset characteristics

Dataset	Size	Nodes	Density	Attributes
Web-Google [22]	2733	1299	0.00328924	Web pages and hyperlinks
ca-GrQc [23]	14496	5252	0.00155304	author, co-author
Global Superstore	112,628	1609	0.0143	user-id, product category
Clickstream	114,468	12441	0.015	IP address, product category

We run Naive approach and NATBEAN algorithm for  $k = 10$  by increasing the number of nodes on all datasets. The execution time for the above mentioned datasets are shown in Figs. 3, 4, 5 and 6. The running time of NATBEAN on Web-Google is faster than Naive approach at every scale of node size (see Fig. 3). Similarly for Ca-QrGc (Fig. 4), we observe that NATBEAN performs better than naive approach when the number of nodes are greater than 300. For Global Superstore NATBEAN performs faster than Naive approach from  $n \geq 15000$ . However, for  $n \geq 30000$  the running time is almost uniform as the number of nodes are increasing (see Fig. 5). On Clickstream, NATBEAN performs better than Naive approach for  $n \geq 20000$  onwards (see Fig. 6).

For ca-GrQc dataset the value of  $k$  is tested from 100 to 3500 as shown in Fig. 8 and for all values of  $k$  the differences among running times are approximately equal. However, on Global Superstore for  $k = 6, 12, 18$  the variations among running times are negligible and approximately equal when the node size

<sup>1</sup> <https://community.tableau.com/thread/194200>.

<sup>2</sup> <http://hortonworks.com/hadoop-tutorial/loading-data-into-the-hortonworks-sandbox>.

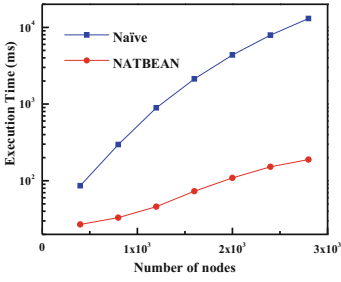


Fig. 3. Execution time on web-google

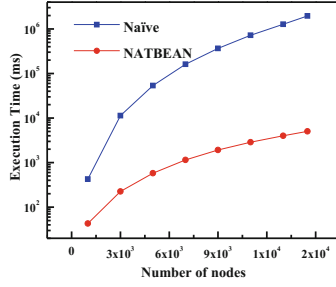


Fig. 4. Execution time on ca-GrQc

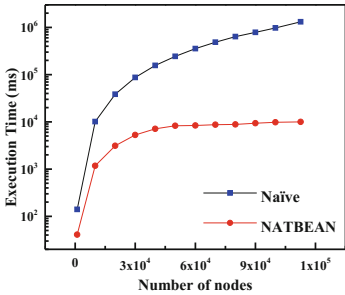


Fig. 5. Execution time on Superstore

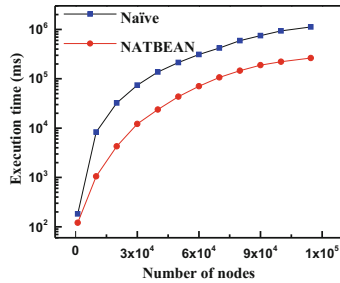


Fig. 6. Execution time on Clickstream

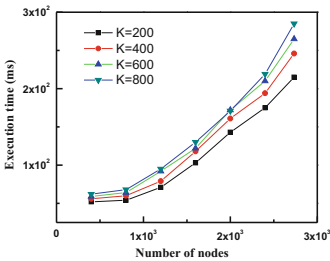


Fig. 7. Execution time on web-google

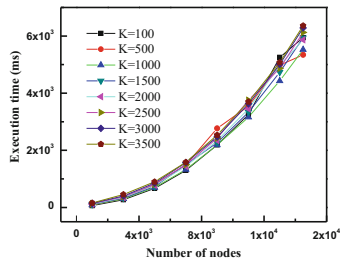
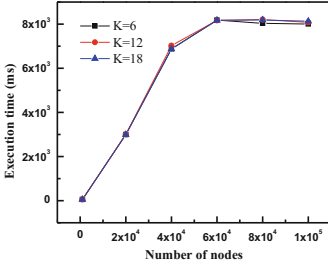
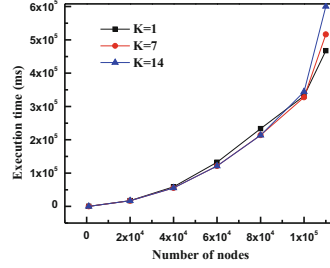


Fig. 8. Execution time on ca-GrQc

is greater than 60000 (see Fig. 9). This is due to the reason that all the 18 products exist below 60000 nodes for Superstore dataset. Similarly on Clickstream dataset for  $k = 1, 7, 14$  the running times are almost equal as  $n$  is increasing (see Fig. 10). From Figs. 7, 8, 9 and 10 we conclude that the algorithm NATBEAN performs faster for the smaller and as well as for larger increment of  $k$  values. The above figures illustrates that NATBEAN performs well than that of Naive approach for finding top- $k$  categories for an IP address-Product network.



**Fig. 9.** Execution time on Superstore



**Fig. 10.** Execution time on Click-stream

### Time and Space Complexity

In Naive approach, the computational cost for finding distinct categories is  $O(n)$  time. Then for each category finding corresponding IP addresses takes  $O(n)$  time. So that for all categories  $C_j$ ,  $j = 1, 2, \dots, n$ , it takes  $O(n^2)$  time. i.e., to find the number of similar IP addresses for all categories, the running time is  $O(n^2)$ . Finally, to retrieve top- $k$  categories with most similar IP addresses in a decreasing manner requires  $O(n)$  time. Hence, the Naive approach of finding top- $k$  categories with most similar IP addresses in a descending order takes  $O(n^2)$  time. The NATBEAN algorithm works faster than Naive approach in terms of running time.

The procedure **Category\_Identifier** finds distinct categories in  $O(n)$  time. The run time for **Store-Count** is less than  $O(n)$ . This depends upon the number of categories obtained from category identifier. The procedure **User-Location** replace the distinct categories by respective users that takes  $O(n)$  time. The number of IP addresses corresponding to distinct categories are collected by the procedure **Early-Associates** and it takes  $O(n)$  time. The total running time to retrieve top- $k$  categories is  $O(kn)$ . For all the IP addresses and category pairs, the space complexity for Naive approach requires  $O(|V|^2)$  space. The space complexity for NATBEAN algorithm is  $O(m)$ , where  $m$  is the number of edges. Comparison of running times and space complexities of proposed algorithms are shown in Table 6. The advantages of NATBEAN algorithm are (i) It removes redundant categories while finding total categories (ii) It eliminates redundant IP addresses while associating the similar IP addresses in Early-Associates approach with out reading the data repeatedly and (iii) Using top- $k$  products, most

**Table 6.** Comparison of time complexities of proposed algorithms.  $n$ ,  $m$  denotes number of nodes and number of edges respectively.

Algorithm	Time complexity	Space
Naive (Baseline)	$O(n^2)$	$O(n^2)$
NATBEAN	$O(kn)$	$O(m)$

accessed and less important categories can be identified and the less quality products can be improved through website management.

## 6 Conclusion

Finding the interactions, similar and abnormal behavior of users are useful in the analysis of social networks for improving the quality of products in online shopping. In this paper, we defined similar behavior of users, some propositions for similar behavior of users about distinct categories. In addition, we proposed an efficient algorithm for finding top- $k$  categories with maximum number of IP addresses in a descending order. This algorithm retrieves the categories in such away that the category with maximum number of IP addresses first and next category with next maximum number of IP addresses and so on. We compared NATBEAN with the naive approach in terms of both running time and space. Mining top- $k$  categories with maximum number of IP addresses are useful in identifying most popular and less important products that leads to modification of products according to the need of customers. Finding top- $k$  categories play a significant role in purchase influence mining, recommender systems and social network analysis.

**Acknowledgement.** The authors would like to thank the anonymous reviewers of this paper for their valuable comments and suggestions.

## References

1. Boldi, P., Leonardi, S., Mascolo, C., Vazirgiannis, M.: Web and social graph mining. *IEEE Internet Comput.* **18**(05), 9–10 (2014)
2. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. *J. Am. Soc. Inform. Sci. Technol.* **58**(7), 1019–1031 (2007)
3. Lorrain, F., White, H.C.: Structural equivalence of individuals in social networks. *J. Math. Sociol.* **1**(1), 49–80 (1971)
4. Lin, D.: An information-theoretic definition of similarity. In: *ICML*, vol. 98, pp. 296–304. Citeseer (1998)
5. Jeh, G., Widom, J.: SimRank: a measure of structural-context similarity. In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 538–543. ACM (2002)
6. Ganesan, P., Garcia-Molina, H., Widom, J.: Exploiting hierarchical domain structure to compute similarity. *ACM Trans. Inf. Syst. (TOIS)* **21**(1), 64–93 (2003)
7. Fagin, R., Lotem, A., Naor, M.: Optimal aggregation algorithms for middleware. *J. Comput. Syst. Sci.* **66**(4), 614–656 (2003)
8. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst. (TOIS)* **22**(1), 143–177 (2004)
9. Holme, P., Huss, M.: Role-similarity based functional prediction in networked systems: application to the yeast proteome. *J. R. Soc. Interface* **2**(4), 327–333 (2005)
10. Leicht, E.A., Holme, P., Newman, M.E.: Vertex similarity in networks. *Phys. Rev. E* **73**(2), 026120 (2006)

11. Sun, J., Qu, H., Chakrabarti, D., Faloutsos, C.: Neighborhood formation and anomaly detection in bipartite graphs. In: Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM 2005), 8-p. IEEE (2005)
12. Rossi, R.A., McDowell, L.K., Aha, D.W., Neville, J.: Transforming graph data for statistical relational learning. *J. Artif. Intell. Res.* **45**(1), 363–441 (2012)
13. Cai, J., Yu, S.Z., Wang, Y.: The community analysis of user behaviors network for web traffic. *J. Softw.* **6**(11), 2217–2224 (2011)
14. Zweig, K.A., Kaufmann, M.: A systematic approach to the one-mode projection of bipartite graphs. *Soc. Netw. Anal. Min.* **1**(3), 187–218 (2011)
15. Xu, K., Wang, F., Gu, L.: Network-aware behavior clustering of internet end hosts. In: 2011 Proceedings IEEE INFOCOM, pp. 2078–2086. IEEE (2011)
16. Xu, K., Wang, F., Gu, L.: Behavior analysis of internet traffic via bipartite graphs and one-mode projections. *IEEE/ACM Trans. Netw.* **22**(3), 931–942 (2014)
17. Jakalan, A., Gong, J., Su, Q., Hu, X., Abdelgder, A.M.: Social relationship discovery of IP addresses in the managed IP networks by observing traffic at network boundary. *Comput. Netw.* **100**, 12–27 (2016)
18. Taheri, S.M., Mahyar, H., Firouzi, M., Ghalebi, E., Grosu, R., Movaghar, A.: Hell-Rank: a Hellinger-based centrality measure for bipartite social networks. *Soc. Netw. Anal. Min.* **7**(1), 22 (2017)
19. Rossi, R.A., Ahmed, N.K.: Role discovery in networks. *IEEE Trans. Knowl. Data Eng.* **27**(4), 1112–1131 (2015)
20. Jin, R., Lee, V.E., Li, L.: Scalable and axiomatic ranking of network role similarity. *ACM Trans. Knowl. Discov. Data (TKDD)* **8**(1), 3 (2014)
21. Li, J., Li, H., Soh, D., Wong, L.: A correspondence between maximal complete bipartite subgraphs and closed patterns. In: Jorge, A.M., Torgo, L., Brazdil, P., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 146–156. Springer, Heidelberg (2005). [https://doi.org/10.1007/11564126\\_18](https://doi.org/10.1007/11564126_18)
22. Rossi, R.A., Ahmed, N.K.: Web-Google - Web Graphs (2013)
23. Rossi, R.A., Ahmed, N.K.: ca-GrQc - Miscellaneous Networks (2013)

# Booking Based Smart Parking Management System

Jhanavi Jyothish<sup>1</sup>, Mamatha<sup>1</sup>(✉), Srilakshmi Gorur<sup>1</sup>,  
and M. Dakshayini<sup>2</sup>

<sup>1</sup> BMS College of Engineering, Bengaluru, India  
jhanavicne@gmail.com,  
jukurmamtha@gmail.com

<sup>2</sup> Department of ISE, BMS College of Engineering, Bengaluru, India  
dakshayini.ise@bmsce.ac.in

**Abstract.** Today with the growing number of vehicles in the metropolitan cities, there is high demand for a smart parking management system. When people reach their destination, searching for a parking slot to park their vehicle itself creates lots of traffic congestion in the parking lot/roads taking their precious time. Hence there is a need for a smart parking management system assisting the users with the information about the availability of free parking slots at the entrance of the places like malls, organizations etc. they visit for avoiding congestion, irritation and tension. The proposed solution system in this paper tries to resolve this by detecting the available slots for parking in the parking area using IoT technologies and displaying the same on a webpage/display for user's kind reference. This system also provisions the users to book the free parking slot soon after entering the parking area thus solves the internal congestion inside the parking lot, saves their valuable time and relieves people from tension.

**Keywords:** Smart parking · IoT technology · Parking lot congestion  
Parking slot availability · Slot booking

## 1 Introduction

Internet of Things (IoT) is a global system of IP (Internet Protocol)-connected sensors, actuators, networks, machines and devices. The devices connected in the network to sense and collect data which communicate with each other and are managed by the controlling device [2]. The collected data is then shared across the internet which can be later processed and analyzed. The analyzed information is then used for various applications. It is made possible by the development and proliferation of Internet



Protocol addressable devices connected to the Web. It is quickly going mainstream. By 2020 there will be around 50 billion Internet addressable device, which translates into a trillion business opportunity [13]. IoT is creating opportunity for developing new models, and provides ways of delivering services to customers, meeting their evolving needs for more personalized products.

The devices used in the development will assimilate real world and digital world to progress towards better life in terms of quality and productivity. IoT [6] is being implemented in Smart homes which is the most anticipated feature, where top companies are getting into competition with smart appliances.

Wearable devices are additional features trending on the internet. IoT can be applied in any field of day to day, for example energy meters, wearable devices, connected cars, agriculture and healthcare devices.

To realize a Smart city, Internet of Things is the foundation. Out of the several issues in developing this, vehicle parking plays a major role [8]. To find a parking space in a busy and over populated cities like Bangalore, Mumbai, Delhi etc. is very difficult for the drivers, as each one owns a personal car [11]. This scenario can utilized as opportunity by developing smart parking solution which can enhance the efficiency parking which leads to reduction in time for searching a parking slot and traffic congestion. There are some problems related to parking and traffic congestion, which can be solved if the drivers can book the parking space beforehand at the destination [2, 4]. With the recent advances to create a low-cost, low-power embedded systems using Internet of things is helping to provide better solutions. There is advancement in sensor technology, which is deployed for variety of IoT based solutions. A survey conducted by the International Parking Institute [4] reveals that there is an increase in number of state-of-the-art ideas related to parking systems. The implemented systems need efficient sensors to be deployed in the parking slot for monitoring the occupancy as well as quick data processing in order to provide efficient parking solutions to the user [12]. Ultrasonic sensors are used to measure distance of the object in its line of sight. When the object in front is glossy or in environments where dust and humidity are high, ultrasonic technologies is the only option to mechanical probing [7]. The Ultrasonic Sensor will calculate the time required by the high-frequency sound pulse start from the sensor, hit the object and to reflect back through echo pin.

## 2 System Model and Algorithm

Proposed system aims at providing solution to the today's one of the major problems of India (vehicle parking).

## 2.1 Algorithm

```

[Nomenclature
Pst : parking status
Tr : reserved time period]

User checks the Pst of slots using Web App at the entrance
if (Pst(any slot) == green coloured)
{
  Slot is free
  Book the slot (becomes yellow coloured)
  Same is updated in the database
  Move inside the parking area towards the booked slot to park
}
if (parked at the booked slot on or before Tr)
  Parking slot indication at the layout[database] becomes red coloured[Occupied status].
else
  Tr is exceeded and indicate the slot status with green colour[Free slot]
else
{
  no free slots
  wait for some slot to become free or move out to search for parking }

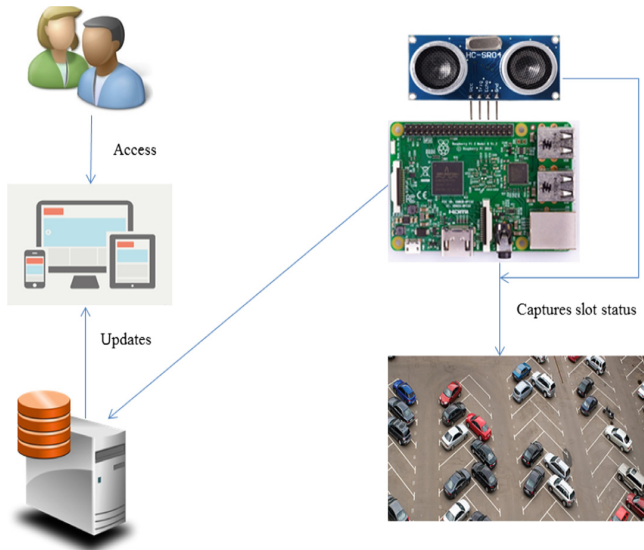
```

## 2.2 System Model

The smart parking system model proposed is as shown in Fig. 1 and consists of the following:

- Users
- Raspberry pi
- MySQL database
- Web application
- Ultrasonic sensor
- Parking lot

Ultrasonic sensors are deployed at all the slots of the parking area connected to the Raspberry pi to know the parking status whether any of the slot in the lot is free or occupied [10]. All these sensors sense and send the parking status information to the



**Fig. 1.** Smart parking system model

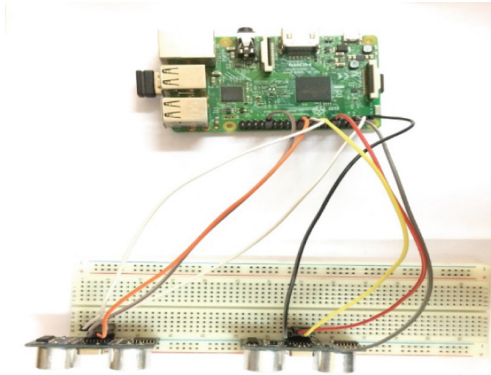
Server, where database gets updated with this information. This parking status ( $P_{st}$ ) information could be displayed along with the layout of parking lot at the entry of the parking area. In this lay-out free and occupied parking slots could be shown with green and red LEDs. At the entrance of the parking lot the user can observe the free slot available in the parking lot. The user then reserves the slot as per their wish. Then the user goes to the reserved slot and parks the vehicle. After parking the vehicle the slot will be occupied by this user and the corresponding LED in turns to red color indicating occupied.

Users are the one who views the parking space on the web page to know the slot status as “occupied or “empty”. Users can also reserve the slot soon after they enter the parking area [1].

Server processes the information and stores the same in the database. Parking lot: The parking area has an Ultrasonic sensor Connected to the Raspberry-pi. The sensor detects the presence of the car by calculating distance and sends the same to the database. Web application: It is the GUI where the user can see the status of the parking slot and can reserve the slot.

### 3 Implementation

The connection set-up of the traffic management system used in the implementation of the proposed solution is as shown in the Fig. 2. This system is implemented and tested with 2 parking slots in the lab. The algorithm implemented is:



**Fig. 2.** Implementation of Smart parking system

- There are 2 pins named trigger and echo in the ultrasonic sensor. On the raspberry-pi one ultrasonic sensor is connected to ports 18 and 24 and another sensor is connected to ports 12 and 23.
- The ultrasonic sensors deployed shown in the Fig. 2 notify whether the slot is occupied or empty. This is done by measuring the distance by the sensor. If the value sensed is within the 4.5 m then it can be claimed that the slot is occupied, else the slot is free.
- This information from the Raspberry pi board would be stored/updated with the database at the server.
- There will be a display at the entrance which shows the free, reserved and occupied slots with the green, yellow and red colored LEDs respectively. By using this information the user will be allowed to see the status, book the free slot and also would be navigated to the reserved slot and park the car.
- When the slot becomes occupied the database will be updated with the status. This is reflected in the display with yellow LED also.
- Once the car is moved from the slot then the corresponding LED turns green on the display indicating free slot.

## Result

In this work, 2 slots in the parking lot are considered. The first case is where both of these are empty. Each of the ultrasonic sensors are connected to the Raspberry Pi as shown in the Fig. 3.

In the terminal we run the program written in python, which displays the measured distance and based on that distance predicts whether the slot is empty or occupied. These values are stored in the database. All these information are updated in the website hosted using Localhost.

As a second case we consider the case when one of the slots gets occupied and the other to be empty shown in the Figs. 4 and 5. The distance measured by the ultrasonic

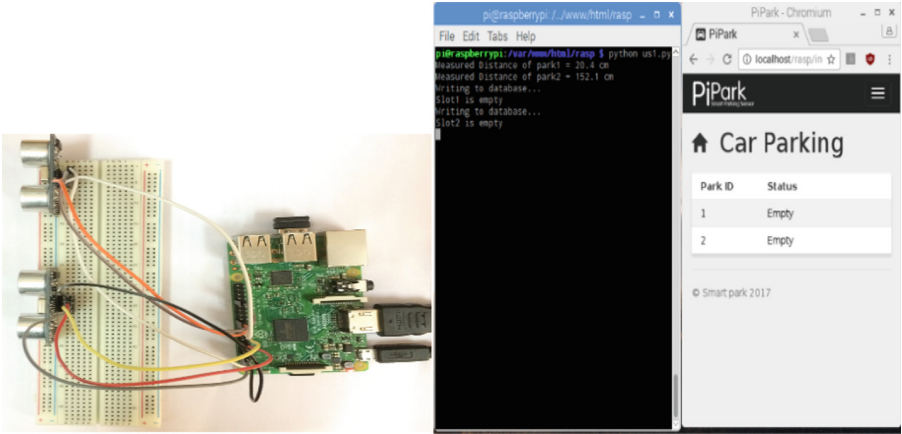


Fig. 3. Connection and output when both slots are empty

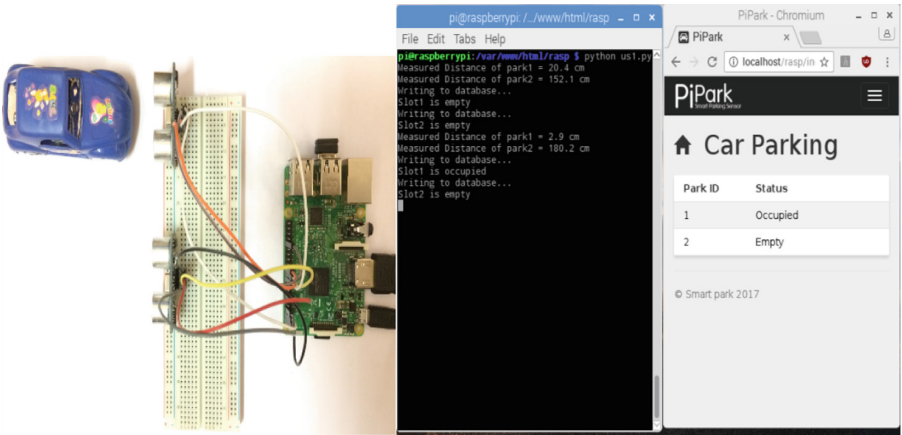


Fig. 4. Connection and output when first slot is occupied and second slot is empty



Fig. 5. Connection and output when first slot is empty and second slot is occupied

sensor where the car is parked will be very small as compared to the other ultrasonic sensor where no car is parked.

Based on this the decision of whether car is parked or not can be known. This information is displayed on the website under the status column which is retrieved from the database.

Under the third case we consider the case when both the slots get occupied as shown in the Fig. 6. This information is displayed on the website under the status column which is retrieved from the database and displayed.

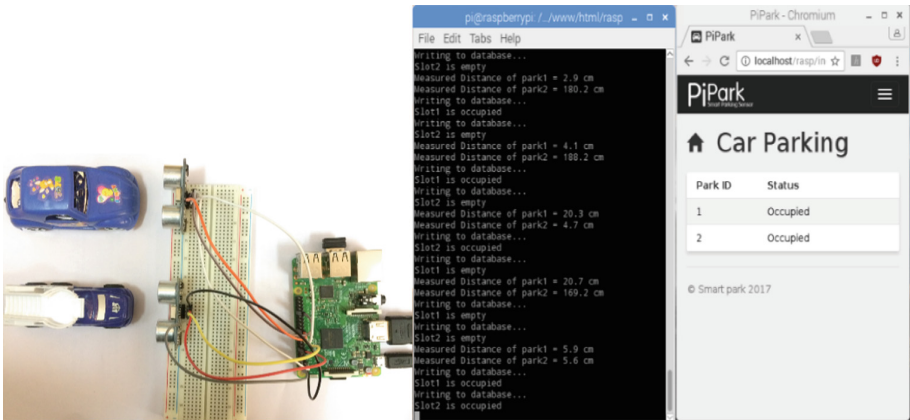


Fig. 6. Connection and output when both slots are occupied.

The result are shown only for two slots but the algorithm holds good for any number of slots in the parking area.

#### 4 Conclusion

The proposed IoT smart parking management solution system assists the people in getting to know about the availability of free parking slots in the entrance of the places like malls, organizations and other private places they visit and booking the free slot for their vehicle parking. So that congestion in the parking lot and the irritation people would experience due to clumsy searching procedure for free parking slot would be avoided. This system also helped the people to be free from the tension of spending time for searching for free parking slot. Thus this Booking based smart IoT smart parking management system successfully allows people to happily park their vehicle and carry on with their work.

#### Future Work

The proposed system uses only ultrasonic sensor. Additionally both pi-camera and ultrasonic sensor can be combined to provide a more efficient and reliable parking system. RFID (Radio Frequency and Identification) - an Automatic Identification and

Data Capture technology uses radio-frequency waves to transfer data between a reader and electronic tag attached to the particular object for tracking and identification. RFID based systems are more fast and reliable. Also it avoids any manipulations and reduces human intervention and is suitable for harsh environment where sensitive sensors cannot be effective. Further RFID provides efficient management, better security and avoids duplication since RFID is always unique.

## References

1. Ahmed, M. et al.: Study on automated car parking system based on microcontroller. 3(1), 256–259 (2014)
2. Basavaraju, S.R.: Automatic smart parking system using Internet of Things (IOT). *Int. J. Sci. Res. Publ.* **5**(12), 629–632 (2015)
3. Ibrahim, F., et al.: Smart parking system based on embedded system and sensor network. *Int. J. Comput. Appl.* **140**(12), 975–8887 (2016)
4. International parking institute: 2012 emerging trends in parking (2012)
5. Kajaree, D., Behera, R.: A survey on web crawler approaches. *Int. J. Innov. Res. Comput. Commun. Eng.* **5**(2), 1302–1309 (2017)
6. Khanna, A., Anand, R.: IoT based smart parking system. In: 2016 International Conference on Internet Things Applications IOTA 2016. January 2016, pp. 266–270 (2016)
7. Maggo, S., Aswani, R.: AUTOPARK: a sensor based, automated, secure and efficient parking guidance system. *IOSR J. Comput. Eng.* **8**(3), 2278–2661 (2013)
8. Moses, N., Chincholkar, Y.D.: Smart parking system for monitoring vacant parking. *Int. J. Adv. Res. Comput. Commun. Eng.* **5**(6), 717–720 (2016)
9. Pandey, S., Singh, K.K.: An innovative approach to smart parking systems : an overview, pp. 93–99 (2012)
10. Shah, K., Chaudhari, M.P.: Arduino based smart parking system. *Int. Res. J. Eng. Technol.* **4**(1), 882–884 (2017)
11. Sulaiman, H.A.B., et al.: Wireless based smart parking system using zigbee. *Int. J. Eng. Technol.* **5**(4), 3282–3300 (2013)
12. Vishwanath, Y., et al.: Survey paper on smart parking system based on internet of things, pp. 156–160 (2016)

# VIBI: A Braille Inspired Password Entry Model to Assist Person with Visual Impairments

V. Balaji<sup>(✉)</sup>, K. S. Kuppusamy, and Shaikh Afzal

Department of Computer Science, School of Engineering and Technology,  
Pondicherry University, Pondicherry, India  
balajipucs@gmail.com, kskuppu@gmail.com, bagbanafzal@gmail.com

**Abstract.** Visually impaired smartphone users are more vulnerable to cyber attacks such as visual/oral eavesdropping and shoulder surfing. One of the prime reasons is that these users are soft-target for the aforementioned attacks as they expose the visual and aural cues easier than the sighted users. Visually impaired people ultimately suffers from privacy threats in smart-phone platform. This paper proposes a braille inspired password entry model termed as VIBI to assist persons with visual impairments. We have examined the security attacks on smart-phones and proposed a multi-model system named VIBI, that provides the secure braille pattern based text input for persons with visual impairment. The user enters the password in an interface designed with inspirations from braille by tapping many times on touch screen with their fingers. Users can enter password in both portrait landscape orientations. We have conducted sessions with visually impaired participants to evaluate the model which shows that VIBI is faster, safer and easier to access than existing authentication models for smart-phones.

**Keywords:** Smartphone security · Visually impaired · Braille input Authentication

## 1 Introduction

Smart-phones have turned into one of the most essential components of our day to day life. Smart phone carries lot of Personal and sensitive data. When authentication of access fails, attackers can easily access the smartphone data. Shoulder surfing is a form of attack where attackers will observe the moves of an user in the physical proximity. Persons with disabilities particularly persons with visual impairments suffers a lot from shoulder surfing. When the visually impaired persons are entering the password in the device [20], attackers can view the graphical password, PIN (Stroke based input). Our primary target is to provide high level accessible authentication for smart-phone users.

Statistical report shows that the total number of mobile phone users is approximately 7.4 billion [7] and the Android operating system has been dominating the others operating in smartphone market which is installed approximately in 82.8% of smart devices globally [9]. The smartphones generally carry



lots of sensitive information including personal and business level details. Hence attackers target smartphone users in order to steal these information [14] in various ways. Authentication is the primary check against the attacks. The authentication mechanism in smart-phones can be provided by passwords, pattern lock etc. The threats confronted by the disabled people are shoulder surfing and visual/aural eavesdropping attacks. So highly secured accessible password protection is a necessary constraint to dispose of the attacks. We haven taken accessibility as major component for providing Accessible security models.

According to the WHO (World Health Organization) report, 15% of the world population is disabled, 285 million people are estimated to be visually impaired where 39 millions are blind and 249 have low vision [21]. Providing assistive smartphone technology to the differently-abled users is essential so that they become an active participants of our society.

## 1.1 Braille

Braille<sup>1</sup> is a pioneer assistive method used by visually impaired users for reading and writing. Braille based smartphone applications have entered into mobile phone platforms [8]. Persons with visual impairments also use mobile specific screen reading services such as Talkback [17], VoiceOver [19] The Braille method is organized with 6 dots, each row consist of 2 dots ( $3 * 2$ ).

Typing in password secretly is a challenging task for person with visually impairments. Screen readers makes aural feedback for each key press when entering password. Avoiding these key sounds makes it hard for visually impaired users to enter the correct password. In our work we focus to achieve secured accessible smart-phone password input mechanism and hence we have proposed a password entry model with an interface inspired by the braille patterns. The proposed model is termed as VIBI (Visually impaired Braille input).

The main contributions of this work are as listed below:

- (a) *Proposing an accessible and secure password entry model titled VIBI.*
- (b) *To utilize the familiarity of the persons with visual impairments in using the Braille by creating an interface with six dots components, for strengthening the password entry mechanism for smart-phones.*

The remainder of this paper is organized as follows: The Sect. 2 provides an overview of the related works. Section 3 deals with the Methodology and Design Goals of VIBI. Section 4 deals with experiments. Conclusion are given in Sect. 5.

## 2 Related Works

As the touch-screen based interfaces are now everywhere which makes interaction easier and provide more engaging and natural user experience. However, the touchscreen interfaces demand constant visual attention which is impossible for

---

<sup>1</sup> <https://en.wikipedia.org/wiki/Braille>.

persons with visual impairments. There are some alternative mechanisms available for them, but in the form of dedicated cumbersome hardware with limited portability and covers limited functionalities. These devices are expensive, Bluetooth based gadgets, but we need to carry additional components like batteries and additionally do manual setup which is time consuming and gives limited result. There are many commercial and freeware assistive software available for visually impaired users such as TalkBack [16] in Android phones and VoiceOver in iOS phones<sup>2</sup>. Many works have been carried out in the field of mobile accessibility for persons with disabilities: Shabnam and Govindarajan [12] proposed Braille-coded gesture patterns for touch-screens which is a character input system called *Eyedroid* in which screen is divided into two columns and three rows, the pattern motivated from 3 \* 2 braille system which is gesture based haptic app. Kane et al. [5] proposed an edge oriented gesture for visually impaired people. Siqueira et al. [13] have developed *BrailleEcran* which is an android application that consist of points on screen protector and Braille symbols on it.

Srivastava and Dawle [15] have proposed *Mudra*: A multimodal interface for braille teaching which consist of mobile phone, raspberry pi with refreshable one character braille and audio headset gives voice output and tactile feedback. Nicolau et al. [10] focused on *HoliBraille*, a system that combines touch input and multi-point vibrotactile output on multi touch screen mobile devices. They are motivated from traditional Perkins Brailier and designed an interface based on Perkins Brailier. The model consists of both software (App) and hardware. The prototype consists of six vibrotactile motors attached to springs and a silicone case. Ludi et al. [6] have shown a framework *AccessBraille*, which is an iOS framework that suggested to provide a Braille keyboard to an iOS application. The *AccessBraille* keyboard enables blind users to enter text via Braille easily onto iPads display. The keyboard framework is designed to enable steadiness in Braille input within and across through iOS apps. To initialize the keyboard, the user must swipe 6 fingers (three on each hand) upwards on the screen. The dots of braille represented as 6 columns here. The issue with this *AccessBraille* keyboard is each letter is spoke out by voice over inbuilt accessibility feature which may slow down the over all performance of the system.

*TeslaTouch* [22] uses an instrumented touch surface to provide tactile sensations on users fingers tips through electrostatic friction. However, feedback is restricted to a single point of contact. Azenkot et al. [1] proposed a non visual authentication model called *passchord*. This approach has proposed a prevention mechanism of aural and visual eavesdropping. A preliminary study investigated with 13 visually impaired participants is done to evaluate the model. To measure the password strength of the model, entropy methods has been used and metrics calculated, *passchord* model produces 15 possible combinations in 4 fingers tapping. Eiband et al. [2] presented a survey about shoulder surfing, with the participation of 174 users. He has observed variety of copying strategies. The work contributes the empirical evidence of shoulder surfing in the real world and reveal the necessity to enhance the security aspects. Visually impaired people

<sup>2</sup> <http://www.apple.com/in/accessibility/iphone/vision/>.

use smartphones with the help of Assistive technologies like screen readers. The visual and aural output of smartphone usage is a powerful cue for the attackers in carrying out proximity based attacks.

There are several methods available to provide authentication in smartphone platform such as PIN (personal identification number), password authentication, pattern lock, biometric and others. Although, these technique are providing significant contribution to protect from unauthorized people, but differently-abled persons are highly vulnerable because accessibility issue is not well addressed. In this section, we have listed some of the existing screen lock techniques.

- *Slide lock*: This method is used to lock mobile screen in an Android Operating System, but it does not provides high level security. Unauthorized persons can easily access and slide to unlock the screen without permission [11].
- *PIN lock*: This method is used to lock mobile screen, by setting numbers 0 to 9. Users are required to enter a precise pin to unlock mobile device. Drawback in this method for visually impaired persons is the easier shoulder surfing [3].
- *Password*: Users Need to provide alphabets, numbers and special characters. Drawback in this method is less secured, more complex input required. Password will take much time to give input, hackers can easily trace by shoulder surfing and eavesdropper [3].
- *Pattern lock*: This lock method is widely used to draw various graphical passwords in nine dots for normal users, which is hard to access for visually impaired users. Attackers can easily perform shoulder surfing and observe patterns [18].

### 3 Methodology and Design Goals of VIBI

The design goals of the proposed VIBI model are as listed below:

- Visually impaired persons can enter the password quickly without significant barrier.
- Visually impaired users can be protected from aural eavesdropping.
- Highly secured braille password strength with the combination of pressure gradient techniques.

The architecture of the secure user friendly Braille inspired input screen locking system is shown Fig. 1. Authentication is provided through three different inputs 1. Character passwords (visible), 2. Hidden password, 3. Pressure Gradient Input.

#### 3.1 Interface

The proposed model entitled as “VIBI” (Visually Impaired Braille Input), provides a secured accessible braille input for persons with visually impairments. The home menu of VIBI model is shown in Fig. 2.

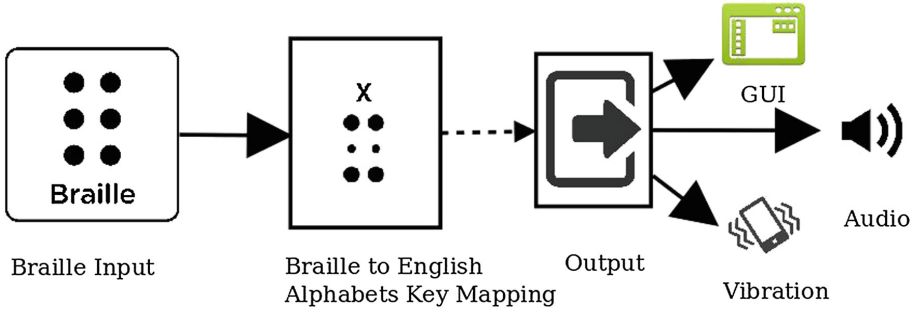


Fig. 1. VIBI architecture diagram

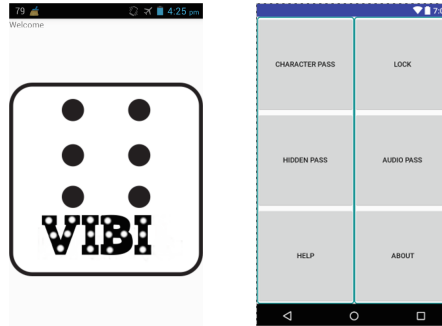


Fig. 2. Home menu of VIBI

- Braille text input - This module controls the touch input of users. It will translate the given touch input into braille alphabet. This module checks whether the touches in grid are matched or not. If the touch inputs match then produce alphabets characters, if it is not matching the grid input, then it will throw the re-enter message.
- Pressure module - This module is an additional feature of VIBI to provide high level authentication for smart phone users. Pressure module is responsible for pressure sensitivity input.

## 4 Experiments

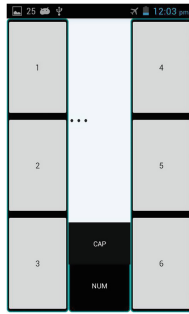
In order to calculate the performance of VIBI model, we have conducted various experimental sessions to observe performance, we included six visually impaired participants. Demographic details of the participants are shown in the Table 1. Participants were asked to use the developed VIBI Andorid app and asked to provide their feedback<sup>3</sup>. The performance is evaluated with the following

<sup>3</sup> <https://github.com/BALAJIPUCS/Braille-Pattern-Password.git>.

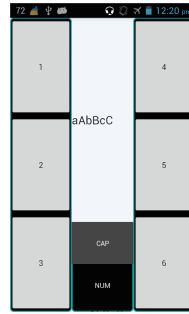
**Table 1.** Details of participants and experimental tools

Participants	Age	Impairments	Screen Readers Experiencing in mobile usage	Mobile phone	OS version
6	20 to 24 = 2	Blindness = 3 Low vision = 3	Talkback	Expert User= 1	Lenovo K3 note= 2
	25 to 35 = 2			Advanced user = 3	Samsung J7= 2
	36 to 55 = 2			Average User= 2	Lenovo A6000 = 2
					Android V5.0.2 Lollipop Android V5.1 Lollipop Android V5.0 Lollipop

components: (a) Comparison of Time Taken to lock and unlock braille pattern (b) Ease-of-Use (EoU) which is measured in Likert scale [4] of 1 to 5. 1 means least satisfaction and 5 means maximum satisfaction in our VIBI model (Figs. 3 and 4).

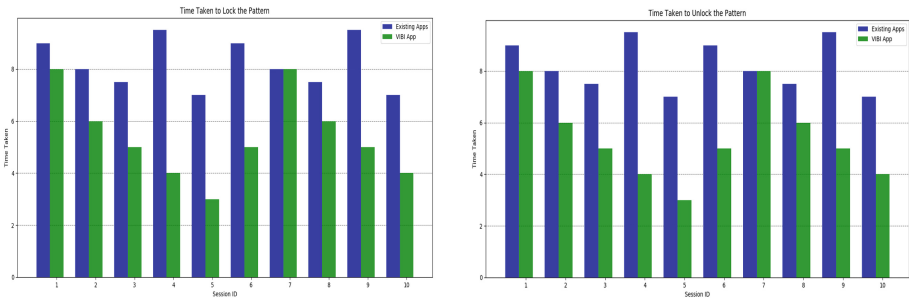


**Fig. 3.** Hidden password

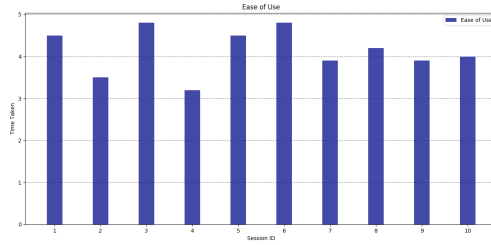


**Fig. 4.** Character password

Figure 5 Shows the mean value of Time (No of seconds) Taken to unlock pattern. Ease-of-Use metric is illustrated in Fig. 6.



**Fig. 5.** Time taken to lock and unlock the pattern



**Fig. 6.** Ease of use

## 5 Conclusions

In this paper, we have proposed a secure accessible password entry mechanism inspired by braille to assist persons with visual impairments. We have adopted this Braille grid based graphical user interface for security lock techniques. A prototype Android application was designed. We incorporated 6 visually impaired participants to evaluate the preliminary model. We have conducted experimental sessions and asked participants to use the VIBI Braille model, then collected their valuable feedback in the dimensions of security and time-taken. We have received a positive feedback from visually impaired participants. VIBI model was rated user friendly by the participants of the experiment. In future VIBI model will incorporate Bio-authentication techniques to provide multilayer security to special users.

## References

1. Azenkot, S., Rector, K., Ladner, R., Wobbrock, J.: PassChords: secure multi-touch authentication for blind people. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 159–166. ACM (2012)
2. Eiband, M., Khamis, M., von Zezschwitz, E., Hussmann, H., Alt, F.: Understanding shoulder surfing in the wild: stories from users and observers. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 4254–4265. ACM (2017)
3. Helkala, K.: Disabilities and authentication methods: usability and security. In: 2012 Seventh International Conference on Availability, Reliability and Security (ARES), pp. 327–334. IEEE (2012)
4. Jamieson, S., et al.: Likert scales: how to (ab)use them. *Med. Educ.* **38**(12), 1217–1218 (2004)
5. Kane, S.K., Wobbrock, J.O., Ladner, R.E.: Usable gestures for blind people: understanding preference and performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 413–422. ACM (2011)
6. Ludi, S., Timbrook, M., Chester, P.: AccessBraille: tablet-based Braille entry. In: Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility, pp. 341–342. ACM (2014)

7. Smartphone Market: Smartphone market. <http://www.idc.com/promo/smartphone-market-share/vendor>. Accessed 12 Apr 2017
8. Milne, L.R., Bennett, C.L., Ladner, R.E., Azenkot, S.: BraillePlay: educational smartphone games for blind children. In: Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility, pp. 137–144. ACM (2014)
9. MobileStatistics: Mobilestatistics. <https://deviceatlas.com/blog/16-mobile-market-statistics-you-should-know-2016>. Accessed 15 May 2017
10. Nicolau, H., Montague, K., Guerreiro, J., Marques, D., Guerreiro, T., Stewart, C., Hanson, V.: Augmenting Braille input through multitouch feedback. In: Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software and Technology, pp. 81–82. ACM (2013)
11. Rempel, J.: Comparing the accessibility and screen enhancement features of Google Android Lollipop 5.0 and Apple iOS 8.1.1 for people with low vision (2015)
12. Shabnam, M., Govindarajan, S.: Braille-coded gesture patterns for touch-screens a character input method for differently enabled persons using mobile devices. In: Proceedings on International Conference on Communication, Computing and Information Technology ICCCOMIT, vol. 1, pp. 1–5. CiteSeer (2014)
13. Siqueira, J., de Melo Nunes, F.A.A., Silva, C.R.G., de Oliveira Berretta, L., Ferreira, C.B.R., Félix, I.M., Luna, M.M., et al.: BrailleÉcran: a Braille approach to text entry on smartphones. In: 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), vol. 2, pp. 608–609. IEEE (2016)
14. Sonowal, G., Kuppusamy, K.: MASPHEID: a model to assist screen reader users for detecting phishing sites using aural and visual similarity measures. In: Proceedings of the International Conference on Informatics and Analytics, p. 87. ACM (2016)
15. Srivastava, A., Dawle, S.: Mudra: a multimodal interface for Braille teaching. In: Proceedings of the 6th Augmented Human International Conference, pp. 169–170. ACM (2015)
16. TalkBack: Talkback. <https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=en>. Accessed 21 Apr 2017
17. Google TalkBack: Google talkback. <https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=en>. Accessed 5 Mar 2017
18. Thaler, L.: Echolocation may have real-life advantages for blind people: an analysis of survey data (2013)
19. VoiceOver: Voiceover. <https://www.apple.com/in/accessibility/iphone/vision/>. Accessed 16 Mar 2017
20. Wakabayashi, N., Kuriyama, M., Kanai, A.: Personal authentication method against shoulder-surfing attacks for smartphone. In: 2017 IEEE International Conference on Consumer Electronics (ICCE), pp. 153–155. IEEE (2017)
21. WHO: Who. <http://www.un.org/disabilities/default.asp?id=18>. Accessed 04 Apr 2017
22. Xu, C., Israr, A., Poupyrev, I., Bau, O., Harrison, C.: Tactile display for the visually impaired using teslaTouch. In: CHI 2011 Extended Abstracts on Human Factors in Computing Systems, pp. 317–322. ACM (2011)

# A Rehabilitation Therapy for Autism Spectrum Disorder Using Virtual Reality

T. Manju<sup>1</sup>(✉), S. Padmavathi<sup>2</sup>, and D. Tamilselvi<sup>1</sup>

<sup>1</sup> Department of IT, Thiagarajar College of Engineering, Madurai, India  
{tmanju, dtamilselvi}@tce.edu

<sup>2</sup> Department of CSE, Thiagarajar College of Engineering, Madurai, India  
spmcse@tce.edu

**Abstract.** Virtual reality (VR) is a technology that simulates 3D image or environment which allows user to interact with a real or virtual environment. Using VR we can artificially create sensory experience such as sight, touch, hearing, and smell. The immersive environment can be like the real world in order to create a lifelike experience. It can be created using head mounted display, Projection, Monitors, Haptic devices, etc. Its applications widen its wings to various fields such as Education, Medicine, Military, Aerospace, etc. Autism Spectrum Disorder is a neuro-developmental disorder that highly affects normal people's peace of mind. The symptoms include lack of attention, interaction, social behaviors and so on. To overcome these, we propose a virtual environment based therapy to enhance the social skills, emotions and attention of the Autism child. The virtual environment includes various levels. First level focuses on attention grasping using color lights and sounds. Second Level focuses on increasing social interactions touching a ball, throwing it and bursting same color balloons, etc. Third Level focuses on decision making. The proposed virtual reality therapy produces positive results over repetition and it also notices at what stage the autism kids become panic, frustrated and enthusiastic.

**Keywords:** Virtual Reality · Virtual environment (VE)  
Rehabilitation therapy · Autism Spectrum Disorder · Social interaction

## 1 Introduction

### 1.1 Virtual Reality

Virtual Reality (VR) is to immerse a user within a computer generated, virtual environment that should be visually identical to the real one. The basic concept is to receive the sensory input from the outside world and use the visual and auditory cues to give a feeling of reality to the virtual world. VR has its applications in a large spectrum of fields such as Military, Education, Healthcare, Entertainment, Fashion, Construction, Business and the arts. Virtual reality applications are Augmented Reality, Virtual Worlds and Kinect. For making more realism and increase Human Computer Interaction, AI plays a major role. It is also used to design good haptic interfaces. An insightful study of typical VR systems is done. All components of VR application such



as input, output and software and interrelations between the components are thoroughly examined. VR has its own hardware issues such as motion sickness, vomiting, etc. It has its own advantage also. Generally, any of the concept can make understandable by pictorial representation or animation or with real time examples video. Like that VR has the prominent role in Education field for better understanding of the concepts. This is the key factor for us to include this in our proposed work. Additionally, human factors and their implication on the design issues of Virtual Environment (VE) are also included.

## 1.2 Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder [13] is a neuro-developmental disorder. ASD ranges as a spectrum of symptoms. So, it is named as Autism Spectrum Disorder (ASD). Asperger Syndrome comes under this umbrella. This disorder has its own spectrum of deficiencies. Autism affected children lacks in attention, social communication, emotions and interactions. ASD has various set of conventional therapies. Repeated use of therapies leads to a good result. Repeated training by man power leads to tiredness and frighten. So, the proposed work planned to implement all the conventional therapies into virtual environments.

There are two types of conventional therapies. They are discrete trial teaching and stimulus-response reward technique. Discrete trial teaching is a method that breaks down tasks into smaller components called trials, and stimulus-response - reward techniques that use physical objects to teach basic skills such as attention management, compliance, and imitation. However, most children with autism find task repetition boring and frustrating, and the objects used don't appeal to them. Consequently, children with autism often spend a lot of time off-task and have difficulty in sustaining their selective attention. Caretakers use a variety of strategies to help such children stay on task and have a more positive experience, such as annotating text on top of physical objects, using verbal and physical prompts, and offering rewards.

These therapeutic interventions cannot be made common for autism treatment, because of its spectrum nature. It means it varies with each individual affected by autism. Autism is characterized by deficits in social interaction and communication, and unusual and repetitive behavior. Cognitive abilities in people with autism vary between those with average to above average intelligence, and others who function within the moderate to profoundly mentally retarded range. Mostly, autism manifests at birth or within the first two-and-a-half years of life. Many autistic children are perfectly normal in appearance, but spend their time engaged in puzzling and disturbing behaviors that are different from those of typically developing children. They may show little or no interest in people including their parents, and pursue repetitive activities with no apparent purpose. The prevalence of autism is estimated as 1 to 2 per 1000 children, and close to 6 per 1000 children. According to the Centers for Disease Control and Prevention (CDC) around 1 in 68 children has been identified with some form of ASD in 2012 than in 2002 (1 in 150).

Many conventional therapies produce better results over a long period of time. To overcome this delay, a promising technology virtual reality technology has been introduced. In Virtual Reality same conventional therapies has been converted to

virtual environments and make the autistic child to focus, interact and react over the environment as in the real world. This technology doesn't make the child distress. With high end processing units, we can have very smooth virtual environments.

This paper is organized as follows: Sect. 2 focuses on the related work. Section 3 focuses on Methodology of the proposed Work. Section 4 deals with outcome of the proposed work and Sect. 5 concludes with the Work.

## 2 Related Works

Cordeil et al. [1] made a comparative analysis on CAVE and HMD for immersive Collaboration in network connectivity. After few analyses, they have identified that participants using HMD where faster than CAVE. Affordances for collaborative data analysis using both HMD and CAVE are the same. Thus concluded that in near future latest HMD will be used by massive users. For analysis they have used 3D network visualization because of its abstract nature. They have analyzed the impact of VR platform on task performance, collaboration and user experience.

Spatio-Temporal Based Learning Method [8] used for learning and reasoning the interactions among the objects. Through Learning, we can gain the experience of how to achieve the goals in gaming environment.

Mindful Meditation (MM) [4] leads to heavy psychophysical effects. In real time, the study has been done as during and after the MM, unable to predict the psychophysical effects. So it has been implemented using VR. It leads to positive results. The existing findings on MM only interpret the hypo metabolic state of mind and physiological parameters. So it does not produce the correct results. VR considers ecological factors, Parasympathetic nervous activity and sympathetic activity.

Realism [5] of virtual surfaces can be evaluated using haptic models constructed from data recorded during interactions with real surfaces. This model has 3 components. They are surface friction, tapping transients, texture vibrations. Following a perceptual discrepancy paradigm, intensity of surface property such as slipperiness, hardness can be avoided.

ClinicaVR [2], is a VE classroom tool for accessing attention and inhibition in children and adolescents. It aims at investigating certain validity and reliability aspects of tool, examine the relationship between performance in the virtual test and the attendant sense of presence and cyber sickness experienced by participants and assess potential effects of gender and age on performance in the test. Results of this tool support both concurrent and construct validity as well as temporal stability. Genders will not lead to performance degradation, but age does. This tool did not cause much cyber sickness.

Treatment for Online Gaming Addiction [6] can be done by improving the functional connectivity of the cortico-striatal-limbic circuit by simulating the limbic system using virtual reality therapy (VRT). After treatment, Young's Internet Addiction Scale (YIAS) scores were reduced. Connectivity to posterior cingulate cortex and bilateral temporal lobe gets increased. From results VRT seems to reduce the severity of online gaming addiction.

FACE [15] is an interactive life like facial display developed in android platform that helps the children with autism to learn, identify, interpret, and use emotional information and extend these skills in a socially appropriate, flexible, and adaptive context. Therapist will help the student to interact with FACE. The treatment scheme is based on a series of therapist-guided sessions in which a patient communicates with FACE through an interactive console. If the student is not interested, the image will fade off. The architecture of the facial automaton consists of an anthropomorphic head and a facial tracking and expression recognition device. This module will track the actions of the autism child and detects the face expression. From this they can easily identify the deficiencies. FACE is an application which has face like appearance which is able to express and modulate the basic emotions in a repeatable and flexible way, to quantitatively analyze the emotional reactions of individuals through optical analysis of facial expression, to track a human face over time, and to automatically store all data. FACE's control can be performed by an external supervisor or by an algorithm which implements a predefined design. The skeletal structure has been constructed using CAD/CAM. Soft tissues of the head were fabricated from materials used for facial reconstruction in the world of animatronics and archeology.

Virtual Dolphinarium [11] has been developed for potential autism intervention. Instead of having Dolphin Assisted Autism therapy they provide IDM- Enabled Autism Therapy. As Dolphin is an endangered species Dolphin Assisted Therapy is not advisable. To overcome these virtual dolphins were made and same Dolphin assisted therapy was made by the therapists. In Dolphin assisted therapy the children are to spend their initial times in pool activities to encounter the real dolphins. All these activities are enriched through Virtual dolphin interaction program. It allows children with autism to act as dolphin trainers at the poolside and to learn nonverbal communication through hand gestures with the virtual dolphins. Immersive visualization and gesture-based interaction are implemented to engage children with autism within an immersive room equipped with a curved screen spanning a 320° and a high-end five-panel projection system. It will promote learning skills and positive behavior.

CAVE Automated Virtual Environment (CAVE) is the virtual environment widely used in various research implementations. A workflow has been designed [3] that facilitates annotation operation such as creation, review and modification. Using this CAVE environment, the above said workflow has been implemented and obtained good results in task performance and an experience. This data annotation is used in immersive VR application to support data analysis.

The Virtual Reality Social Cognition Training (VR-SCT) [14] intervention was developed to increase social skills, social cognition and social functioning in providing effective treatments for adults with HFA. Primarily, this pilot study investigated the feasibility of a 10-session VR-SCT intervention in adults with High Functioning Autism (HFA). A secondary aim was to quantify social change over time using social performance and skill measures, and a functional questionnaire.

Bekele et al. [10] proposed a facial emotion expression recognition system. It monitors eye gaze and physiological signals related to emotions. Using these data, we can able to know how the autism affected adolescents respond to the facial expressions. The facial expressions of 10 adolescents were taken and processed with the eye gaze and physiological signals to identify the type of facial expression.

Wang and Sourina [12] proposed a novel method for multi-fractal analysis of EEG signals named generalized Higuchi fractal dimension spectrum (GHFDS) and applied in mental arithmetic task recognition from EEG signals. Electroencephalography (EEG) is used to monitor the brain's functioning and using neuro feedback technique we can train back the brain through audio or video or tactile cues. This multi fractal analysis technique produces improved result in both single channel and multi-channel subject dependent algorithms.

### 3 Methodology

Usually, ASD evolves from the age group of 18 months to adults. Many technologies are providing various solutions to the adults suffering from Autism. If we provide solutions to the age group starting from 2 years then it will be useful to the society. But providing solution to children than adults is a tedious process. We have proposed the work considering the age group starting from 2 years. The proposed framework deals with development of Virtual Environment that treats the autistic children. Figure 1 depicts the virtual scene. An autistic child is left in the Virtual room. Make the child to interact with the virtual environment by sequence of activities. First is grasping attention. This is enriched by sounds, highlighting colors and so on. This may or not make the child attentive. If not, the autism kids will be noticed their emotions that what type of expression is the child expressing. If panic, change the environment to make the kid cool. If not train them in the same environment till reaches the standard score of normal child. After making attention, make them to socially interact with the environment. Interaction means performing the tasks included in the environment. Also make the kid to travel along the path with friends. This will increase the social interactions. By repeated tasks the autistic kid can able to interact with strangers in the real world. After interactions, certain tasks are given to test their decision-making skill,



**Fig. 1.** Proposed model of virtual scene

concentration, and reasoning skill. The entire system is encircled with Cameras, Tracking System, Gesture recording system, and Audio-Visual Recording System. Expressions have been tracked and identified the mood of the child using Luxand Face SDK. Later on gestures were tracked and find out whether they are interested to interact with the environment are not. These data are stored in higher end database. Cameras are used to track the entire body actions of the child. The Visual tracking system is used to grasp their emotions or expressions from faces through eye ball tracking and eye gaze movement tracking. All these recorded data will be processed until it meets the standard score. Figure 2 depicts the flow of virtual training.

After the treatment, the performance of the Autism child is monitored. Also, their emotions, expressions, social interaction and interest towards the environment are monitored using Survey Questionnaire and analyzed with the standard Likert Scale range. Table 1 depicts the IQ range and its significance. The total score of each child is likely to fall in the specific range as in Likert scale after this therapy. These results will

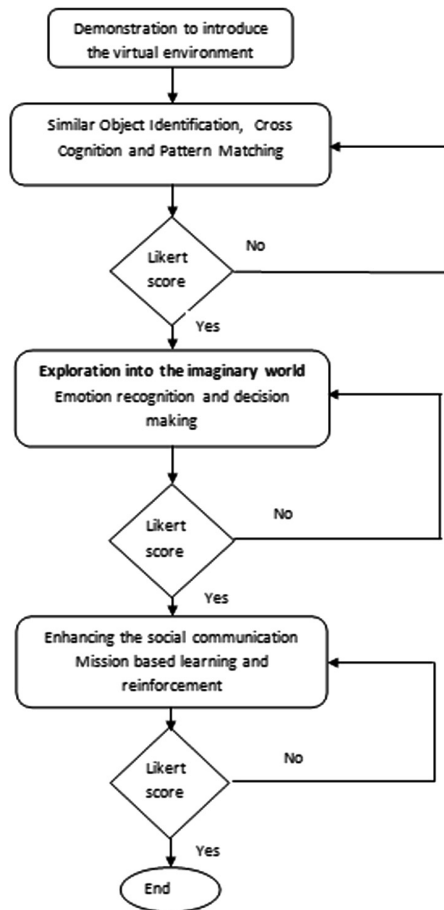


Fig. 2. Flow of virtual training

**Table 1.** IQ levels and its significance

S. no	IQ level	Significance	Population (%)
1	Below 70	Problem in IQ	4
2	90 to 110	Average	95
3	Above 130	Intellectual giftedness	1

be compared with the standard IQ, BQ and Likert Scale of a normal child to prove the results. Iterate process till it reaches the target (i.e., IQ - 100). To reach the target, the interactive module can be re-designed into an interesting and creative mode. To achieve smoothness of this interactive module 3D rendering can be increased.

#### 4 Outcome of the Proposed Work

Conventional autism therapies are achieved by repetition. But it will be enhanced by therapist. Each Child must be monitored by a single therapist. Some of these therapies include hectic devices to control the autism kids. Even with VR, we can go with devices. But the only reason we omit devices is the distressing nature to the child. To overcome this, we have come up with the proposed model. In the above proposed model, the results were obtained and compared with the standard scale. Repetition of therapy will be made until it reaches the specific standard range. VR treatments highly produce positive results by increasing the IQ, BQ and Social Skills. The Verbal and Non-Verbal Emotion Recognition is achieved by the standard scales. The new Advanced Clinical Solutions for WAIS-IV and WMS-IV Social Perception Subtest [21] was utilized to measure social perception abilities. This measure yields four scaled scores derived from various tasks. First, the SP-Affect Naming assessed the ability to match photographs of faces and people interacting to basic words of emotions. First level of our game is assessed using this SP-Affect Naming and its specific scale falls under 0 to 60. Second, SP-Prosody is a similar assessment but with auditory stimuli. Second level of game is assessed using SP-Prosody. The range of Likert scale for this falls under 0 to 30. These two subtests combine into the third Social Perception Total Score (SP-Total). Fourth, the SP-Pairs score reflects a combination of abilities in deciphering non-literal language, such as sarcasm, and the intention of the speaker. Two measures assessed Theory on Mind (ToM), or the ability to generate inferences about the thoughts and feelings of others. Our last level of game is measured using the above-mentioned metric and the results are analyzed for the specific range to be 17 to 80.

Participants are asked to interact with the virtual environment and their emotions, social interactions and attention has been noticed. Responses were recorded, compared with the standard scores. If not achieved the standard specific range, then repetition of same treatment takes place. The scoring criteria are based on the 6-point Likert scale method. Participants describing higher levels of intentional and mental states of the stimuli are awarded higher scores with the raw score range of 0–30. The results of the above proposed work will range in the scale given below in Table 2.

**Table 2.** Metrics for analysis

S. no	Measure	Specific range (Likert scale-6 pt)
1.	ACS-SP	65–110
2.	Ekman60	0–60
3.	Theory on mind – eyes	0–36
4.	Triangles	0–30

The target is to train and treat 4% of population to reach the above-mentioned scores. The proposed methodology not only improvises the social skills and attention but also induces positive behavior among the autism children. Table 3 depicts the result of autism kids that undergone training in virtual environment. 5 kids were taken into consideration for the initial stage of the proposed work and made them to interact with the virtual environment. Level 1 is attention grasping and Level 2 is increasing social interactions. Based on the Theory on Mind metrics, the standard range for a normal child falls between 0–36. Without repetition we have obtained the below positive results. By repetition over a time, it will produce good results and leads to progress to other levels.

Pretest results are obtained from conventional therapies such as occupational therapy and speech therapy. The post test results are obtained after the proposed VR training. From the Table 3 the scores have been improved among the autism kid. Results can also be improved by increasing the impressiveness of the virtual environment through high rendering and having tactile haptics feedback.

**Table 3.** Result analysis

Child	Severity	Age	Level 1		Level 2	
			Pre test	Post test	Pre test	Post test
Kid 1	Average	4	4	7	6	7
Kid 2	Mild	5	6	10	7	10
Kid 3	Average	4	4	6	6	8
Kid 4	Average	6	4	7	6	7
Kid 5	High	4	2	7	4	7

## 5 Conclusion

The therapy using VR is highly admissible and responsive even if it is expensive. It produces a very high percentage of positive results. Autistic children normally expects the smooth environment, our proposed model provides such smooth environment with high rendering. Our proposed therapy addresses lack of attention, social interaction and emotional value. As per our study VR based therapy over a time period of repetition will produce better results. Future scope of research is to purely immerse the child in virtual environment and addressing other deficiencies.

## References

1. Cordeil, M., Dwyer, T., Klein, K., Laha, B., Marriott, K., Thomas, B.K.: Immersive collaborative analysis of network connectivity: CAVE-style or head-mounted display? *IEEE Trans. Vis. Comput. Graph.* **23**(1), 441–450 (2017)
2. Nolin, P., Stipanovic, A., Henry, M., Lachapelle, Y., Lussier-Desrochers, D., Rizzo, A., Allain, P.: ClinicaVR: classroom-CPT: a virtual reality tool for assessing attention and inhibition in children and adolescents. *J. Comput. Hum. Behav.* **59**, 327–333 (2016). Elsevier
3. Pick, S., Weywers, B., Hentschel, B., Kuhlen, T.W.: Design and evaluation of data annotation workflows for cave-like virtual environments. *IEEE Trans. Vis. Comput. Graph.* **22**(4), 1452–1461 (2016)
4. Cresentini, C., Chittaro, L., Capurso, V., Sioni, R., Fabbro, F.: Psychological and physiological responses to stressful situations in immersive VR differences between users who practice mindful meditation and controls. *J. Comput. Hum. Behav.* **59**, 304–316 (2016). Elsevier
5. Culberston, H., Kuchenbecker, K.: Importance of matching physical friction, hardness and texture in creating realistic haptic virtual surfaces. *IEEE Trans. Haptics* **10**, 63–74 (2016)
6. Park, S.K., Kim, S.M., Roh, S., et al.: The effects of virtual reality treatment program for online gaming addiction. *Comput. Methods Programs. Biomed.* **40**(1), 63–74 (2016). Elsevier
7. Tentori, M., Escobedo, L., Balderas, G.: A smart environment for children with autism. Published by IEEE Conference on Pervasive Computing, vol. 14, no. 2, pp. 42–50. IEEE (2015)
8. Ersen, M., Sariel, S.: Learning behaviors of and interactions among objects through spatio-temporal reasoning. *IEEE Trans. Comput. Intell. AI Games* **7**(1), 75–87 (2015)
9. Escobedo, L., Tentori, M., Quintana, E., Favela, J., Garcia-Rosas, D.: Using augmented reality to help children with autism stay focused. Published by the IEEE Conference on Pervasive Computing, vol. 13, no. 1, pp. 38–46 (2014)
10. Bekele, E., Zheng, Z., Swanson, A., Crittendon, J., Warren, Z., Sarkar, N.: Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *IEEE Trans. Vis. Comput. Graph.* **19**(4), 711–720 (2013)
11. Cai, Y., Chia, N.K.H., Thalmann, D., Kee, N.K.N., Zheng, J., Thalmann, N.M.: Design and development of a virtual dolphinarium for children with autism. *IEEE Trans. Neural Syst. Rehabil. Eng.* **21**(2), 208–217 (2013)
12. Wang, Q., Sourina, O.: Real-time mental arithmetic task recognition from EEG signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* **21**(2), 225–232 (2013)
13. Munson, J., Pasqual, P.: Using technology in autism research: the promise and the perils. *IEEE Comput. Soc.* **45**, 95–97 (2012)
14. Kandalaft, M.R., Didehbani, N., Krawczyk, D.C., Allen, T.T., Chapman, S.B.: Virtual reality social cognition training for young adults with high-functioning autism. *J. Autism Dev. Disord.* **43**(1), 34–44 (2012)
15. Pioggia, G., Iglizzo, R., Ferro, M., Ahluwalia, A., Muratori, F., De Rossi, D.: An android for enhancing social skills and emotion recognition in people with autism. *IEEE Trans. Neural Syst. Rehabil. Eng.* **13**(4), 507–515 (2005)
16. Gobetti, E., Scateni, R.: Virtual reality: past, present, and future. Center for Advanced Studies, Research and Development in Sardinia Cagliari, Italy, vol. 58, pp. 3–20 (1998)



## Author Index

- Achuthan, Geetha 153  
Adate, Amit 287  
Adrijit, Goswami 296  
Afzal, Shaikh 320  
Akshya Kaveri, B. 224  
Alkuhlani, Ahmed Mohammed Ibrahim 108  
Anitha Vijayalakshmi, B. 171  
Arun Kumar, A. 262
- Balaji, V. 320  
Bannore, Aparna 126
- Chandrasekaran, K. 3
- Dakshayini, M. 312  
Devane, Satish 126  
Don.S 287
- Easwarakumar, K. S. 209  
Elias, Susan 209
- Gauthama Raman, M. R. 224  
Gayathri, K. S. 209  
Gireesha, O. 224  
Gorur, Srilakshmi 312
- Hema, M. S. 23
- Janani, S. 62  
Jyothish, Jhanavi 312
- Kanagaraj, K. 239  
Kandasamy, A. 3  
Kandasamy, Selvaradjou 194  
Keerthana, S. 273  
Kuppusamy, K. S. 320
- Mahaveerakannan, R. 93  
Maheshprabhu, R. 23  
Maivizhi, Radhakrishnan 139  
Mamatha 312  
Manju, T. 328  
Manohar, Hansa Lysander 12  
Meerja, Akhil Jabbar 81  
Micheal, A. Ancy 34
- Mitra, Pabitra 296  
Muthu Nagappan, M. 262
- Nageswara Guptha, M. 23  
Nesa Sudha, M. 171  
Niladuri, Sreenath 153
- P., Vimala Rani 179  
Padma, V. 47  
Padmavathi, S. 328  
Palanichamy, Yogesh 47  
Pattanayak, Binod Kumar 255
- R., Manoranjani 179  
Ramaswamy, M. 62  
Rath, Mamata 255  
Ravi, Nagarathna 179  
Reuban Gnana Asir, T. 12
- S., Mercy Shalinie 179  
Samreen, Shirina 81  
Samuel Manoharan, J. 62  
Santhosh, Gayathri 47  
Sareeka, A. G. 153  
Sasirekha, S. 273  
Saxena, Rishabh 287  
Seshadri, Karthick 179  
Shankar Sriram, V. S. 224  
Shishira, S. R. 3  
Somu, Nivethitha 224  
Soni, Gulshan 194  
Sudha, S. 262  
Suresh Gnana Dhas, C. 93  
Suresh, R. 262  
Swamynathan, S. 239, 273
- Tamilselvi, D. 328  
Thorat, S. B. 108  
Thottempudi, Ramalingeswara Rao 296
- Vani, K. 34
- Yogesh, Palanichamy 139