Navin Kumar
Arpita Thakre (Eds.)

LNICST

218

Ubiquitous Communications and Network Computing

First International Conference, UBICNET 2017
Bangalore, India, August 3–5, 2017
Proceedings

EAI

Springer

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 218

More information about this series at http://www.springer.com/series/8197

Navin Kumar · Arpita Thakre (Eds.)

Ubiquitous Communications and Network Computing

First International Conference, UBICNET 2017 Bangalore, India, August 3–5, 2017 Proceedings

Springer

Editors

Navin Kumar
Amrita University Bangalore
Bangalore
India

Arpita Thakre
Amrita University
Bangalore
India

ISSN 1867-8211 ISSN 1867-822X (electronic)
Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering
ISBN 978-3-319-73422-4 ISBN 978-3-319-73423-1 (eBook)
https://doi.org/10.1007/978-3-319-73423-1

Library of Congress Control Number: 2017962900

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 International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

We are delighted to introduce the proceedings of the very first edition of the 2017 European Alliance for Innovation (EAI) International Conference on Ubiquitous Communications and Network Computing (UBICNET). This conference brings together researchers, developers, and practitioners on one platform to discuss advances in communication such as 5G and interconnected systems. The theme of the conference was the "Internet of Things and Connected Society."

The technical program of UBICNET 2017 comprised 23 full papers in oral presentations in the main conference tracks. The tracks were arranged in the following sessions: Safety and Energy Efficient Computing; Cloud Computing and Mobile Commerce; Advanced and Software-Defined Networks and the Advanced Communication Systems and Networks. Beside the high-quality technical paper presentations, the technical program also featured six keynote speeches and a panel discussion on "The Impact of 5G-IoT and Wearables and India's Efforts Toward Standardization/ Development." The excellent keynotes speeches by experts from industry focusing on the highly challenging objectives of the country to built 100 smart cities in the next four years were highlighted. Various challenges on safety, security, and the time frame were also discussed. However, converting these challenges into opportunities was the key point of discussion to motivate the audience and encourage them to start working toward this goal. Similarly, the keynote speeches on mission-critical communication solution with 5G and interference of radio signal converted into opportunities for ubiquitous communication were also interesting talks. In addition, the conference also had three tutorials; the tutorials and a workshop on security in IoT, IoT protocols, and artificial intelligence and machine learning were equally attended by many participants. Indeed, the very first edition of the conference was very successful.

The success of the conference relied on the structured coordination with the steering chair, Imrich Chlamtac, and the general chair, Navin Kumar, as well as the Technical Program Committee (TPC) co-chair, Arpita Thakre. The conference management and EAI teams were quick in responding to queries, which was another reason for the success of the conference. We sincerely appreciate their constant support and guidance. It was also a great pleasure to work with such an excellent Organizing Committee and we thank them for their hard work in organizing and supporting the conference. In particular, the TPC, led by our TPC co-chairs, Dr. Arpita Thakre, who ensured timely review of all the papers and the selection of only high-quality of papers. We also sincerely thank the Organizing Committee co-chairs and other members, in particular the local arrangement co-chairs, Sagar B. and Ms. Sreebha, who worked tirelessly to ensure the event ran smooth and as per the plan. We are also grateful to the conference managers, Lenka, Monika Szabova, Ivana Allen, and Dominika Belisova, for their continuous support. In addition, we are very grateful to all the authors who submitted their papers to the UBICNET 2017 conference.

We strongly believe that the UBICNET conference provided a good forum for all researchers, developers, and practitioners to discuss the relevant technology, research, and development issues in this field. We hope future editions of UBICNET will be as successful and stimulating as indicated by the contributions presented in this volume.

December 2017

Navin Kumar
Arpita Thakre

Organization

Steering Committee

Steering Committee Chair

Imrich Chlamtac	CREATE-NET, Italy

Steering Committee

Navin Kumar	Amrita Vishwa Vidyapeetham (AVV University), India

Organizing Committee

General Chair

Navin Kumar	Amrita Vishwa Vidyapeetham (AVV University), India

General Co-chair

Sudarshan T. S. B.	ASE Bangalore, India

Program Chairs

Shikha Tripathi	ASE Bangalore, India
Dilip Krishnaswamy	IBM Inc., India

Technical Program Committee Chairs

Kumar Padmanabh	Robert Bosch, India
Venkatesha Prasad	TU Delft, The Netherlands

Workshops Chair

Syam Madanapalli	DELL Inc. India, ASE, Amritapuri, India

Web Chair

Rajesh M.	ASE Bangalore, India

Publicity and Social Media Chairs

Kartinkeyan R.	ASE Bangalore, India
Nippun Kumaar A. A.	ASE Bangalore, India

Sponsorship and Exhibits Chair

Shekar Babu	Amrita University, Bangalore, India

Finance Chair

Rakesh N.

Publications Chairs

Arpita Thakre	ASE Bangalore, India
Kirthiga S.	ASE Coimbatore, Tamilnadu, India

Panels Chair

Murty N. S.	ASE Bangalore, India

Tutorials Chairs

Vamsi Krishna T.	PESIT University, Bangalore, India
Kaustav Bhowmick	ASE Bangalore, India

Demos Chair

Kishore A.	UTL Technology, India

Posters and PhD Track Chairs

Seshaiah P.	NEC Inc., UK
Balaji Hariharan	ASE Amritapuri, Kerala, India

Local Chair

Ramesh T. K.	ASE Bangalore, India

Conference Manager

Monika Szabova	EAI - European Alliance for Innovation

Technical Program Committee

Amod Anandkumar	Mathworks Inc., India
Kiran Kuchi	IIT Hyderabad, India
Claudio Sacchi	UNITN, Italy
Debu Nayak	Huawei, India
Mayur Dave	Reliance Telecom, India
Dharma P. Agrawal	University of Cincinnati, USA
Indranil Saha	IIT Kanpur, India
Suvra Sekhar Das	IIT Kharagpur, India
Niranth Amogh	Huawei, India
Vladimir Poulkov	Technical University, Sofia, Bulgaria
Preetam Kumar	IIT Patna, India
Ashutosh Dutta	AT&T, New Jersey, USA
Kalyan Sundaram	Sai Technologies, India

Sanjay Kumar	BIT, Mesra, India
T. V. Prabhakar	IIT Kanpur, India
Eduardo R.	University of Aveiro, Portugal
Abyayananda Maiti	IIT Patna, India
Everesto Logota	Cisco, UK
Saravanan Kandaswamy	University of Porto, Portugal
Sumeet Agarwal	IIT Delhi, India
Sweta Sarkar M.	University of California, USA
Saif K. Mohammed	IIT Delhi, India
Joongheon Kim	Intel Corporation, USA
Jun Bae Seo	IIT Delhi, India
Prasant Misra	Tata Consultancy Services, India
Sunil Kumar	University of California, USA
Dileep P.	Intel Inc., India
Tjo Afullo	University of Kwazulu Natal, South Africa
Neelesh B. Mehta	Indian Institute of Science, India
Jamil Khan	University of Newcastle, Australia
Vandana R.	Trinity College Mumbai, India
Yoan Shin	Soongsil University, South Korea
Vivek Deshpandey S.	MIT, India
Rohit Gupta	EUROCOM, France
Akos Lakatos	University of Debrecen, Hungary
M. M. Deshmukh	Trinity College, Pune, India
Suman Kumar Maji	IIT Patna, India
Maroun Jneid	Antonine University, Lebanon
Ravi Pandurangan	Chaitanya Bharathi Institute of Technology, India
Loc Nguyen	Vietnam Sunflower Soft Company, Vietnam
Eswaran P.	SRM University, India
David Koilpillai	IIT Madras, India
Asif Ekbal	IIT Patna, India
Şaban Gülcü	Necmettin Erbakan University, Turkey
Shibo He	Zhejiang University, China
Arijit Mondal	IIT Patna, India
Vishal Satpute	VNIT, India
Walid Saad	Virginia Tech, USA
Dhanesh Kr. Sambariya	Rajasthan Technical University, India
Cong Wang	City University of Hong Kong, SAR China
Mahesh K. Marina	University of Edinburgh, UK
Shahanawaj Ahamad	University of Hail, Saudi Arabia
Rajiv Misra	IIT Patna, India
Sriparna Saha	IIT Patna, India
Ioannis Papapanagiotou	Netflix, USA
Shruti Jain	Jaypee University of Information Technology, Solan, India
Jianwei Niu	Beihang University, Beijing, China
Sachin Ruikar	Walchand College of Engineering, India

Changqiao Xu	Beijing University of Posts and Telecommunications, China
Xu Huang	University of Canberra, Australia
Huan Xuan Nguyen	Middlesex University, UK
Deepa Kundur	University of Toronto, Canada
Ramadan Elaiess	University of Benghazi, Libya
Jaime Lloret Mauri	Polytechnic University of Valencia, Spain
Abhishek Shukla	R.D. Engineering College Technical Campus, India
Nilanjan Banerjee	University of Maryland, USA
Paolo Bellavista	University of Bologna, Italy
Sunghyun Choi	Seoul National University, South Korea
Swades De	Indian Institute of Technology, India
James Gross	KTH, Sweden
Pan Hui	Hong Kong University of Science and Technology, China
Dimitrios Koutsonikolas	University at Buffalo, USA
Huadong Ma	Beijing University of Posts and Telecommunications, China
Jorge Sa Silva	University of Coimbra, Portugal
Anand Seetharam	California State University, USA
Shamik Sengupta	University of Nevada, USA
Salil Kanhere	The University of New South Wales, Australia
Reema Sharma	Oxford College of Engineering, India
Rahul Bhattacharyya	MIT, USA
A. Chokalingam	IISc Bangalore, India
Ekram Hossain	University of Manitoba, Canada
Mehdi Rast	Amirkabir University of Technology, Iran
Bharat B. N.	PESIT Bangalore, India
Fouzi Lezzar	Abdelhamid Mehri-Constantine University, Algeria
Ashish Kr. Luhach	Lovely Professional University, India
Mastaneh Mokayef	UCSI University, Malaysia
Samad Kolahi	Unitec Institute of Technology, New Zealand
Mario Henrique Souza Pardo	University of São Paulo, Brazil
Ranjitha Kumar	University of Illinois at Urbana-Champaign, USA
Sourav Bhattacharya	Bell Labs, USA
Sudarshan Rao	BigSolv Lab Pvt Ltd., India
K. Zahedi	University Teknologyi Malaysia
Y. Zahedi	University Teknologyi Malaysia
Muhammad R. Kamarudin	University Teknologyi Malaysia
Mohd H. Jamaluddin	University Teknologyi Malaysia
Dhaval Vyas	Queensland University of Technology, Australia
Kellie Vella	Queensland University of Technology, Australia
Jinglan Zhang	Queensland University of Technology, Australia
Ross Brown	Queensland University of Technology, Australia
Jeedigunta Venkateswar	Samsung

Prithiviraj Venkatapathy	SRM University, India
Ajay Chakrabarty	BIT Mesra, India
Raja Srinivas	Tata Tele, India
Aloknath De	Samsung
Dilip Pangavhane	Tech Somaiya University, India
Wangwei Wilson	Huawei
Seshadri Mohan	UALR, USA
V. Sinha	IIT Kanpur, India
R. V. R. Kumar	IIT Kharagpur, India
K. Varya	IIT Madras, India
Peter Lindgren	Alborg University, Denmark

Contents

'MobAware'-Harnessing Context Awareness, Sensors and Cloud for Spontaneous Personal Safety Emergency Help Requests

V. G. Sujadevi, Aravind Ashok, Shivsubramani Krishnamoorthy, P. Prabaharan(✉), Prem Shankar, Mani Bharataraju, Sai Keerti, and D. Khyati

Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University,
Kollam, Kerala, India
{sujap,aravindashok,praba,premshankar}@am.amrita.edu,
br.ramanand@gmail.com, mani.bharatarajuk@gmail.com,
saikeerthipanchagnula@gmail.com, khyatidm@yahoo.co.in

Abstract. Significant increase of crimes against women in recent years and the advent of smart phone and wearable technologies have accelerated the need for personal safety devices and applications. These systems can be used to summon for help during the emergency situations. While several mobile applications that sends emergency help requests are available they need to be manually activated by the victim. In most of the personal emergency situations the victim might not be in a position to reach out for the Smart phone for summoning help. In this research we address this issue by implementing a system that automatically senses certain personal emergency situations, that summons for help with minimal or no user intervention. Summoning of help gets triggered when the smartphone sensors senses an abnormal events such as unusual movement and voice. This system also profiles the spatial information using the crawled web data and provides the contextual information about the risks score of the location. By using sensors and context awareness our system summons for emergency help with minimal/no intervention by the user.

Keywords: Context aware system · Smartphone · Cloud service · Personal safety Emergency help · Decision tree

1 Introduction

Hundreds of physical harassment incidents are reported throughout the world every year. The increased number of violence against the women and children significantly contributes to these incidences [1]. It is not feasible to employ police force and security personnel to prevent these crimes, as covering all the locations is difficult. The 'Nirbhaya' attack in Delhi [1] in which a 23-year-old Indian citizen was brutally harassed in a bus has generated a widespread of fear among women all over the country. This has ushered the era of personal safety systems that can send emergency help requests to police and the caretakers in the event of any dangerous personal safety situation. Several personal safety applications are available for the

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 1–12, 2018.
https://doi.org/10.1007/978-3-319-73423-1_1

smartphones that can send emergency information to the predefined list of contacts whenever in danger. The situation of the crime scene might not allow the victim to reach out to the smart phone and unlock them to be able to reach out to help. This could be due to the time and physical limitations. This might also due to the fact that the victim might not be in a situation to respond properly when encountered with a surprise attack. These triggering of the alert procedures can be automated and simplified. This might help the victim to reach out and summon the help fast. These can be achieved by harnessing the sensors and actuators in the smartphones and by providing context aware response to the victim in real time. Context aware systems are computing systems that provide relevant services and information to users based on their situational conditions [2]. A context can be best explained with the help of these three features: (1) Location- Where you are; (2) Neighborhood- Who are you with and (3) Environment- What resources are you around [3].

Using Context Aware Systems, this paper proposes a novel application - 'MobAware' in A system that runs on a smartphone that is capable of sensing danger for the user from its environment and automatically sends emergency requests to nearby social network friends, relatives, police stations and Non-Governmental Organizations who specialize in these emergency response procedures. The system avoids the need to make the user manually trigger the application for summoning the help. The proposed system also notifies the user on the risk level of the present location. This context is calculated based on the history of attacks occurred in that area. The application runs as an Android Service [4] in the background, which makes using the Smartphone with the Personal safety application and unobtrusive one. This also helps the user switch to any other applications without having the need to interrupt it. The system also provides a web based interface that can provide real time information such as real-time user tracking and monitoring.

The next section discusses the literature survey followed by our proposed system. In the fourth section the system architecture and implementation details are discussed. In the last section Conclusion and Future Work are discussed.

2 Literature Survey

Context-aware computing has been an interesting research area for more than a decade. It was first mentioned and discussed [5]. The work in [5] describes about an active map service, with the help of context-aware computing, the system was able to provide the context aware information to their clients, about their located-objects and how those objects location changed over time. A detailed survey of the existing context-aware systems and services are discussed [6]. An automated context-aware application using decision trees has been created [7]. The system was used to learn user's preferences to provide personalized services based on the user's context history. A context aware service platform called 'Synapse' [8] has been created to predict the most relevant services a user will use in a particular situation, based on the users habits. The system uses Hidden Markov Model to provide this personalized service [8]. Inferring the user's activity by analyzing the data obtained from a single x-y accelerometer using clustering algorithm and neural networks has been performed [9]. Pre-processing techniques, which can be implemented in mobile devices for extracting user activity from accelerometer data has been proposed [10]. Employing the social networks for sending emergency requests and responses has been implemented [11]. Obtaining of users activities from various mobile and external sensors were used to publicize the users activity in Social Networking sites like Twitter and Hi5 has been proposed [12]. A next generation public safety system designed to be fully context aware to initiate an emergency call to summon response has been developed and deployed at the university campus [13]. There are several mobile applications with the focus on addressing violent crimes like sexual assault, rape, robbery and domestic violence. Some of the popular ones are Circle of 6 [14], Sentinel [15], bSafe [16], Fightback [17] etc. However these applications requires the user to manually trigger by the touch of a button. Dialling of voice calls to the list of contacts by vigorously shaking the mobile 3 times in 5 s [18].

3 Proposed System

The proposed 'MobAware', system (Fig. 1.) senses the environment of the user and automatically sends emergency requests based on the level of dangerous circumstance or scenario. The architecture of the application is given below. The system has three major subsystems. They are described in the following sections.

1. **Cloud Service:** The cloud service consists of crawlers for information acquisition from various news feeds and social networks. A crawler is a program that visits

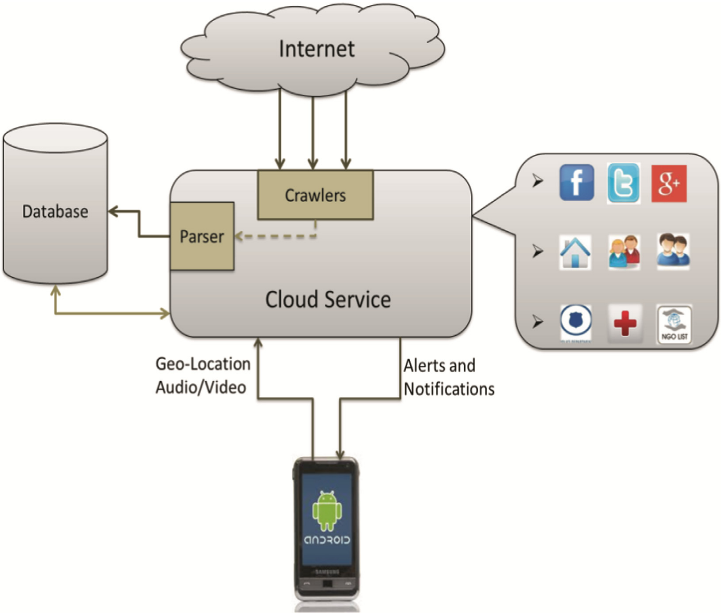


Fig. 1. Overall architecture diagram

several news media web portals and information database systems, acquires information from those sources. This data is then parsed and further processing is done on this data to extract the physical security incidences like, theft, accidents, robbery, assaults, specific attacks against women, children and other vulnerable citizens etc. The data crawled is categorized based on the location, based on the available GIS data and stored in a database which contains all the previous histories. Based on the number of incidents and severity of the incidence which occurred a score is provided. The higher the score for a location the risky the location is categorised as. This cloud service also has interface to the social networking systems including Facebook, Twitter and Google+. Based on the configuration of the individual settings the system sends emergency help requests to the friends list in the social network platforms. The system also interfaces with a short messaging service (SMS) to inform the users, systems and organizations which uses voice/SMS services for the incident response. The system also contains the information on nearby police stations, Non-governmental organizations specializing in first aid and responses, based on the user's current location. This ensures the quick response times to respond to incidents. The system can also be queried by the user for the risk levels of the location before visiting the location. This feature helps the user to take necessary precaution/avoid the visit including the avoiding the visits in the night time which is deemed unsafe etc. This information is available in the web portal that is dedicated for the personal safety.

2. **Smartphone Application:** A smartphone application has been designed and developed for the Android operating system. The mobile application consists of four sensors illustrated in Fig. 2. With the help of these four sensors the mobile application gathers data about the user's location and the users surrounding environment. The data obtained by the sensors is sent to a Decision tree where, based on the input given it determines whether the user is in unsafe condition or not. The application also has a SOS feature in which the user can manually trigger the alert service by pressing a button in the event of need for the manual intervention.

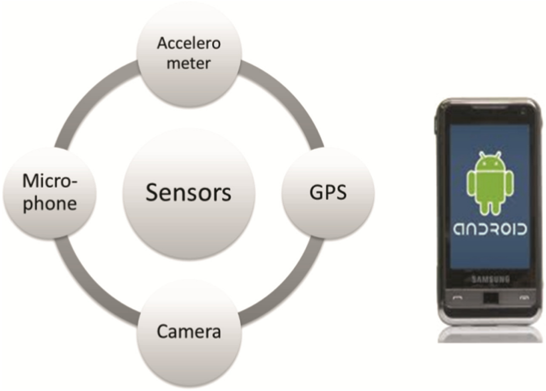


Fig. 2. Different sensors used by the system

3. **Web Portal:** When the "Mobaware" application is installed in the user's mobile for the first time, the system provides a facility to register user information etc. Once the account is created, the user is provided with the credentials to log in and use the web portal [21], which is created for managing the personal safety of the individual. The web portal can be used for monitoring and tracking the user. The web portal contains the real-time location of the user marked in a map. The map also shows the availability of nearby friends, police stations, NGOs and hospitals and their respective distance from the user. This web portal can also be used to monitor the places where any alert request is made in real time. The map also shows the number of attacks and physical violence, which occurred in the past. This information updated on a daily basis. Figure 3 shows city-wise distribution of registered rape cases, which was collected using the systems crawlers and Analytics subsystems in the month of April, 2015.

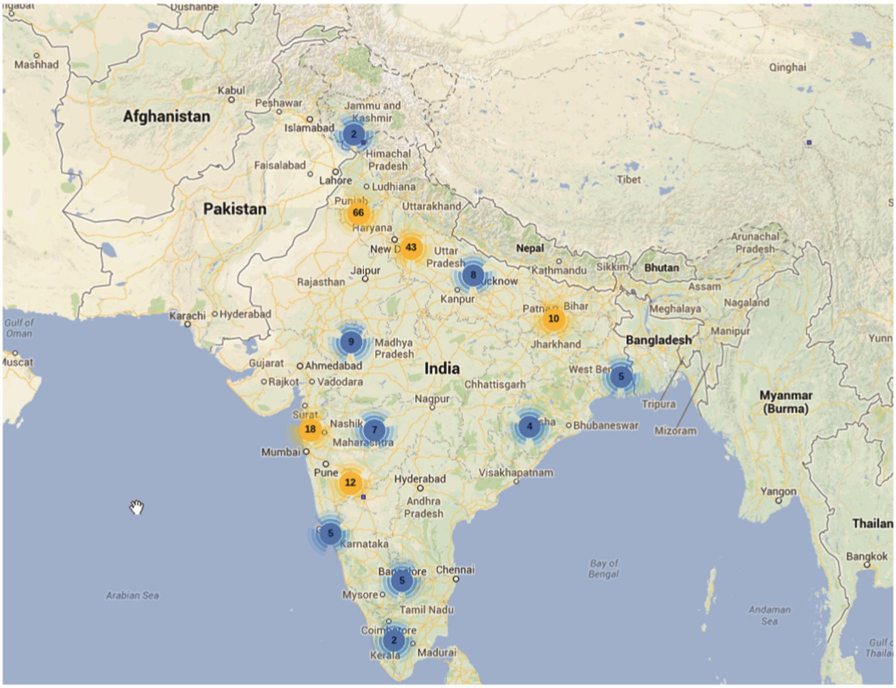


Fig. 3. Places with registered rape cases across India in month of April 2015

4 Implementation Details

The three main system functionalities of this system are listed below:

1. **Context Finding and Auto-Alerting:** Several smartphone applications [14–17] are available at the application market store. Most of these applications have a software

based SOS button that requires the user to manually trigger the alert button for sending emergency requests while in danger. However in an adverse situation of physical assaults, the time and freedom to take the smartphone application and unlocking it to perform the trigger is limited. Hence, we proposed a cloud based system that is aware of the context of the environment and surroundings and initiates the help requests automatically in adversarial circumstances. The sensors present in the smartphone helps to achieve this by acquiring the context aware data. This data is pre-processed and fed in to the analysis engine which follows the algorithm described below. Figure 4 is the pictorial representation of the algorithm.

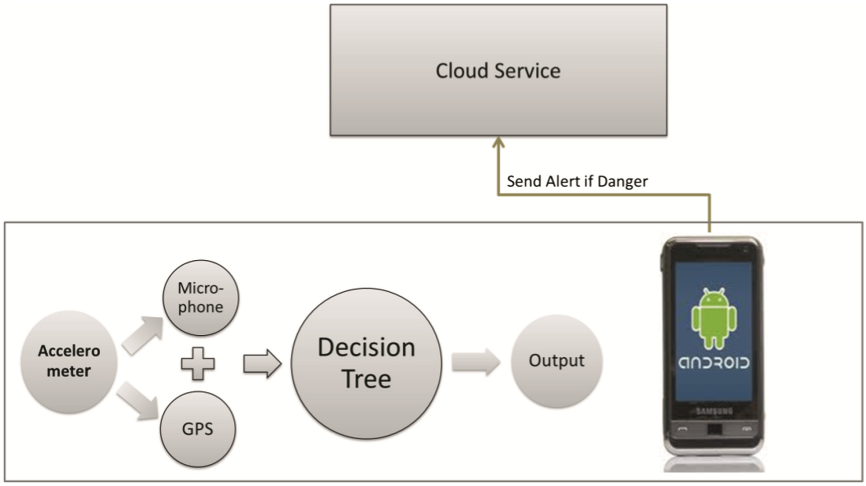


Fig. 4. Context identification and alerting

Algorithm:

- Listen for any unusual shake/motion by using accelerometer.
  – Invoke Microphone and GPS sensors.
- Microphone and GPS Sensors starts working in parallel.
  – Listens for any audio signal with the help of Microphone.
    Find the maximum, minimum and average amplitude of the audio signal.
    Perform offline voice to text conversion of the audio signal.
  – Identify the location of user with the help of GPS Sensor.
    Find out the User Activity.
    Compute the risk factor of travelling in that area.
  – Provide the obtained results to a Decision Tree.
- Identify the situation of the user based on the input of Decision tree and store it.
- Repeat steps 1–3 two more times.
- If the output of the decision tree suggests adversarial situation more than once i.e. minimum of 2 out of the 3 outcomes then send alert message.

The mobile uses three sensors for the working of this feature:

1. *Accelerometer*: An accelerometer is a component device that measures proper acceleration [19]. It is one of the motion sensors used by smartphones and other wearables to detect and monitor motion or vibration. In this system accelerometer is used to observe any shake or vibration. When it observes shake, the system invokes the Microphone and GPS sensors.
2. *Microphone*: Whenever the microphone is invoked it listens for some audio signal. It then finds the maximum, minimum and the average amplitude of the signal. It also performs a voice to text conversion and checks whether words like 'Help', 'Save' are in it or not which results in further action.
3. *GPS*: Using the GPS facility the system identifies the current location of the user. Based on the speed with which the user is travelling it estimates the current activity of the user; i.e. it tries to identify whether the user is walking, exercising, idle or on a vehicle. Once the location data is fed into the system from the GPS sensor signal, the next step is to assess the risk score of travelling in that location. The risk score of a particular location is calculated based on factors like:
   - Number of past occurrences of known assault incidents in that area and its surrounding locations.
   - Proximity to First aid places, Law enforcement offices, Hospitals, NGOs etc.
   - Time of travel.

The values obtained by microphone and GPS is given as input to decision tree. The decision tree based on its input decides whether the user is in adversarial circumstance.

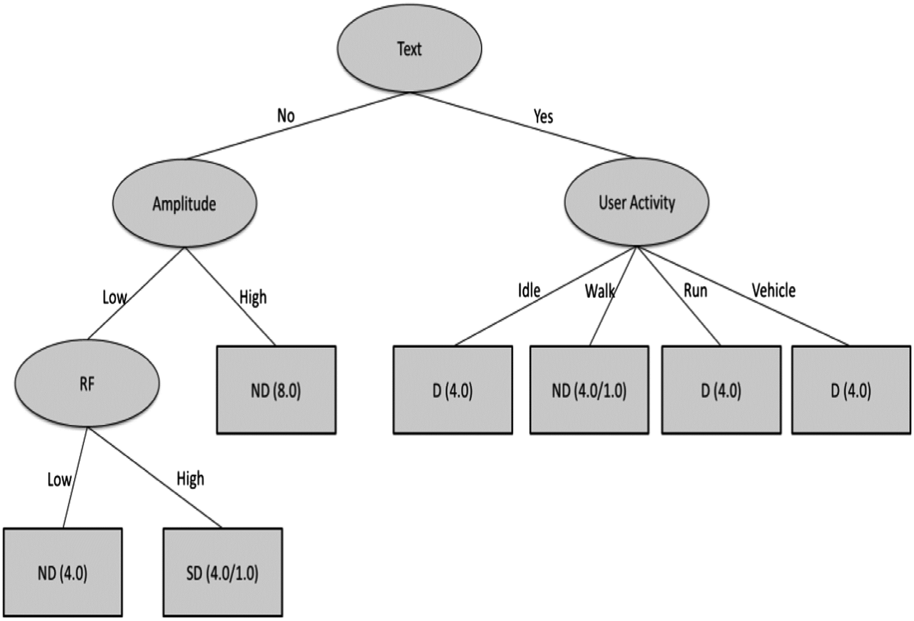
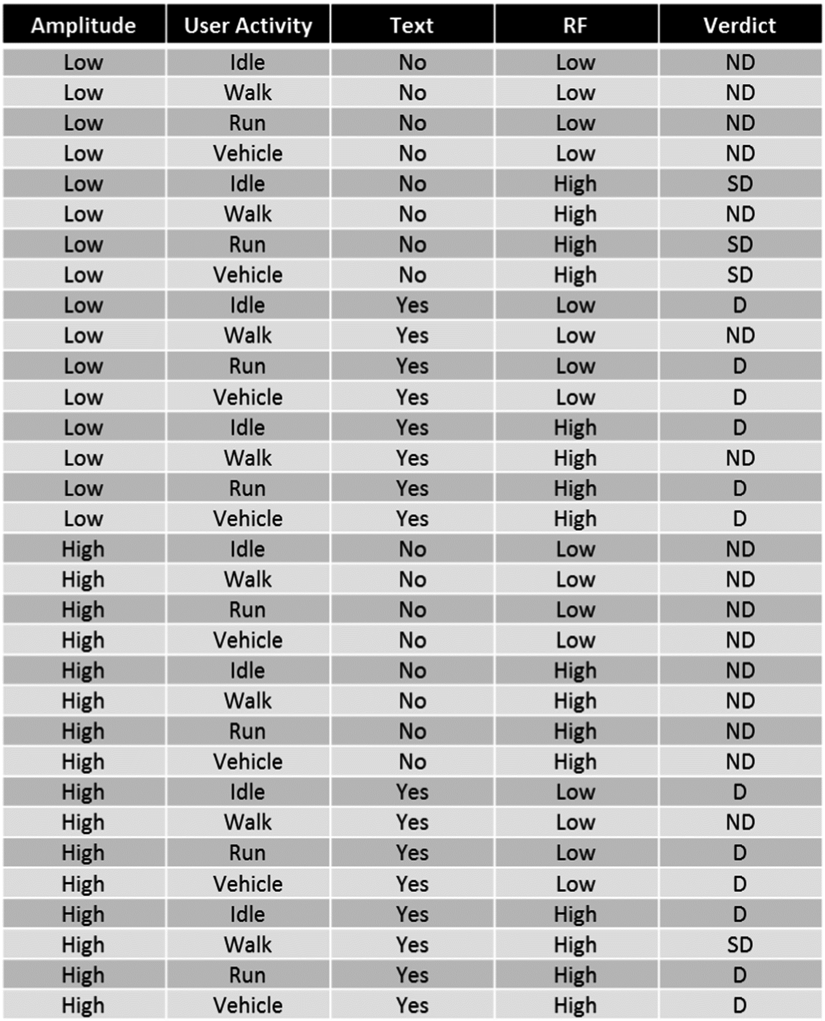


Fig. 5. Pruned decision tree

The entire process is repeated three times. If the output of the decision tree suggests adversarial circumstance for the user at least for 2 times out of the three instances or semi dangerous for all three cases, the mobile sends a signal to the cloud service a request to ask for help immediately. The pruned decision tree and the training data are shown in Figs. 5 and 6 respectively. One of the biggest concerns for the first responder system is the number of false positive calls or the emergency requests. These have detrimental effect in the first responder systems. Firstly false positive calls/emergency requests take away the precious times of the first responders who otherwise could be helping the actual needy people. Secondly, it can overwhelm the entire first responder system, in such a way that they might even start ignoring the true emergency calls. Thirdly the end user



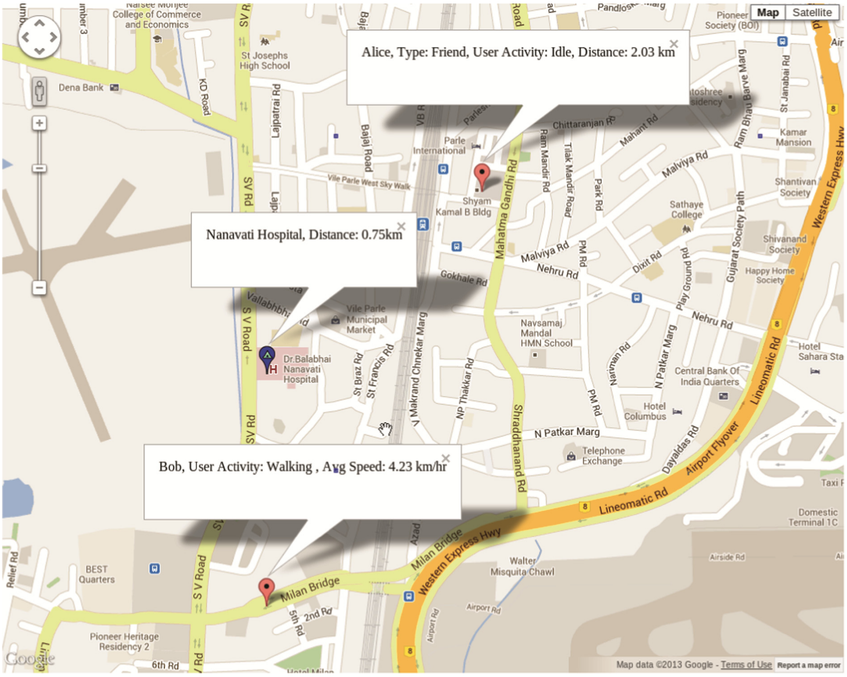
Amplitude	User Activity	Text	RF	Verdict
Low	Idle	No	Low	ND
Low	Walk	No	Low	ND
Low	Run	No	Low	ND
Low	Vehicle	No	Low	ND
Low	Idle	No	High	SD
Low	Walk	No	High	ND
Low	Run	No	High	SD
Low	Vehicle	No	High	SD
Low	Idle	Yes	Low	D
Low	Walk	Yes	Low	ND
Low	Run	Yes	Low	D
Low	Vehicle	Yes	Low	D
Low	Idle	Yes	High	D
Low	Walk	Yes	High	ND
Low	Run	Yes	High	D
Low	Vehicle	Yes	High	D
High	Idle	No	Low	ND
High	Walk	No	Low	ND
High	Run	No	Low	ND
High	Vehicle	No	Low	ND
High	Idle	No	High	ND
High	Walk	No	High	ND
High	Run	No	High	ND
High	Vehicle	No	High	ND
High	Idle	Yes	Low	D
High	Walk	Yes	Low	ND
High	Run	Yes	Low	D
High	Vehicle	Yes	Low	D
High	Idle	Yes	High	D
High	Walk	Yes	High	SD
High	Run	Yes	High	D
High	Vehicle	Yes	High	D

Fig. 6. Training data for decision tree

of the emergency application him/herself might be annoyed and might lose the faith in the system. This is the reason for repeating the process three times is to increase the efficiency of the system by removing the false positives due to user negligence and unrelated triggers.

Here the value in text is yes if any keywords like 'Help', 'Save' etc. are present and vice versa. Also RF is the risk factor calculated. The output of the decision tree can be any of: dangerous (D), semi-dangerous (SD) and non-dangerous (ND).

2. **Real-Time User Tracking:** As mentioned in the previous sections, MobAware system has a dedicated web portal that can be used for tracking and monitoring the user in real time. For tracking a particular user, the user has to go thru a registration process, which is a mandatory. The user also needs to provide an explicit consent for the permission to tract the user. The web portal currently leverages Google maps GIS system [20] for displaying the current location of the user. By default the map is shown as zoomed in to make it easy and visually convenient for the people to track and monitor a user. While privacy could be thought of it as a concern, but due to the fact that the application helps to protect a user from the adversarial circumstances outweighs the privacy concern. The map is updated every 5 s and current information is provided for the accurate tracking purposes. The map also indicates several landmarks that includes nearest police stations, hospitals and a list of friends whose current geo-location is available along with the distance and direction from them.


Alice, Type: Friend, User Activity: Idle, Distance: 2.03 km

Nanavati Hospital, Distance: 0.75km

Bob, User Activity: Walking , Avg Speed: 4.23 km/hr

Fig. 7. Tracking Bob in real-time in Google maps

Figure 7 gives an example in which a friend of Bob is tracking him. The person, tracking Bob gets to see the exact location of Bob, his current activity and the speed with which he is moving. In the map one can also see nearby police stations, hospitals, and online friends and the distance from them.

3. **SOS Report Abuse:** SOS or Report abuse is another important feature provided by MobAware system as shown in Fig. 8. This is used when there is a need to manually trigger the process of summoning help or to send emergency help requests. This facility could be used in two scenarios:
   (i) User is in adversarial circumstance, but is able to manually trigger the SOS button. For example, suppose Alice is walking alone in an area, which is not densely populated. She observes some anti-social elements have been following her for some time. In this case, she could press the SOS or Report Abuse button in the mobile application to call for help.
   (ii) A user is summoning for the help on the behalf of others who was assaulted or is in adversarial circumstances. In this scenario suppose Bob just went through a nearby lane. He notices Alice (stranger to Bob) is being assaulted by a group of anti-social element. Now when Bob wants to help Alice he could use the mobile application and trigger the SOS or Report Abuse button calling for help.

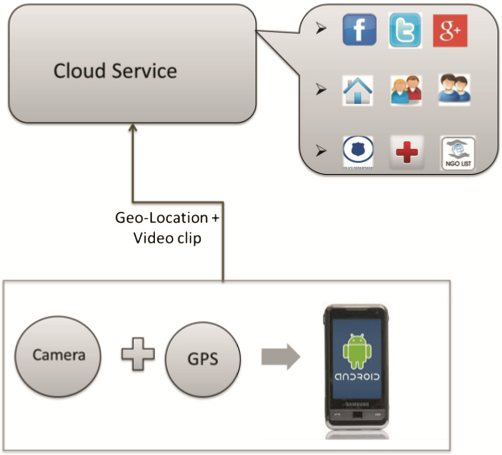


Fig. 8. Report abuse flow diagram

The advantage of the SOS Report Abuse button is that it uses video and GPS features of the mobile. Whenever a SOS or report Abuse button is pressed, the camera of the mobile automatically starts to record a video. In this way the user could convey the exact situation of that area very easily. The video along with GPS location is sent to the cloud for sending alert messages.

5 Conclusion

The ability to gather user's context and determining whether he is in an adversarial situation in real-time could be very useful for many people in combatting physical violence. This is especially useful for the people who travel frequently. In this paper, the authors have proposed an automated context aware application, which could be used for women, children and other vulnerable people to ensure the safety and security. A prototype and a product has been built and being used as a limited pilot testing. With the integration of context aware systems, mobile technology and social networks and by automating the emergency request sending procedure, significantly reduces the time taken by the responders which is crucial to helping the victim. Currently pilot testing is being conducted, the results of which will be published in the future work in addition to extending the features in dedicated wearable systems with a focus on personal safety.

References

1. Delhi Rape Case Incident. Wikipedia (2012). http://en.wikipedia.org/wiki/2012_Delhi_gang_rape_case
2. Dey, A.K.: Providing architectural support for building context-aware applications. Ph.D. thesis, Georgia Institute of Technology (2000)
3. Schilit, B.N., Adams, N.I., Gold, R., Peterson, K., Goldberg, D., Ellis, J.R., Weiser, M.: An overview of the PARCTAB ubiquitous computing experiment. Pers. Commun. IEEE **2**(6), 28–43 (1995)
4. Services. Android Developers. http://developer.android.com/guide/components/services.html
5. Schilit, B.N., Theimer, M.M.: Disseminating active map information to mobile hosts. IEEE Netw. **8**(5), 22–32 (1994)
6. Baldauf, D.S., Rosenberg, M.F.: A survey on context-aware systems. Int. J. Ad Hoc Ubiquit. Comput. **2**(4), 263–277 (2007)
7. Byun, H.E., Cheverst, K.: Utilizing context history to provide dynamic adaptations. Appl. Artif. Intell. **18**, 533–548 (2004)
8. Si, H., Kawahara, Y., Aoyoma, T.: Stochastic approach for creating context aware services based on context histories in smart home. In: Proceedings of Exploiting Context Histories in Smart Environments (ECHISE 2005), pp. 3480–3495 (2005)
9. Randell, C., Muller, H.: Context awareness by analyzing accelerometer data. In: Proceedings of the 4th IEEE International Symposium on Wearable Computers (ISWC 2000), pp. 175–176. IEEE Computer Society, Washington, D.C. (2000)
10. Figo, D., Diniz, P.C., Ferreira, D.R., Cardoso, J.M.P.: Preprocessing techniques for context recognition from accelerometer data. Pers. Ubiquit. Comput. **14**(7), 645–662 (2010)
11. Santos, A.C., Cardoso, J.M.P., Ferreira, D.R., Diniz, P.C.: Mobile context provider for social networking. In: Meersman, R., Herrero, P., Dillon, T. (eds.) OTM 2009. LNCS, vol. 5872, pp. 464–473. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-05290-3_59
12. White, C., Plotnick, L., Kushma, J., Hiltz, S.R., Turoff, M.: An online social network for emergency management. Int. J. Emerg. Manag. **6**, 269–382 (2009)
13. Krishnamoorthy, S., Agrawala, A.: M-urgency: a next generation, context-aware public safety application. In: MobileHCI 2011 Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 647–652. ACM, New York (2011)

14. Circle of 6. Tech 4 Good Inc. (n.d). http://www.circleof6app.com/. Accessed 14 Jan 2017
15. Sentinel. Mindhelix (n.d). http://sentinel.mindhelix.com/. Accessed 02 Jan 2017
16. bSafe. Bipper, Inc. (n.d). http://www.bsafeapp.com/. Accessed 14 Feb 2017
17. Fightback. CanvasM, TechMahindra (n.d). http://www.fightbackmobile.com/welcome. Accessed 14 Jan 2017
18. Accelerometer. Wikipedia (n.d). http://en.wikipedia.org/wiki/Accelerometer. Accessed 26 Jan 2017
19. Google maps Api v3. Google Inc. (n.d). https://developers.google.com/maps/documentation/javascript/. Accessed 19 Jan 2017
20. Amrita University. Amrita Mitra: connecting you to the needed help (n.d). http://personalsafety.in/apss/. Accessed 14 Jan 2017

A Comprehensive Crowd-Sourcing Approach to Urban Flood Management

Ramesh Guntha[1(✉)], Sethuraman Rao[1], Maik Benndorf[2], and Thomas Haenselmann[2]

[1] Amrtia Center for Wireless Networks & Applications (AmritaWNA),
Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham, Amrita University,
Coimbatore, India
{rameshg,sethuramanrao}@am.amrita.edu

[2] Department of Computer Science, University of Applied Sciences Mittweida,
Technikumplatz 17, 09648 Mittweida, Germany
{benndorf,haenselm}@hs-mittweida.de

Abstract. Urban flooding is a common occurrence these days due to many reasons. Providing timely and adequate help to the victims is challenging. Enlisting the citizens to help themselves using their smartphones to provide real-time status updates and ensure timely delivery of needed help is a winning proposition. This paper describes a novel crowd-sourcing approach to urban flood management addressing the inherent challenges using smartphone applications and services that can be deployed by a variety of entities – government agencies, NGOs, social networks, etc. It enables sharing of real time information on the status of flooding, rescue and relief requests and responses, etc. It also collects data from various smartphone sensors in the background which is analyzed and synthesized to track the location and movement of people and to assess the integrity of structures such as bridges. It can also be a valuable resource for future city planning.

Keywords: Urban flood relief · Crowd-sourcing · Smartphone sensors
Sensor fusion · People movement detection · Structural integrity detection

1 Introduction

Urban flooding is a common occurrence across the world. It is estimated that the overall annual cost of floods in Asia alone runs upwards of USD 16 billion [13]. As a result of continuous heavy rains, the city streets get flooded, especially when the planned drainage capacity is insufficient or it has been damaged. The drainage capacity could have been damaged or blocked due to some vegetation growth or due to some construction work leading to improper dumping of debris, etc. Storms, mudslides, etc. could also cause damage to the drainage system. In some cities adjacent to rivers, flooding could also be due to the swollen river. Coastal cities could also experience flooding due to storm surges.

In all these cases, when the flooding occurs, the authorities may not have timely and accurate information and updates on the nature and extent of flooding at various locations, the amount of damage to life and property, the rescue and relief needs of the

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 13–24, 2018.
https://doi.org/10.1007/978-3-319-73423-1_2

people, volunteer help available, etc. They will have to rely on the feedback from the rescue personnel who reach the location or on the calls for help received from the victims. This results in substantial delays in the estimation of damages and the allocation and dispatch of relevant relief and rescue efforts to the affected citizens. This could in turn cause added suffering and result in additional losses to life, health and property. In this paper we propose a solution to gather accurate and timely information through the crowd-sourcing method powered by internet-enabled smartphones and computers. With the help of this real-time information, we believe the people's suffering can be greatly reduced.

Smartphones are ubiquitous in today's world especially in the urban areas. Many people possess more than one smartphone. Use of smartphone apps for various tasks has become a way of life. This behavior presents a unique opportunity to unleash the power of these smartphones and their usage patterns towards solving the problems associated with the way urban flood management is done today. If the citizens of a metro can be motivated to get involved in crowd-sourcing initiatives to help the authorities or other relief agencies in flood management, it has the potential to revolutionize the approach to flood management. The citizens need to realize that they are actually helping themselves by participating in crowd-sourcing. When this realization dawns, they will want to help spontaneously without the need for any incentives. The participation of citizens in crowd-sourcing will enable real-time status updates on flooding and will ensure timely delivery of appropriate rescue and relief supplies to the victims. It will also encourage citizens to participate in the relief efforts by donating their time and money towards it.

This paper presents a novel crowd-sourcing approach to urban flood management. This solution will address the inherent problems in conventional urban flood management with the help of citizens. A suite of smartphone applications and services that can be deployed by a variety of entities – government agencies, NGOs, social networks, etc., is developed. This suite enables sharing of real time information on the status of flooding, rescue and relief requests and responses, etc., among the citizens, rescue personnel and government authorities. This information is also made available to the citizens in real time thereby benefitting them directly. This suite also collects data from various smartphone sensors in the background. This data is analyzed and synthesized to track the location and movement of people during a flooding event and also to assess the integrity and load carrying capacity of bridges. This is done by collecting data from the smartphones of people walking and driving on the bridges. The safety information of the bridges is used to calculate various safe evacuation routes and these routes are presented through a map interface. In addition to the citizens as end users, the app-suite provides detailed, relevant and up-to-date information to the relief and rescue providers. The administrative authorities can also use the data collected from this suite for matching demand and supply, to generate custom reports and as a valuable resource for future city planning. By conducting root cause analysis of issues based on the data collected, they can identify suitable remedies towards improving drainage systems and other infrastructure to mitigate future floods.

2 Related Work

Crowd-sourcing has been attempted in the past fairly successfully towards solving specific problems related to flood management. In the framework of the FP7 SPACE Project GEO-PICTURES, AnsuR and United Nations (UNOSAT) collaborated on using a smartphone App for crowd-sourcing geo-referenced insitu images for the purpose of improving flood assessment from Radar EO Images. This was successfully deployed during the 2001 monsoon season in Thailand when severe flooding occurred (The European Association of Remote Sensing Companies 2012) [1].

During the 2012 floods in Philippines [2] and Beijing, China [3], successful crowd-sourcing initiatives were launched. In Philippines, the initiative was to track the places and people in need of help the most using a spreadsheet in Google docs. Google Person Finder app was also used. The participation from the public was robust and enthusiastic both for using the spreadsheet for tracking as well as for updating the spreadsheet based on their knowledge. In Beijing, China, users of the Guokr.com social network launched a campaign to create a live crisis map of the flood's impact using Google Maps. Up-to-date real time status information was generated by crowd-sourcing hours before similar information was released by the government agencies. The success of the above initiatives bears clear testimony to the power and feasibility of crowd-sourcing as an urban flood management tool.

Our goal is to take this to the next level by providing an integrated and comprehensive suite of flood management applications and services based on the smartphone. This suite will enable the end users to share their knowledge about the flood and the victims. The suite will also provide useful and current information about the flood situation, rescue and relief service needs and availability information to the public. The suite will also provide detailed and relevant information to the rescue personnel on the ground to help in their rescue and relief operations. Additionally, it will empower the administrative officials to quickly and efficiently match the demand and supply for rescue and relief, and provide searching and filtering options to generate custom reports and tables. It will also help in the city planning exercise for the future by helping the officials conduct root cause analysis of various problems encountered based on the data collected.

Jha et al. [4] provide operational guidance to government policy makers, NGOs and technical specialists on how to manage the risk of floods in a rapidly transforming urban environment and changeable climate. Reference [5] is an article in the Intellecap publication, Searchlight discusses the issues posed by urban flooding in India. The technique of data fusion from smartphone sensors has been used in several applications [6, 8, 10].

3 Research Challenges

There are several challenges in urban flood management and the application of crowd-sourcing for urban flood monitoring. The families living in cities tend to be nuclear and isolated with very little social interaction with the neighbors. They also tend to live in multi-storied multi-tenant buildings consisting of hundreds of housing units. Identifying the location of trapped victims who need help and identifying the type and quantum of help needed becomes a big challenge in such situations. In addition, there may be buildings in the city that are poorly planned and constructed in the low-lying areas which were water bodies at one time. The slums and the other underprivileged populations in the city tend to live in such areas. The roads leading to such areas may also be narrow and very poorly maintained. However, the price points at which smartphones are available today have made them affordable to practically all strata of society. Therefore, introduction of smartphone based crowd-sourcing of flood management and relief will mitigate the challenges mentioned above to a great extent.

In addition, there are certain challenges that arise when crowd-sourcing is applied. The veracity and reliability of the data obtained needs to be ascertained. Spurious and malformed data may be supplied simply due to callousness on the part of participants or with specific malicious intent by some rogue elements in the society. Such bad data needs to be identified and weeded out. The application can dynamically build and maintain the trust profiles of end users. The users can be rated based on their trust profiles and those below a threshold can be discarded. There is also the chance of inadvertent duplication of rescue or relief requests coming in either from the same source or from multiple sources. We need to have suitable mechanisms to identify and eliminate such duplicates.

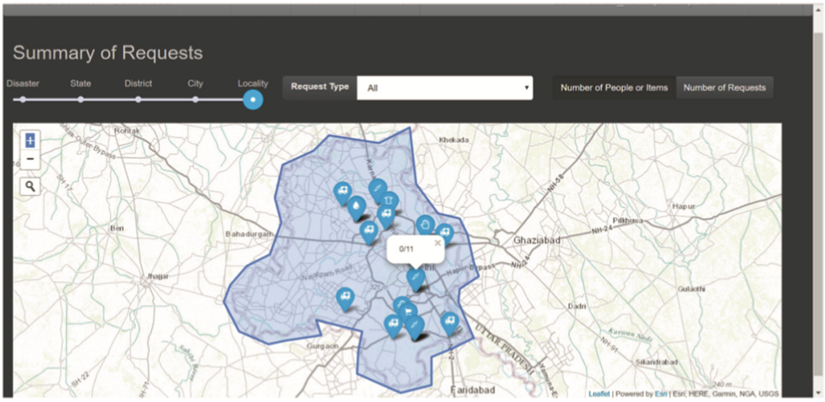
Motivating the end users to participate actively in crowd-sourcing is also a challenge. General display of apathy by the citizens towards the call for participation is a likely scenario. This needs to be handled by raising the awareness about the benefits that will accrue to the society which will directly improve their own quality of life. The younger generation in schools and colleges is the ideal target audience for creating this awareness by running campaigns. In addition, incentivizing the end users by providing free data or SMS service or discount coupons at shopping malls, etc., are potential ways to improve participation.

There is also the likelihood of the flooding affecting the functioning and stability of the communication network in the city. This is more likely in a rural scenario than in an urban scenario. When this happens, ad hoc networks may be provisioned to provide alternate channels of communication to the residents. Flooding may also affect the availability of power and the users may not be able to charge their phones. Vending or providing battery packs or other sources of power will alleviate this situation.

4 Our Approach

We are developing mobile and web applications targeted to citizens, rescuers and administrators. These applications are supported by a high performance, scalable, fault tolerant server architecture. The mobile and web applications feature light-weight, high performance, and simple-to-use interactive graphical user interfaces in multiple languages. The registration process for the citizens is kept simple and quick with only email verification. Whereas for the rescuers it will require some more background checking. For the administrators, the system allows more flexible/configurable registration process which suits the respective authorities who use the system.

The mobile application allows citizens & rescuers to upload information, request for and respond with rescue and relief. In addition the mobile application also captures the various sensor data automatically and sends to the server. All the information and requests are geo-tagged and time-stamped for accurate analysis and representation. The web application, in-addition to above features, also has reporting interfaces for summary and data visualization and analysis. A sample screenshot of requests from a locality overlaid on a map is shown in Fig. 1.


Summary of Requests

Request Type: All

Fig. 1. A sample screenshot from the web app

The users can upload images and videos of any flood and hazard situation. These images and videos are automatically analyzed using machine learning techniques and computer vision algorithms to determine the extent of flooding, damages to property and lives, and various hazard situations like fallen trees, washed out roads, collapsed electric poles, etc. Users also can fill some simple and intuitive forms explaining the situation. Users are also allowed to enter free-form text. The free-form text is analyzed using natural language processing techniques to extract relevant information. All such information about the extent and depth of flooding, damages to property and lives is stored with geo-tag and time stamp information. The information is further summarized and presented in various reports.

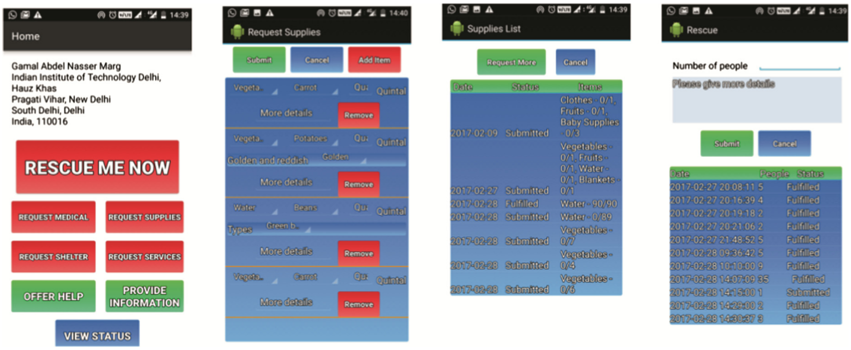
The mobile application also automatically captures and sends various sensor data such as data from GPS, accelerometer, light sensor, etc., to the server. This data is analyzed to infer location of people and their movement characteristics which is in turn used to estimate the count of stranded people.

During the flooding situation, people might use various bridges in the city to evacuate to safe places. But because of the flooding, the bridge's structural integrity might be affected. There is a critical need to ensure that the bridges are safe to use during the evacuation process. Our approach is to extend the mobile application such that the built-in movement sensors of a smartphone can capture vibrations of the bridge which naturally emerge during its usage. The acceleration sensors in most off-the-shelf smartphones have been proven to be sensitive enough to capture vibrations caused by moving vehicles or even by pedestrians. These vibration patterns can be used to draw conclusions about the remaining integrity of the construction. Eventually this data can be used to suggest to which degree the bridge can still be used.

Safe evacuation routes are automatically determined based on the integrity of bridges thus calculated, along with various hazard situations on the roads. This information will be continuously revised based on the latest available information. This work is being done by one of the partners in the consortium.

Apart from sharing the information related to flooding and damages, users can request or offer help through these applications. Users can request for rescue of stranded people and animals, request for relief such as food, clothing, water, blankets and other essentials, and services like medical help, power supply, water supply, cleanup of damaged property, fallen trees, fallen electric poles, dead bodies of animals and people. Citizens can request for or offer shelters. The status of each shelter such as free capacity, timings, restrictions, etc., will be available on the summary page. These requests for help are also geo-tagged and time-stamped. This allows for aggregation of the amount of help needed by various regions, so that the authorities can properly allocate and dispatch efforts and resources promptly and accurately. This ensures that citizens receive adequate help in a timely manner.

The application also allows citizens to pledge and respond with help or supplies. All the responses for help by either authorities or citizens will be entered into the application, allowing the requesters to see that help is on its way. This can lead to rapid and efficient micro-level matching of demand and supply in real time which will in turn result in a highly responsive and effective system of mitigation and relief for the victims of flooding. Figure 2 shows some sample screens from the mobile app.



Gamal Abdel Nasser Marg
Indian Institute of Technology Delhi,
Hauz Khas
Pragati Vihar, New Delhi
South Delhi, Delhi
India, 110016

Vegeta.. Carrot Qu Quintal
Vegeta.. Potatoes Qu Quintal
Golden and reddish Golden
Water Beans Qu Quintal
Types Green b..
Vegeta.. Carrot Qu Quintal

Date	Status	Items
2017-02-09	Submitted	Clothes - 0/1, Fruits - 0/1, Baby Supplies - 0/3
2017-02-27	Submitted	Vegetables - 0/1, Fruits - 0/1, Water - 0/1, Blankets - 0/1
2017-02-28	Fulfilled	Water - 90/90
2017-02-28	Submitted	Water - 0/89
2017-02-28	Submitted	Vegetables - 0/7
2017-02-28	Submitted	Vegetables - 0/4
2017-02-28	Submitted	Vegetables - 0/6

Number of people
Please give more details

Date	People	Status
2017-02-27 20:08:11	5	Fulfilled
2017-02-27 20:16:39	4	Fulfilled
2017-02-27 20:19:18	2	Fulfilled
2017-02-27 20:21:06	2	Fulfilled
2017-02-27 21:48:52	5	Fulfilled
2017-02-28 09:36:42	5	Fulfilled
2017-02-28 10:10:00	9	Fulfilled
2017-02-28 14:07:09	35	Fulfilled
2017-02-28 14:15:00	1	Submitted
2017-02-28 14:29:00	2	Fulfilled
2017-02-28 14:30:37	3	Fulfilled

Fig. 2. Sample screenshots from the mobile app

The citizens can view the latest flooding information such as extent of flooding, help required, help being provided, hazards, integrity of bridges, suggested evacuation routes, along with the pictures sent by people. This information page will be automatically updated in real-time. All the information will be presented both on a map and as spreadsheets. The authorities have more interactive data visualization tools to slice and dice the information along the spatial and temporal grains for better relief planning. These data tools are also quite useful long after the flooding has abated, in future planning of city infrastructure.

All the collected data related to flooding, help requests and responses are geo-tagged and time-stamped. The images and text are analyzed using machine learning techniques and natural language processing techniques to extract information about flooding, damages and hazards. All the information is correlated with multiple sources when possible to ensure accuracy. The duplicate information is removed using geo-location and by correlating with multiple sources. Where possible, the call-center process can be used to confirm the exact rescue and relief requirements. The data is then, on a periodic basis, aggregated into multiple levels of summaries along the spatial and temporal granularities. This data also can be used to automatically match available resources to the help requests in an optimal manner. This data can be used to plan for and provide required help, targeting it to the areas where it is needed the most. This allows for historical and geographical analysis of flooding, damages to infrastructure, help and shelters required and provided, etc. This data can be a valuable resource for planning future infrastructure work in the cities to avoid further flooding scenarios.

The system provides both mobile and web interfaces, supported by scalable, high performance, event-driven, and robust server architecture (Fig. 3). The entire system is based on open source packages, which will reduce the deployment costs considerably. The data is exchanged with server in the light-weight and universal JSON format. All the server commands are exposed in the form of APIs, which can be accessed by these mobile and web interfaces. The same APIs can also be accessed by third party systems as well. All the interfaces are designed to be intuitive, light weight, highly interactive and responsive to suit the screen size and would let users achieve what they want with minimum clicks.

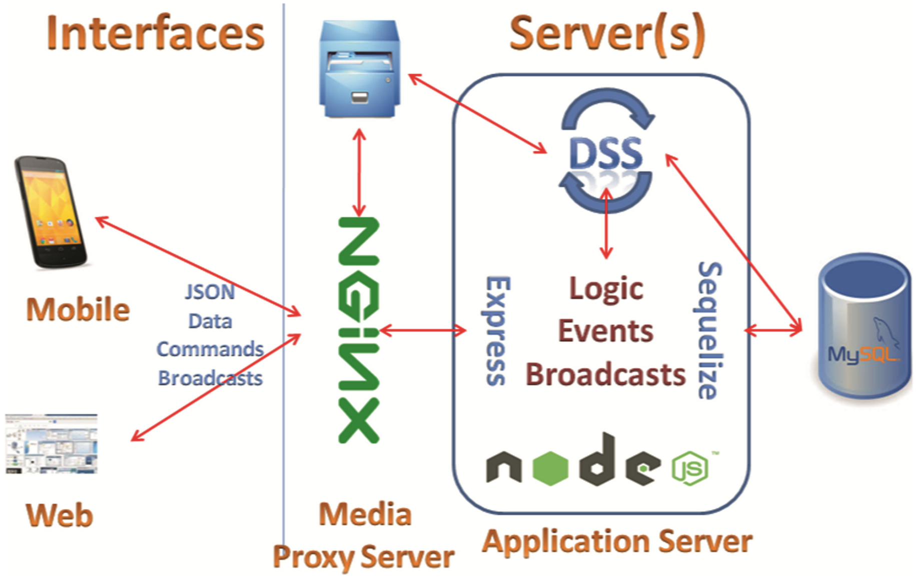


Fig. 3. System architecture

On the server side we use a high performance web server called NGINX, which handles all the media such as images, audio and video. The web server is backed by the Application server powered by high performance, highly flexible, extensible, robust, and heavily adopted NodeJS server. We use Express package as command router inside NodeJS, and Sequelize as the ORM (Object Relational Mapping) package to interact with MySQL database.

The code which comprises the application logic is written in a re-usable modular fashion. The logic consists of periodic event code to run any scheduled tasks and broadcasting code to push any relevant updates to user interfaces based on what's being viewed by a given user at that time. That way, the user would get updates on only what he/she is currently viewing, thereby reducing bandwidth and resource consumption.

The DSS module is the brains of the system, which employs several machine learning and natural language processing algorithms to analyze image and text data. It also ensures data accuracy by correlating data from multiple sources. It employs data mining techniques to predict the extent and duration of floods and finally aggregates data into multiple spatial and temporal granularities so that it can power the data visualization interfaces.

The data flow diagram (Fig. 4) shows how and what data is exchanged between various users, processes and data stores. All the users like citizens, rescuers and administrators can upload images, audio and video files, which are handled by NGINX server and these media files are stored in the file system. The users can also query for latest summary of the flooding event which has information about extent of flooding, damages to life and property, hazards, status of rescue and relief efforts etc. All the queries are handled by NodeJS application server and the application logic hosted on NodeJS, which in turn queries the data from MySQL database. The users can submit requests for rescue and relief and they can also respond with offer of efforts and resources towards rescue and relief. All these requests and responses are handled by NodeJS application server and the data is stored in MySQL database. In addition, the smartphone users automatically transmit various sensor data. All this data is also stored in MySQL database through NodeJS application logic. Apart from these data exchanges, the administrators can manage the rescue and relief efforts and can also query for various statistical data based on dates, geo locations, etc., to analyze the various aspects of flooding. All these data requests are handled by NodeJS and MySQL database. All the users would get automatic broadcasts whenever any data related to what they are currently viewing in their applications or data related to requests and responses they submitted is updated. All these broadcasts are determined and triggered by NodeJS application server.

The Decision support system triggers various sub processes either periodically or based on any data events. The media analyzer process would analyze images, audio, video and free-form text to extract information related to extent of flooding, hazards, rescue and relief requests and responses, whenever such information is submitted by users. The data aggregation process summarizes all the statistical data related to flooding and rescue and relief operations into various spatial and temporal granularities, and these summaries are provided to users through summary pages and data visualization tools. The data cleanup process removes duplicate and inaccurate data by confirming with multiple data sources and/or by confirming with the users through call-center process.

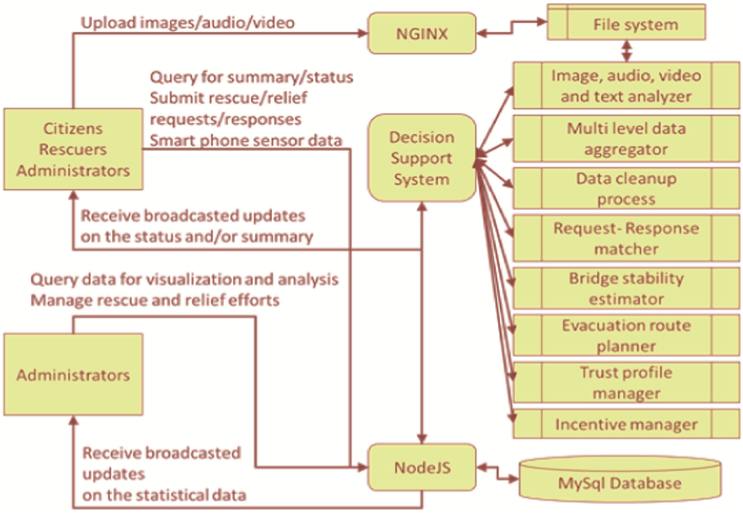


Fig. 4. Data flow diagram

The matching process optimally matches the offers for rescue and relief with requests for rescue and relief on the basis of location, urgency, and any other factors which are configured in the system. The Bridge stability estimator estimates the structural integrity of the bridges based on the data collected by user's smartphone sensors and a few other factors not discussed here. The Evacuation route planner uses the bridge stability estimations and hazard conditions analyzed by other modules to suggest several safe evacuation routes out of the flood zones to various shelters. The trust profile manager updates the trust score of the user based on authenticity of information provided by user. The incentive manager updates incentive score of the users based on the number of authentic and unique information provided and based on the rescue and relief offerings. Except the media analyzer process, all the other sub processes are triggered on a periodic basis. All these sub processes query data from and update the processed data to the MySQL database through NodeJS application logic. Whenever any sub process updates the data in the database, the NodeJS application logic determines the relevance of this data to the currently logged in users and broadcasts these updates automatically to their interfaces so users can view these updates in real-time.

4.1 Deployment Models

We envisage the following deployment models which allow for faster deployment and also provide flexibility if the users choose so. We offer it to the world as Software as a Service (SaaS) deployed in the cloud; all that the users need to do is to register and start using it. When a disaster occurs, the relevant authorities would create a disaster event in the system and assign relevant geographical regions to the disaster and all the users who live in those areas can start sharing information related to the disaster. Since it is a SaaS implementation, many people in different regions would be using the same system simultaneously for different flood scenarios if needed. All the information across many such events is stored in the same logical location, which becomes a valuable resource for intense data analysis and planning. Apart from SaaS, we also provide enterprise deployment options for required Governments or NGOs. Both types of deployment come with detailed APIs, which allow for any other systems to interact and extract data and present or use it in multiple ways.

A third model would consist of two parallel deployments, one as SaaS and one as an enterprise deployment. While the enterprise version will get updated based on authentic information by an administrative body such as a Government agency or an NGO, the SaaS version can be updated by the general public. In this model, the SaaS version will have faster updates in near real time. The administrative agency can use this information to verify the authenticity and update the enterprise version with authentic information based on their finding in addition to initiating the necessary action. This model was suggested by an officer of the Indian National Disaster Management Agency (NDMA) when we got our prototype reviewed by him.

4.2 Current Status and Adoption Strategy

We are in the process of releasing a beta version of this system. Our beta users will consist of Government agencies such as NDMA as well as NGOs. We also plan to run a mock drill within our University campus to stress test the system under heavy load. Stress testing using canned test cases has already been done successfully.

The computer vision algorithms are being tested and verified using some flooded pictures obtained from Kulmbach, a town in Germany chosen by our German partners for prototyping their work. These pictures have the water level marked along with the dimensions of bridges which are used as reference objects. The water level arrived at using computer vision are compared to the measured values.

Usually crowd-sourcing models could suffer from lack of adoption from users as there is no tangible incentive to users. However, once it is shown to work well in a couple of instances, adoption for subsequent deployments becomes that much easier. We plan to use the following approach to help in the adoption. Amrita University's parent NGO, M.A. Math has done relief work during flooding events in many cities in the past through its voluntary organization, "Embracing the World" [12]. We plan to choose a couple of such cities for our initial deployment as we can leverage the connections M.A. Math has with the local authorities and people in those cities. We will then run advertising campaigns in the local schools and colleges about this application and how they can benefit and serve their community at the same time by using it. We will work with the local data providers to provide some form of credit for data usage by this application so that users won't incur data charges for sharing images, videos etc. We will also introduce some incentive points for every unique piece of information shared by the users. These incentive points could be exchanged for shopping vouchers or free mobile talk time.

5 Conclusion

The widespread adoption of smartphones and their common usage patterns among the urban population can be very effectively leveraged to improve the efficiency of urban flood management and provide relief in near real time to the victims. We take a comprehensive approach with applications that target multiple user types – citizens, relief providers and administrative authorities. It allows them all to provide relevant and useful information towards flood management while at the same time getting the latest status information from the system. The system is used to both request and volunteer for relief and rescue operations thereby providing micro-level matching of supply and demand in real time. In addition, sensor data from smartphones is collected automatically and synthesized to estimate locations and mobility patterns of victims and also to estimate the stability of bridges. Safe evacuation routes are calculated based on all the available information and displayed on a map.

By providing both a mobile app and a web app, we also cater to operations that require more real estate on the screen, especially various types of maps and reports that may need to be generated by the administrative authorities. A scalable and cost-effective architecture based on open source software packages, light-weight communication protocols to make the application highly responsive and a simple and intuitive graphical user interface to ensure ease of use are some of the salient features of the solution. We will support both SaaS and enterprise deployment models. We have also outlined our adoption strategy that addresses the inherent challenges in crowd-sourcing. Overall, our proposed comprehensive approach to crowd-sourcing of urban flood management has tremendous potential to revolutionize the conventional methods of flood management.

This is presented as an urban flood management system as the smartphones are widely used and internet is ubiquitous in urban areas. This system is equally effective in rural areas as long as smartphones or computers with internet access are available.

Acknowledgments. This project is part of an Indo-German collaborative research program in civil security and is partly funded by Department of Science and Technology, Government of India and Federal Ministry of Education and Research, Germany. The project name is "Vulnerability of Transportation Structures, Warning and Evacuation in Case of Major Inland Flooding" with the acronym, FloodEvac. Several German and Indian educational institutions and relief agencies are involved in this program.

We would also like to acknowledge the support and encouragement from Amrita University, especially the Chancellor, Mata Amritanandamayi Devi, popularly known as Amma, for this project.

References

1. The European Association of Remote Sensing Companies (EARSC) Newsletter: Validating Space Observations for Flooding with Crowd-sourcing In-Situ Observations by ANSUR (2012). http://eomag.eu/articles/1856/validating-space-observations-for-flooding-withcrowd-sourcing-in-situ-observations-by-ansur

2. iRevolution - From innovation to Revolution: Crowd-sourcing Crisis Response Following Philippine Floods (2012). http://irevolution.net/2012/08/08/crowd-sourcing-philippinefloods/
3. iRevolution - From innovation to Revolution: Crowd-sourcing a Crisis Map of the Beijing Floods: Volunteers vs Government (2012). http://irevolution.net/2012/08/01/crisis-mapbeijing-floods/
4. Jha, A.K., Bloch, R., Lamond, J.: Cities and Flooding - A Guide to Integrated Urban Flood Risk Management for the 21st Century. The World Bank (2011). http://www.gfdrr.org/sites/gfdrr.org/files/urbanfloods/pdf/Cities%20and%20Flooding%20Guidebook.pdf
5. Carr, C.: Environmental Degradation and Urban Flooding. Searchlight South Asia (2012). http://urbanpoverty.intellecap.com/?p=472
6. Haenselmann, T., et al.: Scriptable sensor network based home-automation. Embedded and Ubiquitous Computing-EUC 2007, Taipei, Taiwan (2007)
7. Haenselmann, T., King, T., Effelsbeg, W., Fuchs, M.: Skriptbasierte drahtlose Gebäudeautomation mit Sensornetzen. Embedded and Ubiquitous Computing-EUC 2007, Praxis der Informationsverarbeitung und Kommunikation, vol. 30. Jahrgang (2007)
8. Bicocchi, N., Castelli, G., Mamei, M., Zambonelli, F.: Improving situation recognition via commonsense sensor fusion. In: DEXA 2011 - 22nd International Conference on Database and Expert Systems Applications, Toulouse, France, August 2011
9. Hristidis, V., Chen, S., Li, T., Deng, Y.: Survey of data management and analysis in disaster situations. J. Syst. Softw. **83**, 1701–1714 (2010). Elsevier
10. Jotshia, A., Gongb, Q., Battac, R.: Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. Socio-Econ. Plan. Sci. **43**(1), 1–24 (2009). Elsevier
11. Laituri, M., Kodrich, K.: On line disaster response community: people as sensors of high magnitude disasters using internet GIS. Sensors **8**(5), 3037–3055 (2008). Open Access Journal
12. Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. J. Mol. Biol. **147**, 195–197 (1981). Embracing the World - by Mata Amritanandamayi Math. www.embracingtheworld.org/
13. http://floodlist.com/asia/report-asia-pacific-region-floods-cost-us16-billion-2014

Aggregation Using the Concept of Dynamic-Sized Data Packet for Effective Energy Saving in Wireless Sensor Network

Smitha N. Pai[1](✉), H. S. Mruthyunjaya[2], Aparna Nayak[1], and A. Smitha[1]

[1] Department of Information and Communication Technology, M.I.T., Manipal University, Manipal, India
{smitha.pai,aparna.nayak,smitha.a}@manipal.edu

[2] Department of Electronics and Communication Technology, M.I.T., Manipal University, Manipal, India
mruthyu.hs@manipal.edu

Abstract. Data aggregation process can extract relevant information from raw data obtained from various sources using certain mathematical functions. Aggregation reduces the transmission of redundant data. A protocol named DP_AODV is implemented in this paper. Aggregator nodes (cluster head) are identified using the positional information. Routes are established between these aggregator nodes using efficient routing techniques. Data is aggregated along the path to the destination conserving additional energy. The aggregation process involves averaging the data if it is within the threshold range, else, only the data part along with the positional information is appended to the payload. Size of the Data packet varies dynamically based on the number of nodes having co-related data at that particular instance. The common header occupies a substantial part of the packet. Avoiding multiple transmission of common part of the header saves energy.

Keywords: Aggregation · Wireless sensor network · Energy · Data packet

1 Introduction

Most applications using sensors are used to monitor, measure continuously varying physical parameter like humidity, temperature, light intensity, etc. Sensors require power to run the electronic circuitry. The source of power can be from the battery, solar panel or electrical grid lines. Applications like irrigation in agriculture need batteries running for one crop season of nearly six months. In the agricultural field, the moisture content in the soil, humidity and temperature is measured continuously. The main consumption of energy in this network is during transmission and reception of data. Measured data has to be sent from the location where it has sensed (source) to the main collection center called the base station (sink). If the distance between the source and sink is larger than the transmission range of the sensor, data is sent using multiple hops. Efficient route between source and sink is essential. Energy consumption is further reduced by aggregating data at some strategic location. DP_AODV, a Dynamic-Sized

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 25–36, 2018.
https://doi.org/10.1007/978-3-319-73423-1_3

Data Packet Protocol is proposed in the current paper and it addresses all these issues. The ns2 simulator is used to strengthen the result. Co-related data is aggregated at the aggregator node based on the threshold value. If the difference in the collected data is beyond the threshold value, the data is appended to the payload part of the existing packet. The second section that follows substantiates the research work that is carried out in the relevant area. The third section addresses the methodology used to attain the required result. This is followed by the experimental results to reinforce the idea. Elaboration on the result follows. The methodology incorporated in this paper involves the process of clustering, routing, time synchronization between various clusters and packet handling.

2 Related Work

The current work is based on COMMON-Sense, a project which is still going-on at Pavagadh district, India [1]. The temperature and humidity of the surrounding area along with the moisture content in the soil is measured. In the paper emphasis is given to conserving energy at the MAC layer. Finally, data is sent via the gateway to the monitoring station. This paper [1], forms the basis for selecting the environmental parameters for the current project.

Literature survey is carried out along the lines of routing, co-relation, aggregation, data handling, energy consumption, connectivity, and coverage. Literature [2, 3], addresses intensively the various Routing techniques in wireless sensor networks. A delay-aware network structure for WSNs with in-network data fusion is proposed [4]. Study on optimized transmission and fusion cost is emphasized in work [5]. Spatial co-relation awareness in the dynamic and scalable tree is available in work [6]. Best response dynamics to local data is discussed in the literature [7]. Adaptive clustering without relying on exact sensor location information is dealt with in reference [8]. Whole integrated network situation for head node selection is studied in the paper [9]. Not only balancing the energy expenditure among sensors but also extending the network lifetime by equal usage of multiple optimal intermediate routers is addressed in the literature [10]. Route based on less time to reach the base station with aggregation taking place at the first level of the tree is dealt with [11]. Hierarchical Agglomerative Clustering where in, repeated merging of small clusters are carried out until all clusters scale to the satisfied threshold is accounted for [12]. Network division into unequal sized grid-shaped cluster is handled [13]. In [14], the grid whose cluster head consumes more energy takes part in cluster head rotation, shares energy load, balancing the energy dissipation. Some nodes send data in the form of a chain to the cluster head which again aggregates data at the sink. A method of clustering and sending data, based on the prediction of the value obtained is emphasized [15]. Data aggregation is carried along the spokes of the wheel [16]. Ring based data gathering is addressed [17]. In [18], the transmission range and the k-neighbor topology control strategy is used to prolong the network life. In [19], the network coverage and connectivity are achieved by dividing the sensors into sets of equivalent nodes and finding a better path. Paper [20] addresses that the connectivity is maintained even when some nodes have failed by dynamically adjusting the transmission power of the nodes. In [21], layered space-time directed graph is used to reduce the energy consumption and delay in the network. Reduction of energy consumption and data collection delay increases the accuracy [22]. In [23], the optimal size of the data packet is computed. This literature helps in identifying the maximum size of the data packet that can be supported in the network. In [24], the connectivity, is expressed by the probability that a node lies on a path to the sink, as a function of the probability that adjacent cells in a grid are connected. Survey paper [25], shows various ways of routing based on communication model, reliability, topology, and network structure for energy savings. Most of the literature surveyed gives the only emphasis on one or two parameters like routing, aggregation, low energy consumption, etc. In the current paper, actual value of the data is used for computation (though ns2 does not support data handling). Data transmission is efficiently handled by using techniques like routing and aggregation. The maximum packet size of 1024 bytes can be used for transmission.

3 Clustering and Routing

Clustering a set of sensor nodes help in localizing the co-related data in a randomly deployed network. Within each cluster, every node is in the hearing range of each other. Figure 1 depicts a network of 75 nodes. Selection of aggregator node is based on the positional information. Set of nodes within the cluster associate itself with the cluster head. Figure 2 shows the process of routing within and across the cluster. The green path shows the route followed to send data across the aggregator node. Blue lines shows the path to send data within the cluster. The number written in red, close to each of the node specifies the maximum number of hops that are required for any node to reach that aggregator node. Algorithm 1 shows the process of selecting the cluster head and Algorithm 2 to find a neighbor within the cluster.

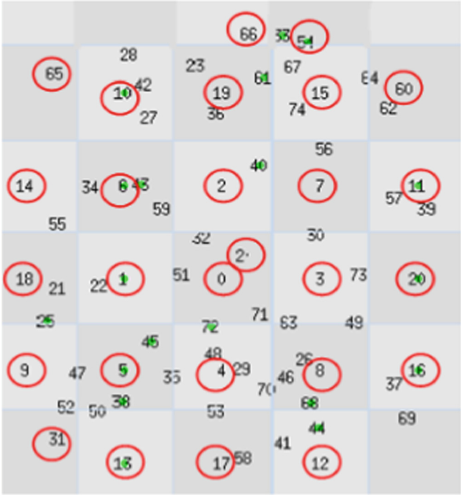


Fig. 1. Selection of aggregator node

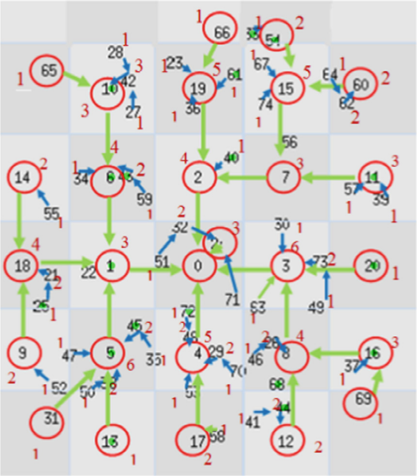


Fig. 2. Routing within and across clusters (Color figure online)

The Algorithm 1 finds the node closest to the center of a given rectangular cluster area. This node is assigned the status of cluster head.

Algorithm 1: To identify the cluster head in each cluster

```
Input: N is the Node in the current cluster C
       Xi, Yi are the Coordinate of the node Ni in C
       Xc, Yc are the Coordinates of the topological
       Centre of the cluster C
Output: Aggregator Node H
begin
 H := Ø; MinDist := ∞; m := ∞; i:= 0
 While (Ni exists) repeat
     MinDist := DistanceBetween ((Xc, Yc), (Xi, Yi);
     If   (m > MinDist) m := MinDist;    H:= Ni
     i=i+1
 EndWhile.
End.
```

In Algorithm 2, every node tries to find out which node is closest to itself within the given cluster. These nodes are associated with each other such that the node which sends the data is far from the center of the cluster and the one which receives data is closer to the center.

Algorithm 2: To find the neighbor during the routing process within the cluster.

```
Input: Ni is the Node id for Node i
       Na is the aggregator node (cluster head)
       Distia is the Distance between node Ni and Na
       Neighbor Nij  // Neighbor of i is j
Output: Neighbor N
Begin
  Distia: = ∞; Distja:= ∞; i:= 0; j:=1;
  For all Neighbor j, Vj= 0 // Node j is not visited
  If (Ni) N← Nia // If Ni exists Neighbor N of i is a
  While (Nj & !Vj) repeat
  If (Distia < Distij))
          N← Nia         // Neighbor N of i is j
    Else
          N← Nji         //Neighbor N of Nj is Ni
   End If
   Vj=1;
   j:=j+1
  End While
End.
```

4 Time Synchronization Among Nodes to Send Data to Next Node

Synchronization is essential to carry out lose-less in-time data transmission between clusters. Algorithm 3 computes the waiting period that is required before forwarding the packet to the next hop neighbor. Figure 3 shows how many hop counts are required to reach the destination from any node in the network.

```
Algorithm 3: Waiting period and Elapsed Time
computation for forwarding of packets at any node N
Input: H is the Hop count to reach the current node from
the location where it has sensed the data
M :=1   // Hop count to itself set to 1
N is the No of nodes transferring data to that node
Output: Waiting period T
begin
   M:= Max of (H) from all neighbors    //for node N
   T:=The Time required to receive data from one hop
                neighbor x M
   ElapsedTime:= 0
   InitialTime:= CurrentMeasuredTime()
   While (M ≠ 1 && T > ElapsedTime) repeat
         If (packet received)
                M:= M-1
         EndIf
         ElapsedTime:=CurrentMeasuredTime()-InitialTime
     EndWhile
  Forward the Packet
 End.
```

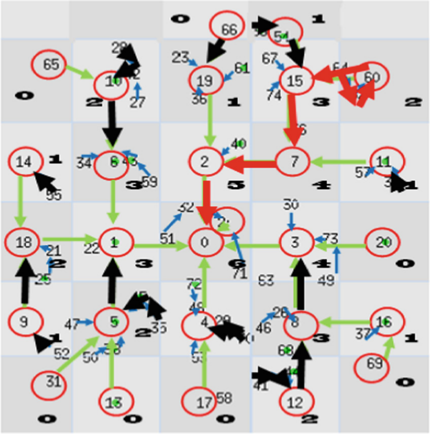


Fig. 3. Hop count computation to reach sink from any node in the network (Color figure online)

The black lines in the diagram represents how many hops are required to reach the base station from any of the node along its aggregating path. The red line represents the maximum path that is being taken to reach by any of the node in the whole network. In the current example this path length is six.

All the nodes sense the data and this information is forwarded to its neighbor. In order to carry out the aggregation process, all the data have to be available at the same time at the aggregator node. Some nodes take longer path (hops) to reach the aggregating cluster head nodes. In order to achieve the time synchronization, the path length for any node's data to reach the aggregator node is computed before-hand. At each aggregating location, the maximum among all the path length, for any of the nodes sending data to that node is computed. The waiting period is computed based on this maximum path length. The number of nodes which are aggregating their data at each aggregator is computed beforehand. As each data is received at the aggregator node the counter is decremented by one. As soon as the counter becomes one, data is sent to the next node without waiting. The Algorithm 3 computes the maximum hop counts required to reach the aggregator cluster head.

5 Aggregation of the Data at the Payload

Algorithm 4 is used to aggregate the data in the packet. If there is co-relation between data in the packet is within the threshold value (0.001 in this paper) then the data is aggregated else data is added to the payload along with the positional information. The choice of 0.001 is as per user requirement. The number of nodes contributing to the aggregation is based on this value. Smaller the value, smaller number of nodes will contribute to the aggregation. In a scenario like agriculture, the relative difference in the moisture content of the soil is low. Hence large area can be covered in one go.

Algorithm 4: Packet creation and forwarding towards Aggregator Node.

```
Input: D is the Data
       Ni is the Node I; Na the aggregator node a
Output: Packet P forwarded to next aggregator node
Count is the No of nodes whose data matches with the
current node's data computed so far.
begin
    Count:= 1
    Dp:= ∅; //Previous nodes data
    Dc:= Current node data
    StartTime := 0
While Ni ≠ Na repeat
ElapsedTime:= CurrentTime - StartTime
WaitTime := Max no of nodes sending data to this node
            multiplied by the time required for one hop
   For each Dp within the cluster
     If   (ElapsedTime > WaitTime)
         If | Dp - Dc | <= Threshold value
```

```
// computes average of the similar data computed so far
           Dp← (Dp x Count + Dc) / (Count +1)
           Count←Count +1
         Else
            Dp= Dp followed by Dc
 //New unexisting data which is beyond the threshold
             Count ←1
            EndIf
       EndIf
   EndForEach
 // Update in the packet format the values of Dp and the
 count corresponding to that Dp
 P=CMN Header appended with Dp, Count and current time
 Packet P forwarded to next hop node
 Endwhile
 End.   //The packet P available at the aggregator node
```

The algorithm for aggregation involves sensing the data and forwarding it to its next hop neighbor. At each of the nodes, there is a waiting period before aggregating and forwarding the data. This time is dependent on any of the nodes with a longest path (hops) which is incident at the aggregator node. After the waiting period all the data which have arrived are analyzed for co-relation. If there is a difference of less than 0.001 cm, then the average of the data which is incident at this node is appended to the current packet. Information on number of nodes having the similar data is available in the packet. Other data which is not co-related is simply appended to the current packet. The size of the data packet varies dynamically based on the co-related value. If the co-relation is large, smaller packet size is sent.

6 Simulation

Ns2 is used for carrying out the simulation. The agricultural application is taken up for designing the parameters [26]. The soil moisture content is noted once in every 5 min which is the input data to the simulator. The rate at which data is acquired is based on the rate in which the data changes at a specific location in the field. If it is raining, (or if the field is getting irrigated) then the moisture content value changes rapidly. During frequent changes the readings are taken often. One reading per day is taken during dry weather. Table 1 shows the simulation parameters that is used for the current work.

Table 2 shows the readings as obtained at the sink. *Time Now* specifies the time interval at which the data packet is obtained at the sink. *Cnt* refers to the number of nodes having *that specific* co-related value of *Data*. *Id* refers to the node which first sensed *that* value of data at the interval *Data time*. Data which was not able to synchronize within the specified time arrive late.

The first line in the Table 2 indicates that data is received at the sink node at the time interval 8.30 ms. The data at each of the nodes were sensed at the time 7.50 ms. The node with Id 51 has data of 2.094 cm. The count (Cnt) of *one* signifies only one neighbor exist with that value. This data is appended to the packet containing data obtained by node with id 32 and data 2.096 cm. This is follows data with Id 71 with data 2.098 cm. Here the count value is *two* signifying that there are two nodes which are close by and both are having their readings which are within a difference of 0.001. The average of the two readings is 2.098 cm. Each line represents one packet. The size of each of the packet is varying as noticed in Table 2.

Table 1. Simulation parameters used for carrying out the experiment

Radio parameters		**Simulation parameter**	
Radio frequency	868 MHz | 915MHz	Simulation interval	300secs.
Antenna Height	1m (min. reqd. height 0.0819m)	Topology	Random deployment
Antenna Type	Omnidirectional –Quarter wave	Channel	Wireless
Transmit Power	3.16 mW =5dBm	Network interface	Wireless Physical
Receive Power	−104dBm@ 5dBm=3.98e-14W	Queue length	50
Carrier Sense Threshold	−104dBm@ 5dBm	Network size	75 nodes
Capture Threshold	10 dBm	Transport protocol	UDP
Gain of transmitting/ receiving antenna	1	Application Traffic	CBR
Path loss	1	Data acquisition interval	150 | 300 | 600
Sensor parameters (Tiny Node)		Topology	Random deployment
Receive Power	0.042W=16.23dBm	Channel	Wireless
Idle Power	0.006W=7.78dBm	Network interface	Wireless Physical
Transmit Power	0.099W=19.95dBm	Queue length	50
Sleep Power	0.000003W=-25.2 dBm	Network size	75 nodes
		Transport protocol	UDP
		Application Traffic	CBR
		Data acquisition interval	150 | 300 | 600

Table 2. Readings as measured at the sink

Time now	Data time	Id	Data	Cnt	Data time	Id	Data	Cnt	Data time	Id	Data	Cnt	Data time	Id	Data	Cnt
8.60	7.50	51	2.094	1	7.50	32	2.096	1	7.50	71	2.098	2				
8.83	7.50	72	2.093	1	7.50	48	2.096	1	7.72	17	2.096	2	7.84	4	2.097	4
9.50	7.50	28	2.100	1	7.50	42	2.098	1	7.72	10	2.096	2	7.90	6	2.096	4
	7.65	14	2.098	2	7.52	13	2.098	1	7.74	9	2.095	2	7.56	18	2.099	3
	7.98	5	2.097	7	7.8	1	2.097	2								
10.00	7.52	19	2.098	5	7.74	60	2.095	3	7.72	54	2.098	2	7.56	15	2.099	3
	7.88	7	2.096	2	7.56	11	2.099	3	7.88	2	2.096	2				
11.0	7.5	63	2.097	1	7.86	3	2.095	4	7.50	12	2.100	3	7.50	8	2.098	4
	7.74	16	2.095	3	7.5	20	2.097	1								
11.16	7.50	27	2.100	1												

All timing measurements are in seconds. Data readings are in centimeters. *Cnt refers to the count.*

7 Result Analysis

The current protocol DP_AODV is compared with the protocol M_AODV [27]. M_AODV takes the efficient route to the destination with minimum hop count. Unnecessary broadcasting of path search messages are reduced in M_ADOV as compare to AODV protocol [28]. The result shows a relative comparison of the energy consumed by both the protocol for the same scenario.

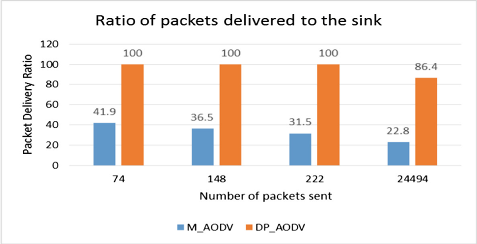
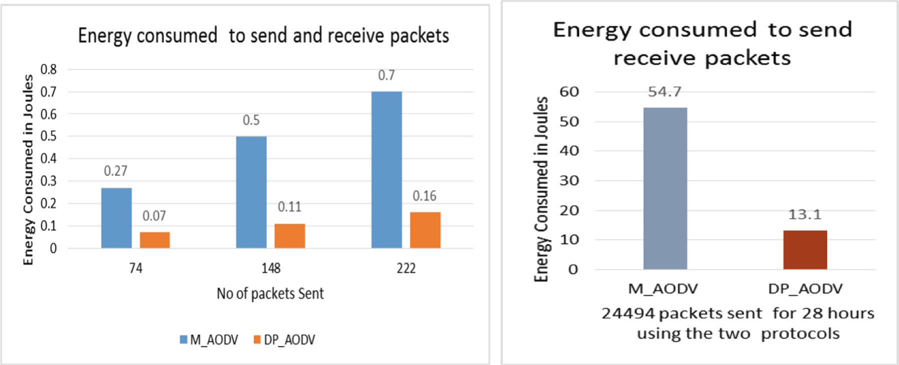


Fig. 4. Packet delivery ratio



a) Small number of packets delivered b) large number of packets delivered

Fig. 5. Energy consumption to send and receive packets

Figure 4 shows the packet delivery ratio for varying number of packets (74, 148, 222 and 24494) that are sent to the base station. Figure 5a shows the amount of energy consumed to send (74, 148, 222) packets and Fig. 5b to send 24494 packets.

The packet delivery ratio is high in the case of DP_AODV as compared to M_AODV. The dropping of packet is maximum in the case of M_AODV. This is because of huge traffic flow in the network. Packet delivery ratio is high in DP_AODV. When few (up to 222) packets are sent, the packet drop is low in M_AODV and almost nil in DP_AODV. When large number of packets are sent, due to non-synchronization of time between packets, queue size increases leading to dropping of packets in DP_AODV.

The energy consumption during transmitting and receiving of data is high in the case of M_AODV as compared to DP_AODV. With DP_AODV, packets are locally aggregated with regard to co-related data. If the difference in data is within of 0.001, data is aggregated, else the Id of the node with the measured data is appended to the existing payload.

8 Discussion and Conclusion

Wireless sensor network has limited energy. Hence energy conservation is a critical issue in the system design of WSN. In applications involving environmental, physical parameters measurement it is essential that the batteries run continuously for long intervals of time. To accomplish this, the number of transmissions and receptions should be reduced. Communication is the main cause for heavy energy consumption. In-network aggregation, clustering and combining of co-related data and combining data from multiple sensors and sending it as a combined packet, saves energy.

In the current paper, routing techniques are incorporated such that the route follows the co-related data within the clusters. Data from many packets are aggregated together and are sent simultaneously. Aggregation here refers to appending data to existing packet (if data is un-related) or averaging, if data is related. This results in Dynamic-Sized Data Packet. The common part of the header is retained, and only the data part is appended to the common header along the path. Data is sent across the aggregator nodes (cluster head) using minimum hop count to reach the base station.

Efficiency is improved using DP_AODV due to the following reasons. Firstly, number of packets reaching the base station is reduced. Due to this reason, the processing overhead at base station is reduced. Traffic is reduced. Packet collision and dropping of packets is reduced. Secondly, all data are sent to the same base station. This redundant information need not be specified for each data that is sent. Thirdly, the size of the data packet changes dynamically based on how many data are co-related. Common part of data is sent only once resulting in energy savings. The problem associated with aggregation is that the aggregator node has to wait for all nodes to send data. A large amount of delay is introduced in obtaining the value at the base station. Synchronization is of utmost importance. Loss of single packet can result in loss of large amount of information.

The current paper emphasis on how energy is saved using proper routing, clustering, aggregation and data combining techniques. This is especially useful for application as in agriculture where the soil moisture content measurement could be carried out for nearly six months when the information collected is highly related. The proposed method can also be incorporated for other applications as well. Monitoring the health of the coral reef, toxicity in water (for marine life), humidity and luminosity information for growth of crop are examples of some other application.

References

1. Panchard, J., Rao, S., Prabhakar, T.V., Jamadagni, H.S., Hubaux, J.: COMMON-sense net: improved water management for resource-poor farmers via sensor networks. In: International Conference on Information and Communication Technologies and Development (2006)
2. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirc, E.: Wireless sensor networks: a survey. J. Comput. Netw. **34**(4), 393–422 (2002)
3. Al-Karaki, J.N., Kamal, A.E., et al.: Routing techniques in wireless sensor networks: a survey. J. IEEE Wirel. Commun. **11**(6), 6–28 (2004)
4. Aschenbruck, N., Fuchs, C.: An integrated simulation environment for sensor data fusion applications in wireless mesh networks. In: Military Communication Conference, pp. 1778–1783 (2011)
5. Luo, H., Liu, Y., Das, S.K.: Routing correlated data with fusion cost in wireless sensor networks. IEEE Trans. Mob. Comput. **5**(11), 1620–1632 (2006)
6. Villas, L.A., et al.: A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks. Ad Hoc Netw. (2011)
7. Zeydan, E., et al.: Energy-efficient routing for correlated data in wireless sensor networks. Ad Hoc Netw. **10**(6), 962–975 (2012)
8. Ci, S., et al.: Adaptive clustering in wireless sensor networks by mining sensor energy data. Comput. Commun. **30**(14), 2968–2975 (2007)
9. Sha, C., Wang, R., Huang, H., Sun, L.: Energy efficient clustering algorithm for data aggregation in wireless sensor networks. J. China Univ. Posts Telecommun. **17**(2), 104–109 (2010)
10. Liu, T., Li, Q., Liang, P.: An energy-balancing clustering approach for gradient-based routing in wireless sensor networks. J. Comput. Commun. **35**, 2150–2161 (2012)
11. Misra, S., Thomasinous, D.: A simple, least-time, energy-efficient routing protocol with one-level data aggregation for wireless sensor networks. J. Syst. Softw. **83**(5), 852–860 (2010)
12. Du, T., Qu, S., Liu, F., Wang, Q.: An energy efficiency semi-static routing algorithm for WSNs based on HAC clustering method. J. Inf. Fusion **21**, 18–29 (2013)
13. Yuea, J., et al.: Energy efficient and balanced cluster-based data aggregation algorithm for wireless sensor networks. In: Proceedings of the International Workshop on Information and Electronics Engineering, Harbin, China, vol. 29, pp. 2009–2015 (2012)
14. Tang, F., et al.: A chain-cluster based routing algorithm for wireless sensor networks. J. Intell. Manuf. **23**(4), 1305–1313 (2012)
15. Jiang, H., Jin, S., Wang, C.: Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor networks. IEEE Trans. Parallel Distrib. Syst. **22**(6), 1064–1071 (2011)
16. Sutagundar, A.V., Manvi, S.S.: Wheel based event-triggered data aggregation and routing in wireless sensor networks: agent-based approach. J. Wirel. Pers. Commun. **71**(1), 491–517 (2013)
17. Lu, K.H., et al.: Hierarchical ring-based data gathering for dense wireless sensor networks. J. Wirel. Pers. Commun. **64**(2), 347–367 (2012)
18. Guo, D., et al.: Prolonging network lifetime and balancing network energy in multi-domain WSNs. Int. J. Sens. Netw. **20**(4), 209–218 (2016)
19. Boucetta, C., et al.: Tree-based modeling of redundancy and paths in wireless sensor network. Int. J. Inf. Commun. Technol. **8**(2), 212–234 (2016)
20. Deniz, F., et al.: An adaptive energy-aware and distributed fault-tolerant topology-control algorithm for heterogeneous wireless sensor networks. J. Ad Hoc Netw. **44**, 104–117 (2016)

21. Chen, H., et al.: Spanning tree-based topology control for data collecting in predictable delay-tolerant networks. J. Ad Hoc Netw. **46**, 48–60 (2016)
22. Ardakani, S.P., et al.: CBA: a cluster-based client/server data aggregation routing protocol. J. Ad Hoc Netw. **5**, 1–33 (2016)
23. Sankarasubramaniam, Y., Akyildiz, I.E., Mchughlin, S.W.: Energy efficiency based packet size optimization in wireless sensor networks, pp. 1–8. IEEE (2013)
24. Legakis, H., Mehmet-Ali, M., Hayes, J.F.: Lifetime analysis for wireless sensor networks. J. Sens. Netw. **17**(1), 1–16 (2015)
25. Nikolaos, A., et al.: Energy efficient routing protocols in wireless sensor networks: a survey. IEEE Commun. Surv. Tutor. **15**(2), 551–591 (2013)
26. Pai, S.N., Shet, K.C., Mruthyunjaya, H.S.: Simulation environment for sensors placed in agricultural field. In: International Conference and Workshop on Emerging Trends in Technology, pp. 752–755 (2012)
27. Pai, S.N., Shet, K.C., Mruthyunjaya, H.S.: Efficient path finding algorithm for transmission of data in agricultural field using wireless sensor network. In: Das, V.V., Chaba, Y. (eds.) Mobile Communication and Power Engineering. CCIS, vol. 296, pp. 345–348. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-35864-7_50
28. Perkins, C., Belding-Royer, E., Das, S.: Ad hoc on-demand distance vector (AODV) routing. No. RFC 3561 (2003)

TVAKSHAS - An Energy Consumption and Utilization Measuring System for Green Computing Environment

Tada Naren(✉) and Barai Dishita

V.V.P. Engineering College, Gujarat Technological University,
Chandkheda, Ahmedabad, India
naren.tada@gmail.com, dishi.dvm@gmail.com

Abstract. There is a large difference between the power draw from the mains and the actual utilization of commercial devices like CPU, as some amount of the power is bound to be wasted. In this paper, the technique that can be implemented for measuring the difference between the actual power draw and the utilized power draw, the steps to reduce the amount of power draw and hence saving the energy costs is mentioned. TVAKSHAS is an energy analytical device that measures the efficiency and performance of CPU and consists of a real time power measurement circuit that reads the real time power drawn from the mains by the CPU and sends it to the wireless sensing and communicating device, i.e. sensor node, TelosB in our project. It also consists of a software based power measurement tool that measures the performance of the CPU in terms of Watts. Hence by analysis of these two data, the difference between power obtained and power utilized can be obtained on base station in the wireless sensor network.

Keywords: Power consumption · Energy consumption · Utilization
Wireless sensor network

1 Introduction

The use of commercial devices like desktop computers, LCD displays, etc. and also the home appliances like oven, refrigerators, etc. are increasing day by day. Hence the energy utilization keeps on increasing day by day which also effects on the electricity bills every month. But the matter of concern is that the amount of energy consumed is not totally utilized for the devices. Some amount of energy is bound to be wasted in some other form of energy. Hence it should be made aware to the users about how much energy is wasted, how much amount of energy is actually utilized. The difference between the values of the amount of energy drawn and amount of energy utilized by any commercial device defines the efficiency of the commercial device, in terms of the power draw. In this paper, various techniques that have been implemented to measure this difference are being focused upon. PowerNet [3], a hybrid sensor network, monitors the power and utilization of the computing systems on the basis of a large scale deployment. SmartMeter.KOM [1] combines the abilities of taking measurements of the electric current flow, switching the mains connection of the attached load, etc.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 37–45, 2018.
https://doi.org/10.1007/978-3-319-73423-1_4

1.1 PowerNet Project

The PowerNet project [3] is a hybrid sensor network developed for computing the power consumption and utilization of various systems. It comprised of approximately 140 wired and wireless meters and almost 23 software sensors that monitored PCs, laptops, LCD screens, etc. This project was active for 14 months and the wireless meters for 3 months. The deployment environment had a large number of diverse set of devices that had large variations in workload and configuration. To improve the efficiency of this type of computing system needs the detailed and accurate data of energy consumption and energy waste. The overall design of the PowerNet deployment that measures the power usage and utilization of individual devices and also transmitting the data over the network to store on a central server is as shown in Fig. 1.

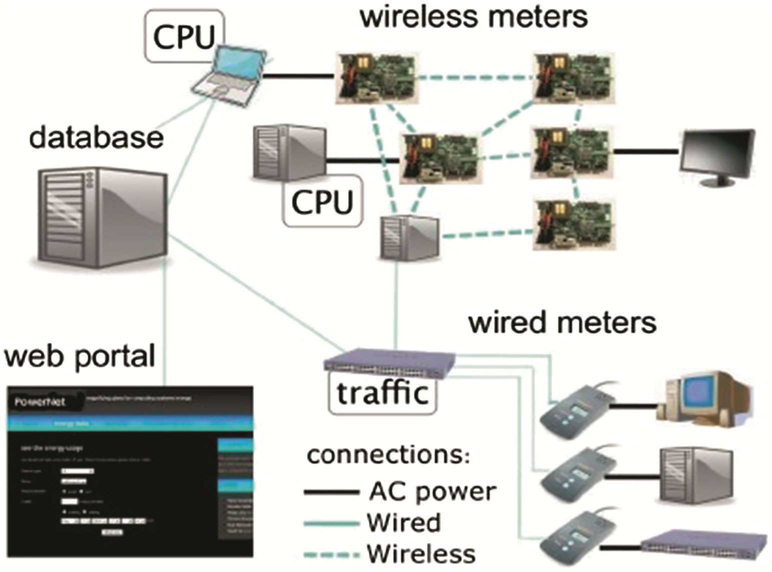


Fig. 1. Deployment of PowerNet project [3].

1.2 The SmartMeter.KOM Project

The SmartMeter.KOM project [1] is a combination of the abilities to take high resolution measurements of the electric current flow, switching the mains connection of the attached load, and a wireless communication device to exchange readings with other nodes in the sensor networks. These devices have the capability to deactivate the devices when the users are absent, the generation of events when operating modes change, or the automatic disconnection of faulty devices. To deploy this system, the sensor used is Allegro Microsystems, AC5712, a precise, low offset current sensor based on Hall Effect. As a communication device, MicroChip MRP24J4DMA is used as it combines a radio transceiver chip with an onboard antenna and all required external components on a single circuit board. These devices are located in always a position where connection to the mains is available. The architecture of the Smartmeter.KOM device is as shown in Fig. 2.

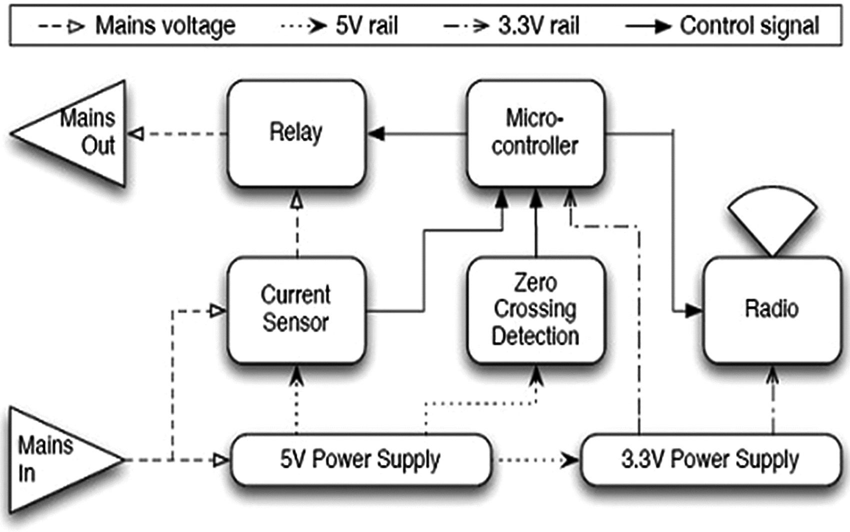


Fig. 2. Deployment of SmartMeter.KOM project [1].

2 The Project TVAKSHAS

The project TVAKSHAS is meant to be developed for the purpose of the quantitative as well as qualitative analysis of the difference between the amount of energy consumed by the commercial device and the actual amount of energy utilized by that device. It is a general tendency that the device does not utilize the amount of energy it is consuming at its whole. Some amount of energy is bound to be wasted and hence there is a variation between the amount of energy consumed by the device and its actual utilization. Also, the project TVAKSHAS aims to give the qualitative analysis like learning about how much power is drawn by which computer at the micro level i.e. measuring the CPU utilization in terms of the active CPU cycle, number of active processes, the amount of CPU consumed by each process in terms of power draw, and then taking the sum of these values to find out the total CPU utilization in terms of power draw. Based on these values, the difference between the consumed power and the utilized power can be derived. These observations can be useful to create awareness to the users for the purchase of different devices.

The deployment of this project is to be done using the wireless sensor network technology. The sensor node, TelosB, can be attached to the CPU whose utilization is to be measured. The measurements can be taken using the TelosB mote that collects these data and sends it to the base station node that can be again be a TelosB mote.

This TelosB mote can be connected to the laptop or other PC to store the data, to perform various analyses through graphs, etc. The TelosB motes need to be configured with the CPU for measuring the data and also receiving the measured data at other PC. The configuration and implementation of algorithm are to be done using the latest version of TinyOS and programming in the nesC language.

The routing protocol used can be the Collection Tree Protocol that routes the measured data from the TelosB mote to the base station TelosB mote. This protocol is considered to be efficient and robust protocol for TinyOs in Wireless sensor network deployment. To collect the data from the CPU, we need to configure the TelosB mote with external power measuring amplifier circuit that takes the readings of the power drawn by the device from the mains and also the power utilized by the CPU.

Using the ADC and DAC convertors in TelosB, the data can be sent to the mote and then to the base station using the CTP routing protocol.

2.1 Architecture of TVAKSHAS

Figure 3 given below describes the overall architecture of Project TVAKSHAS.

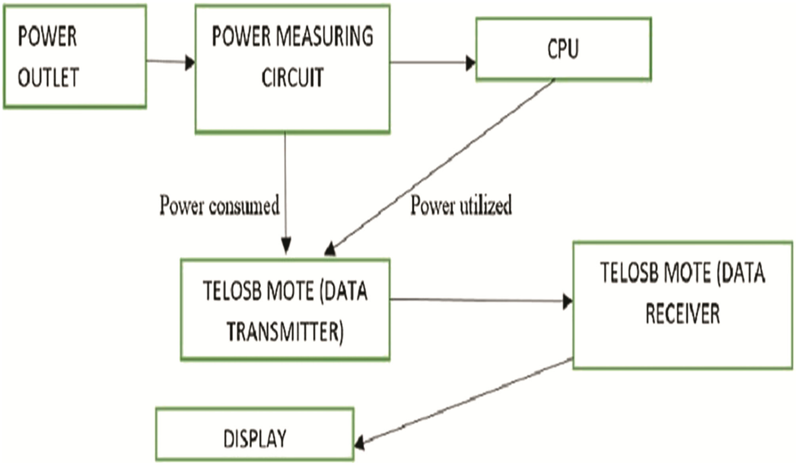


Fig. 3. General architecture of project TVAKSHAS.

As shown in the above figure, the power measuring circuit that measures the real time power consumed by the CPU from the mains and is been connected between the CPU and the mains supply. The circuit has also been connected with the TelosB mote so a\that the power measurements can be read by the TelosB mote and sent through the network to the base station. Hence, the real time values for power consumption can be obtained by this circuit and segregated to the TelosB mote. The next module in this project is development of the tool for power utilization measurement of the CPU in nesC language.

This code or we can say API collects the micro level data of CPU like CPU cycles per second, number of interrupts per second, etc. and based on these observations the amount of CPU utilized in Watts can be obtained which can again be sent through the telosB mote and then to the sensor network to the base station using the CTP protocol.

At the base station, once the data from the telosB mote having power consumption. And utilization values are received, using appropriate GUI, we have the analysis of the difference between the power consumption and the utilization of the CPU on real time basis.

2.2 Working Module of TVAKSHAS

The project TVAKSHAS has been divided into sub modules as follows:

1. Module 1- Hardware based power consumption measurement
2. Module 2-Software based power utilization measurement
3. Module 3- Assembling above data at base station and analysing it using graphs and tables. basis. the attached load, etc.

2.2.1 Hardware Based Power Consumption Measurement

The hardware based power consumption measurement of the project TVAKSHAS includes designing of a power sensor on a circuit board and attaching it to the telosB mote so that the mote can be treated as a power sensor.

The design of the following circuit is influenced from [4].The circuit diagram of this circuit is as shown in the following Fig. 4.

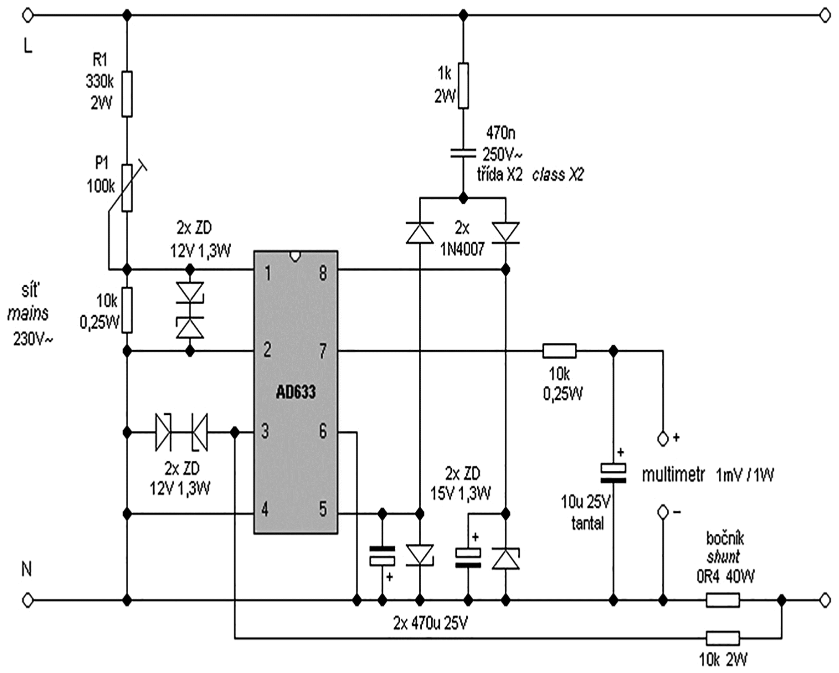


Fig. 4. Power consumption measuring circuit.

The above circuit is directly connected to the TelosB mote which communicates with the base station using the CTP [2] protocol to send the data sensed by the above circuit.

2.2.2 Software Based Power Utilization Measurement

The Software based power utilization measurement module consists of coding of an API that measures the actual power utilization of the CPU by magnifying the micro level consumption of power in different processes like context switches, interrupts, etc. and finally total power utilized can be calculated.

The actual flow of the second module is as shown in the following Fig. 5.

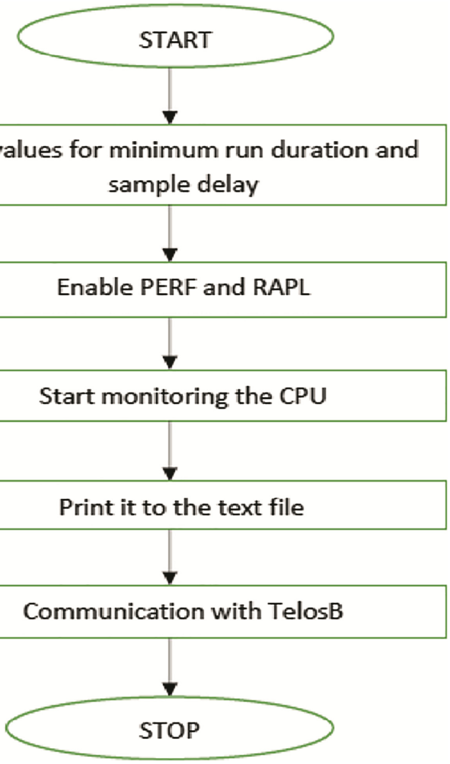
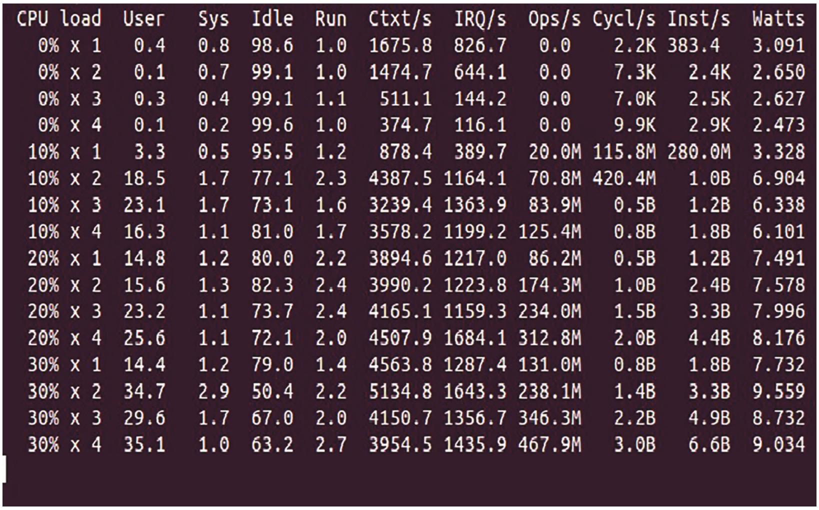

START
Set values for minimum run duration and sample delay
Enable PERF and RAPL
Start monitoring the CPU
Print it to the text file
Communication with TelosB
STOP

Fig. 5. Flowchart of TVAKSHAS power utilization tool.

2.3 Assembling Above Data at Base Station and Analyzing It Using Graphs and Tables

The results obtained from the above tool are as shown in the following screen shot i.e. Fig. 6. As we can see from the snapshot given below, the tool measures the number of CPU working, the number of users working parallel, the amount of seconds the system remains idle, number of context switches per second, number of interrupts per second, number of active CPU cycles per second, etc.



```
CPU load  User   Sys  Idle  Run  Ctxt/s  IRQ/s   Ops/s Cycl/s Inst/s  Watts
  0% x 1   0.4   0.8  98.6  1.0  1675.8  826.7    0.0    2.2K 383.4   3.091
  0% x 2   0.1   0.7  99.1  1.0  1474.7  644.1    0.0    7.3K   2.4K  2.650
  0% x 3   0.3   0.4  99.1  1.1   511.1  144.2    0.0    7.0K   2.5K  2.627
  0% x 4   0.1   0.2  99.6  1.0   374.7  116.1    0.0    9.9K   2.9K  2.473
 10% x 1   3.3   0.5  95.5  1.2   878.4  389.7   20.0M 115.8M 280.0M  3.328
 10% x 2  18.5   1.7  77.1  2.3  4387.5 1164.1   70.8M 420.4M   1.0B  6.904
 10% x 3  23.1   1.7  73.1  1.6  3239.4 1363.9   83.9M   0.5B   1.2B  6.338
 10% x 4  16.3   1.1  81.0  1.7  3578.2 1199.2 125.4M    0.8B   1.8B  6.101
 20% x 1  14.8   1.2  80.0  2.2  3894.6 1217.0   86.2M   0.5B   1.2B  7.491
 20% x 2  15.6   1.3  82.3  2.4  3990.2 1223.8 174.3M    1.0B   2.4B  7.578
 20% x 3  23.2   1.1  73.7  2.4  4165.1 1159.3 234.0M    1.5B   3.3B  7.996
 20% x 4  25.6   1.1  72.1  2.0  4507.9 1684.1 312.8M    2.0B   4.4B  8.176
 30% x 1  14.4   1.2  79.0  1.4  4563.8 1287.4 131.0M    0.8B   1.8B  7.732
 30% x 2  34.7   2.9  50.4  2.2  5134.8 1643.3 238.1M    1.4B   3.3B  9.559
 30% x 3  29.6   1.7  67.0  2.0  4150.7 1356.7 346.3M    2.2B   4.9B  8.732
 30% x 4  35.1   1.0  63.2  2.7  3954.5 1435.9 467.9M    3.0B   6.6B  9.034
```

Fig. 6. Power utilization of the processes when downloading process is ON.

Total power consumed by all these micro level processes are assembled, summed u and finally the amount of power utilized by the CPU is obtained as a magnifier of utilization of these micro level processes.

2.3.1 Analysis Obtained from Project TVAKSHAS

The analysis results obtained using this tool and the power measuring circuit for a computer lab in the department of computer engineering in an engineering college, consisted of similar configurations of CPUs and almost similar workload is as follows.

The following analysis is done using the core i5 7B generation computer for temporary basis. The final outcome of this project will be the readings of the high end PCs of the whole computer department that consists of 6 labs, containing 30 PCs in each. It is assumed that all the PCs in a lab are in ideal conditions and all PCs are running the same program during lab hours According to the lab manual provided online [91], the power ratings for the core i5, 7th generation quad core processor desktop PC is 91 W per hour.

The analysis done using the project TVAKSHAS are as follows:

Consumed Power according to manual available: 91 W
Power Consumption Calculated using the Current Sensor Circuit per hour: 96
Utilization calculated using TVAKSHAS-Util tool: 49.6 W
1-h consumption 96 W
1-h utilization = 49.6 W
Therefore, on working hours (8 h)
8 h consumption = 768 W
8 h utilization = 396.8 W

By extrapolating, we get following results (Table 1):

Table 1. Analysis obtained in project TVAKSHAS.

Property	Example i5	Core 2	Difference
Total consumption	4140 units	2808 units	1310 units
Utilization	2160 units	2160 units	–
Consumption cost	Rs. 33,120	Rs. 22640	Rs. 10,480
Utilization cost	Rs, 17,280	Rs. 17,280	–
Total cost of system	Rs. 40,000	Rs. 30,000	Rs. 10,000

The above results are extrapolated for an hour. The per minute results obtained in TVAKSHAS and sent through the network end then opened the data in an excel sheet and then the analysis is as shown in Fig. 7.

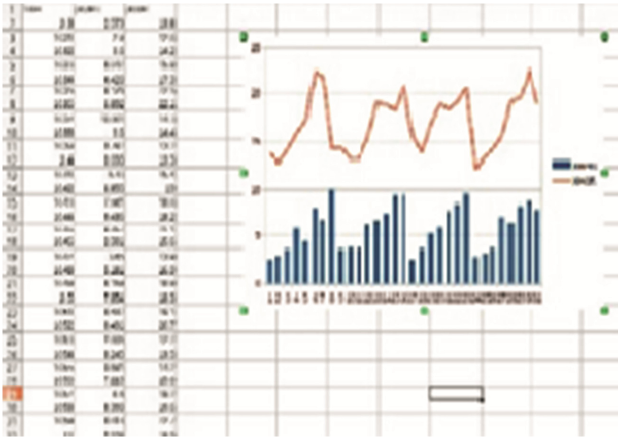


Fig. 7. Output of project TVAKSHAS.

3 Advantages

1. It measures the CPU utilization along with its power consumption.
2. It calculates and analyses of the difference between energy utilization and energy consumption.
3. Qualitative analysis at the process level.
4. To measure how much our conventional methods to reduce energy utilization are useful.
5. To find the best suited CPU depending on the user inputs.
6. Real time measurements.

4 Future Work

From this work, we came to know about the differences in the power consumed by the CPU and actual power utilized by it. In the future, we focus on the reason where the extra power is being wasted and how to overcome that difference using various techniques.

5 Conclusion

The project TVAKSHAS is meant to be developed for the purpose of the quantitative as well as qualitative analysis of the difference between the amount of energy consumed by the commercial device and the actual amount of energy utilized by that device. It is a general tendency that the device does not utilize the amount of energy it is consuming at its whole. Some amount of energy is bound to be wasted and hence there is a variation between the amount of energy consumed by the device and its actual utilization. Also, the project TVAKSHAS aims to give the qualitative analysis like learning about how much power is drawn by which computer at the micro level i.e. measuring the CPU utilization in terms of the active CPU cycle, number of active processes, the amount of CPU consumed by each process in terms of power draw, and then taking the sum of these values to find out the total CPU utilization in terms of power draw. Based on these values, the difference between the consumed power and the utilized power can be derived. These observations can be useful to create awareness to the users for the purchase of different devices.

References

1. Reinhardt, A., Burkhardt, D., Mogre, P.S., Zaheer, M., Steinmetz, R.: SmartMeter.KOM: a low-cost wireless sensor for distributed power metering. In: 6th IEEE International Workshop on Practical Issues in Building Sensor Network Applications, SenseApp (2011)
2. Gnawali, O., Fonseca, R., Jamieson, K., Moss, D., Levis, P.: Collection tree protocol. In: SenSys 2009, Berkeley, CA, USA (2009)
3. Kazandjieva, M., Gnawali, O., Heller, B., Levis, P., Kozyrakis, C.: Identifying energy waste through dense power sensing and utilization monitoring. Technical report, CSTR. Stanford University (2010)
4. http://danyk.cz/wmetr_en.html

Challenges to Developing a Secure, Cloud-Based Offline Mobile Application

Sambit Kumar Patra[1](✉) and Navin Kumar[2]

[1] Accenture Technology, Bangalore, India
sambit_sai@yahoo.com
[2] Amrita School of Engineering Bengaluru, Amrita Vishwa Vidyapeetham Amrita University, Bangalore, India
navinkumar@ieee.org

Abstract. The alliance with mobile device cloud computing technology promises new ways of developing business application. Using web and cloud technology, it is possible to transfer a small part of secured business data from cloud to the small storage mobile devices. However, it is challenging to keep secure data in offline mode when the web application is unable to connect to the cloud and sync those offline data at online mode. The paper discusses architecture of cloud-based mobile application. The proposed architecture helps us to develop cloud-based mobile application. Special consideration of low memory and network connectivity on the mobile device is taken into account. We present various challenges to design a cloud-based mobile application, store the data in a secured manner at offline mode and sync those at online mode. This would greatly improve enterprise productivity even when users are working offline.

Keywords: Mobile application development · Secure data transmission Offline data · Cloud

1 Introduction

The fast evolution of commercial mobile devices has made the technology an essential requirement for the government, enterprise and commercial end users. Mobile technologies are transforming enterprises, industries, and the entire world [1]. Also, the growing need to produce new and innovative mobile applications which provide enhanced business capable government workforce, together with common capabilities like secure email, has led to the huge challenge of providing standardized solutions. The technology today offers the opportunities to drive business transformation through mobility.

Mobility is more invasive than ever before with increasingly rapid evolution of new devices and technology, placing demands and creating opportunities for enterprises and consumers around the globe [2]. Mobility provides better and faster decision making through improved access to key data and analytics capabilities anytime, anywhere. It also enables access to workflow tools on the job, reducing manual processes, supporting on-the-go secure operations, services and management. However, the future of mobility

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 46–53, 2018.
https://doi.org/10.1007/978-3-319-73423-1_5

is in the cloud [3], but when a connection to the cloud is not available, the mobile user is out of the loop.

Now-a-days, many large and small businesses use cloud computing either directly (for example, Google or Amazon) or indirectly (Twitter) instead of traditional on-site alternatives because of various reasons such as cost benefits, universal access, flexibility, and so on [4]. But the predicament with mobile devices is that they are constantly mobile; consequently they tend to lose connectivity with the cloud. For example, if a customer is travelling, most of the time s/he would be out of the coverage for one or other reasons e.g., remote location or cell is down. So, s/he won't be able to update information such as editing an opportunity or updating a case because of the poor connectivity.

Modern businesses, their information systems and mobile devices cannot be expected to hold the data that users need at any moment of time. Any disconnection from network because of any problems such as high latency, low bandwidth or even presence or absence of the network may result in reduced productivity. To successfully moving and promoting the usage of cloud centric enterprise architecture, there would be important and necessary requirements to have support for offline data transfer [5]. Offline support for mobile applications is the ability for the application to react elegantly to the lack of stability in the network connection, high latency and even low bandwidth.

However, providing offline access is not easy: There are significant technical challenges and one of the reasons cloud-only has been pushed as the future is that offline was seen as being a bit complicated. There are many challenges to think of when we consider providing offline access to applications. The challenge is to determine what degree of offline operations is possible at any given time, maximizing productivity in working offline. The challenge is in security: what data is stored on the client, and how this can be compromised. Any offline access requires storage on the device. This means that the data could be read by malware or if the device is stolen, how mobile device and application management that is mobile device management (MDM) and mobile application management (MAM) will address this issue.

In this paper some of the challenges in offering offline support are discussed. We present them in brief and we also discuss our approach in addressing these challenges. A simple architecture is presented to describe the approach which would support offline data transfer in the situations like network unavailability, low bandwidth and so on. It is a new concept and idea. To the best of our knowledge, we have not found this in any literature.

The contents of the rest of the paper are as follows. Section 2 presents the challenges in brief. The architecture and proposed solution is discussed in Sect. 3. Section 4 presents the best practices which can be followed. The conclusion is given in Sect. 5.

2 Challenges

There are number of technical challenges in providing offline support. There are even challenges if developing application to support offline data transfer. It is possible for a browser to store resources on the local device, such as caching. But caching is normally used for speeding up page loads and reducing the bandwidth usage. Furthermore, on the mobile device, we have limitation of data/cache. Therefore, there seems a no. of issues in offline support. We discuss some of them in this section and try to understand the development constraint.

2.1 Responsive Design

Irrespective of different mobile size and operating system, the user interface of the application should remain the same across the various platforms like Android, IOS, Windows, and Blackberry [6]. Also, the application should be responsive enough to manage both portrait and landscape mode for small devices like phone, tablet to bigger screen like TV.

2.2 Developing a Dynamic Scalable User Interface (UI)

Looking at the business needs, the application flow may change constantly and consistently, however, it should not impact to the application architecture. As redeveloping and redeploying process of the application on respective application store like Android, IOS and Windows takes some considerable time and are more expensive, the application architecture should be capable of handling the future business changes. A business based mobile application should be robust enough to create the pages dynamically, handle events, throw errors and manage huge number of data within the limited memory and battery constraints.

2.3 Accessing Native Functionality of Mobile

As the business applications are developing for the mobile devices, the applications are trying to enhance the functionality of the mobile device such as Call, SMS, Camera, Calendar, Notes. Irrespective of device constraints and OS, the application should be integrated with native mobile OS in such a way that it should be merged with those functionality when required. It will be more challengeable.

2.4 Storing Secure Data

The business data is sensitive. Transaction is normally conducted in online mode which is a big challenge to run such application in a mobile device due to the poor network connectivity especially while roaming and moving from one page to another. The cloud based mobile application should be designed such a way that it has the ability to store the data when the device is offline and upload the same at online mode.

2.5 Syncing Offline Data

Synching is another important challenge. During syncing, the device uploads the updated data into server and retrieves data from the server. In a large business application, sometimes same data of some module of the application are handled by multiple users. The main challenge is updating the offline data into server. Again, due to poor network connectivity the updated data are stored in the device for a longer time. During this time, some other devices update the data into server. When the device becomes online, it uploads the offline data into server which is older than the data from the server.

2.6 Connect the Devices with Same Application Flavor

User can able to access the same application and data using his/her multiple devices from remote locations. The application architecture should maintain the same flavor across all the devices. It should consider the devices are connected to each other.

2.7 Personalize the Device

Application should allow user to customize the application with personalized data with his/her own device. It should not be shared to other devices. To personalize the device, user may like to change the theme, add his/her own photo, contact and reminders to the application.

2.8 Keep Personalize Data During Upgrade

Sometimes user may reinstall the application to the same device either he might have deleted the application accidentally or the newer version of the application has been upgraded. During this process, user should not lose his/her personalize data. After upgradation, the new application should maintain those personalize data.

2.9 Embedded with Data and Location Visualization

To take better decision with analytics, the data of the business application should be represented in a chart rather than tabular form. For presenting the data in a better way including touch functionality, the chart library [7] should support zooming and tool tips with bar, line and pie charts. Similarly, data for location visualization application should be embedded with a Maps library which shall provide the distances, travel road guides, time to cover to a location and must visit location near to the business location.

3 Our Approach for Offline Support

There are different approaches to address the above challenges. For example, cloud computing company Salesforce.com, the world's first platform as a service (PaaS) [8], which is best known for its customer relationship management (CRM) product. It provides Salesforce mobile software development kit (SDK), which is an open-source suite of familiar technologies allowing developers to rapidly build HTML5 application that connect to the Salesforce platform. Using this SDK, we can develop a single page mobile application using JavaScript libraries like JQM, Sencha, AngularJS for different

OS platforms like Android, IOS and Windows. Based on this concept, the generalized architecture and framework is given as shown in Fig. 1.

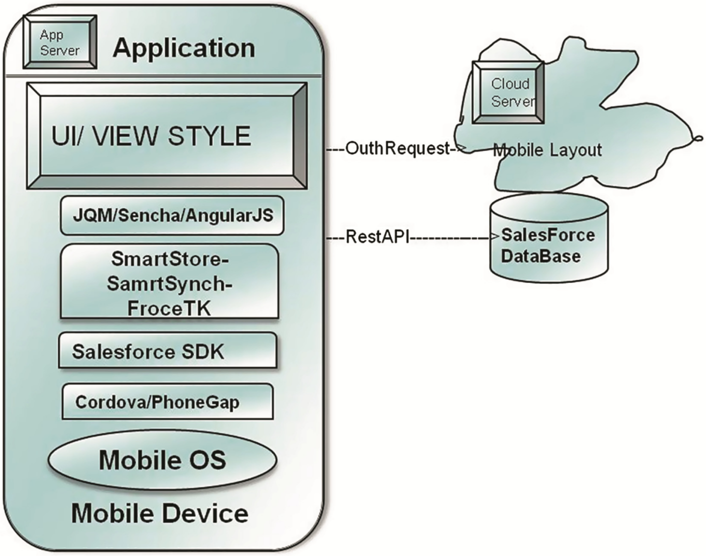


Fig. 1. Development framework

3.1 JQM (jQuery Mobile)

jQuery Mobile framework [9] follows the “write less, do more” approach and take it to the next level. Instead of writing unique applications for each mobile device or OS, the jQuery mobile framework allows us to design a single highly-branded responsive web site or application that will is expected to work on all popular smart phones, tablets, and desktop platforms.

3.2 Sencha ExtJS

Sencha ExtJS [10] is the most comprehensive MVC/MVVM JavaScript framework for building large features cross-platform web applications targeting desktops, tablets, and smart phones.

3.3 AngularJS

AngularJS [11] is a development platform for creating applications using modern web standards. Angular includes many essential features such as mobile gestures, animations, filtering, routing, data binding, security, internationalization, and beautiful UI components. It’s extremely modular, lightweight, and easy to understand.

3.4 SalesforceSDK

Salesforce [8] Inc. is a global cloud computing company best known for its customer relationship management (CRM) product. Salesforce Mobile SDK3.0 is an open-source suite of familiar technologies—like a REST API and OAuth 2.0—that we can use to build great mobile apps. The Salesforce Mobile SDK supports three development approaches for building mobile apps: native, HTML5 and hybrid.

Hybrid development combines the best (or worst) of both the native and HTML5 worlds [8]. We define hybrid as a web app, primarily built using HTML5 and JavaScript that is then wrapped inside a thin native container that provides access to native platform features. Phone Gap is an example of the most popular container for creating hybrid mobile apps.

Smartstore, is an encrypted NoSQL-style JSON document data store. It is the cross-platform encrypted NoSQL mobile database in the market that works with both hybrid and native development models. "NoSQL" databases are a class of database that unlike their "relational" brethren, are designed to be inherently schema less. They take care of storing data, and let the application worry about how to structure it. Some of the benefits of this style of database are their raw speed, a dynamic schema where fields and tables can be added by the application at will, and ease of use for the developer. JavaScript Object Notation (JSON) is a lightweight industry-standard way to encode data for transfer between systems. Some NoSQL databases are designed to be simple key/value stores, but more advanced systems like Smartstore allow for storing and indexing full JSON documents. The Salesforce mobile services provides the Smartstore tools needed to build enterprise mobile apps that allow us to securely transfer and store data on our mobile device for highly performant offline access [6].

SmartSync. The SmartSync library is a collection of APIs that make it easy for the developers to sync data between Salesforce databases and their mobile apps. It provides the means for getting and posting data to a server endpoint, caching data on a device, and reading cached data. For sync operations, SmartSync predefines cache policies for fine-tuning interactions between cached data and server data in offline and online scenarios. A set of SmartSync convenience methods automate common network activities, such as fetching sObject metadata, fetching a list of most recently used objects, and building SOQL and SOSL queries.

ForceTk. JavaScript is a popular programming language for building Web-based applications. The most common use of JavaScript is client-side in a Web browser, for implementing pages with enhanced, responsive user interfaces. Client-side JavaScript is also useful for calling XMLHttpRequest methods (GET, POST, etc.) to work with data managed by the remote Web server. ForceTK provides a convenient, thin JavaScript abstraction of the Force.com REST API, making the API more accessible to JavaScript code running in Visualforce, in hybrid mobile apps, and elsewhere.

3.5 Cordova/PhoneGap

Apache Cordova is an open-source mobile development framework. It allows us to use standard web technologies such as HTML5, CSS3, and JavaScript for cross-platform development, avoiding each mobile platforms' native development language. Applications execute within wrappers targeted to each platform, and rely on standards-compliant API bindings to access each device's sensors, data, and network status.

4 Best Practices

Though there are many different approaches, one can follow to develop the apps for efficient offline support. Above Sect. 3, we have used the concept from sales force and develop a mobile application using Hybrid approach. However, some of the best practices in general would be considered.

4.1 Less is More

Significantly important to be selective about what the users really should view in a mobile app. For instance, instead of displaying all the data, display the selective of the information that needs to be displayed and is useful.

4.2 Limit Data Usage

A cloud based mobile application generally displays data that it retrieves from the cloud. This could mean a lot of Ajax calls; which ultimately will be heavy on the data usage. But, since we have the availability of a secured offline storage, we can store the data and display the stored data when synchronization undertakes.

4.3 Limit Stored Data

It is recommended to limit the data storage to avoid over utilization of device memory. If excessive amounts of data needs to be stored then, a limit for the number of items or records received from the cloud must be maintained and stored in the secure offline storage. Additionally, we can also query records with only the required fields instead of all the fields. This would not only optimize memory consumption but will also help in all performance of the application and reduce data usage.

4.4 Make Sure Long Strings Don't Break the Layout

Naturally, we are discussing about the data itself, but the surrounding UI might also suffer. For example, consider a command bar with buttons. When the text on the buttons is translated, they might grow large and overflow the bar.

4.5 Make the UI “Lazy Load”

Make the UI lazy load some of the info and consider using pagination. For example, it might take too long to load at once thousands of items - make sure the UI doesn’t look frozen or broken while it loads, or that you load one page and then load the rest while the user can also start interacting with the UI.

5 Conclusion

The cloud is one of the more significant shifts that computing has gone through. As we move towards the cloud, we will discover a new service-based world, and being the developers, we will have to serve the collaboration and security needs of all customers – from those working in small team to those working in huge enterprises. With the knowledge and features discussed in this paper, we can deliver flexible and powerful applications that can be managed on the cloud and will provide the customer with both offline and real-time updates.

References

1. Kendrick, J.: Mobile technology: the amazing impact on our lives. ZDNet Mobile Technology, 30 April 2013
2. Armano, D.: The future isn’t about moible: its about mobiliety. Harvard Bus. Rev., 18 July, 2012
3. Guntur, S.: Cloud mobility: the next big disruptive technology. HCL Technology, 18 July 2013
4. Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., Ghalsasi, A.: Cloud computing: business perspective. Sci. Directo Decis. Support Syst. **51**(1), 176–189 (2011)
5. Kim, K.-H., Lee, S.-J., Congdon, P.: On cloud-centric network architecture for multi-dimensional mobility. In: Proceedings of the SIGCOM, 17 August 2012
6. Kadlec, T.: Implementing Responsive Design: Building Sites for an Anywhere, Everywhere Web, 1st edn. New Riders, Berkeley (2012)
7. Patra, S.K.: Data visualization for hybrid application: the challenges in choosing an optimal library for line chart. Int. J. Emerg. Sci. Eng. (IJESE) **2**(10), 12–15 (2014)
8. http://www.salesforce.com/in/. Accessed 12 November 2015
9. https://jquerymobile.com/
10. https://www.sencha.com/products/extjs/
11. https://angularjs.org/

Mobile Commerce Business Model for Customer Oriented Business Transactions

P. V. Pushpa(✉)

Electronics and Communication Engineering Department,
P.D.A. College of Engineering, Glubarga, India
alladpushpa@gmail.com

Abstract. Mobile commerce environment pertains to multiple transactions with the purpose of providing goods and services to customers. The customers are using wireless devices like smartphone, mobile phone and PDA to shop for a broad range of products and information related services. The success of commercial business models for mobile commerce environment depends on the method of providing value to the customers. Modeling the specific characteristics of each business participant is a challenging task in dynamic business environment. In this paper, we propose a formal description of business model by identifying the specific characteristics of mobile participant and illustrate with examples in a trading scenario. The proposed analytical concepts assist in developing innovative user friendly interfaces and also to model e-business domain.

Keywords: Mobile commerce · Customer · Business model
Transaction · Product · Participant

1 Introduction

A mobile commerce business involves activities like purchasing and sale of commodities, banking and foreign exchange activities which takes place through mobile device. The environment supports exchange of huge data during selling, buying and financial related mobile transactions. The business environment includes mobile customers, mobile vendors, technological developments, demographics, social and economic trends. As the priority of each customer changes, it is required to seggregate individual and group preferences based on their profile to provide tailored services and products. The m-commerce has focused mainly on designing innovative applications and mobile transactions. Issues like analyzing the transactions and modeling the business entities (e.g., customer, vendor, broker, banker) need to be addressed to develop innovative business models for commercial purpose.

Due to the ever changing nature of mobile business environment and also to cater to the customer demands, there is a need to develop pragmatic models adapting to new business situation. To enable transactions quickly among business entities in such complex environments, the challenge is to understand the

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 54–66, 2018.
https://doi.org/10.1007/978-3-319-73423-1_6

customer behavior, goals and responsibilities based on educational, economical, social and professional background.

The users can access M-commerce services anywhere, anytime by using small computing devices and also the IOT paradigm allows *'people and things to be connected anytime, anyplace, with anything and anyone, ideally using any path/network and any service'* [1]. The proposed model in [2] describes the significance of deducing context based beliefs about a business customer or vendor and which can be utilized by applications to execute business transaction according to current situation.

In this paper, we describe the method of developing mobile commerce business environment by identifying the characteristics of electronic products, transactions and business participants. We have proposed customer behavior as an important parameter in understanding their new purchase requirements. We have proposed mobile commerce businesss model as a specific case of [3] which facilitates in defining the professional skilss, goals, responsibilities and financial criteria of mobile customer, mobile vendor, mobile broker, and mobile bankers in e-business domain.

The rest of the paper is organized as follows. In Sect. 2, we briefly describe some of the related works, Sect. 3 describes the mobile commerce environment, Sect. 4, provides the formal representation of mobile commerce business model. In Sect. 5, we describe the mobile commerce business system, and lastly the conclusion.

2 Related Works

The rapid development of mobile commerce technology has enabled the establishment of new services. The customers can now conduct on line transactions anytime, anywhere using smart devices. A business model [4,5] deals with information flow between business participants and thereby delivering a product or service to the customers. The transaction management technology, [6] provides data consistency and service reliability in mobile commerce environment. The changing business environment [7] has influence on each participant like government, business organization and customers who are involved in market exchanges. The works in [8] determine user location obtained from GSM base stations and bluetooth sources for personalization using hybrid method.

The paper [9] proposes an analytical framework for the issues like coordination, cooperation, customer value and core competence. The research [10] on m-commerce usage activities are highly influenced by demographic and motivation variables. The m-commerce service providers study the customer's behavior to formulate appropriate marketing strategies. The business model [11] refers to strategies of multiple players, m-commerce services, and also revenue generation. The work [12] discusses how the mobile users have to carry out day-to-day m-commerce transactions with minimal effort in trusted and secured manner. A viable business model evaluation framework based on the VISOR model is presented in [13] which helps to determine the sustainable capabilities of a mobile commerce business model. The business model [14] represents a concept on how the business functions and includes defining the goals, visions and other factors.

The research [15] summarizes the guidelines and factors to create business models based on context information. The use of context-awareness [2] enhances the satisfaction level of business participants in marketing approach. The proposed framework in [16] improves business interoperability through context-based ontology reconciliation. However, the above models dealt with customer demographic variables, security issues and service aspects without defining clearly the goals and responsibilities of individual business participants (e.g.,vendor, broker, banker, wholesaler and retailer) to understand the real business world.

3 Mobile Commerce Business Environment

The proposed business environment mainly consists of mobile participants who are involved in various business activities. Figure 1 shows the commercial environment which has been setup for electronic goods sales and purchase. The customers are using wireless devices like smartphone, mobile phone and PDA to shop for a broad range of products and information related services. The database is established for storing the sales history of products, product benchmarking features and also the transactions. The benchmarking features helps to identify an efficient product to a customer. The description of mobile commerce service provider, mobile customer, vendor and banker are given below in the following sections.

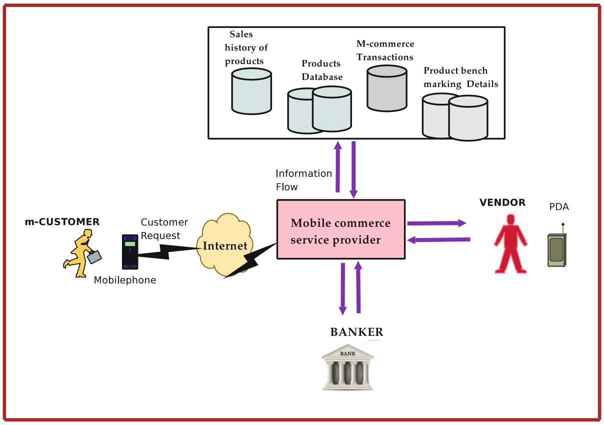


Fig. 1. Mobile commerce business environment

3.1 Mobile Commerce Service Provider

The function of MCSP is to provide services to customers either directly or through gateway of another company. They act as mediators between customers and vendors and offer services to mobile customers.

3.2 Electronic Product

A product is a thing that can be offered to a customer and satisfy a want or need of mobile participant. Each electronic product can be described with three kinds of attributes like general attribute (i.e., quality, brand, price), technical attributes(i.e., memory capacity, display size etc.) and commercial attributes (i.e., product rating, number of products sold, etc.) Let PR represents a product, and each product belong to a broader category of electronic goods with certain technical features. The electronic product description is given by the tuple,

$$PR = \langle GA, TA, CA \rangle \tag{1}$$

where GA, TA, CA represents set of general, technical and commercial attributes respectively, for example,

$$\begin{aligned} GA = \{ga_1 = \text{is ModelID},\ ga_2 = \text{Product Code},\ ga_3 = \text{Product Name}, \\ ga_4 = \text{Display Resolution},\ ga_5 = \text{Size},\ ga_6 = \text{Product Weight}, \\ ga_7 = \text{Product Brand},\ ga_8 = \text{Product Color}\} \end{aligned} \tag{2}$$

$$\begin{aligned} TA = \{ta_1 = \text{Operating Frequency},\ ta_2 = \text{Processor Speed}, \\ ta_3 = \text{MemorySize}, ta_4 = \text{Battery Time}\} \end{aligned} \tag{3}$$

$$CA = \{ca_1 = \text{Best Selling Rank},\ ca_2 = \text{Customer Review}\} \tag{4}$$

The Best selling rank gives the ranking given to products based on the number of units sold. The customer review is the average "score" or ratings as submitted by reviewers. The electronic products which can be offered to a mobile customer in M-commerce are Television, Personal Digital Assistant, Calculators, VCRs, Digital Cameras, Audio Devices, headphones, Hard Disks, Pendrives, Camcoders, Clocks, Printers and many other products.

3.3 M-Commerce Transactions

M-commerce transactions represent a set of activities or interactions involved between mobile customers and vendors for buying or selling of goods or services. The transactions can be generated by mobile device and is defined as any kind of activity or exchange of information between business participants. Activity represents in general the action performed by the mobile participant during any kind of transaction. It is defined by the following tuple,

$$T = \langle TID, Dt, PR, Prc, P_i, P_j \rangle \tag{5}$$

where, TID represents Transaction Identifier, Dt represents the Date of transaction, PR is the type of electronic product, Prc represents the price of the product and P_i, P_j represents the roles performed by i^{th} and j^{th} business participants. Some of actions involved during transaction could be sending *purchase order, viewing bank details, sending quotation, transferring of goods, account statements, balance enquiry, direct debits, bill payments and so on.* Tables 1 and 2 gives some of the examples and description of mobile commerce transactions.

Table 1. Examples of M-commerce transaction

Transaction no	Example of transaction
T_1	Request for purchase order using mobile phone
T_2	Request for money transfer for the product purchased using smartphone
T_3	Request for after sales service
T_4	Request for business calculator general information during business hours.
T_5	Issuing a quotation for the product

Table 2. Description of transaction

Transaction no	Description of transaction
T_1	⟨0001, 11-11-2011, Calculator, 750 Rs, Mobile Customer, Vendor⟩
T_2	⟨0015, 01-01-2012, Camera, 15500 Rs, Mobile Customer, Banker⟩
T_3	⟨0017, 01-01-2013, Laptop, —, Mobile Customer, Vendor⟩
T_4	⟨0025, 01-05-2015, Business Calculator, –, Vendor, Mobile Customer⟩
T_5	⟨0055, 05-05-2014, Camcoder, —, Vendor, Mobile Customer⟩

3.4 Business Participants

A mobile participant is a person or an entity dealing with any kind of transaction and has different roles in business domain. A specific behavior is exhibited by each participant during any business transaction (e.g., buying, selling). The transactions are established by mobile phone and thereby initiating a process with the system (e.g., smart phone or mobile phone), when purchasing electronic goods. Business participant is an important entity and we describe briefly about each participant in the following sections.

Mobile customer: The customers represent an important class of business participants. A mobile customer is a person who is using mobile devices like Smart phone or PDA for making commerce related transactions. Customer interacts with his device for getting desired service from M-commerce service provider. They use mobile phones or smart phones to browse information, purchase products, manipulate price comparisons, read product reviews or interact wirelessly with service providers to get specific service. Some customers are using multiple mobile devices for personal and professional life. Text message is the most common method of receiving mobile advertisements and few smartphones also provide location-based services. The mobile customer behavior refers to the buying decision using mobile phones to satisfy his want or need. It involves study of how customers buy, what they buy, when and why they buy. Moreover, there is a strong influence of groups like family, friends and colleagues on the customer purchasing behavior. Table 3 gives description of mobile customer behavior for purchasing activity based on several factors.

Table 3. Mobile customer behavior for purchasing activity

Classification of customers	Sources of information	Influence of advertisement	Price comparison	Payment mode	Social network
High school student	Friends	High	Low	Cash	Medium
College student	Friends	High	Low	Cash	High
Employed full time	Colleagues	Low	Medium	Credit	Low
Old age	Family	Low	High	Cash	Low

Mobile Vendor: Vendor represents an entity or a person who is responsible for conducting transaction to the customers anytime, anywhere. They perform the complete transaction cycle with mobile customers directly or via their mobile phones during their working business hours. The vendors send alerts, receive payment, issue coupons and track customers. The sales history of vendors contains the number of products sold, the price of product, the date of selling, the name of customer and so on.

Mobile Broker: A person or an entity acting as an agent for arranging transaction between mobile customers and vendors. The function of broker is to provide timely information about market conditions to a moving customer based on his/her location. The broker provides information on product price, product commercial information, market index and so on to a mobile customer.

Mobile Banker: A Banker is an entity who establishes financial and banking services with hand held devices. Some of the functions include, opening of account for mobile customer, money transfer service using mobile devices.

3.5 Benchmark Testing

The benchmark testing procedure provides information about how an electronic product performs better than other in real world environment. The parameters typically used and measured are cost per unit of product, time and quality of product. These methods quantify the user experience to do business, take decisions and thereby recommending for the products. Some of the functional capabilities of the electronic products being compared are:

- Product weight, strength, durability and size
- Data transfer rates and capacity of storage
- Ease of setup, configuration, assembly and usage
- Completeness of user documentation
- Evaluation of the product technical feature set
- The value in terms of Function vs Cost
- Ratings of ease of use and user satisfaction

4 Formal Representation of Mobile Commerce Business Model

A Mobile Commerce Business model is constituted with mobile participants, mobile transactions, mobile devices and the type of products involved in commercial business activity. The MCBM is described by

$$MCBM = \langle MP, T, A, PR, D \rangle \tag{6}$$

where, MP = $\{mp_i,\ where\ i = 1, 2, .., m\}$ is a finite set of mobile participants, T = $\{t_j,\ where\ j = 1, 2, .., n\}$ is a finite set of transactions executed by participants, A = $\{a_1, a_2, .., a_q\}$ where 'q' is the number of activities or actions initiated by mobile participants (e.g., *balance inquiry, downloading account statement, ordering cheque books, viewing recent transactions*, *browsing for catalogs* etc.), PR represents the electronic product involved in commercial business transaction, D = $\{d_i, i = 1, 2, .., k\}$ is a set of participant mobile devices (e.g., Mobile phone, smartphone, PDA). Each participant $mp_i \in$ 'MP' has distinct role 'r' in business domain is identified and characterized by two sections namely *identification section* and *specification section* which we describe in the following sections. The works in [17,18] present high level conceptual templates based on ontology design patterns to understand the business world.

1. *Identification section* (IdntSec): The identification section, contains information such as name, unique id and email id of each business participant
   IdntSec(mp_i) = [name(mp_i), id(mp_i), email id(mp_i)]

- name (mp_i) $\in$ string; mobile participant name.
- id (mp_i) $\in$ integer; identifier of mobile customer or vendor.
- email id (mp_i) $\in$ string; participant email identifier.

2. *Specification section* (SpecSec): The specification section describes the specific characteristics of mobile participants such as Goals (G), Professional Skills (PS), Responsibilities (R) and Financial Criteria (FC) exhibited during mobile transactions. The representation of specification section of each individual participant is as follows.

$$SpecSect(mp_i) = [G(mp_i),\ PS(mp_i),\ R(mp_i),\ FC(mp_i)] \tag{7}$$

Goals: the business objectives laid by customer/vendor and given by

$$G(mp_i) = (l, f), l \in string, f \in string; \tag{8}$$

where, ‘l’ and ‘f’ gives a natural language description of specific characteristics of the participant.

Professional Skills: the set of activities initiated by mobile device during product purchase or selling in a business environment. For example, the professional skills of participant initiates a specific product purchasing action with a role as buyer through mobile device.

$$\begin{aligned} PS(mp_i) = (a, pr, r) \mid (mp_i, a, pr, r), \\ a \in A,\ pr \in PR,\ r \in \{\text{“buyer, mobile customer''}\}; \end{aligned} \tag{9}$$

Responsibilities: actions involved in achieving a business goal with role as mobile customer. The responsibilities of a mobile participant is to place a purchase order with mobile device.

$$\begin{aligned} R(mp_i) = \{(a, r) \mid (mp_i, a, r),\ mp_i \in MP, \\ a \in A,\ r \in \{\text{“mobile vendor''},\ \text{“wholesaler''}\}; \end{aligned} \tag{10}$$

Financial criteria: The financial criteria permits the participant to buy cost efficient products. Financial status of participant indicates whether to buy costlier or cheaper product.

$$\begin{aligned} FC(mp_i) = \{(pr, l) \mid (mp_i, pr, l),\ mp_i \in MP, \\ l \in \{\text{cheaper, costlier}\}; \end{aligned} \tag{11}$$

In the following section, we discuss several business situations with examples for each mobile participant.

4.1 M-CUSTOMER (MC)

A M-customer receives goods, service or product from vendor or wholesaler at his location. Let $MC = \{mc_1, mc_2, .., mc_{m_1}\}$ be a set of customers, where ‘m_1’ represents the total number of customers involved in business activity. The customer specification section is described as follows:

The *goal* is to buy a product with high discount with role as customer.

$$\begin{aligned} G(mc_i) = \{(pr, d, r), \mid (mc_i, pr, d, r),\ mc_i \in MC,\ pr \in PR, \\ d \in \text{very high discount, high discount}\},\ r = \{\text{“customer''}\}; \end{aligned} \tag{12}$$

The *professional skills* initiates the activity of funds transfer for the product purchased with mobile device.

$$\begin{gathered} PS(mc_i) = \{(a, pr, d) \mid (mc_i, a, d), \\ a \in A,\ d \in \{\text{“smartphone”},\ \text{“mobilephone”}\}; \end{gathered} \tag{13}$$

The *responsibilities* of mobile customer is to initiate an activity of placing a purchase order for a product with mobile device.

$$\begin{gathered} R(mc_i) = (a, pr, r) \mid (mc_i, pr, a, d),\ pr \in PR,\ mc_i \in MC, \\ a \in A, d \in \{\text{“mobilephone''},\ \text{“smartphone''},\ \text{“PDA''}\}; \end{gathered} \tag{14}$$

The *financial status* is high to buy good product.

$$\begin{gathered} FC(mc_i) = (l, f), l \in \{\text{low},\ \text{moderate},\ \text{high}\}, \\ f \in \{\text{bad},\ \text{good}\}; \end{gathered} \tag{15}$$

4.2 M-VENDOR (MV)

Mobile vendor sells goods or services to the customers through mobile device. Let $MV = \{mv_1, mv_2, .., mv_{m3}\}$ be a set of vendors, where m_3 is the total number of vendors involved in mobile commerce business.

The *goal* is to initiate selling of products based on location and earn profit.

$$\begin{gathered} G(mv_i) = \{(a, pr, k), \mid (mv_i, a, pr, k), a \in A, \\ pr \in PR,\ k = \text{high dividend}\}; \end{gathered} \tag{16}$$

The *professional skill* is to provide timely information about product prices with role as mobile retailer or wholesaler.

$$\begin{gathered} PS(mv_i) = (a, pr, r) \mid (mv_i, a, pr, r), \\ a \in A,\ pr \in PR,\ r \in \{\text{“seller},\ \text{wholesaler''}\}; \end{gathered} \tag{17}$$

The *responsibility* initiates an activity of sending quotation throug e-mail for the electronic product with role as dealer.

$$\begin{gathered} R(mv_i) = \{(a, r) \mid (mv_i, a, pr, r), \\ mv_i \in V,\ a \in A,\ pr \in PR,\ r \in \{\text{“dealer},\ \text{wholesaler''}\}\}; \end{gathered} \tag{18}$$

The *Financial status* initiates the action of purchasing maximum electronic products.

$$\begin{gathered} FC(mv_i) = \{(a, l),\ a \in A, \\ l \in \{maximum, minimum\}\}; \end{gathered} \tag{19}$$

4.3 M-BROKER (MBR)

The broker has to promote transactions by co-ordinating between potential sellers and buyers in M commerce environment. Let $MBR = \{mbr_1, mbr_2, .., mbr_{m_2}\}$ be a set of brokers, where m_2 represents total number of brokers. The mobile broker does not take the ownership of the product being sold, but receives commission from the mobile customer or mobile vendor or both of them.

The *goal* is to initiate transaction from vendor to mobile customer.

$$G(MBR) = \{(mc_{i1}, mv_{i2}, a) \ \mid (mbr_i,\ mc_{i1},\ mv_{i2}, a),\\ mbr_i \in MBR,\ mc_{i1} \in MC,\ mv_{i2} \in MV,\ a \in A\}; \tag{20}$$

The *professional skill* initiates the activity of collecting the timely information on product availability.

$$PS(MBR) = \{(a, pr) \ \mid (mbr_i, a, pr)\ mbr_i \in MBR\ a \in A, pr \in PR\}; \tag{21}$$

The *responsibility* of the broker is to arrange for money transfer transaction using mobile device with role as mediator.

$$R(MBR) = \{(a, d, r) \ \mid (mbr_i, a, d, r)\ mbr_i \in MBR,\\ a = \langle\text{“money transfer''}\rangle,, d = \langle\text{“mobile device''}\rangle,\ r = \langle\text{“mediator''}\rangle; \tag{22}$$

The *Financial status* of broker initiates the action of sending mobile advertisements to customer mobile device.

$$FC(MBR) = \{(a, d) \mid (mbr_i, a, d),\ mbr_i \in MBR,\ a \in A,\\ d \in \{\text{mobile device, smart phone}\}\}; \tag{23}$$

4.4 M-BANKER (MBK)

The banker performs the business of banking, which is described as payment of cheques, organizing current accounts, and also collection of cheques through mobile phone. A mobile banker also conducts the M-commerce transactions such as deposit accounts, loan amount payment and also exchange of bills for mobile participant. Let $MBK = \{mbk_1, mbk_2, .., mbk_{m_4}\}$, where m_4 represents total number of bankers involved in commercial business.

The *goal* is to conduct timely and secured transactions through mobile devices to build brand of the bank.

$$G(MBK) = \{(a, d) \mid (mbk_i, a, d), mbk_i \in MBK, a \in A,\\ d \in \{\text{mobile device, smart phone}\}\}; \tag{24}$$

The *professional skill* initiates the activity of opening the accounts to multiple mobile customers through the mobile device.

$$PS(MBK) = \{(a, mc_i, d) \ \mid (mbk_i, a, mc_i, d), mbk_i \in MBK\ a \in A,\\ mc_i \in MC,\ d \in \{\text{mobile device, smart phone}\}\}; \tag{25}$$

The *responsibility* is to collect mobile cheques from his customers with role as banker.

$$R(MBK) = \{(a, mbk_i, r) \mid (mbk_i, a, r)\ mbk_i \in MBK,\ a \in A,\ r = \langle\text{“Banker''}\rangle; \tag{26}$$

The *Financial status* allows to arrange for opening accounts of mobile participants at various counters.

$$\begin{aligned} FC(MBK) = \{(a, mp_i) \ &\mid (mbk_i, a, r),\ mbk_i \in MBK,\ a \in A, \\ &mp_i \in \{\text{Customer}, \text{Vendor}, \text{Broker}\}\}; \end{aligned} \tag{27}$$

In the following section, we describe the working of our business model to understand the concepts of business process.

5 Mobile Commerce Business System

A Mobile commerce business system consists of information about mobile participants (e.g., customer, vendor), products, set of roles and actions performed during a commercial transaction. In addition the database is created to store mobile commerce transactions. The relational algebraic model concept can be utilized for information composition at different levels. The information/data and process flow between distinct business entities is realized using relational algebra operators such as (σ, π, $\cup$, $\bowtie$, $\cap$). A transaction of product purchase is analyzed with following steps. The attributes for each entity is given by

LOCATION(<u>LocationID</u>, LocationName, LocationArea)

PRODUCT(<u>ProductID</u>, ProductPrice, ProductLocation)

VENDOR(<u>VendorID</u>, ProductType, ProductPrice, ProductBrand)

CUSTOMER(<u>CustomerID</u>, CustomerLoc, CustomerEd, CustomerFs)

Each customer is identified by ID, qualification and financial status

$$\text{exp}_1 = \sigma_{\text{CustomerID}=55}(\text{CUSTOMER}) \tag{28}$$

$$\begin{aligned} \text{exp}_1 \Rightarrow\ \text{CustomerID} : 55,\ \text{CustomerLoc} :\ \text{shoppingmall}, \\ \text{CustomerEd} : \text{B.E},\ \text{CustomerFs} : \text{medium} \end{aligned} \tag{29}$$

The exp_1 retrieves the record of customer with ID = 55.

$$\text{exp}_2 \leftarrow\ \text{CUSTOMER} \bowtie_{\text{ProductType } = \text{scientific cal}}\ (\text{VENDOR}) \tag{30}$$

The exp_2 gives information about vendors who sells scientific calculators.

$$\text{exp}_3 \leftarrow\ \pi_{\text{ProductPrice, ProductBrand}}\ (\text{exp}_2) \tag{31}$$

The exp_3 gives the price and brand of scientific calculator information to a business customer. We have proposed relational algebraic technique to model a specific business situation. The mobile commerce business model could further be expanded to include the roles and actions executed by participants.

6 Conclusion

In this paper, we propose a formal description of business model by identifying the specific characteristics of mobile participants and illustrate with examples in a particular trading scenario. The proposed analytical concepts assist in developing innovative user friendly interfaces and also to model e-business domain. In future, we want to extend our design aspects to implement context-aware business models by analyzing the context information and thereby assisting business participants to conduct transactions anytime, anywhere.

References

1. Charith, P., Harold, C., Jayawardena, S., Chen, M.: Context-aware computing in the internet of things: a survey on internet of things from industrial market perspective. IEEE Access J. **2**, 1660–1679 (2015)
2. Pushpa. P.V., Venkataram, P.: Context aware M-commerce services: C-IOB model approach. In: IEEE, 8th International Conference on Information, Communications and Signal Processing, pp. 13–16 (2011)
3. Pushpa P.V.: A generic commercial business model for customer oriented business transactions. In: IEEE International Conference on e-Business Engineering, (ICEBE), Macau, November, pp. 238–243 (2016)
4. Poylumenakou, A.K., Doukidis, G.J.: Building e-Business Models: An Analytical Framework and Development Guidlines, pp. 446–464 (2001)
5. Huang, W., Qi, L.Y., Dong, L.H.: Business models and implementations of M-commerce case studies and future research issues, pp. 3637–3640 (2007)
6. Younas, M., Mostefaoui, S.K.: Context-aware mobile service transactions. IEEE International Conference on Advanced Information Networking and Applications (2010)
7. Sievers, R.L.: The business environment, Six forces of influence (2006)
8. Ala-Siuru, P., Rantakokko, T.: Understanding and recognizing usage situations using context data available in mobile phones. In: Proceedings of Ubicomp Applications, pp. 17–21 (2006)
9. Papakiriakopoulos, D.A., Poylumenakou, A.K., Doukidis, G.J.: Building e-business models: an analytical framework and development guidlines. In: 14th Bled Electronic Commerce Conference on e-Everything: e-Commerce, e-Government, e-Household, e-Democracy, pp 25–26, June 2001
10. Chong, A.Y.-L.: Mobile commerce usage activities: the roles of demographic and motivation variables. Int. J. Technol. Forecast. Soc. Change **80**, 1350–1359 (2013). Elsevier
11. Varshney, U.: Business models for mobile commerce services: requirements, design, and the future. IT Prof. **10**, 1520–9202 (2008)
12. Panduranga, S.N.: Simplifying mobile commerce through a trusted transaction broker. In: IEEE ICPWC, pp 267–271 (2005)
13. Sharma, S., Gutierrez, J.A.: An evaluation framework for viable business models for m-commerce in the information technology sector. J. Electron. Markets **20**, 33–52 (2010). Springer
14. Noran, O.S.: Business Modelling: UML vs. IDEF. School of Computing and Information Technology (2000)

15. Al-Qirim, N.: Context-aware mobile business model discovery. J. Procedia Comput. Sci. **10**, 1180–1187 (2012). Elsevier
16. Da Silva, C.F., Hoffmann, P., Ghodous, P.: Improve business interoperability through context-based ontology reconcilation. Int. J. Electron. Bus. Manag. **9**, 281–295 (2011)
17. D'Antonio, F., Missikoff, M., Taglino, F.: Formalizing the OPAL eBusiness ontology design patterns with OWL. In: EKS - Laboratory for Enterprise Knowledge and Systems IASI - CNR, Rome, Italy
18. Missikoff, M., Schiappelli, F.: A method for ontology modeling in the business domain. In: LEKS, IASI-CNR Viale Manzoni, Italy

An Untraceable Identity-Based Blind Signature Scheme without Pairing for E-Cash Payment System

Mahender Kumar(✉), C. P. Katti, and P. C. Saxena

School of Computer and Science System, JawaharLal Nehru University,
New Delhi, India
mahendjnu1989@gmail.com,
{cpkatti,pcsaxena}@mail.jnu.ac.in

Abstract. Blind signature is an interesting cryptographic primitive which allows user to get signature on his document from signatory authority, without leaking any information. Blind signature is useful in many e-commerce applications where user's anonymity is the main concern. Since the Zhang et al., was the first to propose the identity based blind signature, many schemes based on bilinear pairing have been proposed. But the computational cost of pairing operation on elliptic curve is around 20 times the point multiplication on an elliptic curve. In order to save the running time, we present a new Identity-Based Blind Signature (ID-BS) scheme whose security is based on elliptic curve discrete logarithm problem (ECDLP). Performance comparison shows that proposed scheme reduces the cost of computation. Security analysis shows that proposed scheme is secure against the adversary and achieves the property of blindness and Non-forgeabillity. At the end; we propose an e-cash payment system based on our ID-based blind signature scheme.

Keywords: Blind signature · Identity-based encryption
Elliptic curve cryptography · Non-forgeability · Blindness

1 Introduction

Blind signature is an interesting cryptographic primitive which provides the user anonymity. This scheme allows user to get signature from signatory authority (SA) on his document without leaking any information about the document. Since, the notion of blind signature is first posed by Chaum [1, 2], many authors have presented their work on Blind Signature. All schemes are based on the traditional public key cryptosystem where certificate authority issue a digital certificate which binds the user's public key with his unique identity and public key infrastructure (PKI) is required for managing those certificates. In order to address the issue of certificate management and others issues such as key management and public key revocation, Shamir [3] proposed an idea, Identity-Based Cryptosystem (IBC), where user's public key is directly derived from his unique identity. To earn private key correspond to their identity ID, user requests the trusted third party, usually referred as the Private Key Generator (PKG). Nowadays, IBC is becoming very popular as compared to public key cryptosystem

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 67–78, 2018.
https://doi.org/10.1007/978-3-319-73423-1_7

(PKC) and implemented in many areas, e.g., forward encryption [4], delegate decryption [4], key exchange scheme [5], electronic-voting [6], electronic-cash payment system [7–9] etc.

Using the idea of Identity-based cryptosystem, Zhang and Kim's proposals [10, 11] were the first to pose the ID-Based Blind Signature. Later, Gao et al. [12, 13], Elkamchouchi and Abouelseoud [14], Rao et al. [15], Hu and Huang [16], He et al. [17], Dong et al. [18], Kumar et al. [19] presented ID-based blind signature schemes. But due to dependency on elliptic curve pairing operations, none was found efficient because pairing operations are very expensive as compared to the scalar multiplication operation on elliptic curves. Vanstone [20] claimed that system using 128-bit elliptic curve cryptography (ECC) key achieved the same security as using the 1024-bit RSA key. Additionally, ECC takes less power consumption and less storage space which provides strong processing time. In this paper, we mainly concentrate on posing anew ID-based Blind Signature scheme based on solving the difficulty of ECDLP problem. Proposed scheme satisfied the security requirements of blind signature and identity-based cryptosystem. At the end; we propose an e-cash payment system based on our ID-based blind signature scheme.

The remainder of the paper is arranged as follows: we briefly described the preliminaries in Sect. 2. Proposed ID-BS scheme is defined in Sect. 3. Section 4 includes the security analysis and computation comparison of our scheme against with existing schemes. Section 5 includes the e-cash system based on our proposed ID-BS scheme. Finally, conclusion and open problems are made in Sect. 6.

2 Preliminaries

2.1 Elliptic Curve Cryptosystem

Suppose the elliptic curve equation $y^2 = (x^3 + mx + n) modp$, where $x,\ y \in F_p$ and $4m^3 + 27n^2 modp \neq 0$. Formally, the Elliptic Curve is a set of points $(x,\ y)$ which satisfied the above equation and an additive abelian group with point 0 (identity element). The condition $4m^3 + 27n^2 modp \neq 0$ tells that $y^2 = (x^3 + mx + n) modp$ has a finite abelian group that can be defined based on the set of points $E_p(m,\ n)$ on elliptic curve. Consider points $A = (x_A,\ y_A)$ and $B = (x_B,\ y_B)$ over $E_p(m,\ n)$, the addition operation of elliptic curve is represented as $A + B = C = (x_C,\ y_C)$, defined as following: $x_C = (\mu^2 - x_A - x_B) modp$ and $y_C = (\mu(x_A - x_C) - y_A) modp$.

$$\text{Where, } \mu = \begin{cases} \left(\frac{y_B - y_A}{x_B - x_A}\right) modp, if\ A \neq B \\ \left(\frac{3x_A^2 + m}{2y_A}\right) modp,\ if\ A = B \end{cases}$$

Based on elliptic curve, Koblitz [21] and Miller [22] introduced elliptic curve cryptosystem. It is noted that addition operation and multiplication operation in ECC are equivalent to modular multiplication and modular exponentiations in RSA respectively.

Discrete Logarithm problem based on elliptic curve (ECDLP): Consider $B = sA$ where $A, B \in E_p(a, b)$, and $s \in Z_q$, it is computationally easy to compute B from A and s. But it is very difficult to compute s from B and A.

Extended Euclidean Algorithm: Extended Euclidean algorithm finds the modular inverse operation, which widely helpful in public key cryptosystem [23]. In addition to compute the *gcd* of two integer, say x and y, this algorithm express the $gcd(x, y)$ in linear combination of the form $gcd(x, y) = xp + yq$, for some integers p and q.

2.2 Framework of ID-BS Scheme

Definition 1 (Identity-Based Blind Signature): Our ID-Based Blind Signature protocol consists of Four Probabilistic Polynomial-Time (PPT) algorithms, namely, Setup, Extract, BlindSig, and Verifying, run among four entities, namely, Private Key Generator (PKG), Signatory Authority, Requester, and Verifier, where

1. *Setup*: On some security parameter k, PKG computes the system parameter (*PARAM*) and master secret key s. *PARAM* includes the public parameter which is published publically and s is known to PKG only.
2. *Extract*: On given inputs *PARAM*, master key s, and SA's Identity ID_S, PKG computes the private key S_{IDS} corresponding to identity ID_S.
3. *BlindSig:* This algorithm consists of four sub-algorithms, runs between the Requester and SA.
   a. *Commitment:* SA computes public parameters (Q_1, Q_2) against his secret values (n_1, n_2), delivers (Q_1, Q_2) to the Requester and keeps secret values (n_1, n_2).
   b. *Blinding:* Upon receiving the public parameters (Q_1, Q_2) and random chosen secret values (g, h, i, j, k, l), the Requester computes the Blinded Message (b_{M1}, b_{M2}) on given Message M. Now, Requester requests the SA to issue Signature on Blinded Message (b_{M1}, b_{M2}).
   c. *Signing:* For given Blinded Message (b_{M1}, b_{M2}), SA computes the Blind Signature $\left(S'_1, S'_2\right)$ using his private key S_{IDS} and delivers the Blind Signature $\left(S'_1, S'_2\right)$ to the Requester.
   d. *Stripping:* Upon receiving the Blinded Signature $\left(S'_1, S'_2\right)$, Requester strips it against his secret key to outputs the original Signature (S, R). Finally, Requester published the message-signature pair (M, S, R) for verification.
4. *Verifying*: Verifier takes (M, S, R) and SA's Identity ID_S as inputs, runs the verifying algorithm to verifies the Signature.

Two important constraints required against the security of ID-BS scheme are: Blindness property and Non-forgeability of additional signature under parallel chosen message and ID attacks. An Identity- Based Blind Signature is considered as secure if it fulfills the following two conditions:

Definition 2 (Blindness). Blindness property is defined in terms of following game playing between the challenger C and PPT adversary A.

- *Setup*: The challenger C chooses a security parameter k and executes the *Setup* algorithm to compute the published parameter *PARAM* and master key s. Challenger C sends *PARAM* to A.
- *Phase1*: A selects two distinct message M_0 and M_1 and an ID_i, and sends to C.
- *Challenge*: C uniformly chooses a random bit $b \in \{0, 1\}$ and ask A for signature on M_b and $M_{1\text{-}b.}$ Finally, C strips both the Signatures and gives the original signatures $(\sigma_b, \sigma_{1\text{-}b})$ to A.
- *Response*: A guesses bit $b' \in \{0, 1\}$ on tuple $(M_0, M_1, \sigma_b, \sigma_{1\text{-}b})$. A wins the game if $b = b'$ holds with probability $[b = b'] > 1/2 + k^{-n}$.

To define the Non-forgeability, let us introduce the following game playing between the Adversaries A who act as Requester and the Challenger C who act as honest SA.

- *Setup*: On random Security parameter k, the challenger C executes the *Setup* algorithm and computes the parameter *PARAM* and master key s. Challenger C sends *PARAM* to A.
- *Queries*: Adversary A can performs numbers of queries as follows:
  - *Hash function queries*: For requested input, challenger C computes the hash function values and sends it to the attacker A.
  - *Extract queries*: A selects an Identity ID and ask for S_{ID} to A.
  - *BlindSig queries*: A selects an ID and Message M blindly requested the Signature from $C.C$ compute signature on Message M with respect to ID.
- *Forgery*: Game is in favor of A, if against identity ID*, A response with n valid Message-Signature $(M_1, \sigma_1 = (S_1, R_1, r_1)),. (M_2, \sigma_2 = (S_2, R_2, r_2)).... (M_n, \sigma_n = (S_n, R_n, r_n))$ such that
  - Each message M_i is distinct from other Message M_j in given Message-Signature $(M_1, \sigma_1 = (S_1, R_1, r_1)), (M_2, \sigma_2 = (S_2, R_2, r_2))..... (M_n, \sigma_n = (S_n, R_n, r_n))$ set.
  - *Adversary A* is restricted to ask an extract query on Identity ID^*.
  - Execution of BlindSig algorithm is bounded by n.

Definition 3 (Non-forgeability). An ID-BS scheme is break by an Adversary A (t, q_E, q_B, k^{-n}), if A runs no more than t, A make Extract queries no more than q_E and runs *BlindSig* phase no more than q_B, with an advantage more than equal to k^{-n}. Under the adaptive chosen message and ID attacks, our ID-BS scheme is said to secure against one-more forgery, if no adversary A (t, q_E, q_B, k^{-n})-breaks the scheme.

3 Our Scheme: ID-BS Protocol

In this section, we introduce a new ID-BS scheme based on ECDLP. Suppose P be the generator of group G_1 of prime order q. Let the two cryptographic hash function $H_1 : \{0,1\}^* \rightarrow Z_q^*$ and $H_2 : \{0,1\}^* \rightarrow Z_q^*$. *absc(P)* denotes the abscissa of point P on G_1. Our scheme consists of four algorithm, as given in definition 1 in Sect. 2, runs as follows:

Setup: PKG select randomly $s \in Z_q$ and compute public key $P_{Pub} = s.P$. Publishes $PARAMS = \{G, q, P, P_{Pub}, H_1, H_2\}$, and keep secret key s secretly.

Extract: For a String identity $\mathrm{ID_S}$ and his master key s, PKG computes SA's private key $S_{IDS} = s.Q_{IDS} mod n$, where $Q_{IDS} = H_1(ID_S)$ and sends to the SA.

Blind Signature: This algorithm consists of four steps, runs between SA and the Requester as shown in Fig. 1.

Commitment: SA chooses two secret random integer $n_1, n_2 \in Z_q^*$. Computes Q_1, Q_2, q_1 and q_2 and publish them. Where,

$$Q_1 = n_1.P \in G_1 \, and \, q_1 = absc(Q_1) \in Z_q^*$$
$$Q_2 = n_2.P \in G_1 \, and \, q_2 = absc(Q_2) \in Z_q^*$$

Blinding: On given parameters (Q_1, Q_2, q_1, q_2) and Message M, Requester chooses four random numbers $g, h, i, j \in Z_q$ such that $gcd(i, j) = 1$. Selects two random number k and l such that $ki + lj = gcd(i, j)$ (according to the extended Euclidean algorithm). Now, Requester computes R_1, R_2, r_1, and r_2 and requests to the SA for Signature on Blinded Message ($\mathrm{b_{M1}}, \mathrm{b_{M2}}$). Where,

$$R_1 = g.i.Q_1 \in G_1 \, and \, r_1 = absc(R_1) \in Z_q^*$$
$$R_2 = h.j.Q_2 \in G_1 \, and \, r_2 = absc(R_2) \in Z_q^*$$
$$r = r_1.r_2 mod q \in Z_q^*$$
$$b_{M1} = k.H_2(M).q_1.r^{-1}.g^{-1} \in Z_q^*$$
$$b_{M2} = l.H_2(M).q_2.r^{-1}.h^{-1} \in Z_q^*$$

Signing: On given Blinded Message ($\mathrm{b_{M1}}, \mathrm{b_{M2}}$), SA creates the Blind Signature $\left(\mathrm{S}_1', \mathrm{S}_2'\right)$ using their private key $\mathrm{S_{IDS}}$ and sends it to the Requester, where,

$$S_1' = S_{IDS}.b_{M1} - q_1.n_1 \in Z_q^*$$
$$S_2' = S_{IDS}.b_{M2} - q_2.n_2 \in Z_q^*$$

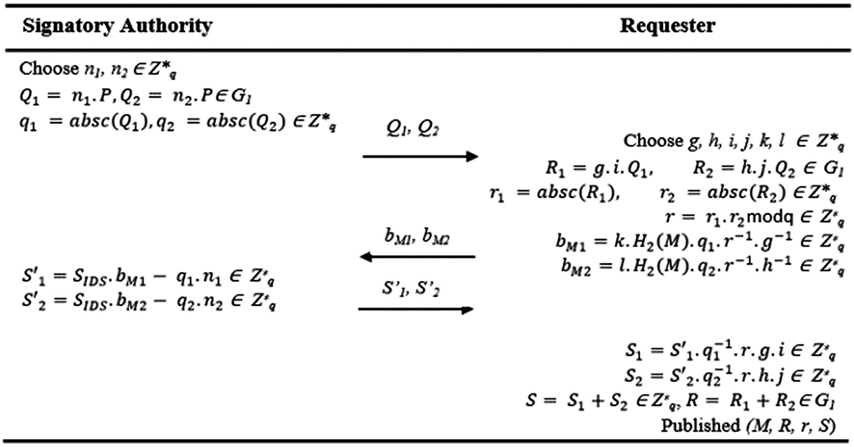
Stripping: On receiving the Blind Signature $\left(\mathrm{S}_1', \mathrm{S}_2'\right)$, the Requester strips and computes the actual signature $\sigma = (S, R, r)$. Where,

$$S_1 = S_1'.q_1^{-1}.r.g.i \in Z_q^*$$
$$S_2 = S_2'.q_2^{-1}.r.h.j \in Z_q^*$$
$$S = (S_1 + S_2) mod q \in Z_q^* \text{ and } R = (R_1 + R_2) mod q \in Z_q^*$$

Finally, requester publishes *(M, σ = (S, R, r))* for verification.

Verify: On given message-signature pair (M, $\sigma = (S, R, r)$), public parameter P_{Pub}, and $Q_{IDS,}$ user accepts the signature if and only if

$$P_{Pub}.Q_{IDS}.H_2(M) = S.P + r.R$$



Signatory Authority

Choose $n_1, n_2 \in Z^*_q$

$Q_1 = n_1.P, Q_2 = n_2.P \in G_1$

$q_1 = absc(Q_1), q_2 = absc(Q_2) \in Z^*_q$

$\xrightarrow{Q_1,\ Q_2}$

Requester

Choose $g, h, i, j, k, l \in Z^*_q$

$R_1 = g.i.Q_1, \quad R_2 = h.j.Q_2 \in G_1$

$r_1 = absc(R_1), \quad r_2 = absc(R_2) \in Z^*_q$

$r = r_1.r_2 \mathrm{mod} q \in Z^*_q$

$b_{M1} = k.H_2(M).q_1.r^{-1}.g^{-1} \in Z^*_q$

$b_{M2} = l.H_2(M).q_2.r^{-1}.h^{-1} \in Z^*_q$

$\xleftarrow{b_{M1},\ b_{M2}}$

$S'_1 = S_{IDS}.b_{M1} - q_1.n_1 \in Z^*_q$

$S'_2 = S_{IDS}.b_{M2} - q_2.n_2 \in Z^*_q$

$\xrightarrow{S'_1,\ S'_2}$

$S_1 = S'_1.q_1^{-1}.r.g.i \in Z^*_q$

$S_2 = S'_2.q_2^{-1}.r.h.j \in Z^*_q$

$S = S_1 + S_2 \in Z^*_q, R = R_1 + R_2 \in G_1$

Published (M, R, r, S)

Fig. 1. BlindSig algorithm of our proposed scheme.

Since, S = S_1 + S_2, then we have,

$$\begin{aligned}
S.P + r.R &= (S_1 + S_2).P + r.R \\
&= (S'_1.q_1^{-1}.r.g.i + S'_2.q_2^{-1}.r.h.j).P + r.R \\
&= ((S_{IDS}.b_{M1} - q_1.n_1).q_1^{-1}.r.g.i + (S_{IDS}.b_{M2} - q_2.n_2).q_2^{-1}.r.h.j).P + r.R \\
&= \left(S_{IDS}.b_{M1}.q_1^{-1}.r.g.i - n_1.r.g.i + S_{IDS}.b_{M2}.q_2^{-1}.r.h.j - n_2.r.h.j\right).P + r.R \\
&= \left(S_{IDS}.r.\left(b_{M1}.q_1^{-1}.g.i + b_{M2}.q_2^{-1}.h.j\right) - r.(n_1.g.i + n_2.h.j)\right).P + r.R \\
&= \left(S_{IDS}.r.\left(k.H_2(M).q_1.r^{-1}.g^{-1}.q_1^{-1}.g.i + l.H_2(M).q_2.r^{-1}.h^{-1}.q_2^{-1}.h.j\right)\right. \\
&\quad \left. - r.(n_1.g.i + n_2.h.j)\right).P + r.R \\
&= (S_{IDS}.H_2(M).(k.i + l.j) - r.(n_1.g.i + n_2.h.j)).P + r.R \\
&= (S_{IDS}.H_2(M) - r.(n_1.g.i + n_2.h.j)).P + r.R \\
&= S_{IDS}.H_2(M).P - r.(n_1.g.i + n_2.h.j).P + r.R \\
&= msk_{Pr}.Q_{IDS}.H_2(M).P - r.(R_1 + R_2) + r.R \\
&= P_{Pub}.Q_{IDS}.H_2(M) - r.R + r.R \\
&= P_{Pub}.Q_{IDS}.H_2(M)
\end{aligned}$$

This proved the correctness of proposed scheme.

4 Analysis of Our Proposed Scheme

4.1 Security Analysis

Theorem 1 (Blindness). *The proposed ID-BS scheme holds the property of blindness.*

Proof. Suppose adversary *A* which acts as SA and challenger *C* which acts as honest Requester, both involves in the *BlindSig* phase. *A* determines bit *b* with probability ½.

Let the information appearing during one of the execution of *BlindSig* phase in the view of *A* be $\left(b_{M1}, b_{M2}, S'_1, S'_2\right)$. Let the Signature be ($R = R_1 + R_2$, $S = S_1 + S_2$). There must be a tuple of random blinding factor (*g, h, i, j, k, l*) that maps the $\left(b_{M1}, b_{M2}, S'_1, S'_2\right)$ to ($R = R_1 + R_2$, $S = S_1 + S_2$).

Let $i = R_1.g.Q_1^{-1}$ and $\mathrm{g} = R_1.h.Q_2^{-1}$ such that there exist a pair of unique blinding factor (*g, h*) which satisfied the equations $\mathrm{S}_1 = \mathrm{S}'_1.\mathrm{q}_1^{-1}.\mathrm{r.g.i}$ and $\mathrm{S}_2 = \mathrm{S}'_2.\mathrm{q}_2^{-1}.\mathrm{r.h.j}$ respectively. However, it is intact to solve the blinding factor (*g, h, i, j*), we only need to exploit the existence of them. Let $k = \mathrm{b}_{\mathrm{M2}}.\mathrm{r.g.}q_1^{-1}.H_2^{-1}(\mathrm{M})$, so there exist a unique factor *l* that satisfied the equation $\mathrm{b}_{\mathrm{M2}} = \mathrm{l}.H_2(m).\mathrm{q}_2.\mathrm{r}^{-1}.\mathrm{h}^{-1}$.

Thus, there exist the blinding factors (*g, h, i, j, k, l*) which leads to the similar relation as in the *BlindSig* phase in Definition 1. Therefore, based on the hardness of ECDLP assumption, a strong adversary *A* determines *b* with probability 1/2 + $\mathrm{k}^{-\mathrm{n}}$.

Theorem 2 (Non-forgeability). *Under the hardness assumption of the ECDLP, our ID-BS Scheme is existential non-forgeable against the adaptive chosen message and identity attacks in the random oracle model.*

Proof. Suppose any PPT-bounded adversary *A* can forge ID-BS scheme under the adaptive chosen message and identity attack. Let a PPT-bounded algorithm *B* which helps *A* to solve the ECDLP problem, i.e. *A* would able to compute *x* from equation $Y = x.X$, where $x \in Z_q$ is unknown to *A*.

Setup: B considers P_{Pub} and gives public parameter $PARAM = \{G, q, P, P_{Pub}, H_1, H_2\}$ to *A*.

Queries: Adversary *A* can performs number of queries as follows:

Hash1 queries: B makes an empty list H_1^{List} having tuple (ID_i, $H_1(ID_i)$, a_i). When A queries to H_1^{List} at an Identity ID_i, *B* response as follows:

- *B* gives $H_1(ID_i)$ to *A*, if ID_i found in the H_1^{List} in tuple of (ID_i, $H_1(ID_i)$, a_i) or (*IDi*, $H_1(IDi)$, *).
- *B* sets $H_1(ID_i) = Q_{ID}$ and gives to *A* and adds the tuple (ID_{i}, $H_1(ID_i)$, *) to list H_1^{List}, if $ID_i = ID^*$.
- *B* chooses randomly $a_i \in Z_q$ and gives $H_1(ID_i) = a_i.P$ to *A* and adds tuple (ID_i, $H_1(ID_i)$, a_i) to list H_1^{List}, otherwise.

Since H1 is random oracle so $H_1(ID)$ gives no information to *A* until he queries H_1 oracle on *ID*.

Hash2 queries: *B* provides $M_j \in G_1$ on applying queries M_j to $H_2(M_j)$ and gives to *A*.

Extract queries: For some unknown $s \in Z_q$. Let $X = sP$, *B* computes $S_{IDi} = sH_1(ID_i) = a_i.X$, i.e. $H_1(ID_i) = a_i.P$. Now *B* sends S_{IDi} to *A*.

BlindSig queries: Suppose A wants to obtain a blind signature on message M_i with identity $\mathrm{ID_i}$. Let $\left(b_{M'1}, b_{M'2}\right)$ be blinded message which *A* gives to *B*. *B* response this queries as follows:

- If $ID_i \neq ID*$, using *IDi* corresponding to H_1^{List}, B finds the private key $S_{IDi} = a_iX$. Using S_{IDi}, *B* finds the blinded signature as in Sign phase in *BlindSig* algorithm.
- If $ID_i = ID*$, *B* sends $(b_{M'1i}, b_{M'2i})$ to A. Let $\sigma_i = (R_i, S_i, r_i)$ be corresponding response.

Forgery: A response with n valid Message-Signature $(M_1, \sigma_1 = (S_1, R_1, r_1)), (M_2, \sigma_2 = (S_2, R_2, r_2))..... (M_n, \sigma_n = (S_n, R_n, r_n))$ against identity $ID*$.

On applying the forking lemma, suppose adversary *A* creates two different valid blind signature (σ_A, σ_B) for message M, where

$$\begin{aligned}
&\sigma_A = (S_A, \mathrm{r}_A)\, and\, \sigma_B = (S_B, \mathrm{r}_B)\\
&S_A = S_{1A} + S_{2A}\\
&= S'_{1A}.q_{1A}^{-1}.r_A.g.i + S'_{2A}.q_{2A}^{-1}.r_A.h.j\\
&= (S_{ID}.b_{m1A} - q_{1A}.n_{1A}).q_{1A}^{-1}.r_A.g.i + (S_{ID}.b_{m2A} - q_{1A}.n_{1A}).q_{2A}^{-1}.r_A.h.j\\
&= S_{ID}.b_{m1A}.q_{1A}^{-1}.r_A.g.i - n_{1A}.r_A.g.i + S_{ID}.b_{m2A}.q_{2A}^{-1}.r_A.h.j - n_{1A}.r_A.h.j\\
&= S_{ID}.r_A.(b_{m1A}.q_{1A}^{-1}.g.i + b_{m2A}.q_{2A}^{-1}.h.j) - n_{1A}.r_A.(g.i + h.j)
\end{aligned}$$

Similarly, $S_B = S_{1B} + S_{2B}$

$$\begin{aligned}
&= S'_{1B}.q_{1B}^{-1}.r_B.g.i + S'_{2B}.q_{2B}^{-1}.r_B.h.j\\
&= (S_{ID}.b_{m1B} - q_{1B}.n_{1B}).q_{1B}^{-1}.r_B.g.i + (S_{ID}.b_{m2B} - q_{2B}.n_{2B}).q_{2B}^{-1}.r_B.h.j\\
&= S_{ID}.b_{m1B}.q_{1B}^{-1}.r_B.g.i - n_{1B}.r_B.g.i + S_{ID}.b_{m2B}.q_{2B}^{-1}.r_B.h.j - n_{1B}.r_B.h.j\\
&= S_{ID}.r_B.(b_{m1B}.q_{1B}^{-1}.g.i + b_{m2B}.q_{2B}^{-1}.h.j) - n_{1B}.r_B.(g.i + h.j)
\end{aligned}$$

Now, we compute

$$\begin{aligned}
S_B - S_A = {} & S_{ID}.g.i\big(b_{m1B}.q_{1B}^{-1}.r_B - b_{m1A}.q_{1A}^{-1}.r_A\big)\\
& + S_{ID}.h.j.\big(b_{m2B}.q_{2B}^{-1}.r_B - b_{m2A}.q_{2A}^{-1}.r_A\big)\\
& -(g.i + h.j).(n_{1B}.r_B - n_{1A}.r_A)\\
S_{ID}.\big(g.i\big(b_{m1B}.q_{1B}^{-1}.r_B & - b_{m1A}.q_{1A}^{-1}.r_A\big)\\
& + h.j.\big(b_{m2B}.q_{2B}^{-1}.r_B - b_{m2A}.q_{2A}^{-1}.r_A\big)\big)\\
& = S_B - S_A + (g.i + h.j).(n_{1B}.r_B - n_{1A}.r_A)
\end{aligned}$$

So, we can compute $\mathrm{S_{ID}}$ as follows:

$$\begin{aligned}
S_{ID} = {} & \big(g.i\big(b_{m1B}.q_{1B}^{-1}.r_B - b_{m1A}.q_{1A}^{-1}.r_A\big)\\
& + h.j.\big(b_{m2B}.q_{2B}^{-1}.r_B - b_{m2A}.q_{2A}^{-1}.r_A\big)\big)^{-1}.(S_B - S_A\\
& + (g.i + h.j).(n_{1B}.r_B - n_{1A}.r_A))
\end{aligned}$$

In order to compute S_{ID}, Adversary *A* should know the value of secret values (n_1, n_2) the Signatory Authority holds. To compute (n_1, n_2) is equivalent to solve the ECDLP problem. Alternatively, on given $(P, Q_{ID} = aP, P_{pub} = sP)$ it is easily to compute

$S_{ID} = sQ_{ID} = saP$ if the master key would not have compromised. But assuming the ECDLP problem is hard to solve, it is very difficult for an adversary A to compute S_{ID}.

4.2 Performance Comparison

In this section, we compared the computational cost of our proposal with other existing scheme. Since, our proposal has the advantages of Blind Signature, ECC, and IBC, the overhead of public key revocation and certificate management is eliminated and most time consuming cryptographic operation such as bilinear pairing on elliptic curve does not affect our proposal.

To achieve 1024-bit RSA level security for pairing-based cryptosystem, we assume the Tate pairing defined over super-singular elliptic curve on a finite field F_q, where $|q| = 512$ bits [24]. Same security level for ECC based scheme, we have to use secure elliptic curve on a finite field F_p, where $|p| = 160$ bits [24]. We assume e, E, M_{ecc} and M_{pair} as pairing, modular exponentiation, ECC-based scalar multiplication and pairing-based scalar multiplication with running time 20.01 ms, 11.20 ms, 0.83 ms and 6.38 ms respectively [24].

As compared to bilinear pairing operations, ECC-based scalar multiplication, pairing-based scalar multiplication and modular exponentiation, the computation cost of hash function operation is very less. Thus, we ignored the computation cost of hash function operation. So, in order to compare the performance, we just focus on the pairing operations, ECC-based scalar multiplication, pairing-based scalar multiplication and modular exponentiation.

Table 1. Comparison of our proposed scheme with existing schemes, in terms of running computational cost (in ms) and signature size (in Bytes).

Proposal	Running cost (in ms)		Size of signature
	BlindSig	Verify	
Zhang et al.'s proposal [10]	$1e + 6M_{pair} \approx$ *58.29*	$2e + 1E \approx$ *51.22*	*148B*
Gao et al.'s proposal [13]	$4e + 3M_{pair} \approx$ *99.18*	$4e \approx$ *80.04*	*384B*
He et al.'s proposal [17]	$5\ M_{ecc} \approx$ *4.15*	$3\ M_{ecc} \approx$ *2.49*	*104B*
Dong et al.'s proposal [18]	$6\ M_{ecc} \approx$ *4.98*	$4\ M_{ecc} \approx$ *3.32*	*104B*
Tian et al.'s proposal [25]	$2e + 6M_{pair} \approx$ *78.30*	$2e + 3_{pair}$ 59.16	*324B*
Our proposal	$4\ M_{ecc} \approx$ *3.32*	$3\ M_{ecc} \approx$ *2.49*	*104B*

Observation and result in [24, 26, 27] shows the running cost of pairing on elliptic curve, modular exponentiation operation and pairing-based multiplication operation is 24, 13 and 8 times the ECC-based multiplication operation. Using their observation, BlindSig algorithm in proposed proposal is 5.69%, 3.34%, 80%, 66.66% and 4.24% of Zhang and Kim's proposal [10], Gao et al.'s proposal [13], He et al.'s proposal [17], Dong et al.'s proposal [18] and Tian et al.'s proposal [25] respectively. The running cost of verify algorithm in proposed proposal is 4.86%, 3.11%, 100%, 75% and 4.20% of that in Zhang et al.'s proposal [10], Gao et al.'s proposal [13], He et al.'s proposal

[17], Dong et al.'s proposal [18] and Tian et al.'s proposal [25] respectively. Additionally, signature size generated in our proposal is 70.27%, 27.08%, 100%, 100% and 32.09% of that in Zhang et al. [10], Gao et al.'s proposal [13], He et al.'s proposal [17], Dong et al.'s proposal [18] and Tian et al.'s proposal [25] respectively, as shown in Table 1. Hence, the proposed ID-based blind signature gives better performance as compared against the previous schemes.

5 Application: E-Cash Payment System

In this section, we are presenting an online e-cash system based on our proposed ID-BS scheme. The proposed e-cash system consists of four entities: *Customers, Bank, Shop and Third Party*, which runs the six algorithms, namely, *Setup, Registration, Account-Opening, Withdrawal, Payment and Deposit*, to complete one transaction as given as follows:

Setup: Third party computes his public key against a random secret key. Third party publishes public parameter and keep secret key.

Registration: Third party registers and computes the bank private key corresponding to their unique identity and gives private key to bank.

Account-Opening: Customer requests for an account number to the Bank and got corresponding to his identity.

Withdrawal: Customer requests for an e-coin of face value f from Bank by providing his account information by running BlindSig sub-algorithm of our proposed ID-BS scheme. Bank verifies customer account by running Verify sub-algorithm, if correct, it releases e-coin (M, f, R, S, r) with face value f to customer.

Spending: With e-coin (M, f, R, S, r), Customer can purchase a product by paying amount f to shop. Shop first verifies the coin using Verify sub-algorithm. If it is valid, shop deposit this coin to the bank, otherwise, informs the customer for invalid coin.

Deposit: On receiving the e-coin (M, f, R, S, r), bank again checks the validity of e-coin by running the verify sub-algorithm. Bank adds this coin to his database, if the received coin is fresh, otherwise sends a warning message to shop for invalid e-cash.

6 Conclusion

In this paper, a new ID-BS scheme has been proposed that incorporates the benefits of Identity-Based Cryptosystem, Blind Signature and Elliptic Curve Cryptosystem whose security is based on the ECDLP. Additionally, under the random oracle model, proposed scheme is non-forgeable against the chosen message and identity attack, and holds the property of blindness. We compared our scheme with some existing schemes and found that our scheme gives better performance. Our proposed is suitable for implementing E-cash payment system. Proposed scheme suffers from key escrow problem which could be solved by using threshold key issuing [28], Hierarchical-Identity Based Encryption [29] techniques, etc.

Acknowledgement. This research work has been partially supported by the Council of Scientific and Industrial Research, a research and development organization in India, with sanctioned no. 09/263(1052)/2015 EMR-I and the UPE-II grant received from JNU. Additionally, the author would like to sincere thanks to the anonymous reviewers for their fruitful comments.

References

1. Chaum, D.L.: Untraceable electronic mail, return addresses, and digital pseudonyms. Commun. ACM **24**(2), 84–90 (1981)
2. Chaum, D.: Blind signatures for untraceable payments. In: Chaum, D., Rivest, R.L., Sherman, A.T. (eds.) Advances in Cryptology, pp. 199–203. Springer, Boston (1983). https://doi.org/10.1007/978-1-4757-0602-4_18
3. Shamir, A.: Identity-based cryptosystems and signature schemes. In: Blakley, G.R., Chaum, D. (eds.) CRYPTO 1984. LNCS, vol. 196, pp. 47–53. Springer, Heidelberg (1985). https://doi.org/10.1007/3-540-39568-7_5
4. Boneh, D., Franklin, M.: Identity-based encryption from the Weil pairing. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 213–229. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44647-8_13
5. Kumar, M., Katti, C.P., Saxena, P.C.: An ID-based authenticated key exchange protocol. Int. J. Adv. Stud. Comput. Sci. Eng. **4**(5), 11–25 (2015)
6. Gray, D., Sheedy, C.: E-voting: a new approach using double-blind identity-based encryption. In: Camenisch, J., Lambrinoudakis, C. (eds.) EuroPKI 2010. LNCS, vol. 6711, pp. 93–108. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22633-5_7
7. Chaum, D., Fiat, A., Naor, M.: Untraceable electronic cash. In: Goldwasser, S. (ed.) CRYPTO 1988. LNCS, vol. 403, pp. 319–327. Springer, New York (1990). https://doi.org/10.1007/0-387-34799-2_25
8. Islam, S.K.H., Amin, R., Biswas, G.P., Obaidat, M.S., Khan, M.K.: Provably secure pairing-free identity-based partially blind signature scheme and its application in online e-cash system. Arab. J. Sci. Eng. 1–14 (2016)
9. Kumar, M., Katti, C.P.: An efficient ID-based partially blind signature scheme and application in electronic-cash payment system. ACCENTS Trans. Inf. Secur. **2**(6) (2017)
10. Zhang, F., Kim, K.: ID-based blind signature and ring signature from pairings. In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 533–547. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-36178-2_33
11. Zhang, F., Kim, K.: Efficient ID-based blind signature and proxy signature from bilinear pairings. In: SN, R., Seberry, J. (eds.) ACISP 2003. LNCS, vol. 2727, pp. 312–323. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45067-X_27
12. Gao, W., Wang, G., Wang, X., Li, F.: One-round ID-based blind signature scheme without ROS assumption. In: Galbraith, S.D., Paterson, K.G. (eds.) Pairing 2008. LNCS, vol. 5209, pp. 316–331. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85538-5_21
13. Gao, W., Wang, G., Wang, X., Li, F.: Round-optimal ID-based blind signature schemes without ROS assumption (2012)
14. Elkamchouchi, H.M., Abouelseoud, Y.: A new blind identity-based signature scheme with message recovery. IACR Cryptol. ePrint Arch. 38 (2008)
15. Rao, B.U., Ajmath, K.A., Reddy, P.V., Gowri, T.: An ID-based blind signature scheme from bilinear pairings. Int. J. Comput. Sci. Secur. **4**(1), 98 (2010)
16. Hu, X.-M., Huang, S.-T.: Secure identity-based blind signature scheme in the standard model. J. Inf. Sci. Eng. **26**(1), 215–230 (2010)

17. He, D., Chen, J., Zhang, R.: An efficient identity-based blind signature scheme without bilinear pairings. Comput. Electr. Eng. **37**(4), 444–450 (2011)
18. Dong, G., Gao, F., Shi, W., Gong, P.: An efficient certificateless blind signature scheme without bilinear pairing. An. Acad. Bras. Cienc. **86**(2), 1003–1011 (2014)
19. Kumar, M., Katti, C.P., Saxena, P.C.: A new blind signature scheme using identity-based technique. Int. J. Control Theor. Appl. **10**(15), 115–124 (2017)
20. Vanstone, S.A.: Elliptic curve cryptosystem—the answer to strong, fast public-key cryptography for securing constrained environments. Inf. Secur. Tech. Rep. **2**(2), 78–87 (1997)
21. Koblitz, N.: Elliptic curve cryptosystems. Math. Comput. **48**(177), 203–209 (1987)
22. Miller, V.S.: Use of elliptic curves in cryptography. In: Williams, H.C. (ed.) CRYPTO 1985. LNCS, vol. 218, pp. 417–426. Springer, Heidelberg (1986). https://doi.org/10.1007/3-540-39799-X_31
23. Paar, C., Pelzl, J.: Understanding Cryptography: A Textbook for Students and Practitioners. Springer, Berlin (2009). https://doi.org/10.1007/978-3-642-04101-3
24. Cao, X., Kou, W., Du, X.: A pairing-free identity-based authenticated key agreement protocol with minimal message exchanges. Inf. Sci. (Ny) **180**(15), 2895–2903 (2010)
25. Tian, X.-X., Li, H.-J., Xu, J.-P., Wang, Y.: A security enforcement ID-based partially blind signature scheme. In: Web Information Systems and Mining, pp. 488–492 (2009)
26. He, D., Chen, J., Hu, J.: A pairing-free certificateless authenticated key agreement protocol. Int. J. Commun Syst **25**(2), 221–230 (2012)
27. Chen, L., Cheng, Z., Smart, N.P.: Identity-based key agreement protocols from pairings. Int. J. Inf. Secur. **6**(4), 213–241 (2007)
28. Shamir, A.: How to share a secret. Commun. ACM **22**(11), 612–613 (1979)
29. Boneh, D., Boyen, X., Goh, E.-J.: Hierarchical identity based encryption with constant size ciphertext. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 440–456. Springer, Heidelberg (2005). https://doi.org/10.1007/11426639_26

Context Information Based FOREX Services

P. V. Pushpa(✉)

Electronics and Communication Engineering Department,
P.D.A. College of Engineering, Glubarga, India
alladpushpa@gmail.com

Abstract. Context-aware computing is one of the key research issues in smart business applications, IoT paradigm and it is evident that it is successful in understanding the data sensed from trading environment. The success of trading depends on providing the proper personalized services to the trader based on his current context. The challenge here is to design and develop a context aware model which extracts meaningful information from raw data generated from underlying physical world. When there exists right information at right time it is possible to make effective decisions. This paper, proposes a C-IOB (Context-Information, Observation, Belief) hierarchical graphical structure model which signifies a causal relationship between low level context and high level context to reason about the customer and vendor context based beliefs. The proposed model provides accurate service with minimum computation and also reduces solution search space since the context information of a business entity is synthesized in the form of beliefs. Our research findings indicate the importance of the context based belief model in developing innovative Forex applications.

Keywords: Forex · Context · Customer · Trader · Belief · Service
Transaction

1 Introduction

A Forex environment deals with the business prospects of an economy in which activities like currency exchange, Traveler's cheque purchase, and other financial services takes place to meet the current needs of traveler's and traders. In this environment, exchange of bulk information takes place to meet the growing needs of customers/vendors. Since the Forex domain is highly dynamic as the new business need of an individual appears, there is a need to develop more accurate realistic context aware models. The challenge here, is to design an approach to analyze the context information of an entity (e.g., trader, transaction, currency, location) and provide timely information or service. Hence, the Forex applications are to be designed and developed to understand the current context to enhance the business experience of a customer.

Context awareness in Forex is about capturing a wide range of context attributes (e.g., location, exchange rate, social status, peak hours, etc.) of an

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 79–90, 2018.
https://doi.org/10.1007/978-3-319-73423-1_8

entity to better understand the business task and the kind of context aware service. It is important that a service must be offered based on a specific context because the kind of service is different according to the context of the user [1]. Hence, the services are provided by efficiently utilizing the context of user [2]. The aim of context aware computing is to extract and infer the context of an entity to provide intelligent Forex services. The context aware Forex applications have to be developed which effectively utilize the context information to adapt their behavior according to the needs of Forex business participant (e.g., customer, vendor, broker etc.). To extend context-aware applications into more cognitive domains includes observing the behaviors of customer environment, business processes, application processes, social events and even emotional and physical states of the customer/vendor.

The Foreign exchange market is the most liquid financial and largest market in the world. Forex is a part of global financial market and is used to invest in other countries, buying or selling of foreign currencies or even to buy foreign products. Forex traders (dealer, broker) are incessantly negotiating prices one among the others to succeed in the market. By building a broad integrated view of context information (physical, system, application and social) the services provided will be better able to meet the needs and aspirations of customers. The work [3,4] discusses the importance of deducing beliefs about a business entity (e.g., customer, vendor) which are primitive in most theories of decision making so that business applications can use these beliefs to develop intelligent user interfaces.

The proposed Context-Information, Observation and Belief (C-IOB) model [3,4] collects the relevant context information, forms observations which are further deduced to beliefs. It has the capability to adapt to real time situations of business needs, thereby enhancing the business customer satisfaction. The context based beliefs capture the relationships between alternative representation of the same type of context information to address multiple situations in Forex environment and they are primitive in most theories of decision making to provide intelligent Forex services. Our research findings indicate the importance of the proposed method in reducing data flow and solution search space in deriving higher level context. The C-IOB model can be further expanded to develop Smart business, e-Healthcare, Smart village and Tourist applications.

The rest of the paper is organized as follows. In Sect. 2, we briefly describe some of works related to context aware systems, Sect. 3, describes the formal representation of C-IOB model. In Sect. 4, we describe the categories of context information with examples and Sect. 5 discusses the application of C-IOB model to Forex system. Section 7, presents simulation environment and results and lastly the conclusion.

2 Related Works

Many researchers highlight the importance of context awareness in the development of context aware systems. A context-aware application is one which adapts its behavior to a continuously changing environment. These are some of the context aware works developed in real time applications for providing context aware services. The relationship between context awareness and user preferences/interests is exploited to adapt and provide the personalized services in [5]. [6] highlights the importance of context modeling (representation of user, the environment and the access mechanism) as the basis to provide personalization within mobile web search. An implementation of wearable system [7] learns context-dependent (location, activity, physiology) personal preferences by identifying individual user states in context-aware mobile phone. The work in [8] proposes a context based foreign exchange system in adhoc environment in which a resource capable node acts as a context manager to select a suitable vendor based on current context.

NAMA [9] considers the context, user profile with preferences to discover current needs and thereby providing personalized services to the user. A context-aware telecommunication service platform (CaTSP) [10] is developed to provide more intelligent personalization services according to user behavior, history, preference and current ambient environment. A multiagent framework [11] considers different contexts like location, device and user information to support personalized shopping-assistance service and multimedia selection service on wireless networks. A prototype system is implemented [12] to provide the personalized shopping services using context history. The paper [13] discusses the importance of using psychological characteristics of user for personalized recommendation in tourist guiding system. [14] proposes a novel context-aware service selection approach using Fuzzy analysis and it uses active context(user's devices, bandwidth of network, location, preferences) and passive context (weather and time) for service selection.

[15] discusses a common conceptual layered architecture for modern context-aware applications to improve extensibility and reusability of system. The situation of the user [2] is modeled through three dimensional space and the space describes the set of possible service access situations and in which identity, access position and time represent the three axis of the situation space. The context considered in [16] plays the role of filtering mechanism, thereby allowing the transmission of relevant data to the required device thus saving bandwidth and reducing query processing time. The paper [17] presents in depth classification of context information and proposes context information modelling technique using fuzzy set theory to incorporate the imprecise sensed context in IoT environment.

Although there have been several efforts to develop context aware applications, the process of acquiring knowledge using cognitive factors, establishing relationships between alternative representations of the same context is not focused. Therefore, it is essential that smart business applications have to be designed to reason about complex situations by deriving higher level context information for intelligent decision making and also to deal with large variety of contex information and uncertainity issues.

3 Formal Representation of C-IOB Model

The context based belief modeling (CBM) problem for context aware service is formalized as a triple

$$\mathrm{CBM} = \langle \mathrm{CI}, \mathrm{O}, \mathrm{B} \rangle \tag{1}$$

1. $\mathrm{CI} = \{\mathrm{C_{py}}, \mathrm{C_{sy}}, \mathrm{C_{ap}}, \mathrm{C_{sc}}\}$, where $\mathrm{C_{py}}$, $\mathrm{C_{sy}}$, $\mathrm{C_{ap}}$, $\mathrm{C_{sc}}$ represents set of physical, system, application and social environment context information of a business entity respectively.
2. $\mathrm{O} = \{\mathrm{ob_1}, \mathrm{ob_2}, \ldots, \mathrm{ob_L}\}$, where $\mathrm{ob_k}(1 \leq \mathrm{k} \leq \mathrm{L})$ is a set of observations formulated by summarizing the context information of an entity (e.g., person, place, thing or object) in a particular environment.
3. $\mathrm{B} = \{\mathrm{bf_1}, \mathrm{bf_2}, \ldots, \mathrm{bf_M}\}$, where $\mathrm{bf_j}(1 \leq \mathrm{j} \leq \mathrm{M})$ is a set of beliefs deduced based on set of observations.

The knowledge is extracted by using intuitive theory to model real world information by cause-effect relationship. The role of intuitive theories in learning and reasoning has been most prominently studied in the context of causal cognition [18]. The C-IOB model is a causal graphical model in which the context information, observation and beliefs are represented by nodes and they are connected by arrows, indicating the direction of causal dependencies from *context information* to *observations* and *observations* to *beliefs*. In the following section, we give the definition of context information, observation and belief.

Context Information: The context information refers to the perception and characterization of current state of a business entity (e.g., customer, vendor, place, transaction). The context information (e.g., business time, festival time etc.) plays significant role and can be utilized in several ways by business applications to model complex situations.

Observation: An observation is formulated by identifying the relevant context information of an entity. Observation enhances the prediction capability by continuously learning and receiving the knowledge of the physical world through perception for example *browsing for exchange rate, sending e-mail, looking for discount* etc.

Belief: The beliefs are deduced based on the observations. It is a descriptive thought that a person has about something and the people acquire beliefs through continuous learning. The notion of belief is considered as an informational attitude obtained from the history of experiences like observations. The beliefs generated are dynamic in nature and theses beliefs vary with the same context. Some of the beliefs about an entity are *competitive trader, trustworthy vendor, day trader, mini Forex trader* etc. The logical operators (AND ($\wedge$), OR($\vee$), NOT ($\neg$) are applied over formulated observations to construct predicates and some of the examples of belief deduction are given below.

1. Day-Customer $\Leftarrow$ High-end-Device $\wedge$ Analyzing-Market $\wedge$ Buying-and-Selling-Quickly

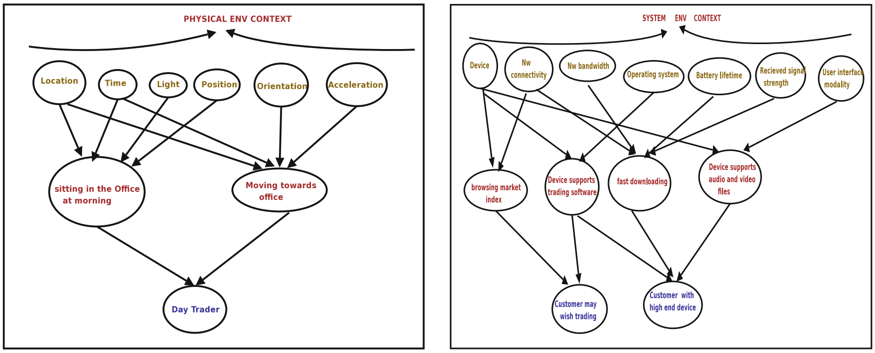
2. Reliable-Vendor ⇐ Provides-high-Quality-news ∨ Offers-Competitive-Spreads ∨ Offers-Optimum-Leverage
3. Neuroticism-customer ⇐ Feeling-Depressed ∧ More-Conscious

We categorize and present four types of context information with examples in the following section.

4 Context Information Types

The broad classification of context information is required, as future IoT/Pervasive computing have to deal with heterogeneous applications and services. The *context information* is defined as the constantly changing status of *physical*, *system*, *application* and *social* environment, when a business entity is executing Forex transactions.

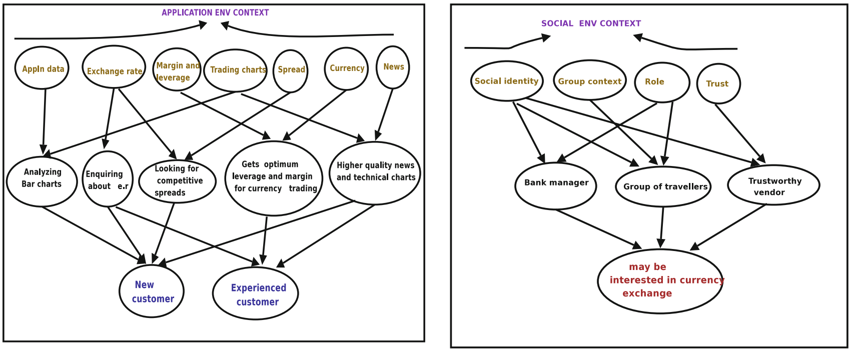
- *Physical environment context information*: It uniquely interprets and characterizes the context space of a physical entity (e.g., buyer, seller, place, dealer). Ex: *peak hours, business time, location of transaction, orientation of device,* etc.
- *System context information:* The system context information deals with computing aspects and finds its significance for the optimization of services in the heterogeneous environment. The system context includes the information about *device-type, battery power, memory, device-usage-history, device-interface and device-modality* of a Forex dealer. Figure 1(a) and (b) shows the examples of belief deduction by utilizing physical and system environment context information.
- *Application context information:* It describes the information related to specific application type and for example, in Forex applications it represents the information about *type of transaction, exchange rate, currency symbol, leverage ratio*and *currency spread.*



(a) Physical Environment (b) System Environment

Fig. 1. Belief deduction using physical and system context information

- *Social context information:* It describes information related to social aspects of an entity (e.g., dealer, social network) who is connected to global network of Internet of Things. Social identity is based on *nature of work*, such as *traveler* or *visitor*, economic status as *poor* or *rich*person, education level as *educated* or *uneducated.* Example of belief deduction by utilizing application and social context information is given in Fig. 2(a) and (b).



(a) Application Environment

(b) Social Environment

Fig. 2. Belief deduction using application and social context information

5 Application of C-IOB Model to Forex System

5.1 Context Based Forex Transactions

Each transaction represent a set of interactions or activities involved between business parties like customers and vendors for buying or selling of currency at particular location and time. Let A = $\{a_1, a_2, .., a_n\}$ represents the number of activities or actions initiated (e.g., *browsing for bar charts, downloading market analysis chart, downloading currency statement and so on).* Some of the examples of transactions are given in Table 1.

Table 1. Context based Forex transactions

Transaction	Description
T_1	Request for foreign currency buying based on specific time
T_2	Request for foreign currency exchange rate at particular location and time
T_3	Request for market analysis information during business hours
T_4	Request for Traveler's cheque purchase rate at particular time
T_5	Request for Traveler's cheque selling rate at particular time
T_6	Enquiring about foreign currency exchange rate during peak hours
T_7	Importing goods at specific location by currency exchange

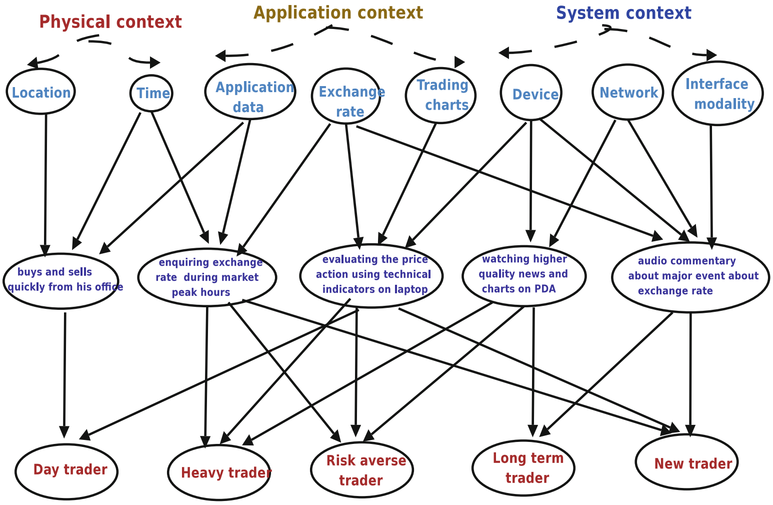

Physical context
Application context
System context
Location
Time
Application data
Exchange rate
Trading charts
Device
Network
Interface modality
buys and sells quickly from his office
enquiring exchange rate during market peak hours
evaluating the price action using technical indicators on laptop
watching higher quality news and charts on PDA
audio commentary about major event about exchange rate
Day trader
Heavy trader
Risk averse trader
Long term trader
New trader

Fig. 3. Example of Causal C-IOB model

5.2 Example of C-IOB Model

Whenever a customer requests for a service the current context information is acquired using CI-Constructs [3] to retrieve the context information from the multi-way datastructure. The thing/object used to acquire context information is given in Table 2. The beliefs are deduced by making suitable combination of four types of context information. An example of C-IOB model based on causal concept for establishing Forex transaction/service is given in Fig. 3.

Table 2. Acquisition of context information

Thing for data acquisition	Acquired context information
System clock	Time is 10 a.m
Proximity sensor, GPS receiver (Passive Infrared Sensors or Capacitive sensors)	Near shopping mall
Logical sensors (operating system, APIs)	PDA, Wi-Fi network
Access point, logical sensors (operating system APIs)	Bandwidth
Age, education, profession	Elder person, M.B.A, Trader
Login/password	Bank manager/new customer
Calendar information	Social event
User feedback/rating	Trustworthiness of vendor

5.3 C-IOB Model Evaluation

The three different causal structures of C-IOB model is shown in Fig. 4. The first structure shows the deduction of beliefs based on Ten context parameters chosen from four environments. The other two structures gives the modification in the deduced belief because of absence of certain context parameters. Therefore, we can address and interpret multiple situations in Forex by extracting specific context in the form of beliefs and thereby we can conclude that certain/relevant context parameters are always required to provide the personalized Forex services. Some of examples of customer context based beliefs and the specific services are given in Table 3.

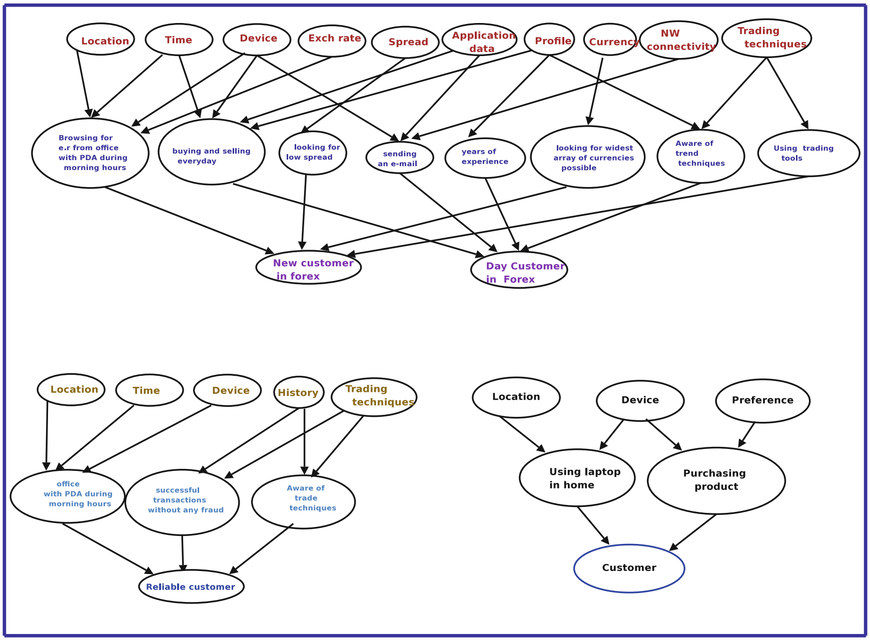


Fig. 4. Alternative structures of C-IOB model

Table 3. Customer context based beliefs and services

Beliefs	Services
Day customer	Providing proper Forex trading startegy
Low currency buyer	Mini account broker service
Customer	Import and export of both goods and services

6 Simulation and Results

6.1 Simulation Environment

The simulation environment is established by considering 550 customers and 100 Forex vendors who are involved in Forex currency buying, selling, traveler's cheque purchase and other transactions. The context information from four environments is collected, for 500 customers who are frequent travelers, novice customers, who wish to exchange currency based on their need. The proposed model has been simulated on hybrid wireless network as shown in Fig. 5. It consists of Samsung grand Smartphone and Mobilephone connected to Linksys access point with 802.11n connectivity and a Laptop with wireless connectivity. The system is very much consistent and competitive to realistic mobile environments. The laptop is used as vendor device, the mobilephone and smartphone are used as customer devices.

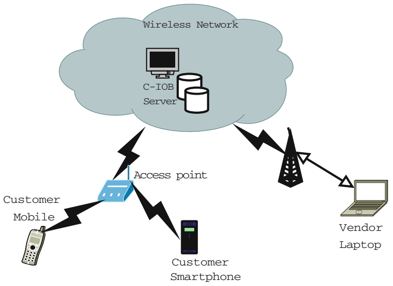


Fig. 5. Hybrid wireless network

The simulation is carried out to test several Forex transactions. The database is created for 1000 customers, 100 vendors and are characterized by their *identity, name, address, phone number, profession and economic status.* The context environment is varied by changing the parameters such as location, time, currency type, exchange rate, spread and so on. When customer device sends a request for a transaction/service, the current context information is acquired, using menu driven programs [3]. The C-IOB server formulates a set of observations, which are further deduced into beliefs and the simulation exhaustively tests the working of C-IOB model under distinct environments.

The Fig. 6 shows the Context Selection Index (CSI) when business participant is either customer, vendor, broker or banker. For a *customer* the physical and social context information is important, because of his frequent change in location and in addition influenced by friends or family in Forex transaction. For a *vendor* and *banker* the percentage of application and social context information contributes more. Lastly, the percentage of social context information is high in belief deduction for a broker as he has to establish a transaction between buyer

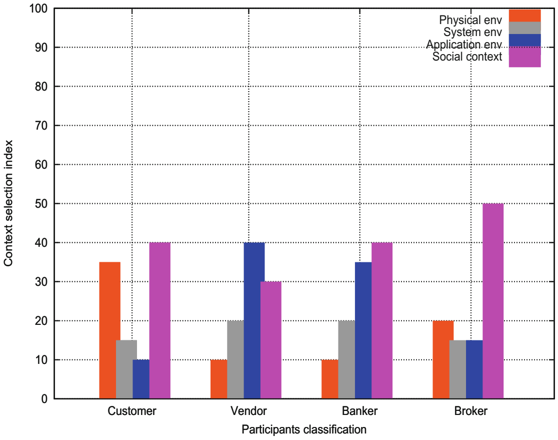


Fig. 6. The CSI of four business participants

Table 4. Context weights chart

Business traders	w_1	w_2	w_3	w_4
Customer	0.33	0.15	0.11	0.41
Vendor	0.11	0.21	0.37	0.31
Banker	0.10	0.21	0.33	0.36
Broker	0.22	0.14	0.15	0.49

and seller by connecting through social network. The context weights chart for business traders are given in Table 4.

The percentage of context information is calculated based on the following formula.

$$\gamma_i = \frac{m_3 \star w_i}{\sum m_3 \star w_i} \tag{2}$$

where, $\gamma_i \mid$ i = {1 = physical, 2 = system, 3 = application, 4 = social}, represents the percentage of context information, m_3 is the number of input context parameters used in belief formation and w_i is the weight associated with each environment and $\sum_{i=1}^{4} w_i = 1$. Therefore *Context Selection Index* assists in designing innovative systems.

The average processing time of context based transaction is given by

$$T_{pt} = N\ (T_{acq}) + T_{of} + T_{bf} + T_{se} \tag{3}$$

where, 'N' represents the number of context parameters used in belief deduction and description of these parameters is given in Table 5.

The graph shown in Fig. 7 gives processing time with variation in number of customer requests. The average processing time for executing a particular Forex

Table 5. Context aware transaction time parameters

Symbol	Description
T_{acq}	Context acquisition time
T_{of}	Observation formulation time
T_{bf}	Belief formulation time
T_{se}	Transaction selection and execution time

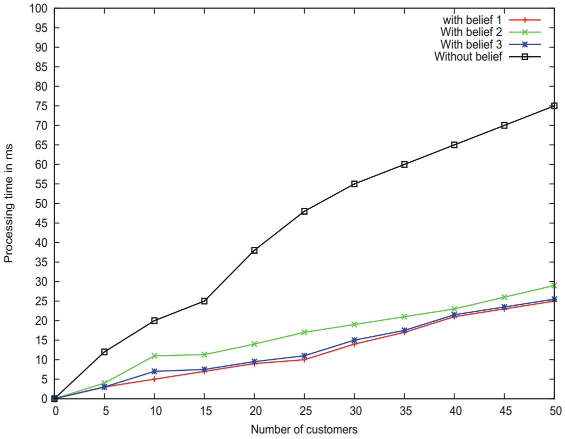


Fig. 7. Processing time vs number of customer requests

transaction is less with belief based request compared to context unaware approach, since the traditional request takes more time in searching the information about the vendor, place and also peak hours.

7 Conclusion

In this paper, we have proposed C-IOB model based on cause-effect relationship by extracting the specific context information of a business entity. The proposed model provides accurate context-aware services and has the capability to adapt to real time situations of business needs by deducing beliefs based on observations, thereby enhancing the customer satisfaction. The simulation results have shown that the time to execute Forex transactions is less with context based beliefs. In our future work, we incorporate qualitative and credible reasoning techniques to address the issues related to imprecise or uncertain context information to make the future computing paradigm more realistic towards real world implementation. As a result, it is possible to obtain more accurate decisions and provide intelligent or new services to the trader.

References

1. Lee, W.P.: Deploying personalized mobile services in an agent-based environment. Expert Syst. Appl. **32**(4), 1194–1207 (2007)
2. Figge, S.: Situation-dependent services? A challenge for mobile network operators. J. Bus. Res. **57**, 1416–1422 (2004)
3. Pushpa, P.V.: Customer context based transactions in mobile commerce business environment. In: 13th International Conference on e-Business Engineering. IEEE, Macau, China, November 2016
4. Pushpa, P.V., Venkataram, P.: Context aware M-commerce services: C-IOB model approach. In: 8th International Conference on Information, Communications and Signal Processing, pp. 13–16. IEEE (2011)
5. Byun, H.E., Cheverst, K.: Exploiting user models & context-awareness to support personal daily activities. Workshop on user modeling for context-aware applications, UM (2001)
6. Arias, M., Cantera, J.M., Vegas, J.: Context-based personalization for mobile web search, pp. 24–30. ACM (2008)
7. Krause, A., Smailagic, A., Siewiorek, D.P.: Context-aware mobile computing: learning context-dependent personal preferences from a wearable sensor array. IEEE Trans. Mob. Comput. **5**, 113–127 (2008)
8. Pushpa, P.V., Venkataram, P.: An implementation of context based foreign exchange system. In: International Conference on Mobile, Ubiquitous and Pervasive computing, Rio De Janerio, Brazil, 29–31 March 2010
9. Kwon, O., Choi, S., Park, G.: NAMA: a context-aware multi-agent based web service approach to proactive need identification for personalized reminder systems. J. Expert Syst. Appl. **29**, 17–32 (2005)
10. Qiao, X., Li, X., Liang, S.: Reference model of future ubiquitous convergent network and context-aware telecommunication service platform. J. CHUPT **13**, 50–56 (2006)
11. Lee, W.P.: Deploying personalized mobile services in an agent-based environment. J. Expert Syst. Appl. **32**, 1194–1207 (2007)
12. Hong, J., Suh, E.H., Kim, J., Kim, S.: Context-aware system for proactive personalized service based on context history. J. Expert Syst. Appl. **36**, 7448–7457 (2009)
13. Bai, Y., Yang, J., Qiu, Y.: FD/I-Based personalized recommendation in context-aware application. In: IEEE International Conference on Multimedia and Ubiquitous Engineering (2007)
14. Long, X., Kuo, G.S.: A novel dynamic fuzzy analysis hierarchy model enabling context-aware service selection in IMS for future next-generation networks, pp. 2814–2818. IEEE (2008)
15. Baldauf, M.: A survey on context-aware systems. Int. J. AdHoc Ubiquitous Comput. **2**(4), 263–277 (2007)
16. Doulkeridis, C., Vazirgiannis, M.: CASD: management of a context-aware service directory. Proc. Pervasive Mob. Comput. **4**, 737–754 (2008)
17. Pushpa, P.V.: Context information modelling for internet of things. In: IEEE International Conference on Contemporary Computing and Informatics (IC3I). Amity University, Noida, December, pp. 425–231 (2016). (Invited Paper)
18. Gopnik, A., Glymour, C.: Causal maps and Bayes nets: a cognitive and computational account of theory-formation. In: The Cognitive Basis of Science, pp. 117–132. Cambridge University Press (2002)

Markov Chain Based Priority Queueing Model for Packet Scheduling and Bandwidth Allocation

Reema Sharma[1], Navin Kumar[2](✉), and T. Srinivas[3]

[1] Department of ECE, The Oxford College of Engineering, Bangalore, India
sharma80reema@gmail.com
[2] Department of ECE, Amrita School of Engineering, Amrita Vishwa Vidyapeetham University, Bangalore, India
navinkumar@ieee.org
[3] Department of ECE, Indian Institute of Science, Bangalore, India
tsrinu@ece.iisc.ernet.in

Abstract. This paper considers classification of diverse traffic types in Internet of Things (IoT) based on importance of data rate, packet size and proposes a priority-based probabilistic packet scheduling strategy for efficient packet transmission. Reduction of peak resource usage, dynamic control of service rate corresponding to arrival rate and QoS buffer management are few main factors considered to develop this strategy. By calculating percentage of link bandwidth required for prioritized traffic in each cycle, we provide quality of service (QoS) to real time traffic in IoT and non-IoT applications. Different experiments including MPEG traffic traces and Poisson traffic are conducted to verify the proposed scheduler. Also, performance of scheduler for both IoT and Non-IoT applications is compared for different data rates. We observe that the proposed packet scheduler satisfies QoS requirements for both IoT and non-IoT traffic.

Keywords: Internet of Things · Average queue length · Quality of service
Packet scheduling algorithm · Delay sensitive applications
Service differentiation

1 Introduction

IoT devices offer numerous novel real world services which result in heterogeneous QoS constraints and brings forward the requirement of a scheduling scheme to achieve overall optimum performance. Current studies on classification and scheduling of IoT services [1–3] rarely considers size of packets, its data rate, type of packets and comparison with Non-IoT data. Also, stringent delay constraints and high bandwidth requirements of multi-user video transmission applications [4], and to provide adequate transmission opportunities to all video/image senders before their tolerable delay deadlines is a longstanding research problem in IoT. This paper considers some of these issues to investigate optimal approach for assigning scheduling priority levels and allocates required bandwidth.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 91–103, 2018.
https://doi.org/10.1007/978-3-319-73423-1_9

IoT has a broad research scope in several areas like healthcare, smart environments, structural health monitoring and transportation, etc. [1, 5]. Applications like structural health checking generally needs unfailing information release from every node to the destination node. Furthermore, the QoS necessity of traffic blocking is comparatively rigorous in terms of throughput and delay because of the association of critical continuous data. In each of these applications, lightweight smart objects are active participants which are capable of sensing different incidents and communicating it to various other devices. Current methods do not offer acceptable solutions for delay sensitive applications. Few slot allocation policies like rate adaptive round robin or round robin provide assured QoS but the allotment of the current slot is not based on the allotment of the previous slots and these policies are hence considered to be stationary.

In this paper, a policy is developed to consider immediate release of delay sensitive data in IoT applications. We are addressing the scheduling scheme for the packets to provide QoS services for different class of services. Policy is actively calculating the number of packets to be scheduled in current cycle from high priority queue based on increase or decrease in its average queue length in consecutive cycles. This results in assuring adequate allotment of bandwidth to each service class by avoiding excess allocation always to the high priority class. After assigning required bandwidth to high priority class, remaining bandwidth can be assigned to low priority class. The major contributions of this paper are:

- Classifying and scheduling packets based on its size, type and data rates to achieve less transmission delay for each priority class
- Analyzing theoretically with Markov Chain model and simulation experiment with MATLAB R2013 to explore and study the waiting time of packets for various priority-queuing schemes and discover out the optimal one
- Comparison of model for IoT and Non-IoT applications.

The rest of the paper is organized as follows. Few recent investigated works are discussed in Sect. 2. Probabilistic model and its assumptions are presented in Sect. 3. Section 4 discusses the detailed analysis. Simulation results are presented and discussed in Sect. 5. Finally, conclusion and future work is included in Sect. 6.

2 Related Work

Recently, IoT has attracted researchers from both industry and academia. Current research explores into various phases of IoT such as service oriented architecture based IoT [6], Web of Things [7], applications and clarifications related to IoT [8] and therefore various issues can be investigated. Many authors [9–13] presented surveys on IoT vision, IoT related projects, IoT enabling technologies, research issues like privacy, trust, energy consumption and resource insufficiency with certain application areas in IoT. Some of them are very important issues which provide useful discussion about QoS requirements but design of QoS models to provide priority to emergency applications are not extensively discussed.

Klepec and Kos [11] proposed a priority model with two queues and presented packet transit time behavior for a delay susceptible application for which bandwidth limit should be the least. The model is straightforward; the exploitation of higher priority was shown at the price of more packet losses for low priority data. Moreover, in order to analyze and monitor energy efficiency [14]; network topologies; issues related to performance of the network [15]; and the accessibility of bandwidth; a number of new methods have been devised. A buffer sharing scheme for resource distribution in wireless local area networks (WLAN's) under diverse traffic conditions is discussed in [16]. The results show that for heterogeneous traffic loads, transmission opportunities are not equally allocated by 802.11. They also showed that large buffer can help in providing this equality but at the expense of increased delay. The solution to delay sensitive applications is not efficient as discussed in the above studies. To guarantee instantaneous communication with no packet loss, queueing delay and specifically to address the delay critical applications, an efficient packet scheduling scheme with service priorities becomes necessary.

3 Probabilistic Model and Its Assumptions

The probabilistic representation and few hypotheses considered are discussed in this section. In the projected model, it is assumed that the data packets are categorized into prioritized (critical data) and non-prioritized (non critical data) traffic and is accumulated in two separate queues. The scheduler calculates the current departure packets using number of departure packets in the previous slot and the number of arrival packets in the current slot. It calculates a weighting coefficient for each queue which represents average number of packets that can be scheduled from the queue before moving to the next queue. The idea is based on weighted round robin scheduler. In this, the system serves each queue in a round robin manner and the calculated dynamic weights are assigned. Scheduling is executed at the beginning of each cycle. The scheduling mechanism differentiates the services based on the priority, by measuring the probability of traffic increased at the current time slot from the previous time slot, predicts the number of packets to be scheduled at the current time slot and the amount of bandwidth to be allocated to each service.

The configuration of the node and scheduling system is given in Fig. 1. Consider that the prioritized traffic queue has a maximum size of *B1* and average buffer size of P_{avg} and non-prioritized queue has a maximum size of *B2*. It is assumed that packets arrive separately for all service classes follow a Poisson procedure with a mean rate of arriving packets per cycle as $\lambda = \sum_{\mathrm{i=0}}^{\mathrm{i_{max}}} \mathrm{i.P(i)}$, the same is depicted in Fig. 1, where P(i) denotes the probability of *i* packets arriving in one round and i_{max} specifies the upper limit to packets arriving. Let $\lambda_1, \lambda_{\mathrm{HQ}}, \lambda_{\mathrm{HL}}$ are the packets arrived, accumulated in the buffer and dropped from the prioritized queue respectively, and $\lambda_2, \lambda_{\mathrm{LQ}}, \lambda_{\mathrm{LL}}$ are quantity of packets arrived, accumulated in the queue and dropped from non-prioritized queue respectively. The system consists of *N* cycles, each of which is further partitioned into different time slots. Here, every time slot carries a packet of variable size. Now, the system model formulation to compute the average quantity of packet serviced

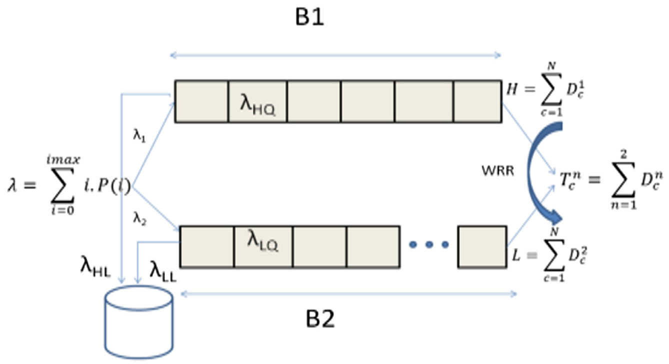

B1
λ_{HQ}
$H = \sum_{c=1}^{N} D_c^1$
WRR
$T_c^n = \sum_{n=1}^{2} D_c^n$
$\lambda = \sum_{i=0}^{imax} i.P(i)$
λ_1
λ_2
λ_{LQ}
$L = \sum_{c=1}^{N} D_c^2$
λ_{HL}
λ_{LL}
B2

Fig. 1. System model

T_c^n in cycle ***c*** for n^th^ class can be designed. The corresponding buffer queue size of the current cycle ***c*** is calculated as:

$$q_c^n = q_{c-1}^n - D_c^n + a_c^n \tag{1}$$

where a_c^n indicates quantity of packets which arrives in the n^th^ queue during cycle c; D_c^n signifies the quantity of packets leaving from n^th^ queue at cycle ***c*** and q_{c-1}^n represents quantity of packets stored in the queue during cycle (c − 1). The quantity of packets served in cycle ***c*** can be given as:

$$D_c^n = \min\left(q_c^n, b_c^n\right) \tag{2}$$

where, b_c^n is the quantity of packets which can be served according to the bandwidth availability and computed based on the anticipated method for n^{th} queue during cycle ***c*** (details given later). The respective queue is picked up in a round robin manner, thus *n* can be computed as:

$$n = (n' \bmod 2) + 1 \tag{3}$$

where, n' represents the previous class selected, n = 1 represents the prioritized service and n = 2 represents the non-prioritized service queue. Probability of average queue length of prioritized queue is calculated as:

$$P_{avg}(c) = 0.01 * P_{avg}(c-1) + (1-0.01) * P(q_c^n) \tag{4}$$

where $P_{avg}(c)$ is the probability of average queue size in current cycle, $P_{avg}(c-1)$ is the probability of average queue size in previous cycle and $P(q_c^n)$ is the probability of instantaneous queue size at current cycle. The sum of the packets serviced in one round can be computed as: $T_c^n = \sum_{n=1}^{2} D_c^n. H = \sum_{c=1}^{N} D_c^1$ and $L = \sum_{c=1}^{N} D_c^2$ are sum of prioritized and non-prioritized packets departed in *N* cycles where c = 1,2 …N.

To dynamically allocates weights with standard scale values, prioritized queue is assigned with two thresholds $T_{min} = 0.083$ and $T_{max} = 0.3667$ which act as indicators to achieve desired and acceptable QoS parameters. At these threshold values, the least blocking probability values for the considered simulation scenario are obtained. Assume *p* is the probability of serving prioritized packets and *q* is the probability of serving the non-prioritized packets. As, only two service queues are taken, the probability of serving the non prioritized packets can be given by $q = 1 - p$. Probability *p* can be further distinguished based on proposed scheduler into three different cases:

$$p = \begin{cases} p_1 & \text{for}\, 0 \le P_{avg} \le T_{min} \\ p_2 & \text{for}\, T_{min} < P_{avg} < T_{max} \\ p_3 & \text{for}\, B1 > P_{avg} \ge T_{max} \end{cases} \tag{5}$$

where, $p_1 = 0.3$, is the probability of weight allocated to prioritized packets when the probability of average queue length P_{avg} is between 0 and T_{min}. This value of p_1 is chosen to provide minimum bandwidth to prioritized queue irrespective of the arrival rate. The p_2 is the probability of weight allocated to prioritized packets when average queue length increases and lies between T_{min} and T_{max} and is calculated by Eq. (6) and $p_3 = 0.7$ to 0.9 is the probability of weight allocated to prioritized packets when the average queue length increases beyond T_{max}. The value of p_3 are chosen to limit maximum bandwidth allotted to prioritized queue and to provide some processing of non-prioritized queue and reduce its blocking probability while providing guaranteed service to prioritized queue in each cycle.

$$p_2 = p' + (P_{avg}(c) - P_{avg}(c-1)).\frac{0.3}{(T_{max} - T_{min})} \tag{6}$$

where p' the probability of weight assigned to prioritized queue in previous cycle. $P_{avg}(c) - P_{avg}(c-1)$ is the change in probability of average queue (increase or decrease) in consecutive cycles. Eq. (5) shows a linear relationship between probability of weights allocated to priority service and probability of average queue size.

4 Model Analysis

The system is depicted by a probabilistic Markov Chain model. Since the investigational process of all transitional nodes is similar, a node is picked up arbitrarily to examine the algorithm because this scheduling algorithm can independently work in each router to schedule packets based on arrival rate. Knowing the scheduling time spent in one node and total nodes existing in the path chosen by the routing algorithm for a particular topology, the processing delay can be found. The blocking probability of prioritized and non-prioritized class can be calculated. The Markov chain model formulation to compute the average packets scheduled or departures T_c^n in cycle ***c*** for n^{th} service class is as follows.

Figures 2(a), (b) and (c) presents the state transition diagrams for state (x, y), where (x, y) represents a state in which *x* non-prioritized packets and *y* average prioritized packets are stored in their respective queues. P(x, y) is the probability of the system being in state (x, y). Four cases are discussed here: (i) when x < B2 and y < B1 (Fig. 2 (a)) (ii) when x $\geq$ B2 and y < B1 (Fig. 2(b)) (iii) when x < B2 & y $\geq$ B1. (Figure 2 (c)) (iv) when x $\geq$ B2 and y $\geq$ B1. Based on Figs. 2(a), (b) and (c), the balance equations for state (x, y) are computed as:

$$\begin{aligned}&(p\lambda_1 + (1-p)\lambda_2 + y\mu_1 + x\mu_2)P_{x,y} - p\lambda_1 P_{x,y-1}(1-p)\lambda_2 P_{x-1,y}\\ &\quad - (y+1)\mu_1 P_{x,y+1} - (x+1)\mu_2 P_{x+1,y} = 0\end{aligned} \tag{7}$$

$$(p\lambda_1 + y\mu_1 + x\mu_2)P_{x,y} - p\lambda_1 P_{x,y-1} - (1-p)\lambda_2 P_{x-1,y} - (y+1)\mu_1 P_{x,y+1} = 0 \tag{8}$$

$$((1-p)\lambda_2 + y\mu_1 + x\mu_2)P_{x,y} - p\lambda_1 P_{x,y-1} - (1-p)\lambda_2 P_{x-1,y} - (x+1)\mu_2 P_{x+1,y} = 0 \tag{9}$$

In Fig. 2(a), when non-prioritized queue is filled, the new coming packets will be dropped which is shown by returning back to the same state. Similarly, as soon as the prioritized queue is filled then those incoming packets are dropped as shown in Fig. 2 (b). To obtain the blocking probabilities of service classes, the above equations need to be solved to obtain state probabilities P(x, y). So, consider a non-complex system to solve blocking probability. The corresponding state transition diagram is shown in Fig. 2(d). To make the computation easier, consider $\mu1 = \mu2 = \mu$ where the service rate μ is taken as the average service rate of two traffic. In the state diagram, if the prioritized queue is full, p(1, 1) state is not considered instead it is shown as loss of λ_2. If non- prioritized queue is full, then p(0, 1) state is considered as p(1, 1) state. The balance equations for the structure are re-written based on Eqs. (7), (8) and (9). The blocking probability for a non-complex system [17] is derived and then the result is extended for a complex system.

$$p_1\lambda_1 P_{(0,0)} + p_3\lambda_1 P_{(1,0)} = \mu_1 P_{(0,1)} \tag{10}$$

$$q_1\lambda_2 P_{(0,0)} = (\mu_2 + p_3\lambda_1)P_{(1,0)} \tag{11}$$

$$P_{(0,0)} + P_{(1,0)} + P_{(0,1)} = 1 \tag{12}$$

From Eq. (10)

$$p_3\lambda_1 P_{(1,0)} = \mu_1 P_{(0,1)} - p_1\lambda_1 P_{(0,0)}$$

Substituting in Eq. (11)

$$q_1\lambda_2 P_{(0,0)} = \mu(P_{(0,1)} + P_{(1,0)}) - p_1\lambda_1 P_{(0,0)}$$

From Eq. (12)

$$P_{(1,0)} + P_{(0,1)} = 1 - P_{(0,0)}$$

Therefore,

$$q_1\lambda_2 P_{(0,0)} = \mu\left(1 - P_{(0,0)}\right) - p_1\lambda_1 P_{(0,0)}$$

$$\mu = P_{(0,0)}\left[q_1\lambda_2 + \mu + p_1\lambda_1\right]$$

$$P_{(0,0)} = \frac{\mu}{\left[p_1\lambda_1 + q_1\lambda_2 + \mu\right]}$$

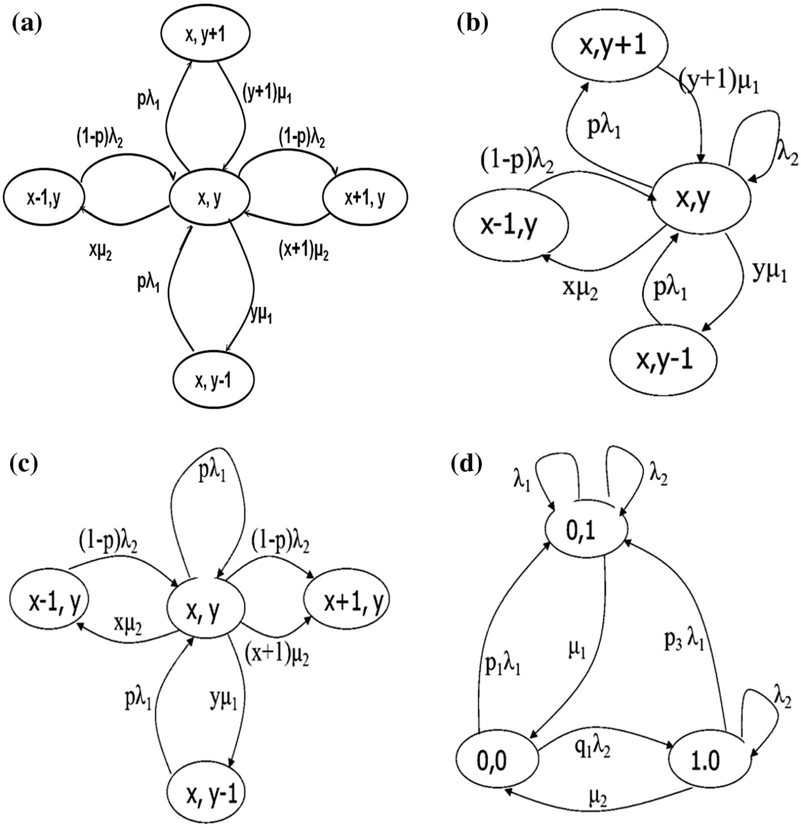


Fig. 2. (a) when x < B2 & y < B1 (b) when x ≥ B2 & y < B1 (c) when x < B2 & y ≥ B1 (d) State transition diagram of a single-channel system for proposed scheme

Substituting $P_{(0,0)}$ in Eq. (11), we get:

$$P_{(1,0)} = \frac{q_1 \lambda_{2*} \mu}{[(p_1 \lambda_1 + q_1 \lambda_2 + \mu)(\mu + p_3 \lambda_1)]}$$

Substituting $P_{(0,0)}$ & $P_{(1,0)}$ in Eq. (10), we have:

$$\mu P_{(0,1)} = \frac{p_1 \lambda_1 * \mu}{(p_1 \lambda_1 + q_1 \lambda_2 + \mu)} + \frac{q_1 \lambda_2 \mu * p_3 \lambda_1}{[(p_1 \lambda_1 + q_1 \lambda_2 + \mu)(\mu + p_3 \lambda_1)]}$$

$$P_{(0,1)} = \frac{\lambda_1 [\mu p_1 + p_1 p_3 \lambda_1 + q_1 p_3 \lambda_2]}{(p_1 \lambda_1 + q_1 \lambda_2 + \mu)(\mu + p_3 \lambda_1)} \quad (13)$$

The blocking probabilities can be calculated as:

1. For the Prioritized queue:

$$\text{Block}_{\text{prob1}} = P_{(0,1)} + (1 - p) P_{(1,0)} \quad (14)$$

which evidently involves two parts: (i) the probability that prioritized packet reaches state (0, 1) and is lost. (ii) the probability that a prioritized packet reaches state (1, 0) but due to probability $(1 - p)$, it is lost.

2. For Non-Prioritized queue:

$$\text{Block}_{\text{prob2}} = \left(P_{(0,1)} + P_{(1,0)}\right) + \frac{\lambda_1}{\lambda_2} p P_{(1,0)} \quad (15)$$

This probability can also be considered as consisting of two parts: (i) the probability that a low priority packet arrives either at state (0, 1) or (1, 0) and is getting dropped; (ii) The probability that a non-prioritized packet arrives at state (1, 0) and is getting lost due to probability p.

5 Simulation Results and Analysis

In this section, we present the simulation results to validate the efficiency of the proposed scheme and to prove that it can support service differentiation. The simulation results are plotted using MatLab R2013b. We have conducted three experiments to investigate the scheduler efficacy. In the first experiment, high priority packets are taken as MPEG traces and low priority packets as Poisson traffic with variable size and under different system loads. In the second experiment, for both high and low priority, Poisson traffic with variable size packets is considered. Third experiment compares IoT and Non-IoT cases and tests the scheduler working when very large packet sizes with high data rate for high priority class arrives as Non-IoT data. Extensive simulation has been conducted to test the scheduler effectiveness for IoT traffic under different data rates. Buffer sizes for both high priority and low priority queues are taken as 10. Details of simulation scenario are given in Table 1.

Table 1. Traffic adopted for conducting experiment

Parameters	IoT traffic (Exp1)		IoT Traffic (Exp2)		Non-IoT traffic (Exp3)	
Service type	High priority	Low priority	High priority	Low priority	High priority	Low priority
Traffic type	MPEG-4	Poisson	Poisson	Poisson	MPEG-4	Poisson
Packet size in bytes	136 to 1000	50 to 702	50 to 702	50 to 702	136 to 424536	50 to 702
Number of flows	2 flows at same time		2 flows at same time		2 flows at the same time	
First datarate	36.8 Kbps to 112.3 Kbps	73.36 Kbps to 203.3 Kbps	0.64 Kbps to 84.3 Kbps	66.8 Kbps to 107.52 Kbps	32.6 Kbps to 50.5 Mbps	358.8 Kbps to 475 Kbps
Second datarate	21.7 Kbps to 89.2 Kbps	64 Kbps to 175 Kbps	0.32 Kbps to 82.96 Kbps	55.3 Kbps to 88.7 Kbps	21.7 Kbps to 40 Mbps	277.5 Kbps to 300 Kbps
Third datarate	16 Kbps to 64 Kbps	45 Kbps to 164 Kbps	0.20 Kbps to 65 Kbps	45.4 Kbps to 64 Kbps	21.7 Kbps to 32 Mbps	160.6 Kbps to 250 Kbps

5.1 Impact of p on Blocking Probability

Figure 3(a) shows blocking probability between two classes against increasing value of probability *p* from p_1 to p_2 and then to p_3, in Exp1 and Exp2 with third data rate. We can observe that if value of *p* increases, blocking probability of the high priority packets decreases and for low priority it increases. So, the required service differentiation can be achieved by adjusting *p* according to the tolerable blocking probability.

It can be observed that due to this scheduling scheme there is a continuous decrease in the high priority blocking probability and simultaneous increase in blocking probability of low priority traffic. This is under the condition where prioritized and non-prioritized packet load is continuously increasing. Prioritized packets priority (in terms of bandwidth allocation) keeps on increasing as probability *p* increases from p_1 to p_2 and then to p_3. From the analysis, for Exp1 and Exp2, it can be easily verified that if bandwidth provided to high priority is greater than or equal to 90%, then in both experiments (Fig. 3(a)), blocking probability is similar for high and low priority services. For bandwidth less than 90% for high priority traffic, blocking probability increases to 1%. Although the packet sizes for high priority packets in Exp1 is of bigger size than Exp2 for IoT but it gets compensated with the higher bandwidth provided by the scheduler.

Figure 3(b) shows impact of probability *p* on blocking probability for Exp3 for non-IoT traffic. We have considered very high data rate and packet sizes in Exp3. We verified our scheduler to test its efficacy under the condition when variable and big packet of sizes of 136 bytes to 424536 bytes is transferred as high priority packets. It can be verified that for high and low priority traffic, blocking probability for non-IoT is slightly increased as compared to IoT traffic due to its high data rate and packet size. As compared to IoT cases, in non-IoT the blocking probability is 0.05% more for the bandwidth greater than 90% for high priority traffic. However, for the bandwidth less than 90% for high priority traffic, the blocking probability increases to 1% even in non-IoT traffic.

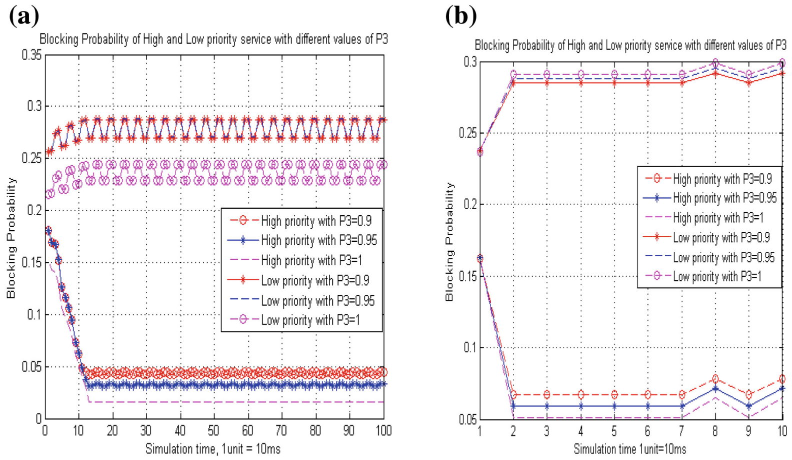


Fig. 3. Blocking probability of high and low priority traffic with different p_3 values in (a) Exp1 & Exp2 (b) Exp3

5.2 Impact of Different Data Rates on Average Blocking Probability

Figures 4(a), (b) and (c) describes average blocking probability of high and low priority traffic with different data rates in Exp1, Exp2 and Exp3 respectively. We observe that with Exp1 blocking probability decreases gradually for high priority traffic as compared to Exp2. In Exp2 blocking probability decreases at a faster rate in the beginning for prioritized packets and then becomes constant. This is due to the reason of comparatively bigger packet sizes and more data rate taken in Exp1 as compared to Exp2. In both experiments, when data rate is less, then average blocking probability is also less. Improvement for non-prioritized packets can be seen for less data rate. Here value of p_3 is 0.9. In Exp3 (Fig. 4(c)) also when data rate for non-IoT traffic is high (first data rate) its blocking probability is more. If data rate is reduced, blocking probability is also reduced for high priority traffic.

5.3 Comparison of Average Blocking Probability in Exp1&2 and Exp1&3

Figures 5(a) and (b) shows the average blocking probability of high and low priority traffic for both Exp1 with Exp2 and Exp1 with Exp3 respectively. For only IoT traffic with different data rates, as shown in Fig. 5(a); we observe that if packet size of prioritized traffic is reduced (Exp2) or if data rate is reduced; the average blocking probability of both prioritized and non-prioritized traffic is reduced. This implies that the length of the packet size can also play an important role in analyzing the performance of any model. Therefore, a smaller size of packet would be considered for better performance. Figure 5(b) compares IoT and non-IoT cases. It is observed that for the non-IoT traffic, the average blocking probability for both high and low priority traffic is more than IoT traffic because of high data rate of non-IoT applications. Table 2 clearly explains impact of increasing probability *p* on blocking probability of high and low

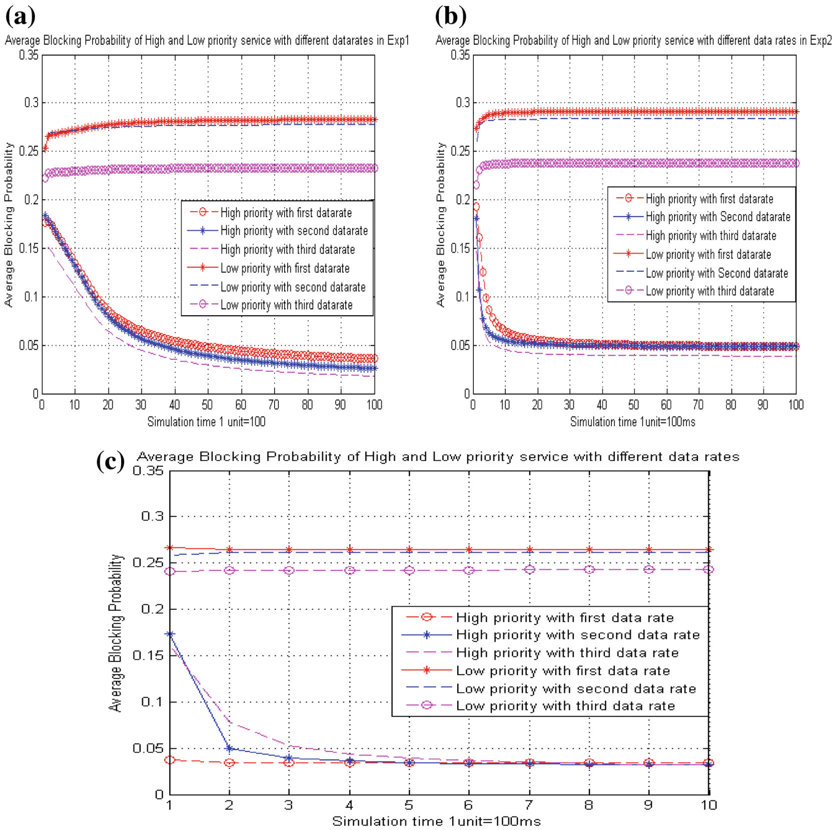


Fig. 4. Average blocking probability of high and low priority traffic for different data rates in (a) Exp1 (b) Exp2 (c) average blocking probability of high and low priority traffic with different data rates in Exp3

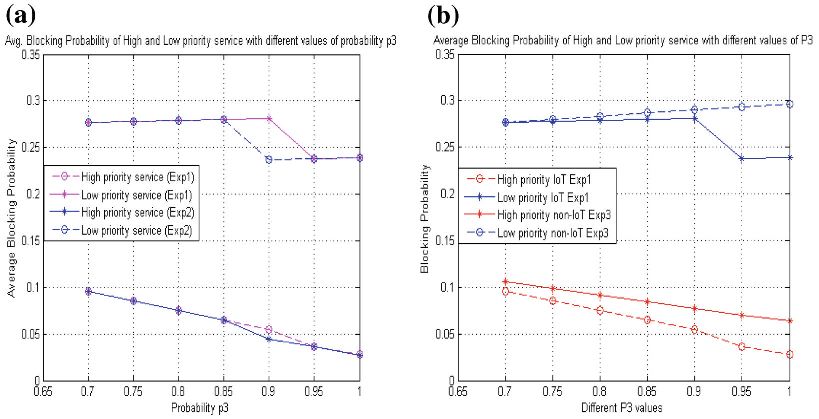


Fig. 5. Average blocking probability of high and low priority traffic with different p_3 values in (a) Exp1 and Exp2 (b) IoT (Exp1) and Non-IoT (Exp3)

Table 2. Impact of increase of probability p_3 on average blocking probability

S.No.	p_3	Blocking probability for IoT traffic (Exp1)		Blocking probability for IoT traffic (Exp 2)		Blocking probability for Non-IoT traffic (Exp 3)	
		High priority	Low priority	High priority	Low priority	High priority	Low priority
1	0.7	0.0956	0.2764	0.0956	0.2765	0.1061	0.2767
2	0.75	0.0852	0.2775	0.0851	0.2777	0.0987	0.2800
3	0.8	0.0750	0.2787	0.0748	0.2789	0.0915	0.2834
4	0.85	0.0649	0.2798	0.0646	0.2800	0.0844	0.2866
5	0.9	0.0551	0.2809	0.0444	0.2374	0.0775	0.2899
6	0.95	0.0366	0.2881	0.0359	0.2384	0.0706	0.2930
7	1.00	0.0284	0.2890	0.0275	0.2393	0.0639	0.2961

priority traffic. In all three cases with decrease in blocking probability of high priority emergency traffic, there is simultaneous increase in low priority blocking probability values which proves that the dynamic scheduling scheme is effective in achieving adjustable service differentiation in IoT and non-IoT applications.

6 Conclusion

A simple and flexible probabilistic scheme has been proposed to offer service differentiation and to provide QoS to emergency applications in IoT. Analytical and simulation results showed that the dynamic scheduling scheme is effective in achieving adjustable service differentiation in IoT and non-IoT applications where large amount of data needs to be transferred continuously at low rate and high rate respectively for long period of time. Also, if any emergency traffic needs to be given priority, this scheduler reduces blocking probability even in the case of congested network. The proposed scheme is tested for variable size packets with different data rates and the expected results are obtained. We also verified that the scheduler satisfies QoS requirements in both IoT and Non-IoT applications.

References

1. Monares, A., Ochoa, S.F., Santos, R., Orozco, J., Meseguer, R.: Modeling IoT-based solutions using human-centric wireless sensor networks. J. Sens. **14**(9), 15687–15713 (2014)
2. Sun, Y., Ma, H., Liu, L.: Traffic scheduling based on queue model with priority for audio/video sensor networks. In: IEEE International Conference on Pervasive Computing and Applications, pp. 709–714 (2007)
3. Dainotti, A., Pescape, A., Claffy, K.C.: Issues and future directions in traffic classification. IEEE Netw. **26**(1), 35–40 (2012)
4. Khalek, A.A., Caramanis, C., Heath, R.W.: Delay-constrained video transmission: quality-driven resource allocation and scheduling. IEEE J. Sel. Top. Sig. Process. **9**(1), 60–75 (2015)

5. Jin, J., Gubbi, J., Luo, T., Palaniswami, M.: Network architecture and QoS issues in the internet of things for a smart city. In: IEEE International Symposium on Communications and Information Technologies (ISCIT), pp. 956–961 (2012)
6. Spiess, P., Karnouskos, S., Guinard, D., Savio, D.: SOA-based integration of the internet of things in enterprise services. In: IEEE International Conference on Web Services, pp. 968–975, July 2009
7. Gupta, V., Poursohi, A., Udupi, P.: Sensor network: an open data exchange for the web of things. In: IEEE Intenational Conference on Pervasive Computing and Communications Workshops, pp. 753–755, April 2010
8. Thoma, M., Meyer, S., Sperner, K., Meissner, S., Braun, T.: On IoT-services: survey, classification and enterprise integration. In: IEEE International Conference on Green Computing and Communications, pp. 257–260, November 2012
9. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Elsevier J. Future Gener. Comput. Syst. **29**, 1645–1660 (2013)
10. Mashal, I., Alsaryrah, O., Chung, T.-Y.: Performance evaluation of recommendation algorithms on Internet of Things services. Elsevier J. Phys. A: Stat. Mech. Appl. **451**, 646–656 (2016)
11. Klepec, B., Kos, A,: Performance of VoIP applications in a simple differentiated services network architecture. In: IEEE International Conference Eurocon 2001 at Bratislava, Slovakia, vol. 1, pp. 214–217, July 2001
12. Said, O., Masud, M.: Towards internet of things: survey and future vision. Int. J. Comput. Netw. **5**(1), 86–94 (2013)
13. Al-Fagih, A.E.: A framework for data delivery in integrated Internet of Things architectures. Ph.D. thesis, Queen's University Kingston, Ontario, Canada, April 2013
14. Liang, J.M., Chen, J.J., Cheng, H.H., Tseng, Y.C.: An energy-efficient sleep scheduling with QoS consideration in 3GPP LTE advanced networks for internet of things. IEEE J. Emerg. Sel. Top. Circ. Syst. **3**(1), 13–22 (2013)
15. Awan, I., Younas, M.: Towards QoS in internet of things for delay sensitive information. In: Matera, M., Rossi, G. (eds.) MobiWIS 2013. CCIS, vol. 183, pp. 86–94. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-03737-0_10
16. Duffy, K., Ganesh, A.J.: Modeling the impact of buffering on 802.11. IEEE Commun. Lett. **11**(2), 219–221 (2007)
17. Yang, L., Jiang, Y., Jiang, S.: A probabilistic preemptive scheme for providing service differentiation in OBS networks. In: IEEE Globecom, vol. 5, pp. 2689–2693 (2003)

A Profound Inquiry of Diversified Application and Trends in Big Data Analytics

Monica Velamuri, Anantha Narayanan, and P. Sini Raj(✉)

Department of Computer Science and Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore, India
{cb.en.u4cse13369,cb.en.u4cse13310}@cb.students.amrita.edu,
p_siniraj@cb.amrita.edu

Abstract. Big Data plays a major role in every field in recent days. Analyzing, storing and visualizing the varied and complex data collected are important tasks for which Big Data tools are used. Big Data handles data in a more efficient manner. So, Big Data is preferably used in enterprises, organizations, companies, business etc. Henceforth, there are various fields such as healthcare/medical, business, sports, education, stock market, web and entertainment etc., which use big data tools. The motive of this paper is to give an insight into different types of analytics that are used in various fields and also to give a detailed study of various organizations and the extent they use the analytics in their respective fields.

Keywords: Big data analytics · Application of big data analytics
Comparison of big data and traditional data storage systems · Hadoop · Spark

1 Introduction

Analytics and analysts are growing in all the work streams. Analytics is used in different ways for different information and hence we have different types of analytics used all over. Analytics is used to improve the performance of an organization to get better profits along with better outcomes of their applications. In order to improve the performance, the past data should be recorded and analysed and prediction should be done. All the fields use analytics in which each field does analytics in a different way and hence gets a different name. Henceforth, we have huge count of analytics. We have started our research by collecting various types of analytics that are available excluding the basic four types; viz., descriptive, diagnostic, predictive and prescriptive analytics. We collected all the analytics in the alphabetical order and we have noticed that, this whole data can be grouped together based on their similarities. In the later sections, we have researched about the companies in that field and implementation of big data analytics to achieve profits.

Before the invention of data processing tools like Hadoop and Spark, traditional file systems, DBMS, RDBMS were used to store and process the data. These could not handle huge amount of semi-structured and unstructured data. Hence, to handle different types of data, we make use of tools such as Hadoop and Spark, which makes the work

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 104–115, 2018.
https://doi.org/10.1007/978-3-319-73423-1_10

easier and very efficient. This is still a developing stream where technologies are getting updated on a large scale.

Size in terms of Big Data is not a fixed quantity. The size of the data has increased to terabytes and petabytes. This will increase in the future, as there is no limit for the streaming data. In 2012, Gartner updated Big Data's definition as, "Big Data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." [1] The characteristics viz., volume, velocity and variety are not sufficient to determine all the characteristics of the data anymore [2]. Hence the number V's are extended to 10 namely, volume, variety, velocity, veracity, validity, value, variability, venue (public and private cloud) [3], vocabulary, vagueness.

2 Data Processing Techniques

Data storage systems play a very essential role. The efficiency of storage is important challenge and hence there are various techniques available. Traditional file systems were used initially and were of great use, but this had many disadvantages and hence to overcome them, Data Base Management Systems (DBMS) came into use [4]. Execution time was reduced, security was increased and memory storage was possible when DBMS was introduced. Moreover instead of checking each and every alphabet just to get the record of a single file, we can use the database system, which helps in an easy retrieval of the files with less efforts (except that the user should know the query language). This reduced the complexity in retrieving data and also the access time.

Though there was a significant improvement in the performance after using database management systems and relational database management systems, there were few disadvantages when it has to deal with huge amount of data [5]. Hence, Big data tools like Hadoop and Spark came into existence, which can handle vast data very efficiently. The disadvantages of DBMS when it has to handle huge amount of data are: Data dependency (it becomes complex when dependencies are high), storage (if the data is huge, space required will be high which is cost ineffective), causation (when filtering of the data is done, important information might be lost which effects decision making), etc.

Since data can be collected from various sources, there are different types of data available. These are majorly from digital sources. Some of the different types of digital data are: Structured data i.e., the data have a high degree of organization and will be in a proper order, which makes search very easy. Semi-structured data means that the data is partially in-order. Examples of semi-structured data are mails, various markup languages (HTML, XML, SGML) etc. Unstructured data is the kind of data that is being collected from various sources. Generally, this data does not have a pre-defined data model or it is not organized in a pre-defined manner. Even if the data has a structure it is considered to be unstructured if it does not follow a predefined structure.

One of the most essential advantages of Big data analytics is that it can handle unstructured data with ease which contains audio, video files along with all the kinds of data available (like text, sensor data etc.) [6].

Figure 1 explains the change in amount of usage of technologies over a period of time. Along with the time, technologies that are used to store and manage the data also got progressed which provides an easy and efficient management. In 1960, only file systems (F) were used whereas in 1965, DBMS (D) was introduced and it overtook the usage of file systems. In 1980, RDBMS (R) was found to be better than DBMS, which resulted in high fall in the usage of file systems. In 2006, Hadoop (H), a big data tool was invented, which was found to be very efficient in handling vast amount of data and then Spark (S) (which is also a big data tool) was introduced in 2012, which overcame the disadvantages of Hadoop.

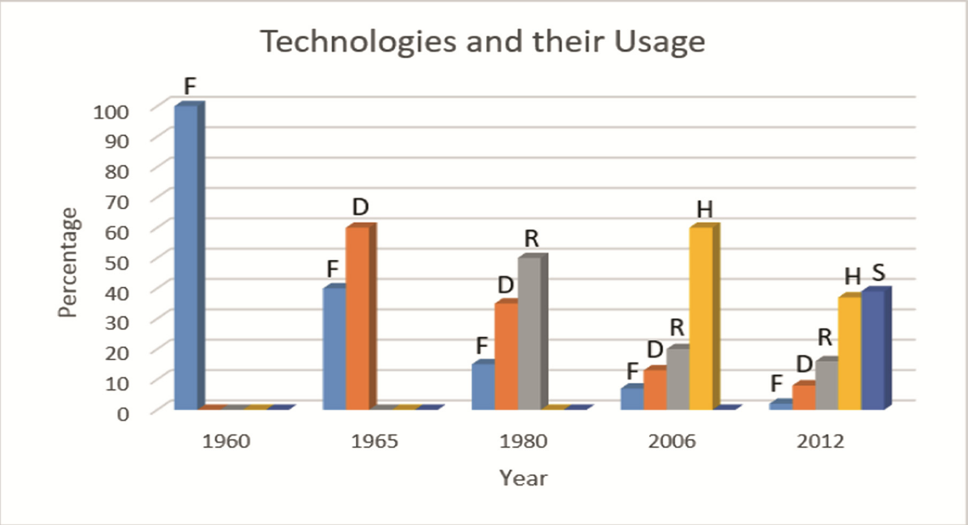


Fig. 1. Extent of usage of different technologies

3 Description of Various Types of Analytics

The four main types of analytics are:

- **Descriptive analytics:** History is an essential factor in the case of analytics and hence, we need to maintain a log of things those occurred in the past which helps us to learn what has happened. As the name suggests, descriptive analytics means a description of an event, or summarization of raw data to make it human interpretable. Data mining algorithms are used to identify patterns from already available information of the past and help us in data visualization. The main application of this is in E-commerce websites, to keep track of the customer's wish list and suggest good products accordingly. It is also used to predict a customer's financial performance [8]. This is considered to be as the preliminary stage for analysis.
- **Diagnostic analytics:** Troubleshooting a problem is a very important task. It helps us to find out the root cause of the problem and helps us to fix it so that the problem does not repeat. Diagnostic analytics is the next phase of Descriptive analytics and

considered to be the second stage of analysis. In this phase we analyze and find out what happened and why did it happened. Like Descriptive analytics, Diagnostic analysis also makes use of data mining algorithms and takes a deeper insight of the data to analyze. This is a very essential step because the same mistakes would cost us a lot and we need to prevent committing same mistakes as done in past. This is applied mostly in Social media to analyze the number of posts and media shared.

- **Predictive analytics:** Interest in future has lead to the introduction of predictive analytics. It uses the historical data and by applying machine learning techniques and statistical algorithms the future outcomes are predicted [9]. The result does not depict of what will happen in the future, instead it tells what might happen along with the reliability factor. Some of the applications are: increase in life expectancy is possible by predicting the disease by analyzing the symptoms [10] in medical field. In the same way, the profits of the business is increased exponentially just by using this predictive analytics on their products. Fraud detection is a very big application of predictive analytics.
- **Prescriptive analytics:** Prescriptive analytics relates to both descriptive and predictive analytics, which means that the outcomes of both these analytics are used to give the best solution for a situation. In a form, it is an extension of predictive analytics in means that, when predictive analytics says what might certainly happen, prescriptive analytics talks about why will it happen and what can be done further along with what might happen (the best possible solutions particularly). Applications of prescriptive analytics are slowly growing and few of them are in healthcare, pricing, travel and transportation optimization, oil production through fracking [11].

The process of analysis should initiate with descriptive followed by diagnostic, predictive and prescriptive. Following the above order gives the best outcome.

4 Application

As analytics are used in various fields, we have done a study on how they are used in different organizations.

4.1 Healthcare/Medical

In the evolution of healthcare practices and research, big data analytics is playing a major role nowadays. The main reason for this is that big data can withstand with all structured, semi-structured and unstructured data. Moreover a large volume of data is being generated due to advances in the technology in medical and clinical research field [12]. By implementing big data analytics in the medical field, doctor's work would be made easier to take decisions within seconds and improve patient's treatment. One of the leading companies that is using big data analytics in the field of health care is ALTEN Calsoft Labs. Their solutions help organizations to get actionable insights to build competitive strategies based on the four fold path of capture, store, and process and analyze. Archimedes, one of the companies that use big data analytics in the healthcare sub-verticals, is best known for its robust disease models and simulations. Kyruus, another such organization using network analytics to identify and understand key social connections that drive commercial, organizational, and health results. Predilytics and Recombinant also belong to this list. Most of the organizations use predictive analytics as it helps us to predict based on the past information which also means it uses Descriptive analytics. OptumHealth is another leading organization in the Global Healthcare Data Analytics market. Its care solutions include health management solutions such as wellness, complex medical support, decision support systems, and physical health programs [13]. Quantzig healthcare analytics solutions help clients in the healthcare industry including pharma, medical devices, diagnostics, and healthcare delivery companies to address key challenges, reduce costs, increase margins, and gain a competitive market advantage. Healthcare analytics solutions include integration of electronic medical records, clinical trials, hospital records, physician notes and pharmacy data to create consolidated and actionable insights, and risk mitigation and resolution of issues for medical cost management, providers' claim processing, and improved payment accuracy. Verisk Analytics is among the leading vendors in the Global Healthcare Data Analytics market. The company provides five types of health solution suites: Enterprise Analytics, Payment Accuracy, Quality & Compliances, Revenue Integrity, and Life and Legal [14]. The Enterprise Analytics suite helps organizations to make informed decisions as well as improve cost containment and quality of care provided to patients. The Payment Accuracy suite helps in fraud detection and real-time claims that uses Predictive analytics. MEDai's health record management solution helps to manage patient records and provides an expert level of care to patients. Its healthcare compliance solution helps discover policies and procedures needed to be followed in healthcare enterprises. MedeAnalytics is also one of the prominent vendors in the market. The company provides a wide range of products for hospitals and individuals. MedeAnalytics's solutions are delivered using the SaaS model. Its product, Clinical Performance Manager, provides a daily insight into clinical cost and the performance of the individual physician [15]. Health care uses mainly of Descriptive analytics by storing the past data of the patients which is useful for predicting the patient's health status and while the patient has to get diagnosed, along with Descriptive analytics, Diagnostic analytics is also used get the insight.

4.2 Business

The concept of big data has been around for years. Applying analytics to the data that flows into the business can provide insane amount of value to the organization. Earlier only few companies understood this and applied basic analytics and later on when the term big data evolved, almost all the companies adopted it. Data analytics offers both speed and efficiency in decision-making based on the past data. Companies use both In-Memory analytics and predictive analytics. In-Memory analytics is by analyzing data from system memory (instead of from your hard disk drive), you can derive immediate insights from your data and act on them quickly [16]. This technology is able to remove data pre-processing and analytical processing latencies to test new scenarios and create models; it's not only an easy way for organizations to stay agile and make better business decisions, it also enables them to run iterative and interactive analytics scenarios.

Predictive analytics technology uses data, statistical algorithms and machine-learning techniques to identify the likelihood of future outcomes based on historical data. It's all about providing a best assessment on what will happen in the future, so organizations can feel more confident that they're making the best possible business decision. Some of the most common applications of predictive analytics include fraud detection, risk, operations and marketing. SAP's best Big Data tool is its HANA in-memory database, by which the company says can run analytics on 80 terabytes of data, integrate with Hadoop, search text content, harness the power of real-time predictive analytics, and more. Oracle has its Big Data Appliance that combines an Intel server with a number of Oracle software products. They include Oracle NoSQL Database, Apache Hadoop, Oracle Data Integrator with Application Adapter for Hadoop, Oracle Loader for Hadoop, Oracle R Enterprise tool, which uses the R programming language and software environment for statistical computing and publication-quality graphics, Oracle Linux and Oracle Java Hotspot Virtual Machine. Google is more of a cloud services company but it is making a push into Big Data analytics by offering BigQuery, a cloud-based Big Data analytics platform for quickly analyzing very large datasets. Unlike most services, you send data up to BigQuery rather than store it in the cloud. Apart from these Tier 1 companies, there are other companies using big data analytics for marketing purpose. Starbucks manages to open new stores in very close proximity with their other stores. It uses big data to determine the potential success of every new location prior to expanding their operations. With location-based data, traffic data, demographic data, and customer data, they're able to estimate the general success rate of each new store, so they can choose locations based on the propensity toward revenue growth, thus decreasing the financial risk of each new store. T-Mobile is using big data to help reduce their customer turnover rate. By analyzing big data, they can determine the core causes for turnover, allowing them to implement effective solutions that will keep more clients on board [17]. As a telecom company, they accrue boundless quantities of data every year, and without big data management, the ability to analyze the data would be greatly inhibited. In business analytics, all the four types of analytics are used and very essential unlike medical/healthcare.

4.3 Sports

Big data in sports industry has been gaining momentum since 2007 and now tracking has become normal. The sports world isn't immune to the impact of big data Statistics-driven sports like Major League Baseball and the National Football League have long crunched numbers to make key decisions; so using big data is a natural progression. Analysts and trainers pored over data to predict performance and develop strategies [18]. Whether motivated by profit or the quest for a win, greater efficiency and increased accuracy, the sporting world is embracing big data to improve performance. Kitman Labs has collaborated with several Olympic teams to prevent future sports injuries by using Big Data and Analytics. The sports and data technology company Kitman Labs uses a unique Athlete Optimization System. This system allows team performance directors, coaches and trainers to understand how athletes are responding physically, as well as mentally to the stresses endured during training and exercise at levels of high competition. Any signs of negative response can trigger the staff to adjust an athlete's training and recovery program to proactively avoid injury. Data analysis is used to identify undervalued players when constructing baseball teams [19]. Predictive analytics looks at patterns in historical data to determine future performance and trends. With algorithmic and biomedical advances, the sports industry has greater confidence in predicting and measuring the success of current and future players. In the NBA, data analysis has provided teams with better ways to measure player efficiency and defensive effectiveness. A player's value can be measured by a number of metrics, including player efficiency rating, win shares and wins above replacement player. Coaches famously watch video to gauge opponents' skills and improve their own player's performance, using knowledge gleaned in the training room, but now they're also using big data analytics to gain an edge. With RFIDs, sensors and GPS trackers, training and coaching staff can capture information, feed the data into analytical engines and use it to influence strategic decision-making. This approach can help them choose exactly the right player for any given play. Ultimately, the goal for athletes, trainers, coaches, broadcasters and others involved in sports decision-making is to leverage real-time data to improve live performance. Like smaller companies that must find an edge to compete successfully with larger enterprises, a team with a smaller budget or more limited pool of athletes can use big data and analytics to gain an advantage. In that way, big data can make sports smarter.

4.4 Education

Online courses and learning systems have been gaining tremendous popularity over the last few years. While their ease of access and availability makes them a very useful medium for knowledge sharing and learning, they do not keep the learners and their learning abilities in mind. Big data allow for very exciting changes in the educational field that will revolutionize the way students learn and teachers teach [20]. Big Data can help to create groups of students that prosper due to the selection of who is in a group. Students often work in groups where the students are not complementary to each other. By using algorithms it will be possible to determine the strengths and weaknesses of each individual student based on the way a student learned online. This will create stronger groups that will allow students to have a steeper learning curve and deliver better group results. It will give students the opportunity to develop their own personalized program, following those classes that they are interested in, working at their own pace, while having the possibility for (offline) guidance by professors. Big Data can give insights in how each student learns at an individualized level. Each student learns differently and the way a student learns affects the final grade of course. Some students learn very efficiently while other may be extremely inefficient. When the course materials are available online, it can be monitored how a student learns. This information can be used to provide a customized program to the student or provide real-time feedback to become more efficient in learning and thus improve their results. Predictive analysis can serve many segments of society as it can reveal hidden relationship that may not be apparent with descriptive modeling. Analytics advancement plays an important role in higher education planning [21]. The descriptive modeling can help to evaluate the teaching staff and their excellence in imparting the education. Using predictive analytics on all the data that is collected can give educational institute insights in future student outcomes. These predictions can be used to change a program if it predicts bad results on a particular program or even run scenario analysis on a program before it is start. Big Data can help provide insights to support student's learning needs. For instance, learning analytics as a fundamental component of Big Data in higher education provide researchers with opportunities to carry out real-time analysis of learning activities. By performing retrospective analysis of student data, predictive models can be created to examine students at risk and provide appropriate intervention, hence enabling instructors to adapt their teaching or initiate tutoring, tailored assignments and continuous assessment. Big Data can afford to shape a modern and dynamic education system, which every individual student can have the maximum benefit from that. Furthermore teachers have valuable tools, were they do not have before, which can make their decisions more specific and are able to choose a big variety of new learning methods. Hence the Big Data are actually involved to change the way of industries including the education. In the new era of data, the traditional difficulties will no longer exists, keeping the good methods. The education system will be enriched with new learning ways, making more efficient and targeted. But the way of this new era, have just began and there are many difficulties such as the lack of experienced personnel on the science of Big Data and Data analytics which can be overcome in the course of time.

4.5 Stock Market

One of the most popular applications of analytics is found in Stock Market analysis. It shares changes every second. If you are able to get good amount of share through stock market investment, that does mean that you have the ability to accurately predict stock behaviour. However, luck factor involves in it but apart from that the knowledge developed through years of experience of dealing in stocks is the important factor [22]. The key point is to analyse data and decipher the relevant patterns. Enormous amount of data is fed into the systems and conventional data mining algorithms are used to process it. Followed by this, big data analytics is applied to the processed data and investments are made. Reliance securities have integrated online trading platform, which uses analytics to provide robotic insights to investors. It scans and captures vast amounts of gathered data, processes it using advanced algorithms, and presents real time analyses. It helps the firm get a macro view of stocks, and provide analytics and optimal risk strategies. Kotak Institutional Equities uses a web-based platform named Consumer Querimetrix that explains and predicts short-term behaviour of Indian investors by analysing vast amounts of data [23]. It uses machine-learning techniques and merges big data analytics to provide consumer insights and capture inflection points. Angel Broking has also incorporated big data analytics in its day-to-day operations to automate processes, speed up activities and enhance customer experience. The firm uses analytics to predict margin-limit multiplier, e-mails and calls classifier, and analysing customer sentiments, queries and complaints. Currently, about 30% of Angel Broking's trades happen online and nearly 80% of its new clients demand online access. HDFC Securities is using a mobile app that helps its clients to trade stocks, track market movements, manage portfolios and analyse industry trends, using big data analytics. The company has also implemented Oracle SuperCluster that supports its increasing customer base and daily transaction load. The platform has increased online trading speed by up to 60% and enabled HDFC to produce reports 67% faster while reducing risk and cutting data center costs. Aditya Birla Financial Group has implemented an online solution, which provides data and insights for its entire business, and supports its customer centricity vision. It helps the firm generate reports and provide information to all sales, marketing, and customer service related units [24]. The solution provides informed insights for faster turnaround, better administration, and better understanding of customers. It has helped the company bring down processing time by 30–40%.

4.6 Web and Entertainment

Web Analytics helps us to keep track of Visitor's behavior, performance of website and data flow. In other words it is the collection, reporting and analysis of website data. The primary concern is to use the website data to determine success or failure of these goals to drive strategy and improve the user's experience. It is not a measuring web traffic but can be used as a tool for business and market research, and to assess and improve the effectiveness of a website. It is often used in customer relationship management analytics. Based on the purchases in the past, the customer is given suggestion, monitoring their likes and frequently searched items to give better recommendation, observe geographic regions from which most of the searches are made, and predicting which products, customers are most and least likely to buy in the future [25]. This can help to improve the ratio of revenue to marketing costs. In addition to these features, Web analytics may include tracking the click through and drill down behavior of customers within the Web site, determining the sites from which customers most often arrive, and communicating with browsers to track and analyze online behavior. Web Analytics follows four basic steps viz., Collection of data, processing the data into information, Developing KPI (Key Performance Indicators) and finally formulating the online strategy. There are two categories of web analytics viz., Off-site and On-site. Off-site web analytics refers to the measurement of a website's audience, share of voice and the comments. It is regardless of whether you own or maintain a website. On-site web analytics measures the performance of your website in a commercial context. This data is typically compared against key performance indicators for performance and used to improve website. Most of the web analytics tools are free. Google offers a free web analytics tool to track number of users visiting your website and to measure traffic sources and goals [26]. It basically generates reports on Audience Analysis, Acquisition Analysis, Behavior Analysis and Conversation Analysis.

The future of this industry is centred on the convergence of digital and analytics solutions. Therefore, in order to gain customer insight, the enterprises are eager to transform their media delivery. Entertainment solely depends on the audience experience. Audience analytics can help organizations continuously capture audience response from multiple sources so that they can deliver the right content to the right person at the right time. Subsequently, each target audience has a different and unique experience, leading to increased revenues from consumers, advertisers, and overall viewer market share.

Media and Entertainment industry demands that content creators and distributors develop a new way to leverage big data to understand and connect with audiences.

As mentioned previously, we have broadly classified the fields as Medical, Business, Sports, Education, Stock market and Web and Entertainment. Figure 2 is a plot that shows the proportion in which these fields are using different types of analytics. From this research we find that Stock Market makes the highest use of analytics whereas Medical and Business fields makes use of analytics almost in an equal proportion. Sports and Education use in moderate amount and Web & Entertainment uses the least. By this we can infer that there will be a great improvement in Education, Sports and Web & Entertainment fields if the use of analytics is significant.

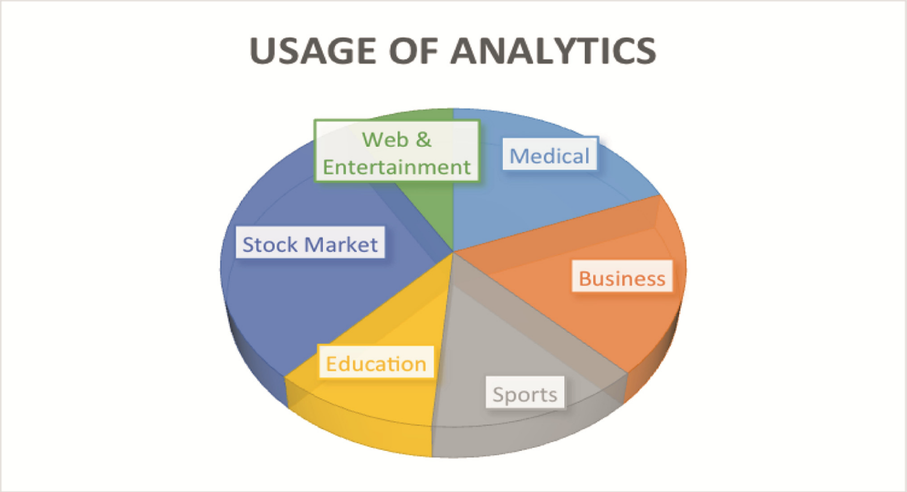


Fig. 2. Usage of big data analytics in the six major fields

5 Conclusion

As the technology is increasing, the size of the data that we get from this is also getting increased proportionally. We get data from many sources like sensors, websites, and electronic gadgets etc., which are contributing terabytes and petabytes of data every day. Every organization produces data in large amount irrespective of it being small or big. This huge amount of data needs efficient organization and management. Analyzing the data that has been well organized, stored and managed is equally important to do computations on that data. Since, there are different types of analytics available, usage of these in an efficient way will fetch better results. Each organization uses these analytics in their own way; depending in which field that organization is working. Therefore, we have given a brief description of the organization's field and their techniques of using these different types of analytics in order to obtain high profits in an efficient manner. This will provide the new comers in this field of big data analytics to understand the concepts of what is big data analytics, types of analytics present and their use in different fields. It also provides an easy path for entrepreneurs in ways of how to achieve better profits by appropriately using analytics in their start-ups. Not only that, but also can be used by other organizations which are in the initial stages of using analytics for a better performance.

References

1. Big Data. En.wikipedia.org. N.p., 2017. Web. 27 Jan. 2017
2. Abarna, K., Rajamani, M., Vasudevan, S.K.: Big data analytics: a detailed gaze and a technical review. Int. J. Appl. Eng. Res. **9**, 1735–1751 (2014)
3. Sangeetha, K.S., Prakash, P.: Big data and cloud: a survey. In: Padma Suresh, L., Dash, S.S., Panigrahi, B.K. (eds.) Artificial Intelligence and Evolutionary Algorithms in Engineering Systems. AISC, vol. 325, pp. 773–778. Springer, New Delhi (2015). https://doi.org/10.1007/978-81-322-2135-7_81
4. Trujillo, G., et al.: Traditional data systems—understanding the big data world—pearson IT certification. Pearsonitcertification.com. N.p., 2017. Web. 5 January 2017
5. Sucharitha, V., Subash, S.R., Prakash, P.: Visualization of big data: its tools and challenges **9**, 5277–5290 (2014). Print
6. What is big data? What are the benefits of big data? Martech. MarTech. N.p., 2017. Web. 24 January 2017
7. Descriptive, predictive, and prescriptive analytics explained. Halo. N.p., 2017. Web. 21 January 2017
8. What is descriptive analytics? Definition from Whatis.Com. WhatIs.com. N.p., 2017. Web. 12 February 2017
9. Predictive analytics. En.wikipedia.org. N.p., 2017. Web. 8 January 2017
10. Predictive analytics: what it is and why it matters. Sas.com. N.p., 2017. Web. 11 January 2017
11. What is prescriptive analytics? Definition from Whatis.Com. SearchCIO. N.p., 2017. Web. 14 January 2017
12. Why analytics alone Won’T bend Healthcare’S cost curve. Health Data Management. N.p., 2017. Web. 21 February 2017
13. Belle, A., et al.: Big data analytics in healthcare. N.p., 2017. Print
14. Raghupathi, W., Viju R.: Big data analytics in healthcare: promise and potential. N.p., 2017. Print
15. Top 5 Examples of big data analytics in healthcare. BI Blog—Data Visualization & Analytics Blog—datapine. N.p., 2017. Web. 29 February 2017
16. Business analytics. En.wikipedia.org. N.p., 2017. Web. 6 February 2017
17. Big data management to fuel their marketing—reachforce. Reachforce.com. N.p., 2017. Web. 2 March 2017
18. Making sports smarter with big data. BetaNews. N.p., 2017. Web. 10 March 2017
19. Nath, T.: How big data has changed sports. Investopedia. N.p., 2017. Web. 8 March 2017
20. Four ways big data will revolutionize education. Datafloq.com. N.p., 2017. Web. 24 March 2017
21. Big data analytics for personalized education - engineering & computer science - ANU. Cecs.anu.edu.au. N.p., 2017. Web. 27 March 2017
22. Ahmar, M.: Want to make big bucks in stock market? Use big data analytics. Analytics India Magazine. N.p., 2017. Web. 20 March 2017
23. How indian brokers use analytics to predict the stock market. ChannelWorld India. N.p., 2017. Web. 22 March 2017

24. The “Big Data” solution for wall street. Stock Forecast Based On a Predictive Algorithm—I Know First—N.p., 2017. Web. 19 March 2017
25. Audience Analysis for Media and Entertainment—IBM Analytics. Ibm.com. N.p., 2017. Web. 24 March 2017
26. Web Analytics - Google Analytics. www.tutorialspoint.com. N.p., 2017. Web. 16 March 2017

SDN Framework for Securing IoT Networks

Prabhakar Krishnan(✉), Jisha S. Najeem, and Krishnashree Achuthan

Center for Cybersecurity Systems and Networks, Amrita University, Amritapuri, India
kprabhakar@am.amrita.edu

Abstract. Internet of Things (IoT) paradigm is the interconnection of machines, intelligent devices and location aware analytics platforms that collectively enable us to have smart world around us. As the billions of already connected devices and newly added devices grow this network, IoT pose the most complex operational and information technology challenges to the way networks are designed and operated. With the emerging technologies like SDN, SD-WAN, NFV, IXP evolving into standards, researchers are proposing new communication platforms to deliver secure and scalable networks for Internet of Things (IoT). In this paper, we discuss major security challenges in IoT networks and present the notion of security architecture for IoT based on programmable and virtualization technologies SDN/NFV, explain the architectural choices and its applications for IoT. We review prior works in this area and discuss our future work to solve security and privacy challenges of heterogeneous systems and networks in IoT.

Keywords: Internet-of-Things (IoT) · Software-Defined-Networking · SDN Network security · Network-Function-Virtualization

1 Introduction

The proliferation of IoT based smart devices, estimated to become a 20 billion interconnected network by 2020, and brings with it several challenges and hard problems with regard to security and privacy of devices, users and the data consumed by applications. With the kind of ubiquity within society predicted, it will create the need for flexible sophisticated methods of integrating these large farms of embedded devices within overall network architectures. This integration will potentially be dynamic, as users and devices roam in and out of the wireless networks, within their context or zones (e.g. fleet with vehicular network sensors that access context aware applications and services from local networks).

IoT architecture can be visualized in 3 tiers; on the top tier are usually well-protected devices, secured servers, personal computers, laptops and smart phones with sophisticated firewall software hardware. The middle tier typically consists of devices of less complex smart appliances and devices such as refrigerators, lights, cameras, televisions, digital screens and luxury HiFi devices. The bottom tier comprises of devices from consumer electronics, mechatronics and lifestyle gadgets such as smart locks, digital doors, perimeter safety and surveillance machines, air-conditioners and wearable, medical implants, geo-sensory equipment, vehicular network devices and so on.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 116–129, 2018.
https://doi.org/10.1007/978-3-319-73423-1_11

None of these three tiers of devices may pose a threat independently restricted to that autonomous homogeneous network. But when we interconnect the devices from across the tiers, the resulting architecture will consist of heterogeneous devices, integrating disparate technologies and communication protocols, Application Program Interfaces (APIs) etc. And this heterogeneous interconnected IoT architecture may pose serious risks and challenges for Quality of Service (QoS), security and privacy. So far we don't have a single one-size solution to address all these challenges.

Thus, ensuring the trust and security of the configuration, topology and integration of all heterogeneous devices into large networks are some major operational challenges. Experimental exploitation of current generation of smart devices or things have demonstrated that breaching and tampering is possible and also established the need for handling IoT devices network security (Fig. 1).

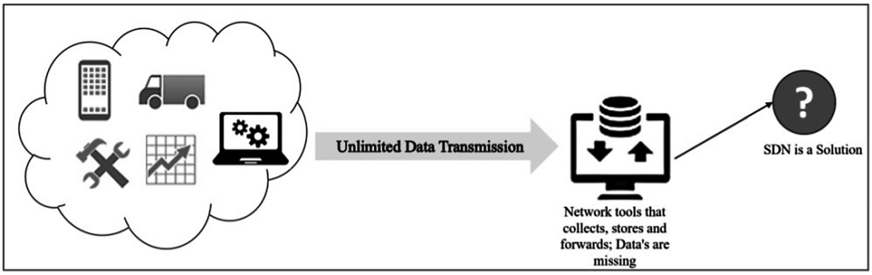

Unlimited Data Transmission
Network tools that collects, stores and forwards; Data's are missing
?
SDN is a Solution

Fig. 1. SDN being a solution

The modern SDN paradigm has initiated a fundamental redesign of how network traffic management, routing control logic, forwarding and network orchestration are architected. The design should also provision flexible, agile device management policies. This design philosophy is implemented through the separation of control logic or brain from the packet forwarding functions. In other words, SDN consists of one centralized control plane that is connected over a standard communication channel to distributed physical data or switching plane.

Some key criteria for evaluating the SDN in IoT network include:

- Ability to securely connect and manage hundreds or even thousands of heterogeneous IoT devices.
- Low latency security monitoring overhead to deliver real-time awareness and operations.
- Scale-out elastic architecture to scale and dynamically load balance/shift workloads, and
- Programmability for enforcing custom policies and applications.

In this paper, we discuss the effectiveness of approaches to design new secured network architecture based on SDN, advanced network virtualization functionalities and clusters.

This paper is organized as follows: Sect. 1 introduces the emerging technologies for the interconnected IoT networks, applications and sets the context for incorporating SDN in IoT architecture, Sect. 2 provides an overview of the threat landscape in IoT and current approaches to IoT Security Sect. 3 gives an overview of the security threats and risks to IoT network, Sect. 4 explains the feasibility and efficacy of SDN architecture in the context of IoT networks and discusses related works in SDN IoT integration. Section 5 articulates some key challenges for this SDN/IoT domain. Section 6 proposes our SDN framework for securing the IoT networks, architecture, design choices and case studies. Section 7 presents our experience from initial experimentation and evaluation, Sect. 8 provides a general outlook of our future work and concludes the paper.

2 Approaches to IoT Security

To implement dependable security architecture for an IoT network, both system characteristics and data centric parameters must be considered. The security framework should combine them to achieve the desired level of privacy, security, risk level, interoperability, recoverability and trustworthiness. Vendor com- munities, business applications, government norms and regulations may drive these factors. Security in IoT network must be implemented at various levels: the manufacturing vendors and supply chain, hardware ASIC or SoC, Operating systems, systems software and application software, middle-box appliances, networking hubs, routers, and switches.

The target IoT environment may have several constraints, For example:

- Real-time infrastructures cannot be brought down for security updates and patching.
- Low-latency, proprietary protocols limit the ability to deploy antivirus and anti-malware software.
- Embedded processors have limited processing power and memory to execute security software.
- IoT devices have a small form factor, limited connectivity and are designed for very low power consumption.
- Attacks toward wireless network infrastructure can cause the unavailabity of network component and data loss.
- Many IoT devices are physically accessible to the attacker.

Despite all these threats, two key areas of IoT security that have not received much attention are:

1. Software integrity: Ensuring the authenticity and integrity of the software on the device. By allowing software that digitally signed, whitelisted and certified by trusted entity, to run and access data.
2. Device authentication: Authentication of the end devices before they can transmit or receive information. Authentication of devices and data is a key success factor for the Internet of Things. A single compromised node can be turned into a malicious one that brings down whole systems or causes disasters with cars, planes, drones, the grid etc.

The known shortcomings of knowledge-based authentication approaches like passwords and PINs must be augmented with standard solutions like Public Key Infrastructure (PKI) in conjunction with new technologies like Physical Unclonable Functions (PUFs). These provide measures to strengthen IoT security from a self-enforced identity perspective. Using a block chain to store data that has been secured with PUF derived keys and attributes provides an immutable assurance that data has not been tampered with, in addition to providing traceability and transparent auditing capabilities.

The majority of proposed security solutions use cryptographic algorithms that normally require high amount of resources. Considering that most IoT devices are associated with low energy and computing resources capabilities, such solutions cannot be implemented to IoT devices with an application of traditional cryptographic mechanisms.

3 Integrating SDN into IoT Networks

This section briefly introduces the area of software defined networking (SDN) and discusses its applicability to both acting as a gateway for IoT devices and as a security controller mechanism.

3.1 Background About SDN

SDN is an open network architecture proposed in recent years to address some of the key shortcomings of traditional networks. The proponents of SDN argued that the control logic of the network and network functions are two separate concepts, and should therefore be separated in different layers. To this end, SDN hence introduced the concepts of control plane and data plane: The centralized control plane (controller) manages the network logic, control traffic engineering functions from the data plane (switches) that just take care of forwarding the packets between the networks. So, the SDN can be considered as a physically distributed switching framework with a logically

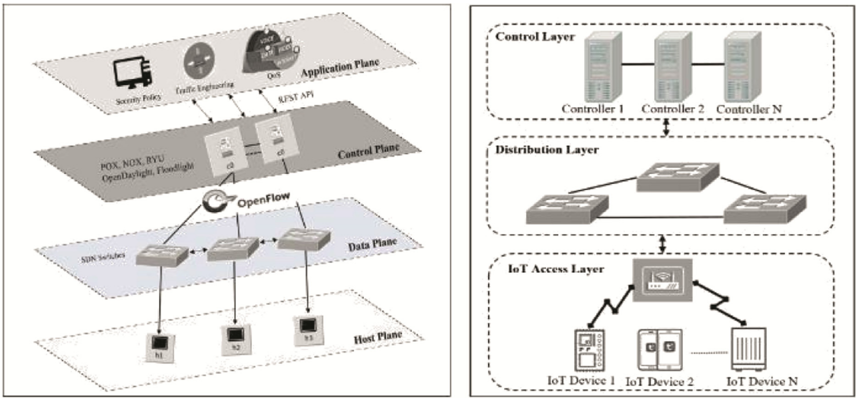


Fig. 2. (a) SDN architecture (b) SDN - IoT integrated architecture

centralized control. SDN is designed for provisioning highly dynamic orchestration and quality of service/security policies (Fig. 2).

3.2 Significant Benefits of SDN-IoT

IoT network environment is a large interconnection of multiple smaller local or adhoc networks or wireless or industrial control networks. The orchestration and visibility of end-to-end traffic and devices is essential for establishing QoS and Security policies. To accurately visualize the operating environment, automated programmed mechanisms are needed and that be provided by SDN. It can validate the addition or deletion or modification of devices/configurations in the monitored network and it can program the policies at various points of the network at run time to react differently depending upon the behavioural characteristics of the devices.

The key feature of the SDN is the dynamic provisioning at run time. The capability can be extended for security monitoring of the network. The SDN applications and elements can be programmed for anomaly detection and diversion of suspicious attack traffic to sandbox or honeypot deception framework for further analysis.

For modern IoT applications which encompass multiple interconnected networks or micro-networks in the Cloud, we can incorporate SDN elements to create a suite of semantic monitoring, fine-grained security analytics, defense mechanisms, software defined perimeter, firewalls at different vantage points or locality or layer boundaries of network.

4 Related Works in SDN-IoT

In this section, we present an overview of the related works that have been proposed in the context of SDN-IoT. Flow based security monitoring mechanism [1] has explained the numerous attacks and mitigation approach. Their infrastructure consists of statistics manager that collects data in real-time from log cache and analyses the flows and mitigation actions such as blacklisting are taken based on the various characteristics of flow.

Fog Computing [2] sets another security feature for IoT devices using SDN. The architecture comprises of gateway edge nodes and servers. Edge gateways and servers are connected via high-speed interfaces that can be either wired or wireless such as 3G, LTE etc. Gateways have their own unique role for master mode that controls the virtual path of gateway function located in slave nodes. Using ClickOS, a virtualized software middle box can concurrently run on a commodity server.

One of the most common and significant security threats deeply researched is that of distributed denial of service (DDoS) attacks. Numerous projects are currently seeking to use SDN based security systems as means of mitigating such attacks. Choi [3] suggested a new framework that discovers generation of new traffic thereby performing DDoS mitigation by limiting the amount of traffic generated for each application. Another technique was to identify malicious flows by developing an anomaly detection technique [4] based on the history of the networks stored in the log cache and then comparing with the real-time traffic generated. Significant effort is also made in wireless network security enhancement by applying SDN in wireless/adhoc networks. An SDN based enterprise solution Odin [5] has built a virtualized multi-layered network architecture that uses abstraction of access points. Another open source project OpenRoads [6], decouples the data plane layer and the network layer providing dynamic control over the network management. In addition to the ongoing SDN-based security research projects, there are a small number of commercially developed security applications that are designed to integrate with SDN controllers in IoT networks.

5 Challenges in SDN and IoT

SDN and IoT integration provides a convincing approach to simplify network management and security control, but SDN has inherent design vulnerabilities that pose serious threats to the integrated IoT network and applications [7]. In the SDN architecture, (a) the switches that maintain the flow tables and its capacity (b) communication channel speed, reliability and bandwidth are the critical points for SDN operation. The following issues could lead to critical point failures:

1. The SDN switches/data-plane evaluates incoming packets, matches with Rule table or Flow tables, which are stored in switch fabric (TCAM) memory, having nite capacity, can be attacked.
2. SDN switches out there in the open, may be compromised and recruited in to the botnets, leading to massive DDoS attack campaign.
3. Control-Data Plane link is vulnerable and if saturated, it may lead to total network breakdown. Network level new flow attacks, DoS attacks such as TCP-SYN/DNS/ICMP Amplification and flooding are common in recent times. Hence the placement of controller and protection of communication channel between controller and switches are critical aspects for security availability of the SDN-based IoT applications.

Though the notion of SDN in the context of IoT applications, is still at an early stage, research is gaining momentum to secure the SDN stack, tackling all the above mentioned critical points of attack and IoT communities are investigating the hybrid SDN/IoT architecture for design choices and implementation trade-offs.

6 SDN Based IoT Network Architecture: Our Proposal

In this section, we discuss our SDN-IoT integration architecture, with two design choices, varying in terms of implementation and modifications to the standard SDN or IoT components. We walk through the building blocks of the architecture following both design choices. The core functions of security analytics, access control, policy decisions and enforcement are implemented in the SDN layer and the data from IoT layer is selectively forwarded through this core framework. By acting on contextual information exposed by the IoT applications sensory network, the gap between the IoT and IT networks are filled by the SDN (Fig. 3).

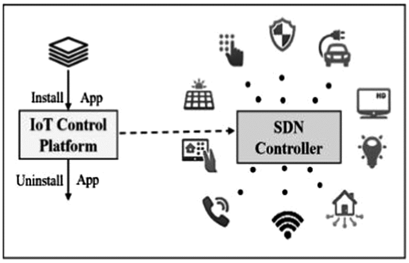


Fig. 3. SDN - IoT control architecture

A systematic design approach to monitor large-scale IoT networks with the SDN gateway and controller, allows for a holistic view of the network and removes the need for additional dedicated hardware. The two design choices for SDN and IoT integration are:

1. Loosely coupled Integration: A flexible flow based monitoring and security mechanisms implemented at SDN control plane as applications. The IoT layer has no modifications and a new layer for SDN is added at the Edge.
2. Tightly coupled Integration: Hybrid Gateway Switch (IoT gateway and SDN Switching) and a security controller, implemented as a sandwich layer between Edge and IoT network. This requires modifications to both SDN IoT layer.

6.1 Design Choice 1 - Loosely Coupled Integration

This design is implemented as a defensive mechanism, attaching the SDN stack to the IoT layer (Gateway) at the Edge security processing. In this framework, we will have SDN applications that monitor the flows and configuration, generate blacklists and whitelists, in the IoT network and analyse packet streams for spatial, temporal and volumetric correlations in their behaviours, protocol violations, and attack signatures (Fig. 4).

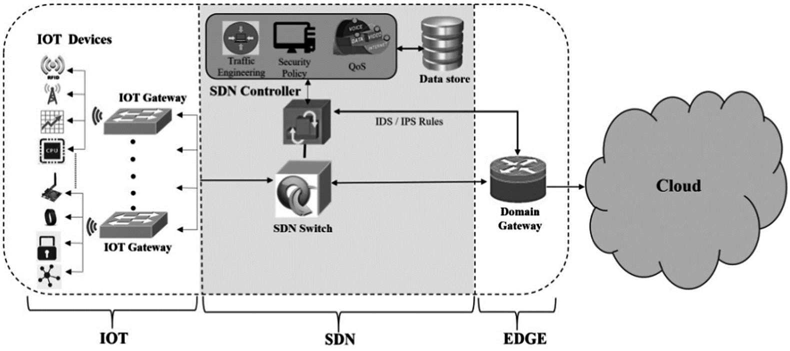


Fig. 4. Loosely coupled SDN-IoT integration

Security threats can be from External or Internal network. The Outside attackers or botnets can target the IoT Gateway or sensor devices or Services, Vulnerable apps installed in the devices in the internal network e.g. Home WiFi-Router/Mode, Webcams. The common indicators of such attacks are: (a) Login access or scanning traffic from the public network, to key IoT gateway or sensor devices, (b) The malicious usage of the IoT devices/apps is beyond their declared functionality.

6.2 Design Choice 2 - Tightly Coupled Integration

This architecture is based on an extending the SDN stack to interface with the IoT stack, specifically the IoT gateway functionalities and protocols. This architecture is built by inheriting the major modules of the SDN Openflow switch stack and by adding new functionalities in the packet processing workflow. This approach is similar to a middle-box device running a modified network operating system that combines IoT gateway and SDN data plane functionalities. This device has extensible architecture and

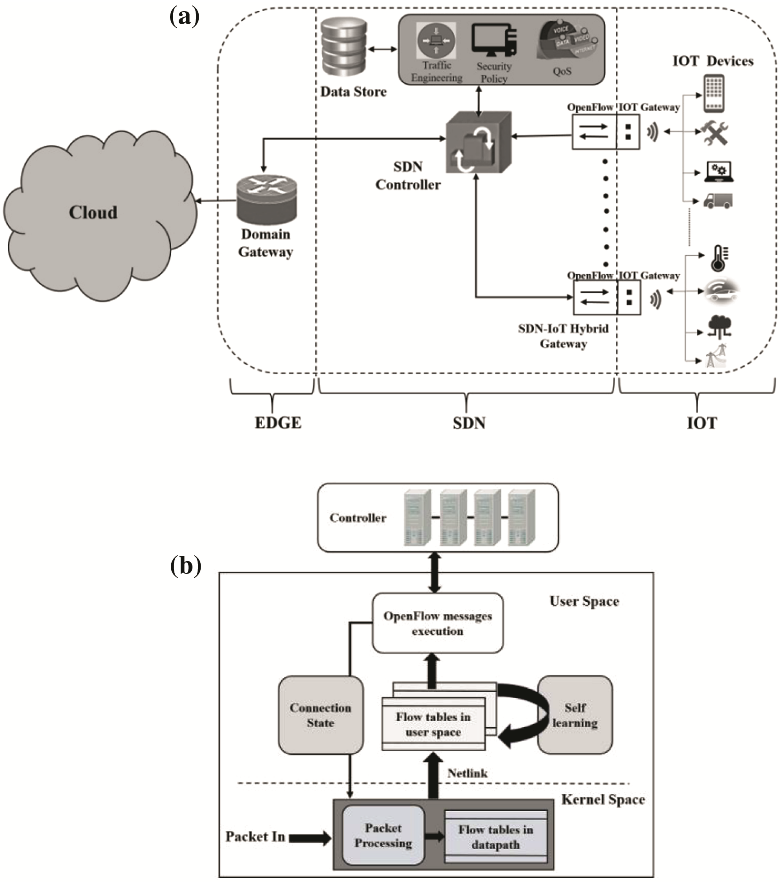


Fig. 5. (a) Tightly coupled SDN-IoT integration, (b) flow analysis in SDN-IoT hybrid device

dynamically loadable modules to support several protocols defined for IoT networks OpenFlow message processing is in charge of receiving and forwarding SDN OF protocol messages in the kernel directly.

In the Fig. 5b, SDN-IoT Hybrid device, Connection-state module performs the connection state tracking, synchronization. Self-Learning module performs flow table analysis and detect anomalies in the network traffic and track down the end points. Initialized with predetermined signatures and correlation rules for well-known attacks, this device has a learning module that learns the features/attributes set, this improving accuracy and granularity of detection. It also supports custom applications and associated libraries for IoT security and monitoring.

6.3 Global Cloud Command and Control

This conceptual global management network encompasses multiple local domains and a central command and control systems are hosted in the cloud. The domain level controllers (i.e. Fog) are interconnected with secure communication (SSL/TLS) channels. It runs a suite of business specific applications to manage enforce end-to-end security policies, traffic QoS, and big data analytics for the IoT network. In a federated architecture, a domain gateway controller negotiates with the global/other domain controller to determine for further processing (Fig. 6).

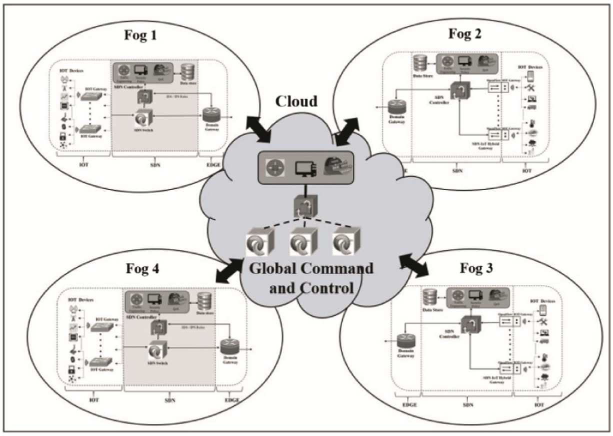


Fig. 6. SDN integrated IoT application

7 Initial Experimentation and Evaluation

In order to design the SDN based detection and mitigation more practical and dependable, we have to face the following challenges:

1. Traditional monitoring mechanisms based on IP entropy and TCP protocol proportion, Blacklisting, signatures are not effective with sophisticated arbitrary packet injection in network and botnets attack flows performing like a normal burst of traffic.
2. The cost of monitoring should be minimal and limited by the link bandwidth, speed and the real time requirements of the applications in target network.
3. The attack detection process should be followed up by mitigation strategies. Once the attacks are detected, it should be mitigated quickly by generating alerts, notifications and defensive rules communicated to SDN controller so that the actionable Rules are installed into the switching plane.

Hence to address these challenges, we have defined some key evaluation criteria in our SDN security controller, especially dealing with DoS attacks: 1. Packet handling rate/response to new connections or new flows and 2. Packet matching efficiency 3. monitoring cost for the new-flow attack. Our implementation strategy included fine-grained monitoring and defense mechanism that has lesser overhead in-terms of: new added modules code foot-print, instruction size in fast path for benign/normal traffic, memory usage for meta-data, extended flow-tables for dynamic security analytics, control protocol overhead and other costs. It can differentiate the DoS flow attack from the normal flow burst which ensures the minimal delay for normal packets to flow through our SDN framework and at the same time the attack/suspicious malware packets are detected at high accuracy and diverted to the self-learning anomaly detection module.

7.1 Experimental Network Topology

We have established a reconfigurable testbed to implement our design choices.

The IoT end-to-end architecture as depicted in this figure is divided in two parts (Fig. 7):

1. *Internal Network*
   - Edge domain gateway, a Linux firewall appliance running SNORT/IDS
   - SDN stack: modified RYU Controller and security/attack detection applications, modified vSwitch (OVS) switching software.
   - IoT stack: gateway running ContikiOS, supporting about 6 network protocols both Wi RF and modified middleware protocol stack, SDN Open flow enabled.
   - IoT Sensor network: 4–6 physical sensors/motes, 2 workstations running a virtual simulation of IoT sensors, running all WiFi/RF protocols
   - IoT Attack: This is a software simulator tool that generates attack traffic, fuzzing protocols and jamming
   - General Internal attacks: We use a set of machines that runs the widely used exploit kits and attack tools.
2. *External Network*
   - We setup legitimate hosts, and users and applications using transport protocols (TCP/IP, MQTT) to gain access into our test IoT network.
   - Attacking hosts users, botnet applications, who gain access through covert channels in TCP/IP, generate DoS attacks and targeted attack to test IoT network to infiltrate malware or steal data.

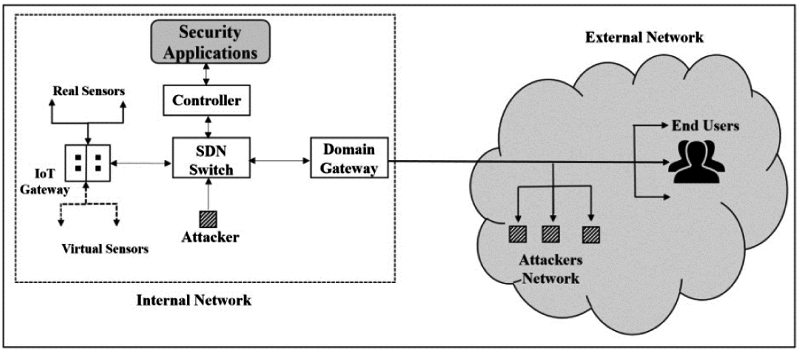


Fig. 7. Experimental network topology

7.2 Case Study: DDoS HTTP BOTNET ATTACK

Attack: A distributed DoS attack is usually mounted by a botnet, which uses a fleet of its victim machines who have legitimate IP address (no spoofing, hence difficult to detect and track and do not exhibit explicit indicators or statistical anomalies).

Detection: Taking the article [8] as reference, we improved upon their work in two aspects: 1. eliminated the need for the HTTP server to inform the SDN application about the botnet. Essentially the botnet detection logic is implemented based on traffic flow analysis at the SDN switch itself. 2. Optimised the detection processing overhead by employing better algorithm implementation approaches. The removal of the back-channel communication overhead between the HTTP server and SDN application (DBA) itself saved us major cycles. We conducted similar experiments and demonstrated that our mechanism has better performance, more portable and with no modifications to the target server environment.

Experiment:
SDN Defence policy: HTTP DoS blocking application runs in the SDN stack.

1. If the number of new connections/rate of new connection attempts exceed a threshold (in this case, 350 connections and 1 connection/second), then it's concluded that a botnet is active from external network. Drop the connections and packets to that destination address of the HTTP server D.
2. Send redirect message with a new destination address D encoded in the HTTP Response, it's assumed that the botnet are not programmed to decode the redirection scheme.
3. The legitimate clients will re-establish new connection to the D (address it is able to decode from the redirect response botnets are expected to continue attacking the original victim address D and are dropped.
4. Any new connections established to HTTP server at D, will be processed through the same detection logic.

Results: The Fig. 8(a) shows the botnet connections reaching the threshold of 350 connections, at which point all connections to destination D are dropped. And at the same time, the SDN application sends a HTTP response with redirection to new HTTP server D'. The genuine HTTP clients then establish new connections to D' which is shown in Fig. 8(b). There is an outage of few seconds (less than 3 s) for the genuine HTTP traffic and it's in acceptable as it's in new connection establishment phase. As we can see from the graphs that the overhead for packet processing by the SDN application at the SDN gateway switch is optimal (less than 3 s) and using dynamic flow rule learning entropy analysis, we can make this botnet detection mechanism responsive and practical for deployment in production IoT network.

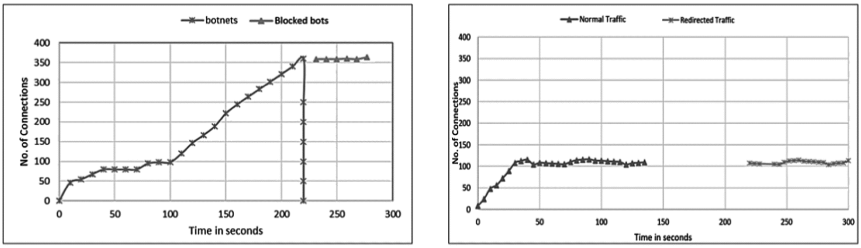


Fig. 8. (a) Botnet attack mitigation dynamics (b) genuine HTTP connections

7.3 Case Study: DoS Flooding Attack

Our experimental consisted of simple setup with IoT Gateway acting as a target of the attack connected behind an SDN gateway in the internal network.

Attack: Attacker nodes are simulated by running LOIC & hPing DoS attack tool from a group of nodes from the external network.

Detection: The DoS detection mechanism running as an SDN application on controller platform, executes a statistical function and analyses each flow based on threshold rates and based on the result, it installs new actionable rules on the data plane SDN switches - to forward or drop packets to the internal network. We based our experiment and compared with the work of [1] and we also ran the Cbench with identical setup, we demonstrated the efficacy of our defense mechanism in terms of implementation approach processing overhead in the SDN gateway.

Experiment: As the IoT link bandwidth is typically constrained by power and speed, we configured a peak link bandwidth of 2.5 Mbps. The genuine TCP traffic is run at 2 Mbps and after a while we ran attack traffic saturating the link at 2.5 Mbps.

Results: Figure 9a shows that at 7 s, the attack traffic kicked in to saturate the link and the genuine traffic was disrupted. But the DoS detection mechanism intercepted those attack traffic and in less than 3 s the bandwidth is recovered for the genuine TCP traffic. The DoS Flooding traffic is dropped at the SDN data plane itself without impacting the

SDN stack. Figure 9b shows the SDN controller performance in terms of number of flow installations per second. About 4.2 average flow installations per second on our DoS attack detecting switch compared to an average 7 flow installations per second on the standard L2 learning switch. So our work has clearly improved the agility of the DoS detection mechanism with SDN, compared to the prior work [1].

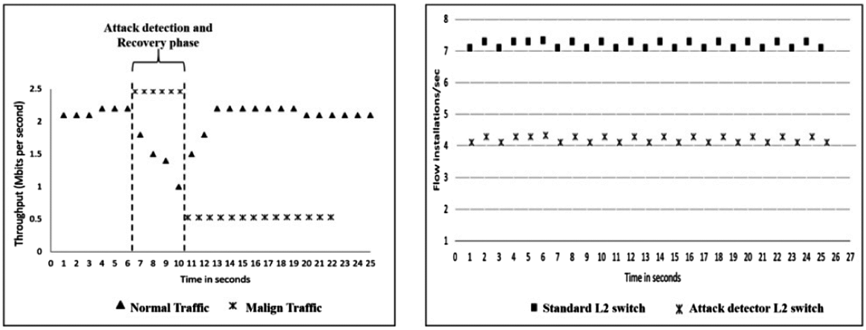


Fig. 9. (a) Link Saturation (b) controller performance

8 Conclusion

In this paper, we have discussed the potential of SDN and its capabilities such as traffic engineering and monitoring dynamic policy enforcement, access control at run time and mobility of devices. We conducted extensive simulations and the results confirm that the SDN-based-IoT applications can detect and mitigate the DoS attacks systematically. We developed reference applications for security policy and access control, in our IoT testbed using Openflow/REST interfaces and the results are proving the feasibility and efficacy of SDN in IoT networks.

Hence we make a strong case for SDN that privacy, trust and security policies can be efficiently enforced in IoT networks. This paper has provided an overview of challenges in IoT security, emphasized the need for flexible and dynamic methods of IoT network security, integration of SDN in IoT network. Our future work will expand these initial experiments to real Industrial IoT networks, fine tune and improve our design choices and position us to develop more efficient implementation to realize a SDN security framework for IoT applications. We believe that, our work has provided a practical proof and direction for applying SDN and other software-defined architectures to tackle extreme proliferation of IoT devices and deploy secure IoT networks for smart applications.

References

1. Bull, P., Austin, R., Popov, E., Sharma, M., Watson, R.: Flow based security for IoT devices using an SDN gateway. In: 2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud), pp. 157–163. IEEE (2016)

2. Lee, W., Nam, K., Roh, H.G., Kim, S.H.: A gateway based fog computing architecture for wireless sensors and actuator networks. In: 2016 18th International Conference on Advanced Communication Technology (ICACT), pp. 210–213. IEEE (2016)
3. Choi, Y.: Implementation of content-oriented networking architecture (CONA): a focus on DDoS countermeasure. In: Proceedings of 1st European NetF-PGA Developers Workshop (2010)
4. Zhang, Y.: An adaptive flow counting method for anomaly detection in SDN. In: Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies, pp. 25–30. ACM (2013)
5. Ezefibe, C.A., Shayan, Y.R.: Towards virtualisation and secured software defined networking for wireless and cellular networks. In: 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1–5. IEEE (2016)
6. Lin, H., Sun, L., Fan, Y., Guo, S.: Apply embedded openflow MPLS technology on wireless openflow–openRoads. In: 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 916–919. IEEE (2012)
7. Flauzac, O., Gonzalez, C., Nolot, F.: Developing a distributed software defined networking testbed for IoT. Procedia Comput. Sci. **83**, 680–684 (2016)
8. Lim, S., Ha, J., Kim, H., Kim, Y., Yang, S.: A SDN-oriented DDoS blocking scheme for botnet-based attacks. In: 2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 63–68. IEEE (2014)
9. Dinesh, M.K., Bhakthavatchalu, R.: Storage memory/NVM based executable memory interface IP for advanced IoT applications. In: 2016 International Conference on Recent Trends in Information Technology (ICRTIT), pp. 1–9. IEEE (2016)
10. Tortonesi, M., Michaelis, J., Morelli, A., Suri, N., Baker, M.A.: SPF: an SDN-based middleware solution to mitigate the IoT information explosion. In: 2016 IEEE Symposium on Computers and Communication (ISCC), pp. 435–442. IEEE (2016)
11. Vandana, C.: Security improvement in IoT based on Software Defined Networking (SDN). Int. J. Eng. Technol. Res. (IJSETR) **5**(1), 291–295 (2016)
12. Xu, T., Gao, D., Dong, P., Zhang, H., Foh, C.H., Chao, H.C.: Defending against new-flow attack in SDN-based internet of things. IEEE Access **5**, 3431–3443 (2017)

Estimation of End-to-End Available Bandwidth and Link Capacity in SDN

Manmeet Singh(✉), Nitin Varyani, Jobanpreet Singh, and K. Haribabu

Department of Computer Science and Information Systems, BITS, Pilani,
Pilani Campus, Pilani, India
{f2012763,f2009586,f2012124,khari}@pilani.bits-pilani.ac.in

Abstract. The traditional networks, with the control and data plane integrated into the same network devices, do not provide a global view of the network performance like degree of congestion, bandwidth utilization, etc. Software defined network (SDN) is an approach towards this problem which separates the control plane of the switch from its data plane and provide a centralized control plane so as to get a global view of the network performance and thus make decisions of how to regulate flows. In SDN, network monitoring can be achieved more efficiently than traditional networks using OpenFlow statistics. SDN controller can keep track of available bandwidth on each link and thus estimate end-to-end available bandwidth of a path simply by composing individual link bandwidths thus avoiding end-to-end probing. We have made two contributions in this paper: (i) proposed and validated a method to estimate end-to-end available bandwidth on any given path by composing link-wise available bandwidths (ii) proposed a method to measure link capacity using OpenFlow protocol. We compared our results to the ones obtained using the state-of-the-art bandwidth measurement tool, Yaz.

Keywords: Link capacity · End-to-end available bandwidth
SDN controller

1 Introduction

Software Defined Networking (SDN) aims to effectively program the network with software running on a central controller. Today's network switches and routers program their forwarding tables locally, which means that network devices make their own decisions internally about how to forward traffic. Traffic-forwarding decisions are informed by distributed control-plane protocols like spanning-tree, OSPF and BGP. But these traditional networking protocols have limited flexibility. In order for them to work, all network devices participating in the forwarding domain have to follow the same rules as defined by the protocol standard. That leaves little room for creativity or unusual business requirements.

M. Singh and N. Varyani—Both authors contributed equally to this work.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 130–141, 2018.
https://doi.org/10.1007/978-3-319-73423-1_12

SDN [1,2] is an emerging networking paradigm that overcomes the limitations of current network infrastructures. In case of traditional networks, we have both the control plane and the data plane integrated into the same network devices. SDN comes up with an approach to break this vertical integration by separating the control logic from the underlying routers and switches. With the separation of the control and data planes, network switches performs only the forwarding operations based on control logic which is implemented in a logically centralized controller (or network operating system), simplifying policy enforcement and network (re)configuration and evolution [3]. A logically centralized programmatic model does not postulate a physically centralized system [4]. The need to achieve adequate levels of performance, scalability and reliability would resist such a solution. Instead, production-level SDN network designs resort to physically distributed control planes [4,5]. To achieve the separation between the control plane and the data plane we have a well implemented programming interface between network devices (like switches and routers) and the SDN controller. The controller maintains the global view of the underlying network and the applications can obtain the statistics of the network state for improved performance. Also, the SDN controller controls the state of the dataplane elements with the help of well-defined communication protocol. OpenFlow [1,2] is a protocol for communication between the control and the forwarding layers of an SDN architecture. OpenFlow enabled switches provide important statistics about the network like port statistics, flow statistics, packets transmitted, packet received and packet loss.

The statistics obtained using SDN Controller can be used to estimate end to end available bandwidth. The end-to-end available bandwidth is defined as the maximum rate that the path can provide to a flow, without reducing the rate of the traffic in that path. This information can be very useful in congestion control, streaming applications and network selection. Our paper suggests an approach to estimate and validate the end-to-end available bandwidth. We have also proposed a method to estimate the capacity of links using statistics obtained from SDN Controller. We have validated the end-to-end available bandwidth values using state-of-the-art tools for bandwidth estimation like Yaz. We have also analyzed the effect of change in the polling time interval, the time interval after which SDN controller requests the network statistics, on the estimated end-to-end available bandwidth.

The reminder of the paper is structured as follows. Section 2 presents a short background on the related work. In Sect. 3 we present our design approach for measuring end-to-end available bandwidth in SDN. Sect. 4 describes the experimental setup we used to validate our method and Sect. 5 discusses about the results of tests. Finally, Sect. 6 ends the paper with concluding remarks.

2 Related Work

There are several end-to-end network performance measurement tools proposed in the literature such as Spruce [1], Pathload [3], IGI/PTR [4], Abing [5], pathChirp [5], DietTopp [6], Yaz [7], and ASSOLO [5]. These end-to-end measurement tools use packet pair techniques to measure performance metrics like bandwidth, delay, latency, etc. These techniques need to send packets from one side in a particular pattern and receive on the other end and analyze the arrival time etc. These approaches are required to be carried out on end-to-end basis. These techniques lead to packet duplication if bandwidth values are estimated for multiple end hosts. This will also cause congestion in network. This may also lead to increase in packets drop rate because of which estimated bandwidth values might not be accurate. All these limitations can be easily overcome by the use of SDN controller which queries network statistics and can estimate end-to-end available bandwidth for any path without packet duplication. Not only this, controller can also set high priority to the packets sent for bandwidth estimation so that these packets are not dropped on the way leading to more accurate bandwidth values than the existing methods.

OpenNetMon [8] is an existing approach in SDN to monitor per-flow metrics like throughput, delay and packet loss between source and destination. This approach measures performance metrics for each flow separately and not for a given path. Our proposed approach measures available bandwidth link-wise and composes on-demand to find out end-to-end available bandwidth for a given path. In [9], the authors measures the end-to-end available bandwidth between any two end-hosts in the network using SDN statistics but their work assumes link capacity to be known in advance. Our approach measures the link capacity of the links dynamically without having to run sender or receiver applications on end hosts. Additionally, we compare our results with well-known tool for bandwidth estimation in order to validate our results.

3 Design Approach

Our approach learns the topology using the controller API. Using this topology, it queries the statistics from the switches in the topology. These statistics are used to estimate the available bandwidth on each link in the network. Instead of using static values for the maximum capacity, we are estimating the capacity dynamically. This is because the links which are visible to the controller are not often the physical links. They may be aggregated links. The Fig. 1 given below illustrates our approach. The module "Create Topology" is storing the network topology fetched from SDN controller and is regularly updated. The module "Store Port Statistics" stores transmitted bytes for each port in a switch and is updated more frequently. This module is also storing time elapsed. The module "Estimate Link Capacity" estimates the link capacity dynamically.

3.1 Learning Topology and Consumed Bandwidth Using Controller

A hashable data structure is used to store the topology and consumed bandwidth/data rate in each link. A pair (S, D) is used as a key which is mapped to the ordered set $(P_1, P_2, TB, T, DR, MB)$ where S and D are source and destination mac addresses respectively of two switches connected directly in the network. Source and destination switches are connected through each other using their

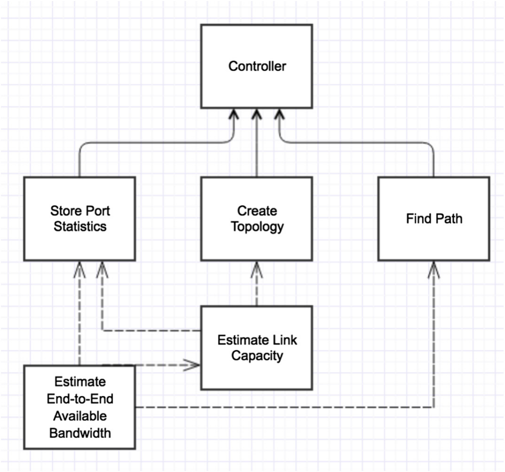


Fig. 1. Diagram illustrating our design approach

respective port P_1 and P_2. DR represents the bandwidth consumed between source and destination switch, T represents the time elapsed, TB is the bytes transmitted by the source S to destination D from beginning to time T and MB is the maximum bandwidth (capacity) of the link connecting the two switches which is estimated dynamically using the method described in Sect. 3.4. We then calculate the consumed bandwidth/datarate(DR) in the path from S to D at time T_{new} using the formula

$$DR = (TB_{new} - TB_{old})/(T_{new} - T_{old}) \tag{1}$$

where TB_{new} denotes the number of bytes transmitted through port P_1 since beginning to time T_{new}, TB_{old} is the number of bytes transmitted through port P_1 since beginning to time T_{old}. Using the Open Daylight statistics, an observation was made that for two switches connected by a link, the received bytes by one switch was much more that the bytes transmitted by the other switch at the ports through which they are connected. Such discrepancy is resolved by replication of packets from one interface of switch to all another interfaces within the switch when it receives a packet. Thus, to have proper calculation of bandwidth available in links, we have used transmitted bytes.

Same data structure is used to store the information regarding which switch is connected to which end host along with the consumed bandwidth in those links.

3.2 Finding Path Between Source and Destination

For a given flow, we are finding its route on the network set by the controller, and then calculating its end-to-end available bandwidth. We use the northbound API provided by the SDN controller for fetching the route of a flow in the network set by the controller according to the policies.

3.3 Estimating Available Bandwidth Between Source and Destination

For each link, we subtract Data Rate (DR) from Maximum Bandwidth (MB) to get available bandwidth of the link. The minimum of the available bandwidths of links along a path is the end-to-end available bandwidth of the path. Our approach estimates end-to-end available bandwidth by composing link-wise bandwidth which avoids generation of redundant traffic if bandwidth values have to be estimated for multiple pair of hosts.

3.4 Estimating Link-Capacity Dynamically

We need the value of the maximum available capacity in order to estimate the available bandwidth. There are many available Opensource tools which can find the capacity of the path. Pathrate [10] estimates the bottleneck capacity by sending packet pairs, called probing packets, back-to-back and measuring the dispersion of the packet pairs. Dispersion varies among different packet pairs and the packet pair bandwidth distribution is analyzed to relate it with capacity.

But with these tools, we need to have receiver and sender applications running on the end hosts. In our approach, the controller itself controls the traffic and uses the port statistics to measure the capacity of a given link. The controller generates UDP packets with a unique destination port number. Since the network topology does not change so frequently, the measurement of capacity need not be so frequent, and thus it will not significantly affect the real traffic. We added flow entries in the ovs-switches to direct the traffic towards the link for which we need to measure the capacity. Consider two ovs-switches s1 and s2 connected with a link l whose capacity needs to be estimated. We add a flow

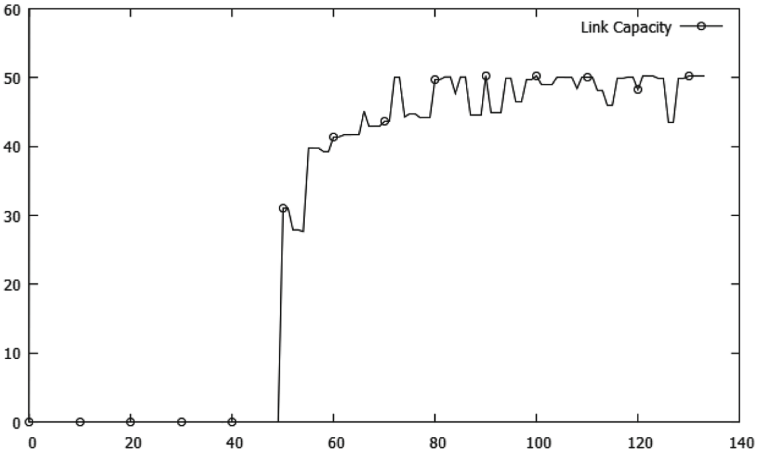


Fig. 2. Estimating link capacity of 50 MB link

entry in s1 to forward all the packets with the predefined unique UDP destination port number P to s2. In s2, we add a flow entry to drop the packet with the unique port number P as these packets are just used for measurement of capacity. We measure the rate at which data is received at s2 using port statistics. As shown in Fig. 2, the traffic is increased in order to estimate the maximum rate at which data is received at the receiver end of the link. We stop iterating once we do not observe any change in the maximum value of received throughput, which is the final link capacity.

4 Experimental Setup

4.1 Bandwidth Estimation Module

As discussed in previous section, we estimate the available bandwidth between the source and the destination by composing link-wise bandwidths using statistics obtained from the controller. The bandwidth values will be used by the validation module to analyze its correctness by comparing it with a well-known bandwidth measurement tool, Yaz.

4.2 Validation Module

This module takes care of analyzing the difference in the bandwidth values obtained from our work and other available tools. The first measurement is the value given by our SDN script for bandwidth measurement. The second measurement is the bandwidth value as given by an open source bandwidth measurement tool, Yaz. While the Yaz adopts heuristic methods for estimating bandwidth by considering the delay between packet streams, our method uses port statistics like transmitted bytes in a given link to estimate the bandwidth.

The basic difference between Yaz and our method is that Yaz uses an experimental approach, sending packet probes and observing delays to measure the available bandwidth but our approach uses a statistical approach, using the data of all switches in the path to find out the end-to-end available bandwidth.

4.3 Traffic Generation Module

The tool used for generating regulated traffic is Distributed Internet traffic generator (D-ITG). D-ITG can be used to generate traffic in a network with controllable parameters like packet rate, packet size, etc. DITG can be used to generate traffic in a particular distribution like normal or Poisson distribution.

4.4 Link Capacity Estimation Module

As discussed in Sect. 3.4, we need to know the capacities of the links in the path. We estimated the capacity of mininet links for various topologies and the results are given in Table 1. We have also estimated the capacity of the physical LAN wire used in our setup in Fig. 5. The actual capacity of the LAN wire used is 100 Mbps. The first row in Table 1 corresponds to the physical LAN wire.

The estimated values (in Table 1) are higher than expected values by a small amount in all the experiments dues to some experimental errors.

Table 1. Expected capacity and estimated capacity

Expected capacity (Mbps)	Estimated capacity (Mbps)
100	101.2344456787
50	50.2809047531
20	21.4398119679
10	10.8275808985

5 Results and Discussion

5.1 Results on Mininet Testbed

In our first setup, OpenDaylight was used as a controller and mininet was used to emulate the topology. A custom topology was created in python for our experimentation (Fig. 3). 10.0.0.11 is the IP address allocated by mininet to the physical computer in which mininet, opendaylight and library is running and can be seen in the Fig. 3. Mininet network was connected to the physical computer using Network Address Translation (–nat).

A cross traffic was generated in the network using iperf tool. At node 10.0.0.4, a user application is connected to our library and is requesting our library end-to-end available bandwidth between 10.0.0.4 and 10.0.0.6. As can be observed

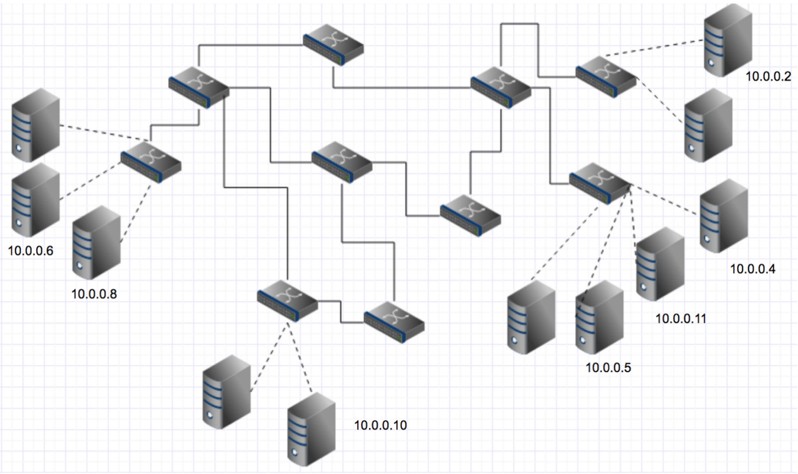


Fig. 3. Mininet testbed

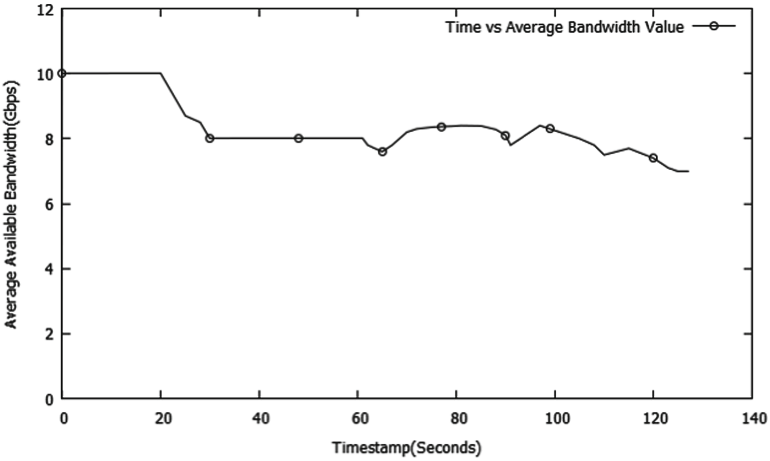


Fig. 4. Time vs. end-to-end available bandwidth between 10.0.0.4 and 10.0.0.6

from Fig. 4, initially there was no traffic in the network and hence we can observe the value of 10 Gps bandwidth available between 10.0.0.4 and 10.0.0.6.

We can observe a dip at t = 20 s. This is due to udp traffic of 350 Mbps sent from 10.0.0.6 to 10.0.0.4. Then a udp traffic of 350 Mbps from node 10.0.0.10 to node 10.0.0.2 was added to the network at around 60 s. We can observe that the traffic has become turbulent. Another udp traffic of 350 Mbps from 10.0.0.8 to 10.0.0.5 was added to the network at around 100 s. We can thus observe a dip in the bandwidth available at 100 s.

5.2 Results on Physical Testbed

We created a second testbed which involved a physical link between the switches to validate our approach in this scenario also. We have created a testbed which consists of two OVS switches, each installed on a separate physical machine (Host-2 and Host-4 in Fig. 5). These two machines, and hence the switches, are connected by a LAN wire of capacity 100 Mbps. Each physical machine has a Virtual Machine (VM) acting as a host (Host-1 and Host-3 in Fig. 5). Each of these VMs are connected to the switches as shown in Fig. 5. Hence we have a linear topology with two switches and four hosts.

We are using D-ITG for the generation of traffic inside the network. We have used ryu controller for this testbed.

5.3 Comparing Results with Yaz

Figure 6 shows the end-to-end available bandwidth values obtained from Yaz and our application plotted against time-stamp. We can observe significant amount of similarity between Yaz results and our work.

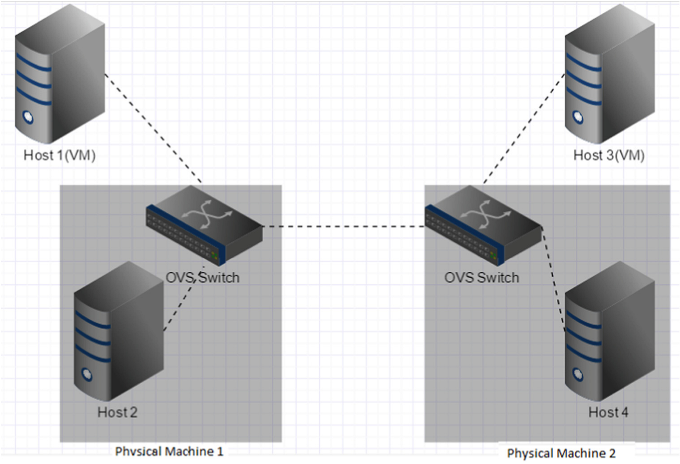


Fig. 5. Testbed with physical link and ovs-switches

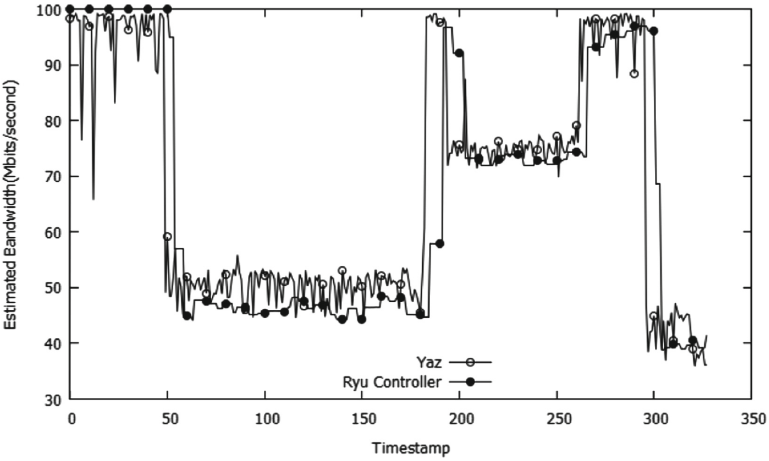


Fig. 6. Comparing the bottleneck bandwidth of the obtained path measured using Yaz and our method

During the experiment, the amount of traffic generated through the link was varied after every few seconds so as to verify that irrespective of varying the traffic through the link, the values given by our Ryu script and Yaz tool are comparable.

The accuracy α in this experiment is computed as:

$$\alpha = 1 - |(B_{yaz} - B_{ryu})/B_{yaz}| \tag{2}$$

where B_{yaz} and B_{ryu} is the bandwidth obtained by yaz and ryu respectively. The graph shows that the bandwidth calculated using our approach is very close to the actual bandwidth available, as calculated by the tool yaz. These two measurements come from two methods which use an entirely different approach from each other, and yet give comparable results and thus provides a ground for the correctness of our approach. The advantage of this statistical method over the experimental method of Yaz is that in this case we have the available bandwidth corresponding to each link in the path which gives us the knowledge of which link in the path has the minimum available bandwidth. Another advantage of this method is that it saves time by storing link wise bandwidth which can be used instantaneously to calculate bottleneck bandwidths for multiple paths. Unlike in heuristic methods for bandwidth estimation, we do not need to do end to end probing in our approach which takes a minimum of round-trip time.

5.4 Estimating Bandwidth Values with Different Polling Time Intervals

One of the important parameters in our approach is the time after which port statistics are requested for every switch from the controller at regular intervals. It is important to note that in a large network, it will put significant pressure/computing delay on the network if we request stats from the controller at a high frequency. Therefore it is important that this time gap between two such requests for statistics is tuned according to the size of the network. We ran different experiments where we varied the polling time interval. We then calculated the cumulative of the end-to-end available bandwidth measured by Yaz and our application across time. In Fig. 7, the absolute of the difference of the two cumulative values was plotted on the y-axis with the time on the x-axis. From the figure, we can infer that the difference increases with increase in polling time. This is because as we increase the polling time, the bandwidth

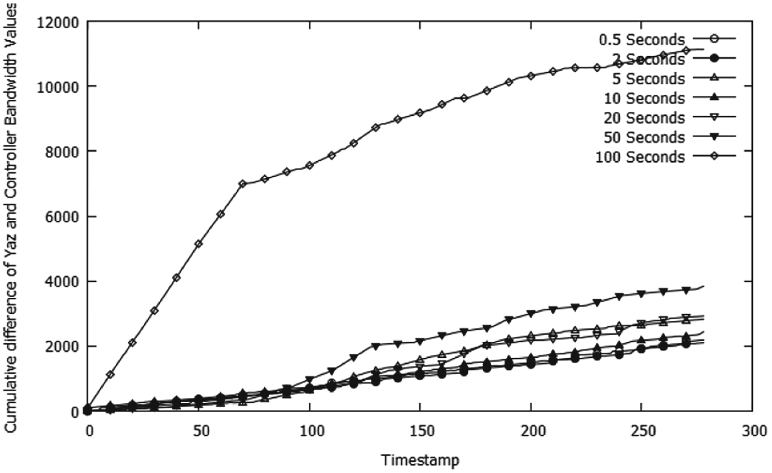


Fig. 7. Cumulative absolute difference between throughput of Yaz and Ryu

estimation period will also be increased and we may not be able to capture the accurate bandwidth with that time delay. But, with the polling time interval of 500 ms, the difference in cumulative bandwidth values is smaller and more closer to the cumulative bandwidth value of Yaz. Thus, for our testbed, we found the polling time of 500 ms to give most accurate measurement of end-to-end available bandwidth.

6 Conclusion

In this paper we present an approach to measure end-to-end available bandwidth in Software Defined Networks (SDN). We have also validated our results with a well known bandwidth estimation tool, Yaz. The bandwidth calculated by our approach is very close to the actual bandwidth available, as calculated by the tool Yaz. We have also presented an approach to find the maximum capacity of the link dynamically rather than taking static values. We also observed the impact of controller polling time interval on the estimation of bandwidth.

There are few areas which are still undiscovered like impact of our approach on the network traffic. SDN controller keeps on querying statistics from all the switches in network after fixed interval which may lead to excessive traffic. We may improve upon this in our future work by querying only the selected switches based on some factors like ignoring stable switches in non-turbulent environment. We may also work on the understanding the scalability of our approach by considering larger networks. The proposed solution doesnt require any application running on end hosts and can be used by various applications to tune the parameters based on the available bottleneck bandwidth.

References

1. McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., Turner, J.: OpenFlow: enabling innovation in campus networks. SIGCOMM Comput. Commun. Rev. **38**(2), 69–74 (2008)
2. ONF: Open networking foundation (2014). https://www.opennetworking.org/
3. Kim, H., Feamster, N.: Improving network management with software defined networking. IEEE Commun. Mag. **51**(2), 114–119 (2013)
4. Koponen, T., Casado, M., Gude, N., Stribling, J., Poutievski, L., Zhu, M., Ramanathan, R., Iwata, Y., Inoue, H., Hama, T., Shenker, S.: Onix: a distributed control platform for large-scale production networks. In: Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation, ser. OSDI 2010, p. 16. USENIX Association, Berkeley (2010)
5. Navratil, J., Cottrell, R.L.: ABwE: a practical approach to available bandwidth. In: Proceedings of the 4th International Workshop on Passive and Active Network Measurement PAM 2003 (2003)
6. Jain, M., Dovrolis, C.: Pathload: a measurement tool for end-to-end available bandwidth. In: Proceedings of the 3th International Workshop on Passive and Active Network Measurement PAM 2002 (2002)
7. Hu, N., Steenkiste, P.: Evaluation and characterization of available bandwidth probing techniques. IEEE JSAC **21**(6), 879–894 (2003)

8. Van Adrichem, N.L.M., Doerr, C., Kuipers, F.A.: OpenNetMon: network monitoring in openflow software-defined networks. In: 2014 IEEE Network Operations and Management Symposium (NOMS). IEEE (2014)
9. Megyesi, P., et al.: Available bandwidth measurement in software defined networks. In: Proceedings of the 31st Annual ACM Symposium on Applied Computing. ACM (2016)
10. Dovrolis, C., Ramanathan, P., Moore, D.: Packet-dispersion techniques and a capacity-estimation methodology. IEEE/ACM Trans. Netw. **12**(6), 963–977 (2004)
11. Rechert, K., McHardy, P., Brown, M.A.: HFSC Scheduling with Linux
12. Strauss, J., Katabi, D., Kaashoek, F.: A measurement study of available bandwidth estimation tools. In: Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement IMC 2003 (2003)
13. Ribeiro, V., Riedi, R., Baraniuk, R., Navratil, J., Cottrell, L.: pathChirp: efficient available bandwidth estimation for network paths. In: Proceedings of the 4th International Workshop on Passive and Active Network Measurement PAM 2003 (2003)
14. Johnsson, A., Melander, B., Bjorkman, M.: DietTopp: a first implementation and evaluation of a simplified bandwidth measurement method. In: Proceedings of the 2nd Swedish National Computer Networking Workshop (2004)
15. Goldoni, E., Rossi, G., Torelli, A.: Assolo, a new method for available bandwidth estimation. In: Proceedings of the Fourth International Conference on Internet Monitoring ICIMP 2009, pp. 130–136, May 2009
16. Luckie, M.J., McGregor, A.J., Braun, H.-W.: Towards Improving Packet Probing Techniques Science (1989)
17. Pakzad, F., Portmann, M., Hayward, J.: Link capacity estimation in wireless software defined networks. In: 2015 International Telecommunication Networks and Applications Conference (ITNAC). IEEE (2015)

Estimation of Raw Packets in SDN

Yash Sinha[1(✉)], Shikhar Vashishth[2], and K. Haribabu[1]

[1] Department of Computer Science and Information Systems, BITS, Pilani,
Pilani Campus, Pilani, India
h2016077@pilani.bits-pilani.ac.in

[2] Department of Computer Science and Automation, Indian Institute of Science,
Bangalore, India

Abstract. In SDN based networks, for network management such as monitoring, performance tuning, enforcing security, configurations, calculating QoS metrics etc. a certain fraction of traffic is responsible. It consists of packets for many network protocols such as DHCP, MLD, MDNS, NDP etc. Most of the time these packets are created and absorbed at midway switches. We refer to these as raw packets. Cumulative statistics of sent and received traffic is sent to the controller by OpenFlow compliant switches that includes these raw packets. Although, not part of the data traffic these packets get counted and leads to noise in the measured statistics and thus, hamper the accuracy of methods that depend on these statistics such as calculation of QoS metrics.

In this paper, we propose a method to estimate the fraction of the network traffic that consists of raw packets in Software Defined Networks. The number of raw packets transferred depends on the number of switches and hosts in the network and it is a periodic function of time. Through experiments on several network topologies, we have estimated a way to find a cap on the generated raw packets in the network, using spanning tree information about the topology.

Keywords: Raw packets · OpenFlow
Software Defined Networks (SDN)

1 Introduction

The traditional networks with their closed and proprietary network devices by their rigidity have narrowed the possibilities of any innovation and improvement. Further, the strong coupling between data and control plane of network devices, especially switches and routers has restricted the development of new functionalities in the existing networks.

To change the current affairs, Software Defined Networks (SDN), an emerging paradigm in networking advocates rifting of control and data plane, for promoting network programmability by splitting network's control logic from the underlying network devices. The state and statistics of the network greatly affect the decision making process while the controller exerts its centralized control.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 142–153, 2018.
https://doi.org/10.1007/978-3-319-73423-1_13

OpenFlow 1.3 [1,2] defines structure and semantics for multipart request and response messages which are used by the controller to query statistics from an OpenFlow compliant switch. The counters of received and transmitted packets for flows as well as ports statistics include count of packets which are responsible for network management such as network monitoring, traffic measurement, pushing configurations etc. We refer to these packets as raw packets. Some of the protocols which generate raw packets are NDP, DHCP etc. Often these packets are generated and absorbed at the intermediate switches and are not part of the end-to-end data traffic. Therefore, accuracy of many of the functionalities of the controller that depend on pure end-to-end packet statistics such as calculation of QoS metrics (delay, packet loss, bandwidth etc.), bandwidth management, congestion avoidance, detection and mitigation etc. is hampered. Thus, an estimate of the fraction of network traffic that consists of raw packets is needed.

In this work, we present the intuitions that can help us to estimate fraction raw packet traffic in the network. We also explain the experiments that we conducted to validate those intuitions. Further, we present an implementation and evaluation of a prototype using Ryu controller [3] running on the top of Open vSwitches [4] emulated using Mininet [5,6].

2 Raw Packets

In this section, we define raw packets and discuss its impact in SDN and the extent to which it affects statistics.

2.1 Definition

We define raw packets as fraction of traffic in the network that is responsible for management of network which includes as network monitoring, performance management, enforcing security policies, traffic measurement, pushing configurations etc., but not a part of the end-to-end data traffic. Simply put, it is control, non-user generated traffic.

For example, the switches and routers perform network discovery, multicast listeners discovery etc. and hosts request networking parameters from DHCP servers etc. Many of these packets generated by protocols like NDP, CDP etc. are generated and absorbed at switches while packets generated by protocols like MDNS, DHCP are absorbed at servers. Hence, they are not part of the end-to-end data traffic.

2.2 Impact of Raw Packets in Statistics

Globally, a lot of network traffic comprises of raw traffic [7]. In our initial explorations, we found out that around 3200–3500 packets are generated every 30 min even in a small emulated network of 2 hosts and 2 switches, which is around 3–9% of the total traffic. Cumulatively, over a period of time, these statistics affect the calculation of metrics such as queue size, bandwidth, packet loss etc.

in a Software Defined Network. For example, OpenNetMon [8] estimates per-flow packet loss by polling each switchs port statistics. But because switches send cumulative statistics (that includes raw packets), the reported statistics is unable to differentiate between raw packets and end-to-end user data traffic.

Further, each of the protocols have separate rate of raw packet generation. For example, MDNS queries are generated quite frequently as compared to ICMP Router Solicitation packets. So, instead of identifying a list of triggers of raw packet generation for so many protocols, (which will be even more in a realistic network), an experimental method to detect a periodic time interval is more practical and deployable (Table 1 and Fig. 1).

Table 1. Packet generated by different protocols

Protocols	Packets
ARP	70
DHCP	528
ICMP	928
IGMP	24
MDNS	1578
Total	3583

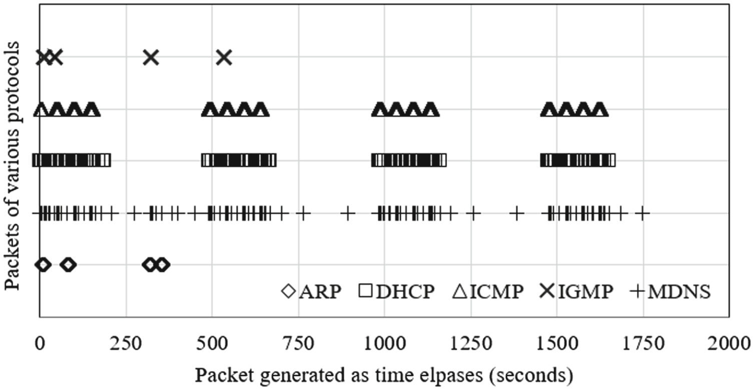


Fig. 1. Rate of raw packet generation of different protocols

2.3 Impact of Raw Packets in SDN

Several background services in SDN controller are responsible for generating a lot of raw packets for several operations like topology discovery, monitoring via packet injection, configuration pushing etc. Therefore, as compared to traditional networks SDN based networks generate more raw packets. Pakzad et al. [9] have shown that even for medium sized topologies such as a tree topology, (d = 4, f = 4, switches = 85 and ports = 424), there can be as many as 424 LLDP packets generated for discovering links part of which are generated and injected by the controller. At first sight, one may argue that the controller already knows about these packets and can separate these counters from received statistics, but many of these packets are generated at switches also such as LLDP Packet-Out message for each port on each switch in OFDP [10] and thus, the controller is unable to estimate the links where these packets are generated.

Because of criticality of load on controller and performance for the scalability of a Software Defined Network [11], authors in [12] have advocated a need to make a trade- off between resource overhead and measurement accuracy. Thus, PayLess [13] proposes a frequency adaptive statistics collection scheduling algorithm and Pakzad et al. propose a new approach to reduce processing cost due to topology discovery in the controller with a minimum reduction of 67% in terms of messages. In their recent work [14], Pakzad et al. have conducted a range of experiments on the OFELIA SDN testbed [15], on a network topology across Italy, Spain, Belgium and Switzerland, the results of which highlight that considerable amount of LLDP packets are generated. For auto configuration in SDN [16], extensions to current protocols such as DHCP-SDN have been proposed that will lead to even greater fraction of raw traffic. These works emphasize that considerable amount of raw traffic is generated that distorts traffic statistics to a greater extent.

3 Intuition

Here, we present the intuitions that can help us to estimate fraction raw packet traffic in the network. We also explain the experiments that we conducted to validate those intuitions.

3.1 Hypothesis

Periodic message exchanges. The message exchanges for network management are periodic in nature. For example, every 15 s a router may send messages to its adjacent routers for network discovery. Therefore, with the help of the time period of the periodic function and number of cycles elapsed, one can estimate roughly the number of messages exchanged.

Raw packets generated proportional to network devices. The number of raw packets transferred in a subnet should be directly proportional to the number of switches and hosts in the network. Thus, the total number of raw packets in the network can be estimated.

Packet flow via spanning tree. Number of packets through each link in the network can be estimated using the spanning tree information about the topology.

3.2 Validation

Experimental Setup. The network is emulated using a Network Emulator, Mininet which emulates any number of virtual end-hosts, routers, switches, and links on a Linux kernel. Furthermore, it allows us to create many custom topologies and emulate some link parameters like a real Ethernet interface, e.g., link speed, packet loss, and delay. We use SDN enabled (i.e. OpenFlow [1] compliant) Open vSwitch Kernel switches and Ryu controller to handle their control plane.

We emulate the network topology using L2 learning switches. These switches memorize source-port mapping by examining each packet. By this mechanism each MAC address gets bound with a port. Afterwards, if the destination address of a new packet is found to be associated with some port then, the packet is pushed to the given port, otherwise it is flooded on all the ports of the switch. No other traffic, except raw traffic is generated in such a network.

Periodic message exchanges. We emulate a small network to detect if messages exchanged for network management are periodic in nature (Fig. 2).

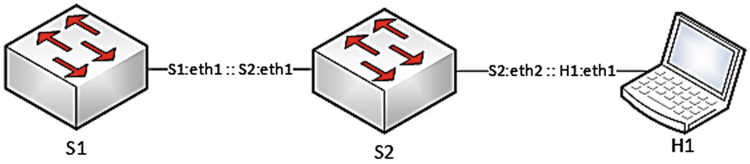


Fig. 2. A network with a host and two switches

The packets exchanged between two switches and between a switch and a host are monitored by sending PortStats query from the controller every 5 s. As shown in Fig. 3, we notice that the number of packets exchanged are periodic in nature. Thus for an experimentally estimated time period, the number of raw packets exchanged between a pair of network devices remains almost constant (doesnt depend on the count of switches and hosts in the network).

Raw packets generated proportional to network devices. As we increase the number of switches and hosts in the network, we see a linear increase in the number of messages exchanged (Fig. 4).

Packet flow via spanning tree. For different topologies such as star, tree etc. while analyzing the PortStats of the switches we find that the number of raw packets exchanged for each port of a switch is proportional to number of switches and hosts connected to the switch through that port in the spanning (Fig. 5).

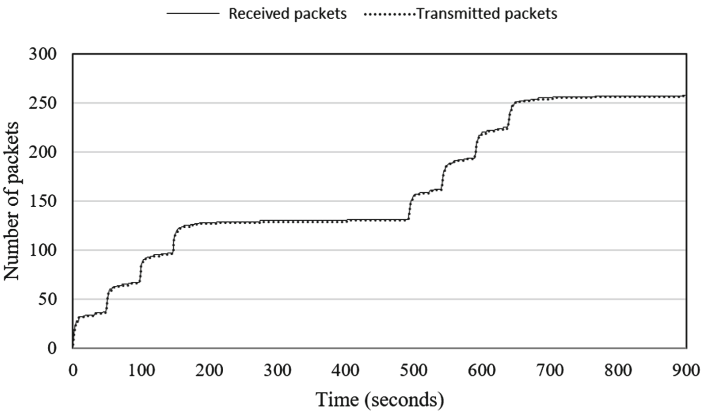


Fig. 3. Exchanged packets between one host and one switch

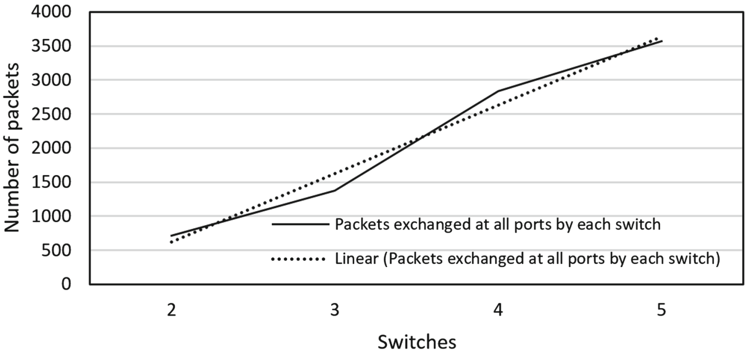


Fig. 4. Total packets exchanged at all ports by each switch as a function of number of hosts in the network

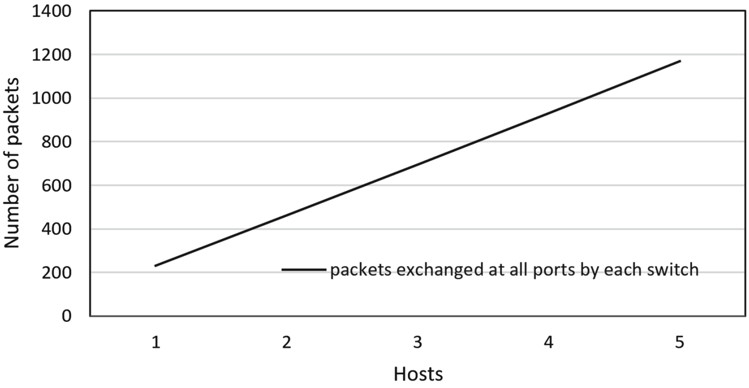


Fig. 5. Packets reported at a port as a function of number of switches connected in the spanning tree

4 Proposed Methodology

In this section, we present the steps of our proposed methodology, the way to get spanning tree information using the controller and the way to estimate constants of the formula.

4.1 Steps

It consists of following steps:

1. Request spanning tree information T of the entire network from the controller.
2. By iterating over each interface of every network device, calculated the number of network devices attached to it.
3. Using the statistics of exchanged packets, timer period (τ) and A_τ and B_τ are estimated. A_τ denotes the number of raw packets exchanged between two switches. And B_τ denotes the number of packets exchanged between a host and a switch in τ time.
4. Then, for a given time t, the total count of the exchanged packets between network devices is calculated using the following formula:

$$N = (A_\tau \times \alpha + B_\tau \times \beta) \times (t/\tau)$$

α: # switches in interface's subnetwork
β: # hosts interface's subnetwork
A_τ: # raw packets exchanged between two switches in τ time,
B_τ: # raw packets exchanged between a host and a switch in τ time

Algorithm 1. Raw packet Estimation

```
procedure RawPackets($T, \alpha, \beta, A_\tau, B_\tau, t, \tau$)
    for ∀ switches $\alpha_i$ in spanning tree $T$ do
        for ∀ switch ports $p_j$ do
            $N(p_j) = (A_\tau \times T\alpha_i p_j n + B_\tau \times T\beta_i p_j n) \times (t/\tau)$
        end for
    end for
end procedure
```

In the above algorithm, $T\alpha_i p_j n$ is the number of switches connected to p_j of α_i and $T\beta_i p_j n$ is the number of hosts connected to p_j of α_i.

4.2 Getting Spanning Tree Information

We run spanning tree protocol at the Ryu controller using OpenFlow 1.3 [2]. By sending a Port Modification message to the switch, it is possible to control the following (Table 2):

Table 2. Port settings allowed in OF 1.3 used for STP implementation

Values	Description
OFPPC PORT DOWN	Status, disabled by service personnel
OFPPC NO RECV	Rejects all packets received
OFPPC NO FWD	No forwarding from the port
OFPPC NO PACKET IN	Packet-In messages, not discharged in case of table-miss

Initially, to receive BPDU packets at the controller, we install flow entry that sends Packet-In of BPDU packets in each switch. To control sending/receiving of BPDU packets, MAC learning is employed. For various STP states the following settings are set up (Table 3):

Table 3. STP state configurations

Status	Port configuration	Flow entry
DISABLE	NO RECV/NO FWD	No setting
BLOCK	NO FWD BPDU	Packet-In, drop packets other than BPDU
LISTEN	No setting	BPDU Packet-In, drop packets other than BPDU
LEARN	No setting	BPDU Packet-In, drop packets other than BPDU
FORWARD	No setting	BPDU Packet-In

When connection between each OpenFlow switch and the controller is completed, exchange of BPDU packets starts and root bridge selection, port role setting, and port state change takes place.

4.3 Estimating Time Period and Other Constants

Time period, τ was estimated by emulating a traffic free network for a long duration of time and analyzing the statistics of packets exchanged. Every switch was polled every 5 s for PortStats for 30 min. The time period of the periodic pattern observed is taken as τ. We take A_τ as average number of packets that are exchanged between two switches and B_τ as the average number packets exchanged between a host and a switch in that time period. If we want to know the number of raw packets exchanged at a time which is not a multiple of v, we can have those values of A_τ and B_τ looked up, e.g. from Fig. 3 at that point of time.

5 Evaluation and Inferences

Here we present the results of the application of above methodologies and the error rate with which we were able to report the raw packets accurately.

5.1 Estimation of Constants

As shown in Fig. 3, we noticed that on an average 345 packets are exchanged between two switches in every 450 s. We take this value as A_τ. Even for a network of varied size and complexity, we found this value to be, in range of 340 to 350 packets, almost constant. Therefore the time period of the network (τ) is 450 s.

Similarly, for calculating the number of packets exchanged between a switch and a host we noticed that on an average 115 packets (τ) are exchanged between a switch and a host in every 450 s.

5.2 Setup I

In the setup we emulate the topology as shown in Fig. 6. For estimating raw packets at port 1 (right side) of switch S1, we see that there are two switches and three hosts connected to the switch via port 1. The time period as calculated above is 15 min. Therefore, the estimated number of raw packets for 30 min is $(345 \times 2 + 115 \times 3)(900/450) = 2070$, which is within 3% experimental error rate. Values for other ports are calculated in a similar way as tabulated in Table 4.

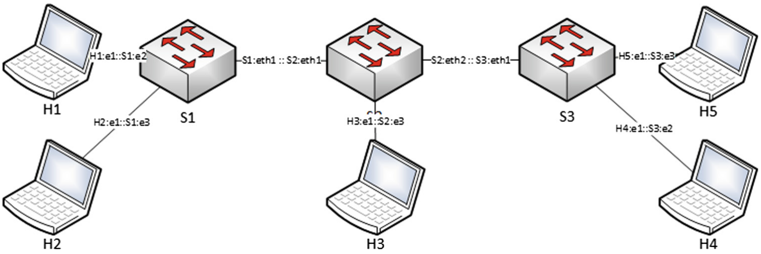


Fig. 6. Experimental Setup 1

Table 4. Raw packet estimation for Setup 1

Switch/port	Switch-1	Switch-2	Switch-3
Port-1	2007	1130	2052
Port-1 estimation	2070	1150	2070
Port-1 error	3.14%	1.77%	0.88%
Port-2	224	1130	223
Port-2 estimation	230	1150	230
Port-2 error	2.67%	1.77%	3.13%
Port-3 225	224	224	
Port-3 estimation	230	230	230
Port-3 error	2.22%	2.67%	2.67%

5.3 Setup II

Here we emulate a tree topology as shown for 900 s. Therefore, number of cycles elapsed is 2. For port 2 of switch S1, we have one switch and three hosts, therefore the number of raw packets exchanged is $(345 \times 1 + 230 \times 3)(900/450) = 2070$, which is within 2.57% experimental error rate (Table 5 and Fig. 7).

Table 5. Raw packet estimation for Setup 1

Switch/Port	Switch-1	Switch-2	Switch-3	Switch-4
Port-1	2851	471	477	474
Port-1 estimation	2070	460	460	460
Port-1 error	3.19%	2.34%	3.56%	2.95%
Port-2	2833	470	473	480
Port-2 estimation	2070	460	460	460
Port-2 error	2.57%	2.12%	2.74%	4.16%
Port-3	2826	470	468	476
Port-3 estimation	2070	460	460	460
Port-3 error	2.34%	2.12%	1.71%	3.36%

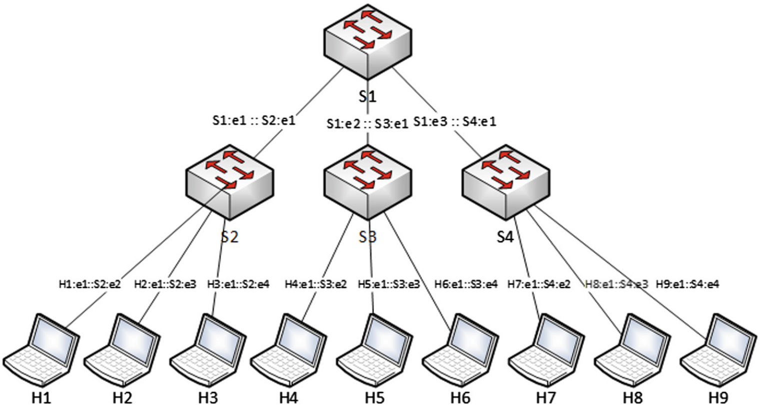


Fig. 7. Experimental Setup 2

6 Conclusion

We have proposed a method to estimate the fraction of raw packets in a given SDN network, which is around 3–9% of the total traffic. The estimation technique proposed in this paper will keep controller informed about the flow of raw packets in the network. This information can be utilized to increase the accuracy of the various techniques like OpenNetMon [8], OpenTM [17] etc. which rely on the cumulative statistics of packet transmitted and received from the switches. The method can be even further generalized for other network devices such as routers as well.

References

1. Mckeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., Turner, J., Louis, S.: OpenFlow: enabling innovation in campus networks. ACM SIGCOMM Comput. Commun. Rev. **38**(2), 69 (2008)
2. Pfaff, B., Lantz, B., Heller, B., Barker, C., Cohn, D., Talayco, D., Erickson, D., Crabbe, E., Gibb, G., Appenzeller, G., Tourrilhes, J., Pettit, J., Yap, K., Poutievski, L., Casado, M., Takahashi, M., Kobayashi, M., McKeown, N., Balland, P., Ramanathan, R., Price, R., Sherwood, R., Das, S., Yabe, T., Yiakoumis, Y., Kis, Z.L.: OpenFlow Switch Specification 1.3 (2012). https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.3.0.pdf
3. Ryu SDN Framework. https://osrg.github.io/ryu/
4. Open vSwitch. http://openvswitch.org/
5. Mininet: An Instant Virtual Network on your Laptop (or other PC) - Mininet. http://mininet.org/
6. Lantz, B., Heller, B., McKeown, N.: A network in a laptop. In: Proceedings of the Ninth ACM SIGCOMM Workshop on Hot Topics in Networks - Hotnets 2010, p. 16 (2010)
7. Official Google Blog: Google Public DNS: 70 billion requests a day and counting. https://googleblog.blogspot.in/2012/02/google-public-dns-70-billion-requests.html
8. Van Adrichem, N.L.M., Doerr, C., Kuipers, F.A.: OpenNetMon: network monitoring in OpenFlow software-defined networks. In: IEEE/IFIP NOMS 2014 - IEEE/IFIP Network Operations and Management Symposium, Management in a Software-defined World (2014)
9. Pakzad, F., Portmann, M., Tan, W.L., Indulska, J.: Efficient topology discovery in software defined networks. In: 2014 8th International Conference on Signal Processing and Communication Systems, ICSPCS 2014, Proceedings, May 2016 (2014)
10. OpenFlowDiscoveryProtocol GENI: geni. http://groups.geni.net/geni/wiki/OpenFlowDiscoveryProtocol
11. Tootoonchian, A., Gorbunov, S., Ganjali, Y., Casado, M., Sherwood, R.: On Controller Performance in Software-Defined Networks
12. Moshref, M., Yu, M., Govindan, R.: Resource/Accuracy Tradeoffs in Software-Defined Measurement
13. Chowdhury, S.R., Bari, M.F., Ahmed, R., Boutaba, R.: PayLess: a low cost network monitoring framework for software defined networks. In: 2014 IEEE Network Operations and Management Symposium, pp. 1–9 (2014)
14. Pakzad, F., Portmann, M., Tan, W.L., Indulska, J.: Efficient topology discovery in OpenFlow-based software defined networks. Comput. Commun. **77**, 52–61 (2016)
15. Su, M., Bergesio, L., Woesner, H., Rothe, T., Kpsel, A., Colle, D., Puype, B., Simeonidou, D., Nejabati, R., Channegowda, M., Kind, M., Dietz, T., Autenrieth, A., Kotronis, V., Salvadori, E., Salsano, S., Krner, M., Sharma, S.: Design and implementation of the OFELIA FP7 facility: the European OpenFlow testbed. Comput. Netw. **61**, 132–150 (2014)

16. Katiyar, R., Pawar, P., Gupta, A., Kataoka, K.: Auto-configuration of SDN switches in SDN/non-SDN hybrid network. In: Proceedings of the Asian Internet Engineering Conference, pp. 48–53 (2015)
17. Tootoonchian, A., Ghobadi, M., Ganjali, Y.: OpenTM: traffic matrix estimator for OpenFlow networks. In: Krishnamurthy, A., Plattner, B. (eds.) PAM 2010. LNCS, vol. 6032, pp. 201–210. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12334-4_21

Real Time Monitoring of Packet Loss in Software Defined Networks

Yash Sinha[1(✉)], Shikhar Vashishth[2], and K. Haribabu[1]

[1] Department of Computer Science and Information Systems, BITS, Pilani,
Pilani Campus, Pilani, India
h2016077@pilani.bits-pilani.ac.in

[2] Department of Computer Science and Automation, Indian Institute of Science,
Bangalore, India

Abstract. In order to meet QoS demands from customers, currently, ISPs over-provision capacity. Networks need to continuously monitor performance metrics, such as bandwidth, packet loss etc., in order to quickly adapt forwarding rules in response to changes in the workload. The packet loss metric is also required by network administrators and ISPs to identify clusters in network that are vulnerable to congestion. However, the existing solutions either require special instrumentation of the network or impose significant measurement overhead.

Software-Defined Networking (SDN), an emerging paradigm in networking advocates separation of the data plane and the control plane, separating the network's control logic from the underlying routers and switches, leaving a logically centralized software program to control the behavior of the entire network, and introducing network programmability. Further, OpenFlow allows to implement fine-grained Traffic Engineering (TE) and provides flexibility to determine and enforce end-to-end QoS parameters.

In this paper, we present an approach for monitoring and measuring online per-flow as well as per-port packet loss statistics in SDN. The controller polls all the switches of the network periodically for port and flow statistics via OpenFlow 1.3 multipart messages. The OpenFlow compliant switches send cumulative statistics of sent and received packets to the controller that includes raw packets (control, non-user generated packets responsible for network management); which, although not being part of the end-to-end data traffic, get counted and act as noise in the statistics. The proposed method takes into account the effect of raw packets and thus, hamper the accuracy of methods.

Other implementations propose approaches for per-flow packet loss only. We also take into account the effect of raw packets (control, non-user generated packets) which makes our packet loss estimation more accurate than other implementations. We also present a study of extrapolation techniques for predicting packet loss within poll interval.

Keywords: Software Defined Networks · Packet loss · OpenFlow
Quality of service (QoS) · Traffic Engineering (TE)

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 154–164, 2018.
https://doi.org/10.1007/978-3-319-73423-1_14

1 Introduction

Software-Defined Networking (SDN) along with OpenFlow [1] has inspired both academia and industry to test new ideas in fields of customized architecture, different algorithms, novel protocols, especially network status monitoring making enhanced Quality of service (QoS) possible.

Packet loss is one of the significant performance metrics used to determine QoS and in network diagnostics. Packet loss due to congestion is a fundamental issue in modern day networks for both network researchers and network operators.

There are various approaches to monitoring the network. The two common approaches are the passive and active approaches. The passive approaches do not increase the traffic on the network for the measurements, but rely on installation of devices that watch the traffic as it passes by. This may require large investments. Flowsense [2] proposes a passive push based monitoring method using control messages sent by switches to the controller to estimate network metrics.

The active approach injects test packets into the network or sends packets to servers and applications to monitor the network. The benefit is that it can be run from virtually anywhere in the network and it gives an end to end perspective of the network behaviour; although it introduces additional network load which affects the network and therefore influences the accuracy of the measurements.

In this paper, we present an approach for monitoring and measuring online per-flow and per-port packet loss and its comparison with some of the novel methods for measuring packet loss proposed for SDN and OpenFlow [1] networks. Other implementations propose approaches for per-flow packet loss only and not for per-port packet loss. This can be utilized by network administrators and ISPs to identify clusters in large network that are vulnerable to congestion.

We compare the accuracy for TCP & UDP packets and for networks having single and multiple flows. We discuss why polling switches for flow and port statistics in linear way is better than other ways such as round robin, last switch, per-flow etc. We also explore how accurate extrapolation techniques are, for predicting packet loss within poll interval.

Rest of the paper is divided into Related Work, Network Model and Architecture, Proposed Approach, Experimental Setups, Results, Future Work and Conclusion.

2 Background and Related Work

Flowsense [2] uses control messages sent by switches to the controller to estimate network metrics. But its estimation is far from the actual value and it works well only when there is a large number of small duration flows. It cannot capture traffic bursts if they do not coincide with another flow's expiry.

OpenNetMon [3] is a OpenFlow controller module that monitors per-flow QoS metrics such as latency, throughput, delay and packet loss by polling flow source and destination switches. Polling is done for every path between every node pairs to be monitored and it is adaptively changed based on new flows' arrivals. But the controller does not know to find new paths based on real-time data. Since the monitoring targets are only limited to edge switches, it is difficult to obtain detailed flow statistics on other switches.

OpenTM [4] presents a heuristics-based monitoring method wherein it uses the routing information learned from the OpenFlow controller to intelligently choose the switches from which to obtain flow statistics, thus reducing the load on switching elements and improving the monitoring accuracy. However, constant polling involves considerable overhead.

OpenSketch [5], proposes a new SDN based monitoring architecture and a new OSDN protocol. This, however, requires an upgrade or replacement of all network nodes. Furthermore, standardization of a new protocol has been a long and tedious task.

PayLess [6] is a query-based monitoring framework for SDN which performs highly accurate information gathering in real-time without incurring significant network overhead. To achieve this goal, instead of letting controller continuously polling switches, an adaptive scheduling algorithm for polling is proposed to achieve the same level of accuracy as continuous polling with much less communication overhead.

3 Network Model and Architecture

First of all we explain the basic network model and the architecture used. The network is emulated using MiniNet Network Emulator [7] which runs a collection of virtual end-hosts, switches, routers, and links on a single Linux kernel. Furthermore, it allows us to create many custom topologies and emulate some link parameters like a real Ethernet interface, e.g., link speed, packet loss, and delay.

We use SDN enabled (i.e. OpenFlow [1] compliant) OpenVSwitch [8] switches and Ryu [9] controller to handle their control plane. We emulate the network topology using L2 learning switches wherein the switches examine each packet and learn the source-port mapping. Thereafter, the source MAC address gets associated with the port. If the destination of the packet is already associated with some port, the packet is sent to the given port, else it is be flooded on all ports of the switch.

Our code sits at the controller and computes the packet loss for the network. We use the controller to request the switches for port and flow statistics via the OpenFlow protocol 1.3. This is done periodically every k seconds by issuing PortStatsRequest and FlowStatsRequest requests. So network overhead involved is 2n messages, where n is the number of switches in the network for every k seconds.

We use tc tool of queueing discipline (also called qdisc) part of Ubuntu kernel's traffic control module to measure actual packet loss from switches emulated using MiniNet. MiniNet also uses qdisc internally to emulate link's properties such as bandwidth, delay etc. and manipulate other traffic settings. Therefore, qdisc can provide true packet loss statistics that can be used for comparison.

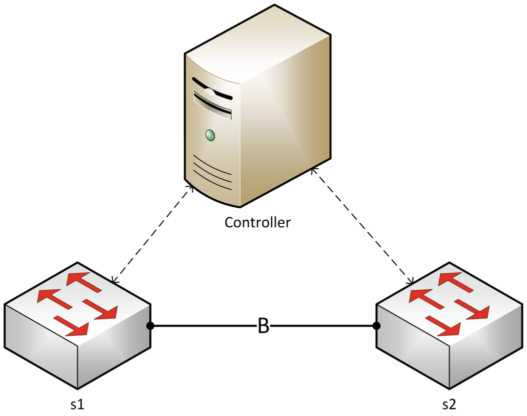


Fig. 1. Base model

We use iperf2 [10] which is a traffic generation tool that allows the user to experiment with different TCP and UDP parameters to see how they affect network performance. We also use Distributed Internet Traffic generator D-ITG [11] to generate bursty traffic to test extrapolation techniques.

We define the base model of our system consisting of two switches s1 and s2, connected by a link of bandwidth B. There is a direct logical link from each switch to the controller Cn. We identify a flow by 4-tuple: in port, out port, source MAC address and destination MAC address (Fig. 1).

Our model network consists of multiple such switches connected together in a various topologies with the link distances equal for all. The end switches are connected to hosts. We assume the switches and the hosts to be homogenous. One of the hosts acts as an iperf server and the other as iperf client, while running the experiment. The bandwidths for the links are variable which is why packet loss is expected to happen at some links in the network.

4 Proposed Approach

In this section we present our approach to estimate online packet loss, which is divided into following stages: 1. estimation of raw packets, 2. gathering port and flow statistics, 3. real time estimation, 4. extrapolation within poll interval.

4.1 Estimation of Raw Packets

We define raw packets as the packets exchanged in the network which are not a part of the active traffic but are used for network management such as for network discovery, multicast listeners discovery, requesting networking parameters from DHCP servers etc. These packets are necessary for many network protocols such as MDNS, NDP, MLD, DHCP etc.

The raw packets are absorbed in the intermediate switches and they are not meant for hosts. The SDN enabled switches send cumulative statistics of sent and received packets to the controller that includes raw packets. Therefore for calculating packet loss accurately, it is necessary to separate data packets from raw packets.

Using Emulation. We emulate the network with no active traffic between hosts for some time and the controller polls the switches queries the switches to accumulate the statistics about the packets exchanged. The network is called raw network. Later this data is used to calculate packet loss.

Using Mathematical Model. We emulated different network topologies for accumulating raw packet statistics and noticed a pattern in which packets are exchanged which was periodic and additive based on the topology of network.

We devised our own algorithm [12] for predicting raw packet flow in any network. The algorithm uses the spanning tree information about network topology and calculates the number of hosts and switches in the subnetwork for each interface of every switch in the network and based on that count, the raw packets are estimated.

4.2 Gathering Port and Flow Statistics

Port Statistics. Other implementations propose approaches for per-flow packet loss only and not for per-port packet loss. Even though OpenTM [4] advocates polling each path's switch randomly or in round robin as most efficient, and OpenNetMon polls each flow-path's last switch; we poll every switch of the network for port statistics.

The reasons are manifold. Firstly, collecting flow based statistics becomes complex and tedious in large networks with multiple flows. Further, non-edge switches generally have a large number of flows to maintain, making the query for flow statistics more expensive. When more flows exist, non-edge switches will be polled more frequently degrading efficiency. For a network with n switches there can be nC2 flows in the worst case. In contrast, in our approach there are at most n port statistics requests and size of query response and polling frequency is independent of number of flows in the network.

Therefore, our approach introduces relatively less additional network load which affects the network and influences the accuracy of the measurements.

Flow Statistics. For flow based statistics collection, the round robin switch selection becomes more complex in larger networks with multiple flows. Therefore, in our implementation we query the switches linearly that does not explode to n_2^C in the worst case.

4.3 Real Time Estimation

For calculating per port packet loss statistics at a given port, we subtract the RX packets (packets received on the port) at the destination port of the destination switch from TX packets (transmitted out of the port) packets at the source port of the source switch. Further, from this value we subtract the raw RX packets at the destination port of the destination switch from raw TX packets at the source port of the source switch.

$$packetLoss(perport) = [TX_a - RX_a] - [X_r - RX_r];$$

where, a: active network, r: raw network.

For calculating per flow statistics, we subtract the packet count at the destination port of the destination switch from the source port of the source switch

$$packetLoss(perflow) = Count_s - Count_d;$$

here, s: source, d: destination.

The raw packet correction cannot be applied to flows because no flows are generated in raw network and hence no data is collected.

4.4 Extrapolation within Poll Interval

Once the controller gets data, it uses two methodologies to predict data: extrapolation based on rate of packet loss in last k seconds; and extrapolation based on rate of packet loss and change in rate of packet loss in last k seconds.

5 Experimental Setups

5.1 Experimental Setup 1: Single Flow

For initial comparison purposes with OpenNetMon, a linear network topology consisting of three connected switches was used, with ends connected with two host systems (Fig. 2).

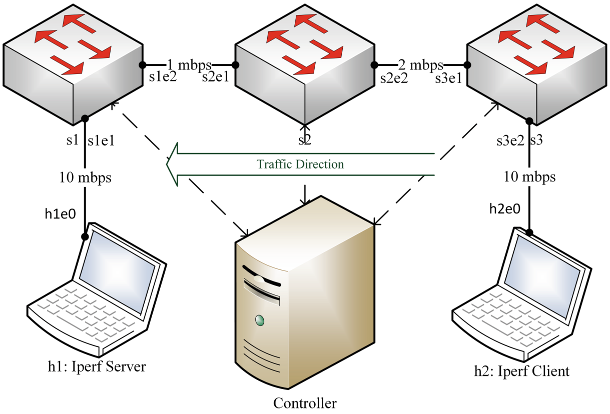


Fig. 2. Experimental setup for single flow

Here, $h\alpha e\beta$ denotes port of host and $s\alpha e\beta$ denotes port of switch, having links with bandwidths 10, 1, 2 and 10 mbps respectively. Iperf was used to send udp/tcp data from $h2$ to $h1$ at 5 mbps, this setting was designed so that approximately 1 mbps loss would occur at link2 and some loss would occur at link 3 as well. The polling frequency (k) was kept as 5 s.

5.2 Experimental Setup 2: Multiple Flows

(See Fig. 3).

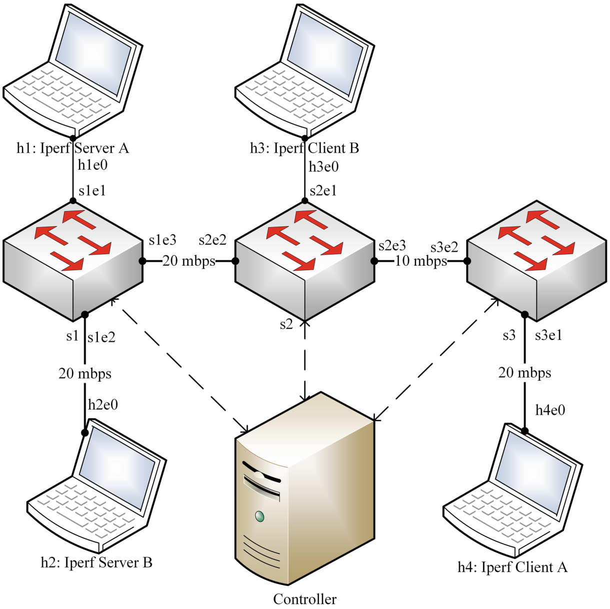


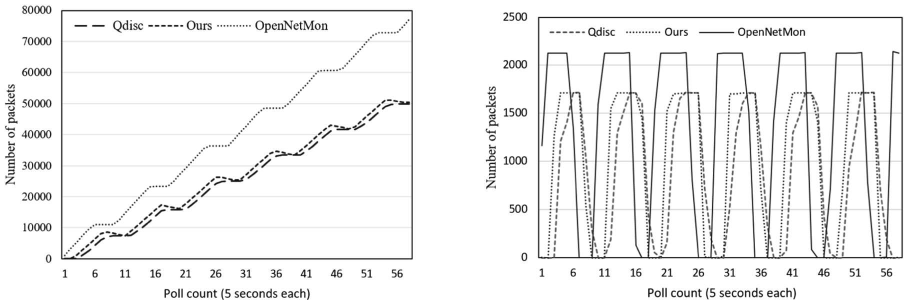
Fig. 3. Experimental setup for multiple flows

A linear network topology consisting of 3 connected switches is used. Switch $s1$ is connected to two host systems and one host each is connected at s2 and s3. Iperf UDP traffic of 6 mbps and 8 mbps is generated for h1–h4 and h2–h3 pair of hosts respectively.

6 Results

6.1 Comparison of Accuracy for Single Flow

We ran experiments on the aforementioned topology for networks transmitting TCP and UDP traffic. On the same network we also ran the OpenNetMon module to compare the accuracy reported. Since there is only one flow, we assume per port loss to be equal to per flow loss, hence comparison is possible. We recorded the actual packet loss data by periodically calling qdisc's tc tool.



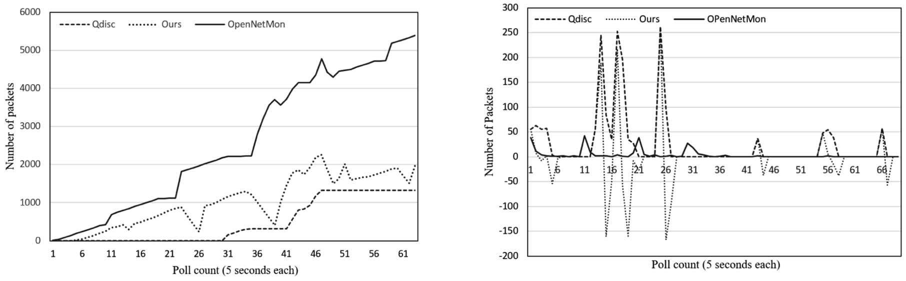
(a) Cumulative Packet Loss

(b) Real time Packet Loss

Fig. 4. UDP traffic

UDP Traffic. UDP traffic of bandwidth 5 mbps is generated using iperf from host h2 to host h1. As the graph shows, our proposed method closely matches with the packet loss reported by qdisc, whereas OpenNetMon reports lags behind as raw packets which are absorbed in the intermediate switches are reported as lost. We present both real time and cumulative statistics here (Fig. 4).

TCP Traffic. TCP traffic is generated using iperf from host h2 to host h1. As the graph shows, our proposed method matches with the packet loss reported by qdisc, whereas OpenNetMon reports lags behind as raw packets which are absorbed in the intermediate switches are reported as lost. We present both realtime and cumulative statistics here (Fig. 5).



(a) Cumulative Packet Loss

(b) Real time Packet Loss

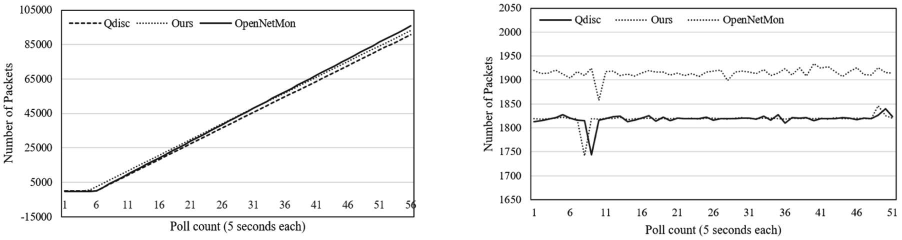
Fig. 5. TCP traffic

As compared to OpenNetMon, this method increases the accuracy of reported real-time packet loss (error is reduced to 2–3% from 9–10%) for aforementioned network.

6.2 Comparison of Accuracy for Multiple Flows

We present the comparison of total packet loss of two flows from h4 to h1 and h3 to h2 calculated by OpenNetMon, qdisc and our approach. As explained in Sect. 4.3, raw packet correction cannot be applied to flows; hence our approach and OpenNetMon's approach gives us identical results. We have skipped those plots.

Since qdisc reports statistics port wise only, for comparison we compare the total statistics of flow 1 and flow 2 combined. The raw packet correction has been applied to total packet loss. Clearly, we can see that OpenNetMon fails to account for the raw packets and thus reports packet loss erroneously (Fig. 6).



(a) Cumulative Packet Loss

(b) Real time Packet Loss

Fig. 6. Multiple flows

6.3 Extrapolation Within Poll Interval

(See Fig. 7).

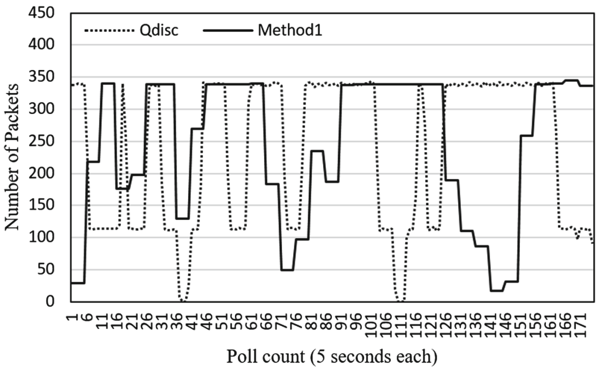


Fig. 7. Packet loss rate based extrapolation

Extrapolation Based on Rate of Packet Loss. For the poll interval we extrapolate the packet loss rate observed during the last 5 s i.e., the packet loss rate is assumed to sustain for the next five seconds. So accuracy largely depends on how dynamically the network traffic is changing and how frequently the controller is polling the switches for the statistics. For relatively stable network, extrapolation gives packet loss with good accuracy but if the network changes drastically within the poll interval, the controller and hence the extrapolation fails to capture it (Fig. 8).

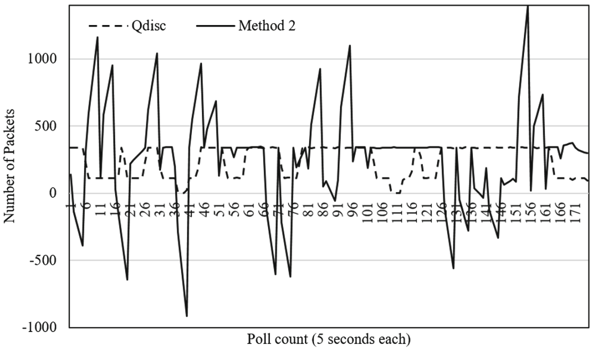


Fig. 8. Packet loss rate based extrapolation and change in packet loss rate

Packet Loss Rate Based Extrapolation and Change in Packet Loss Rate. Here, we extrapolate based on the packet loss rate and also the change in rate of packet loss. The extrapolation is highly sensitive to small changes in rate of change of packet loss, so this proves that the prediction model is not dependent on higher derivatives of time.

7 Future Work

As proposed by PayLess [6], we can reduce the network overhead by using an adaptive scheduling algorithm for polling. For reporting packet loss during the poll interval, we can use history based prediction. More features based on network topology can be added.

8 Conclusion

Network Admins and Internet service providers can utilize per port loss in SDN for identifying clusters in large networks which are congestion vulnerable. By taking into account the packet loss information the proposed method drastically increases the preciseness of obtained real-time packet loss statistics. It is also scalable to large dense networks, as it avoids polling edge switches on per flow basis.

References

1. McKeown, N., et al.: OpenFlow: enabling innovation in campus networks. ACM SIGCOMM Comput. Commun. Rev. **38**(2), 69–74 (2008)
2. Yu, C., Lumezanu, C., Zhang, Y., Singh, V., Jiang, G., Madhyastha, H.V.: FlowSense: monitoring network utilization with zero measurement cost. In: Roughan, M., Chang, R. (eds.) PAM 2013. LNCS, vol. 7799, pp. 31–41. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36516-4_4
3. Van Adrichem, N.L.M., Doerr, C., Kuipers, F.: Opennetmon: network monitoring in openflow software-defined networks. In: 2014 IEEE Network Operations and Management Symposium (NOMS), pp. 1–8, 5 May 2014
4. Tootoonchian, A., Ghobadi, M., Ganjali, Y.: OpenTM: traffic matrix estimator for OpenFlow networks. In: Krishnamurthy, A., Plattner, B. (eds.) PAM 2010. LNCS, vol. 6032, pp. 201–210. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12334-4_21
5. Wellem, T., Lai, Y.-K., Chung, W.-Y.: A software defined sketch system for traffic monitoring. In: Proceedings of the Eleventh ACM/IEEE Symposium on Architectures for Networking and Communications Systems, pp. 197–198, 7 May 2015
6. Chowdhury, S.R., et al.: Payless: a low cost network monitoring framework for software defined networks. In: 2014 IEEE Network Operations and Management Symposium (NOMS), pp. 1–9, 5 May 2014
7. Lantz, B., Heller, B., McKeown, N.: A network in a laptop: rapid prototyping for software-defined networks. In: Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks, p. 19, 20 October 2010
8. Open vSwitch (2015). http://openvswitch.org/download/. Accessed 15 Dec 2015
9. Ryu SDN Framework (2015). http://osrg.github.io/ryu. Accessed 15 Dec 2015
10. Tirumala, A., et al.: iPerf: the TCP/UDP bandwidth measurement tool. https://iperf.fr/
11. Avallone, S., et al.: D-ITG distributed internet traffic generator. In: Proceedings First International Conference on the Quantitative Evaluation of Systems, QEST 2004. IEEE (2004)
12. Sinha, Y., Vashishth, S., Haribabu, K.: Meticulous measurement of control packets in SDN. In: Proceedings of the Symposium on SDN Research. ACM (2017)

Active Home Agent Load Balancing for Next Generation IP Mobility based Distributed Networks

Anshu Khatri(✉) and Senthilkumar Mathi

Department of Computer Science and Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
cb.en.p2cse16004@cb.students.amrita.edu,
m_senthil@cb.amrita.edu

Abstract. Mobile IPv6 is the widely acknowledged technology that supports mobility in networks. A single home agent in the network suffers from the issue of single point of failure and consequently focuses on the deployment of multiple home agents in the network. The load sharing mechanism in most of the exiting methods is passive and centralized in approach. Moreover, the failure detection and recovery mechanism uses the concept of periodic messaging updates which results in signaling overhead. Hence, a new method of active load sharing that is distributed in nature is proposed in this paper. The proposed method contributes a load balancing mechanism at the registration time itself using the concept of preferred home agent. The paper investigates the existing methods and presents the comparative analysis with the proposed method. The advantages of our proposed load sharing are active and distributed approach, less signaling overhead and better throughput.

Keywords: IPv6 mobility · Distributed load balancing · Failure recovery
Optimized routing

1 Introduction

The fundamental communication protocol is the Internet Protocol (IP) that delivers the datagram across the network using the concept of IP address in the packet header. A unique address is assigned to every device that is connected to the internet and it is used in the identification of the devices while sending or receiving data packets [1, 2]. Mobile IPv4 has many limitations such as: optimized routing issue, Home Agent (HA) single point of failure, multiple HAs support and IP security. Mobile IPv6 (MIPv6) provides solution to these problems and came as an acknowledged technology [3]. A specific IP home addressing is associated to the home network to which Mobile Node (MN) connects [4]. A temporary care-of-address is attained by the MN when it moves away from the home network to the foreign network as shown in the Fig. 1.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 165–176, 2018.
https://doi.org/10.1007/978-3-319-73423-1_15

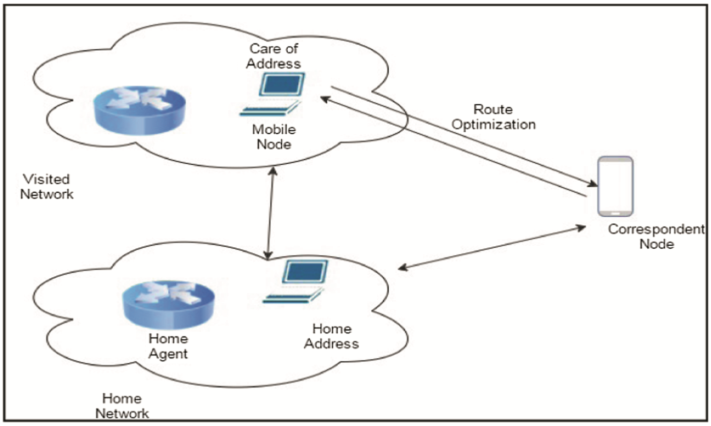


Fig. 1. Mobile IPv6 operation

A registration request is sent by the MN to its home address agent that updates the HA regarding the MN's current location. HA processes the request and provides acknowledgement to the MN. One HA is managing MN registration, cache maintenance and tunnelling of data packets to MN's current location result into the improper load on the HA and produces bad performance report [5].

There are two common approaches: the centralized approach and the distributed approach for HA load balancing. A single HA act as a key entity in the centralized approach. It collects the load sharing information from all of the neighbouring HAs to take decision for the fairly distribution of load among the HAs. The HA act as an administrator for the load sharing but suffer from the problem of single point of failure. This issue is resolved in the distributed approach, where each and every node shares the load information with one another and updates the detail accordingly [6]. The load balancing mechanism mostly uses the concept of "heart beat messages" in order to keep the HAs updated. These messages are basically the router advertisements that every router multicasts at a constant rate. With the reduction in the router advertisement interval, signaling and synchronization overhead occurs [7–11].

This paper focuses on the need of an efficient active load sharing mechanism. The proposed model uses the HA list table and MN list table to keep track of the load measure of each and every node in the network. The information at every node is updated using *Information_Updated* message. If the HA is not in the overloaded state or has at least one HA as the *pref_HA*, then it can address the registration request by the new MN. Failure detection request helps to determine any HA failure in the network and subsequently perform failure recovery. From the comparative results, it is identified that the proposed work outperforms the other existing mechanisms.

The rest of the paper is organized as follows. Section 2 discusses the previous works related to this domain. Section 3 gives the detailed description of the proposed method. The comparative analysis is presented in Sect. 4. Finally Sect. 5 concludes the paper.

2 Related Work

The distributed approach is used in Dynamic Home Agent Address Discovery (DHAAD) protocol follows the concept of distributed approach. This is used by most of the load balancing mechanism for the HA registration in which each HA maintains the list of all the HAs in the network. MN sends an address discovery to the anycast address of a HA and waits for the response. In case of no acknowledgement from the HA, MN resends the registration request. Inter Home Agent (HAHA) protocol also comes under the distributed approach and uses the concept of DHAAD. In this method, whenever HA is in failed state or overloaded state, it sends the HA switch message to the affected MNs. The MN disconnects its current binding and sends the registration request to the preferred HA mentioned in the switch message. If the preferred HA is not specified, it uses DHAAD request message for the HA registration process [12–14].

Home Agent Handoff (HAH) scheme maintains a list of HAs and uses the main features of DHAAD and HAHA methods. Each and every HA shares the information with one another which helps in taking decision of HA re-assignment during HA failure or overloading state. HA sends the HAH switch message to the affected MNs. After receiving the HAH message, the MN follows the same HA registration steps as taken in HAHA method.

The hybrid load balance mechanism comprises of multiple MIPv6 based HAs and MNs. A traffic load table is maintained by each HA which is sorted in descending order of the traffic load field. A timer is associated to each entry of the binding update table and HA reassignment occurs if the timer goes out.

In, Virtual Home Agent Reliability Protocol (VHARP) architecture, one home link contains multiple HAs having different link local IP addresses and one global IP address taken as global HA address [15]. All the communication between the correspondent node and any HA in a home link takes place through this global HA address, representing a single virtual view. There are three states for a HA in this method: active HA, backup HA and inactive HA. The failure detection and recovery mechanism uses the periodic "*Heart Beat Messages*" and does not suffer from any service latency because it is transparent to the MN. Virtual Home Link (VHOL) follows the same architecture and working as VHARP described in [16]. The failure detection mechanism follows the message exchange technique and is transparent to MNs. This method utilizes all the secondary links in addition to the primary link and results in better resource utilization [17].

The addition of more hardware resources and improvement in web server services is suggested in the web services load balancing techniques [18–20]. Distributed and loosely coupled web servers can be deployed to get better solution. This approach is not cost effective. Multiple HA deployment scheme (MHADS) presents dynamic load balancing mechanism and improves the overall performance. The edge router in the home link acts as a Balancer as well as a monitor (BM). It selects the best HA during the registration process and provides active load sharing. Each HA sends update to the BM in regular interval of time. The absence of this update signals the HA failure to the BM, therefore it sends failure detection request to confirm the failure. The ring backup chain is reconstructed for the failure recovery mechanism. The Virtual Private Network based Home Agent Reliability Protocol (VHAHA) contains multiple HAs in a home link that can take any state out of these three states: active HA, backup HA and inactive HA. Global HA address is assigned to the Virtual Private Network (VPN) and each HA shares the status using "*heart beat messages*". When a packet reaches the global HA address, the least loaded HA that is nearest to the MN receives the packet. The "Home Agent Group" (HA Group) method has a main HA that manages all the mobility related tasks and a stand-by HA to take over the responsibility when the main HA fails. It uses the messaging concept in order to identify the HA failure [21–23] which is followed by the destruction of the tunnel with the failed HA and reconstruction of a new tunnel with another HA.

3 Proposed Model

3.1 Notations

Table 1 show the notations used in the proposed scheme.

Table 1. Notations used in the proposed scheme

Symbols	Descriptions
Thres_val	Maximum number of MNs that a HA can provide services
Load_val	Number of the MNs attached to the HA
Pref_HA	Preferred HA for the MN for the registration

3.2 Detailed Descriptions of the Proposed Scheme

In the proposed scheme, each HA maintains a table that keeps tracks of the load measures of the rest of the HAs in the network as shown in Table 2. The HA is assumed to be in overloaded state when the *load_val* of the HA equals the *thres_val*. If it is overloaded then it cannot serve the new registration request and cannot serve as the *pref_HA* as well.

Table 2. HA list table

HA	Load	Thres_val
HA_i	$\mathrm{Load\_val}_i$	$\mathrm{Thres\_val}_i$

Each MN records the current CoA in addition to the *pref_HA* address in MN data table as represented in Table 3. *pref_HA* helps in the reduction of the re-registration of the MN during the HA failure. Consequently, reduces the failure recovery time. The proposed method also provides optimized routing because HA registration request by a MN is always acknowledged by the nearest HA. If the nearest HA is in overloaded state, it examines it's HA list table in order to determine the next preferred HA for the registration and update the same to the MN.

Table 3. MN data table

CoA	HA	Pref_HA

3.2.1 HA Registration and Active Load Balancing

Step 1: When MN is in the Home network, it works as it is in a fixed network.

Step 2: MN enters foreign network and broadcasts HA registration using DHAAD

Step 2.1: Nearest HA sends the acknowledgement along with the *pref_HA* for the future registration process and updates the HA list table.

Step 2.2: If the nearest HA is overloaded

Step 2.2.1: It sends the *pref_HA* to the MN

Step 2.2.2: MN sends the registration request to the *pref_HA*

Step 2.2.3: *pref_HA* sends the acknowledgement to the MN plus the new *pref_HA* after examining its HA list table

Step 3: HA broadcasts the new registration update to the rest of the HAs in the network

Step 3.1: HAs update their HA list table and send *Information_Updated* acknowledgement to the HA (Fig. 2).

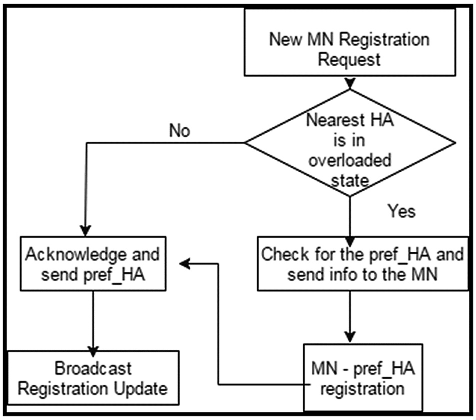

New MN Registration Request
Nearest HA is in overloaded state
No
Yes
Acknowledge and send pref_HA
Check for the pref_HA and send info to the MN
Broadcast Registration Update
MN - pref_HA registration

Fig. 2. Flowchart for active load sharing mechanism

3.2.2 HA Failure Detection

Step 1: HA1 broadcasts the MN registration update to the rest of the HAs in the network

Step 2: HAs reply back with the *Information_Updated* acknowledgement to the HA1

Step 3: If the HA1 receives *no-reply* from any HA in the network named HA2

Step 3.1: HA1 sends failure detection request to the *no-reply* HA2

Step 3.2: Again the *no-reply* from the HA2 is taken as the failure of the HA2 (Fig. 3)

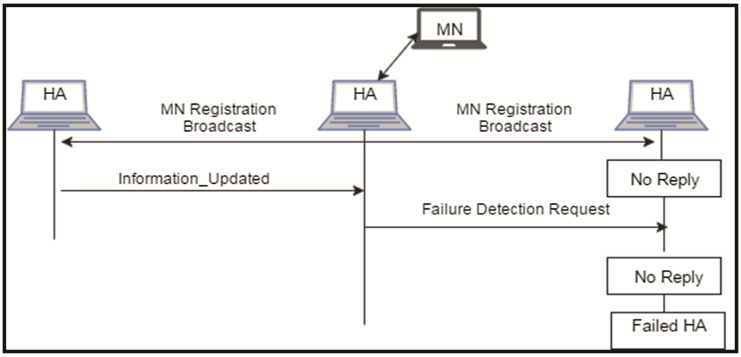

MN
HA
HA
HA
MN Registration Broadcast
MN Registration Broadcast
Information_Updated
No Reply
Failure Detection Request
No Reply
Failed HA

Fig. 3. Failure detection mechanism

3.2.3 HA Failure Recovery

Step 1: The failure of the HA2 is detected using the 3.2.2 Failure Detection

Step 2: HA1 deletes the entry of the HA2 from its HA list table

- **Step 2.1:** HA1 broadcasts this information to the rest of the HAs in the network
- **Step 2.2:** HAs update their HA list table accordingly and reply with the *Information_Updated* acknowledgement to the HA1

Step 3: Affected MNs of the failed HA2 sends binding update to the *pref_HA*

- **Step 3.1:** If *pref_HA* is not overloaded
  - **Step 3.1.1:** It sends the binding acknowledgement to the MN along with the *pref_HA*
  - **Step 3.1.2:** *pref_HA* updates its HA list table and broadcasts it to the other HAs
- **Step 3.2:** If *pref_HA* is overloaded
  - **Step 3.2.1:** It examines its HA list table and return the *pref_HA* to the MN for the registration.

Step 4: MNs establishes the connection and correspondingly update their MN data table

4 Comparison and Analysis

In this section, the performance of the proposed method is compared with the existing methods for load sharing and failure detection mechanism. The most predominantly used mechanism in load sharing is passive in nature, it takes place only after the registration of the MN to the HA. The edge router BM provides active load sharing in MHADS method by selecting the best HA at the time of registration process itself. The load sharing overhead is high in redundant HA method because it is not transparent to the MN and adds to the OTA signaling.

The complicated network architecture in VHARP and VHOL adds to the load sharing overhead. The hybrid method also suffers from the load sharing overhead due to the maintenance and advertisement of the traffic load table. As shown in Table 4, the MHADS has low load sharing overhead as compared to the above discussed method. It is centralised as it uses the edge router that acts as the balancer for the entire network. Every HA sends update messages in regular interval to the BM. The proposed method faces less overhead than MHADS because it is distributed in nature and does not put any overhead on a particular router or a HA.

Table 4. Comparison of load balancing mechanism of existing methods

Metrics	Load sharing mechanism (active/passive)	Load sharing signalling
Redundant HA	Passive	7
Hybrid	Passive	7
MHADS	Active but centralised	5
VHOL	Passive	9
VHARP	Passive	6
Proposed method	Active and distributed	3 or 5

4.1 Comparison of Signaling Overhead

Figure 4 shows that VHARP causes remarkable message exchange. It faces more signaling overhead in comparison to the VHOL because it has less time interval to advertise router messages. Redundant HA method also has significant signaling overhead due to the OTA signaling in addition to the periodic *"Heart Beat Messages"*. MHADS and HA Group also rely on the concept of *"Heart Beat Messages"*. Although VHAHA provides active load sharing, it has higher number of messages exchanged as

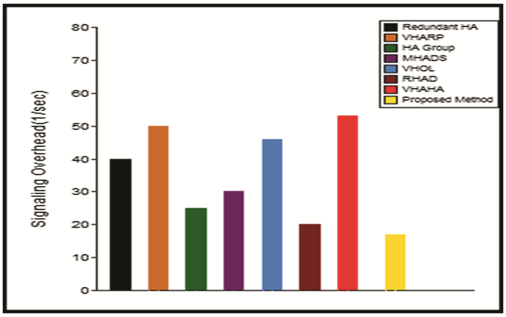


Fig. 4. Comparison of signaling overhead

compared to other methods. The proposed method experiences the least signaling overhead because it does not uses the concept of periodic *"Heart Beat Messages*". The message broadcasting takes place only when an event occurs.

4.2 Failure Recovery Time vs Number of MNs

HA group method takes notable time as compared to the rest of the methods due to the process of tunnel destruction and reconstruction. Figure 5 shows that VHAHA, VHARP and VHOL takes comparable amount of time in failure recovery. The exchange of service takeover request and answer messages followed by the reconstruction of the ring backup chain adds to the failure recovery time in MHADS. Every MN maintains a MN data table in the proposed method. This table keeps track of the *pref_HA* which helps in the reduction of the failure recovery time because each MN knows the next preferred HA.

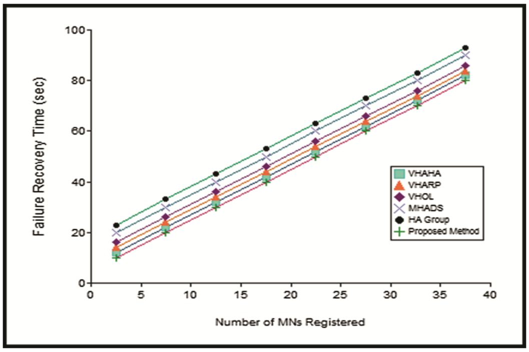


Fig. 5. Failure recovery time vs Number of MNs

4.3 Registration Time vs Number of MNs

It can be depicted from the Fig. 6 that as the number of the MN increases, the time taken for the HA registration for the MNs also increases. VHARP and VHOL have comparable time for the HA registration process. Although VHAHA follows the architecture of the VHARP, it has better registration time than VHARP method. In this, few HAs are taken to build VPN which is addressed using one global HA. In MHADS, the edge router receives the registration request and selects the least loaded HA during the registration process itself. The proposed method uses DHAAD mechanism, which provides less registration time for the MN-HA registration as compared to the other discussed methods.

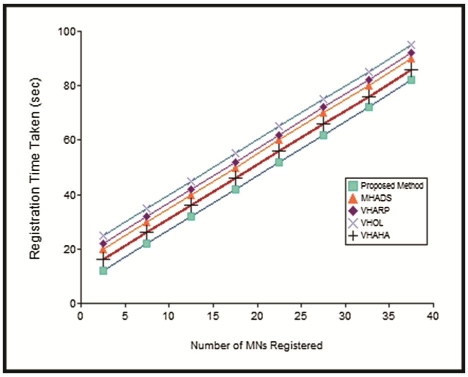


Fig. 6. Registration time vs Number of MNs

4.4 Load Sharing Signaling Overhead

The exchange of periodic *"Heart Beat Messages"* in VHARP and VHOL methods adds to the signaling overhead. Although VHOL solves the issue of the entire home link failure in case of VHARP, it faces more overhead due to the redundancy in the home links architecture. Figure 7 shows that hybrid model has less load sharing signaling overhead in comparison to the previously discussed methods. It uses the concept of traffic load table and each entry in the table has a timer coupled to it. When a HA overloads, re-assignment process get started in which HA does not wait for the ICMP request message and sends the ICMP reply message. Each HA sends update messages to the BM in MHADS method. BM acts as a balancer in load sharing mechanism and selects the least loaded HA for the registration process. The proposed method has the least load sharing signaling overhead because the broadcasting of the messages takes place only if any update or event occurs. It does not suffer from the periodic signaling overhead of *"Heart Beat Messages"* as present in the other methods.

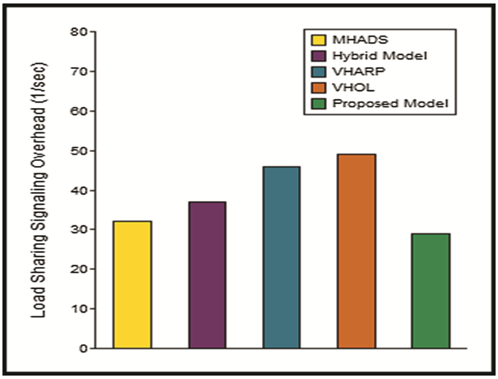


Fig. 7. Comparison of load sharing signaling overhead

4.5 Impact of Number of Registrations of MNs on the Throughput of the HAs

As shown in Fig. 8, the throughput increases with the increase in the number of registered MNs but it starts decreasing, if number of MNs becomes more. HA Group method faces the great fall in the throughput performance because of the tunnelling mechanism. VHOL utilizes all the primary home links as well as the secondary home links and has better throughput than VHARP method. The redundant HA has the least throughput in comparison to the other methods due to the OTA signaling overhead. The MHADS performs active load sharing and provide better throughput. Its performance is comparatively low than hybrid model due to the ring backup chain process in failed HA recovery mechanism. The proposed method has no overhead of ring backup chain process or periodic *"Heart Beat Messages"*. Moreover, the nearest HA sends the acknowledgment to the MN if it is not overloaded, else it updates the MN for the next preferred HA.

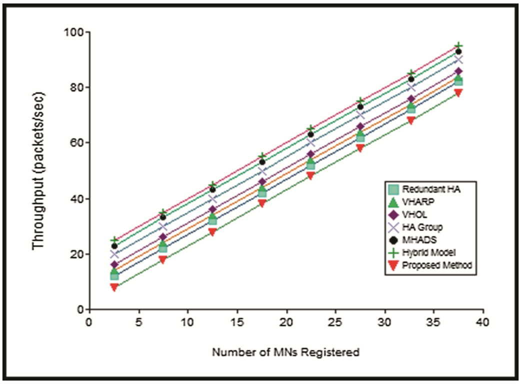


Fig. 8. Throughput Vs Number of MNs

5 Conclusion

In this paper, a new method is proposed for the distributed active load sharing mechanism in which load balancing is taken care during the HA registration process itself. It also describes how failure detection and recovery can be performed effectively under the proposed scheme. The centralized approach is predominantly used in most of the existing methods and faces the issue of single point of failure. The proposed method overcomes this limitation by incorporating distributed approach. Moreover, most of the methods use passive load sharing and concept of *"Heart Beat Messages"* for failure detection and recovery mechanisms. This results in signaling overhead, longer time for failure recovery and poor throughput performance. Although MHADS uses the concept of active load sharing, it is centralized in nature and the edge router acts as a sole point of failure. The comparative analysis of the proposed scheme with the existing methods show that it has better throughput, takes lesser time in failure recovery and has less signaling overhead. Future work can be extended in the field of proactive failure detection and recovery while maintaining less signaling overhead.

References

1. Perkins, C., Johnson, D., Arkko, J.: Mobility Support in IPv6. No. RFC 6275 (2011)
2. Johnson, D., Perkins, C., Arkko, J.: Mobility Support in IPv6. No. RFC 3775 (2004)
3. Terli, V.K.K., Chaganti, S.P., Alla, N.B., Sarab, S., El Taeib, T.: Software implementation of IPv4 to IPv6 migration. In: 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT). IEEE (2016)
4. Yen, Y.-S., Hsu, C.-C., Chao, H.-C.: Distributed balancing with application-layer anycast for home agent discovery on the mobile IPv6. In: 2005 International Conference on Wireless Networks, Communications and Mobile Computing, vol. 2. IEEE (2005)
5. Zhang, H., et al.: A multiple home agent deployment scheme to enhance service availability for MIPv6. In: 11th IEEE Singapore International Conference on Communication Systems, 2008, ICCS 2008. IEEE (2008)
6. Vasilache, A., Li, J., Kameda, H.: Load balancing policies for multiple home agents mobile IP networks. In: 2001. Proceedings of the 2nd International Conference on Web Information Systems Engineering, vol. 2. IEEE (2001)
7. Khan, S., et al.: Home agent load balancing in mobile IPv6 with efficient home agent failure detection and recovery. In: 2006 International Conference on Emerging Technologies. IEEE (2006)
8. Khatri, A., Senthilkumar, M.: Investigation of home agent load balancing, failure detection and recovery in IPv6 network-based mobility. Int. J. Adv. Sci. Eng. Inf. Technol. **7**(2), 632–641 (2017)
9. Kong, R., Feng, J., Zhou, H.: The combination of multiple care-of addresses registration and reverse routing header in nested network mobility. In: 2011 International Conference on Internet Technology and Applications (iTAP). IEEE (2011)
10. Aman, A.H.M., Hashim, A.-H.A., Abdullah, A., Ramli, H.A.M., Islam, S.: Parametric comparison of multicast support for network mobility management: a qualitative approach. Int. J. Multimedia Ubiquit. Eng. **11**(9), 203–210 (2016)
11. Erunika, O., Kaneko, K., Teraoka, F.: Impact of multiple home agents placement in mobile IPv6 environment. IEICE Trans. Commun. **97**(5), 967–980 (2014)
12. Deng, H., et al.: A hybrid load balance mechanism for distributed home agents in mobile IPv6. In: 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003, PIMRC 2003, vol. 3. IEEE (2003)
13. Senthil Kumar, M., Valarmathi, M.L., Ramprasath, G.: A secure and efficient registration for IP mobility. In: Proceedings of the 1st International Conference on Security of Internet of Things. ACM (2012)
14. Goswami, S., Das, C.B.: A survey on various route optimization techniques in network mobility. J. Uncertain Syst. **10**(2), 91–107 (2016)
15. Faizan, J., El-Rewini, H., Khalil, M.: VHARP: virtual home agent reliability protocol for mobile IPv6 based networks. In: 2005 International Conference on Wireless Networks, Communications and Mobile Computing, vol. 2. IEEE (2005)
16. Faizan, J., El-Rewini, H., Khalil, M.: Efficient dynamic load balancing for multiple home agents in mobile IPv6 based networks. In: 2005 Proceedings of the International Conference on Pervasive Services, ICPS 2005. IEEE (2005)
17. Faizan, J., El-Rewini, H., Khalil, M.: Introducing reliability and load balancing in home link of Mobile IPv6 based networks. In: 2006 ACS/IEEE International Conference on Pervasive Services. IEEE (2006)
18. Shyamala, C.K., Ashok, N., Narayanan, B.: Trust-based multi-path security scheme for Ad-hoc networks. Int. J. Control Theor. Appl. **8**(5), 1735–1742 (2015)

19. Cabellos-Aparicio, A., Pascual, J.D.: Load balancing in mobile IPv6's correspondent networks with mobility agents. In: 2007 IEEE International Conference on Communications. IEEE (2007)
20. Lin, J.-W., Yang, M.-F.: Fault-tolerant design for wide-area Mobile IPv6 networks. J. Syst. Softw. **82**(9), 1434–1446 (2009)
21. Rathi, S., Thanuskodi, K.: A secure and fault tolerant framework for Mobile IPv6 based networks. arXiv preprint arXiv:0909.4858 (2009)
22. Rathi, S., Thanushkodi, K.: Design and performance evaluation of an efficient home agent reliability protocol. Int. J. Recent Trends Eng. **2**(1), 2009
23. Sasikumar, R., Ananthanarayanan, V., Rajeswari, A.: An intelligent pico cell range expansion technique for heterogeneous wireless networks. Indian J. Sci. Technol. **9**(9) (2016)

Design of a Real-Time Human Emotion Recognition System

D. V. Ashwin, Abhinav Kumar, and J. Manikandan(✉)

Department of ECE, Crucible of Research and Innovation (CORI), PES University,
100-Feet Ring Road, BSK Stage III, Bangalore 560085, India
ashwindv75@gmail.com, abhinav.kumar685@gmail.com,
manikandanj@pes.edu

Abstract. Emotion recognition systems are in huge demand to understand the emotion of human towards human, animals, computers, machines and systems. This has influenced such systems to be employed for applications in various domains such as website customization, study of audience reaction in theaters, gaming, software engineering, education and many more. In this paper, a real-time emotion recognition system is designed which is capable of recognizing six human emotions of any person in front of the camera, without any prior information about the person. This is achieved using a combination of both geometric and appearance based features. In order to assess and enhance the performance of the proposed design, the system is tested using standard CK+ datasets too. The system designed is evaluated using three different classifiers and their results are reported. Maximum accuracy of 98.73% is achieved by the system. The proposed system is designed using open source software and can be used for various IoT based applications too.

1 Introduction

Human emotions are considered as one of the important mediums through which vital information can be retrieved for various applications. Humans express their emotions in different ways, namely facial expression, body language, tone of voice, words uttered etc. [1, 2]. The facial expressions are very prominent as reported in [3], and the ability to recognize human emotions using a real-time automated system can provide significant impact on several areas like marketing, gaming, e-learning, entertainment and other applications that involve human computer interaction.

Design of emotion recognition systems for various applications is reported in literature for Human Computer Interactions (HCI) in [4–8] and many more. In this paper, a novel attempt is made to detect human emotions using facial expressions that doesn't require any prior information about the person under test. In other words, the system designed is trained only once and it works fine in detecting human emotions of any person sitting in front of the system, without any further training. The proposed system is designed using open source software (OSS) namely OpenCV, dlib and skimage. It employs a combination of geometry-based features proposed in [9] and appearance-based features proposed in [10] for better recognition accuracy.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 177–188, 2018.
https://doi.org/10.1007/978-3-319-73423-1_16

The organization of the paper is as follows. The motivation for proposed work is reported in Sect. 2. The design of proposed real-time human recognition system is explained in Sect. 3. The performance evaluation of system designed is reported in Sect. 4 followed by Conclusion and References.

2 Motivation

The work of Darwin in 1872 [11] attracted the attention of many behavioural scientists towards emotion recognition and nearly after a century in 1978 Suwa made the first attempt to automate the process of analyzing facial expressions in [12]. Lot of work has been carried out in this domain and a detailed survey of the existing work can be found in [13–15]. A few notable limitations of emotion recognition systems include the dependency on prior knowledge such as person specific neutral expression, location of eyes as in [16, 17], use of time and memory intensive algorithm like Gabor-wavelet algorithm in [18], use of algorithms like optical flow used in [19], which are sensitive to factors which cannot be easily controlled like background lighting, use of methods which has limitations on generalizing the system for other datasets [7, 8]. These limitations motivated the design of proposed system, which is completely automated and does not require any prior knowledge. The system is fast and the performance is not affected by factors like background lighting, tilted faces, face size, colour of skin etc. The system works fine with real-time input as well as for standard datasets.

3 Proposed System

The block diagram of proposed real-time human emotion recognition system is shown in Fig. 1 and the working of each block is explained in detail in following subsections. It may be observed from Fig. 1 that two features (Geometrical and Local binary patterns (LBP)) are extracted in parallel for the system designed.

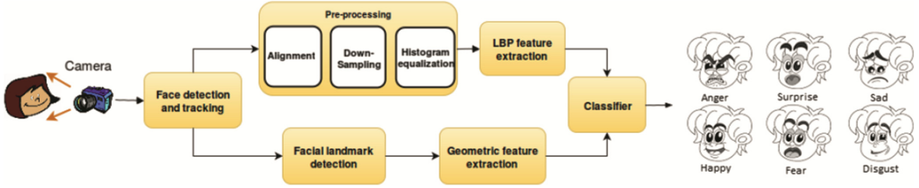
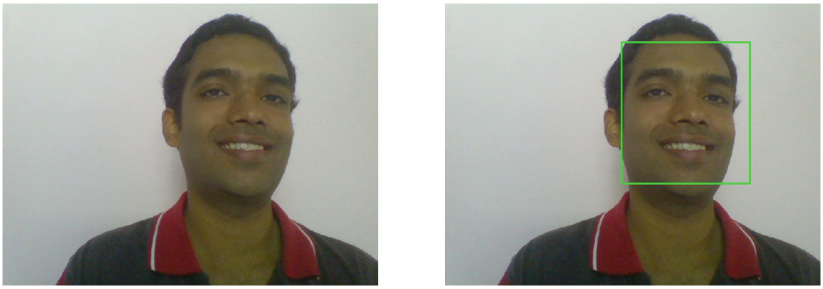


Fig. 1. Block diagram of proposed human emotion recognition system

A Face Detection and Tracking

An image of size 1280 × 720 pixels is first captured using laptop camera and the system detects a face in given image. In case, no face is available in the image, the system will continue to capture images till a face is detected in the image. The face detection is accomplished using Viola-Jones algorithm proposed in [20] and is implemented using OpenCV [21] for proposed system. Viola-Jones algorithm is a Haar feature based cascade classifiers and it comprises of four major steps namely converting image into integral image, extracting haar features from integral image, selecting important features using adaboosting algorithm [22] and finally using the selected features in a cascade form to detect faces in a given image. Viola Jones algorithm detects all the faces in an image, whereas the system focuses on the face which is closest to the camera. The input and output of face detection and tracking block is shown in Fig. 2. The green rectangle tracks the movement of face in the video. It was observed during implementation that the face detection rate is slow and hence correlation tracker proposed in [23] and implemented in dlib [24] is used for the system to enhance the speed of tracking.



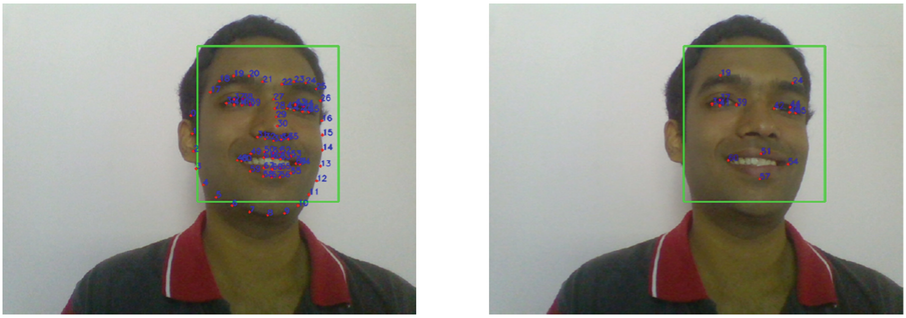
Input Image

Image with Face Detection

Fig. 2. Performance of face detection block

B Facial Landmark Detection

Ensemble of regression trees method proposed in [25] is used to detect the facial landmarks on the filtered face as shown in Fig. 3. The ensemble of regression trees method is implemented in dlib library. The pre-trained model in dlib library provides 68 points as shown in Fig. 3a, whereas only a sub set of 14 points is used in proposed work to extract the geometric features as shown in Fig. 3b.



(a) 68 Facial Landmarks

(b) 14 Facial Landmarks

Fig. 3. Performance of facial landmark detection block

C Geometrical Feature Extraction

The following nine geometric features are extracted from the 14 points obtained from the facial landmark detection block. Let $d(p_i, p_j)$ denote the Euclidean distance between the points with index i and j respectively and g_i be the ith geometrical feature extracted.

Distance between eye lid and eyebrow is computed as

$$g_1 = \frac{d(p_{19}, p_{39}) + d(p_{24}, p_{42})}{2} \tag{1}$$

Distance between eye lid and lip is computed as

$$g_2 = \frac{d(p_{39}, p_{51}) + d(p_{42}, p_{51})}{2} \tag{2}$$

Width of the mouth is computed as

$$g_3 = d(p_{54}, p_{60}) \tag{3}$$

Height of the mouth is computed as

$$g_4 = d(p_{51}, p_{57}) \tag{4}$$

Ratio of mouth width to mouth height is computed as

$$g_5 = \frac{g_3}{g_4} = \frac{d(p_{54}, p_{60})}{d(p_{51}, p_{57})} \tag{5}$$

Distance between eye brows is computed as

$$g_6 = d(p_{19}, p_{24}) \tag{6}$$

Distance between eye lids of an eye is computed as

$$g_7 = \frac{d(p_{37}, p_{41}) + d(p_{44}, p_{46})}{2} \tag{7}$$

Average of half of upper lip length is computed as

$$g_8 = \frac{d(p_{51}, p_{60}) + d(p_{51}, p_{54})}{2} \tag{8}$$

Average of half of lower lip length is computed as

$$g_9 = \frac{d(p_{57}, p_{60}) + d(p_{57}, p_{54})}{2} \tag{9}$$

The geometrical features extracted from 14 facial landmark points are shown in Fig. 4 with white lines connecting the facial landmark points. All the above mentioned geometrical features are sensitive to scale variations and hence an evaluation of normalization methods is carried out using features such as width of the rectangle around face, height of rectangle around face, area of rectangle around face, ratio of width to height of rectangle around face. The normalizing equation is formulated as

$$f_i = \frac{g_i}{D} \tag{10}$$

where i = 1, 2, …, 9. Let us consider H and W to be the height and width of the rectangle around face then the normalizing parameter D in (10) is given as

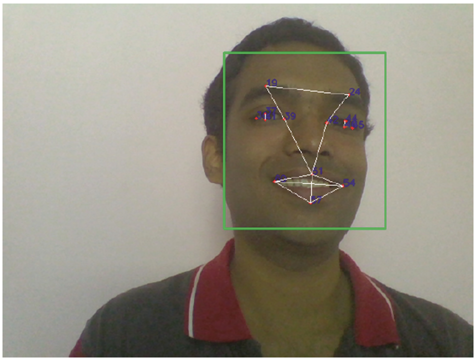


Fig. 4. Extraction of geometrical features from face

$$\mathrm{D} = \mathrm{H}\ \textit{for height normalization} \tag{11}$$

$$\mathrm{D} = \mathrm{W}\ \ \textit{for height normalization} \tag{12}$$

$$\mathrm{D} = \mathrm{H} \times \mathrm{W}\ \textit{for height normalization} \tag{13}$$

$$\mathrm{D} = \mathrm{H/W}\ \ \textit{for height normalization} \tag{14}$$

$$\mathrm{D} = 1\ \textit{for height normalization} \tag{15}$$

In order to assess and finalize among the above mentioned normalizing methods, a performance evaluation experiment was carried out as follows. The experiment involved a person moving towards the web-cam of a laptop from a distance with same expression on the face. It is well known that when the person is away from the camera, the size of detected face in that particular frame will be small and hence the Euclidean distance between the selected facial landmarks will be small. Similarly, when the person moves towards the camera, the size of detected face increases and hence Euclidean distance between the selected facial landmark also increases. During this experiment, all the nine geometrical features were extracted from all the frames of the video sequence. A total of 1474 frames/image were obtained during the process, out of which 266 images were having the emotion of anger, 273 images for disgust, 231 images for fear, 224 images for happy, 252 images for sad and 228 images for surprise. The above mentioned normalizing approaches were individually applied to all the nine features. Average deviation was computed for each case and all the nine features using the equation

$$\delta_{avg} = \frac{1}{9}\sum_{k=1}^{9}\frac{\sigma_k}{\mu_k} \times 100 \tag{16}$$

where σ_k represents standard deviation of k^{th} feature given as

$$\sigma_k = \frac{\sum_{i=1}^{n}(g_k^i - \mu_k)^2}{n} \tag{17}$$

where μ_k represents the mean of k^{th} feature and is given by

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n} g_k^i \tag{18}$$

where n represents number of frames in the video sequence, g_k^i represents value of k^{th} feature in i^{th} frame.

The performance evaluation of normalizing methods was carried out using LDA classifier and Leaving-out-one method to obtain the recognition accuracy and the results are reported in Table 1. It is observed and concluded from Table 1 that Height and Width normalization gave best results with minimum deviation and best recognition accuracy. Hence, width normalization is considered for the proposed system with normalizing parameter D as the width of eye given as

Table 1. Evaluation of different normalization methods

SI No.	Normalizing method	Deviation (δ_{avg}) in %	Accuracy in %
1	UnNormalized	17.65	99.72
2	AreaNormalized	19.75	99.66
3	RatioNormalized	17.66	99.72
4	HeightNormalized	5.89	99.79
5	WidthNormalized	5.87	99.79

$$\mathrm{D} = \frac{d(p_{36}, p_{39}) + d(p_{42}, p_{45})}{2} \tag{19}$$

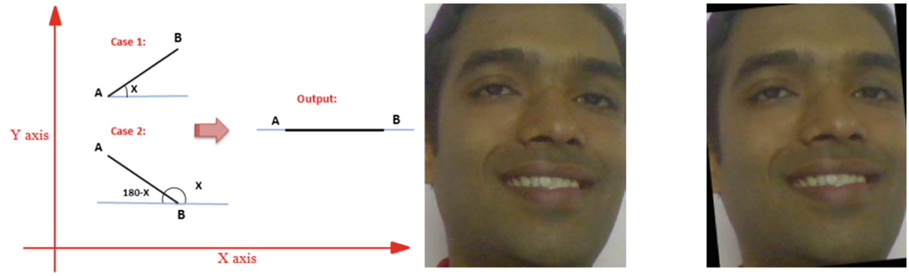
and the nine geometric features are normalized as

$$f_1 = \frac{g_1}{2D}\, f_2 = \frac{g_2}{D}\, f_3 = \frac{g_3}{D}\, f_4 = \frac{g_4}{D}\, f_5 = \frac{g_5}{D}\, f_6 = \frac{g_6}{2D}\, f_7 = \frac{g_7}{0.5D}\, f_8 = \frac{g_8}{D}\, f_9 = \frac{g_9}{D} \tag{20}$$

Different features were scaled by different values to make sure that all the normalized features lie in the same range avoiding feature dominance i.e., it removes the possibility of any feature with a large value dominating other features with relatively smaller values.

D Pre-processing for LBP Feature Extraction

In order to perform pre-processing for LBP feature extraction, the face is cropped and considered as Region of Interest (ROI). After cropping, the face alignment is performed by rotating the cropped image in such a way that the line joining points 36 and 45 are parallel to the horizontal reference line. Illustration of rotating a line to make it parallel to horizontal line is shown in Fig. 5, where point A and B can be considered as landmark points 36 and 45 respectively. The angle of rotation is calculated as



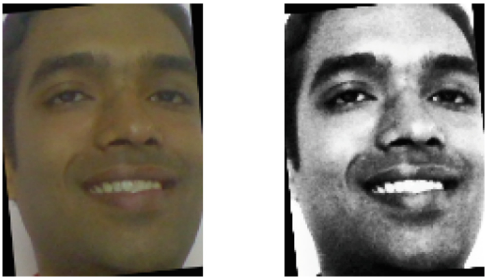
Procedure employed for Face alignment   Cropped Input Face   Aligned Face

Fig. 5. Illustration of face alignment

$$\theta = \tan^{-1}\left(\frac{p_{45}^{y} - p_{36}^{y}}{p_{45}^{x} - p_{36}^{x}}\right) \tag{21}$$

where p_{45}^{x}, p_{45}^{y}, p_{36}^{x}, p_{36}^{y} represents x and y co-ordinates of point p_{45} and p_{36} respectively. The results of input cropped face and its alignment are shown in Fig. 5.

Next step is to down-sample the aligned face to 108 × 147 pixels as shown in Fig. 6. This ensures that different parts of the face share almost the same location irrespective of input face. The resizing is followed by histogram equalization to improve the contrast of the image as shown in Fig. 6.



Resized image   Histogram equalized image

Fig. 6. Preprocessing of cropped input face

E Local Binary Patterns (LBP)

Local Binary Pattern (LBP) operator compares the pixel value at any point with its neighbouring pixel values to generate a binary number representing the pattern. The histogram of LBP image is a robust feature descriptor against variations in illumination. Uniform LBP ($LBP^{u2}_{8,2}$) approach proposed in [26] is used after dividing the 108 × 147 pixels face images into blocks of pixel size 18 × 21, providing a better trade-off between recognition performance and feature vector length. The input face images are divided into 42(6 × 7) blocks as shown in Fig. 7 and LBP features of length 59 are extracted from each block. The features extracted from each block are concatenated to obtain the LBP histogram of length 2478(59 × 42). The LBP feature extraction for proposed system is implemented using LBP function with 'uniform' method available in scikit-image [27] to determine the pattern. The dimensionality reduction of LBP features is performed using PCA (Principal Component Analysis).

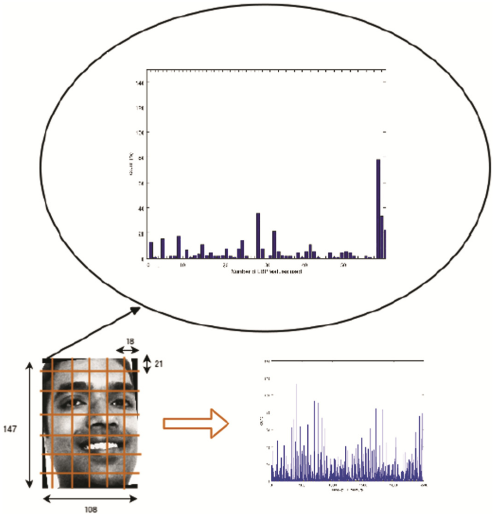


Fig. 7. LBP feature extraction

Implementation of PCA in scikit-learn [28] was used for the proposed work. It was observed during analysis that PCA of feature size 80 gave best results and hence the same is used in this work.

F Classifier Design

The features extracted from input image includes 2478 appearance based features and 9 geometric features. Different combinations of input features are fed to classifiers for assessing the performance of the system designed. The three different classifiers considered are K-Nearest Neighbour (KNN), Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). These classifiers are implemented using scikit-learn library proposed in [28]. The different architectures and combination of inputs considered for evaluation in this work are shown in Fig. 8.

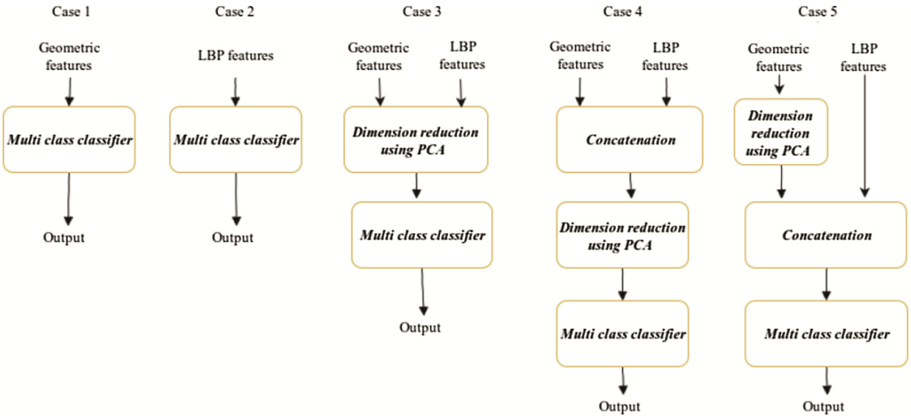


Fig. 8. Different architectures considered for proposed work

4 Experimental Results

The proposed human emotion recognition system consists of a laptop with the software running on it and the system is capable of detecting the emotion of any person sitting in front of it as shown in Fig. 9, without any additional training. The results obtained for different emotions using the system are also shown in Fig. 9 and satisfactory performance was observed on real-time datasets.

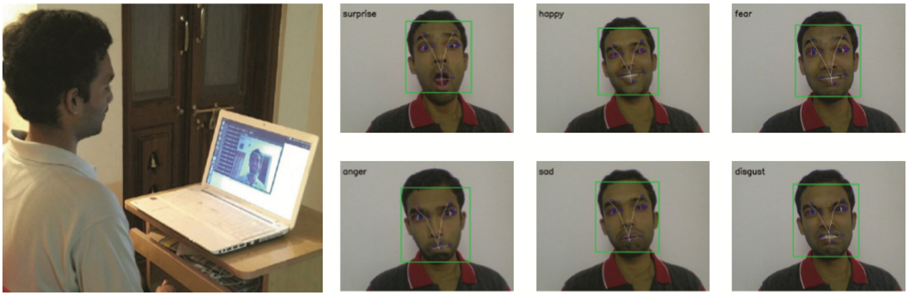


Fig. 9. Hardware setup for testing and the results obtained from the system designed

In order to quantify the performance of the proposed system using different classifiers and architectures, experiments were carried out on extended Cohn-Kanade dataset (CK+) which is one of the most widely used data base for facial expression recognition. The database consists of 593 sequences of 123 subjects. Each image sequence starts with neutral expression and ends with a peak expression. The offered peak expression is fully coded by Facial Action Coding System (FACS) using FACS investigator guide. After applying perceptual judgement to the facial expression labels, only 327 of the sequences were for the human facial expressions: 45 for anger (An), 18 for contempt (Co), 59 for disgust (Di), 25 for fear (Fe), 69 for happiness (Ha), 28 for sadness (Sa) and 83 for surprise (Su). Out of these datasets, only 309 images corresponding to the six basic emotions considered by our system i.e., anger, disgust, fear, happy, sad and surprise are used. The results obtained from these datasets for different architectures are reported in Table 2 using the leaving-one-out method for classification and it is observed that Case 5 using LDA classifier gave best results. It may be noted that this experimentation is not mandatory for real-time system designed and is carried out only to finalize the features and classifier to be used for the real-time system.

Table 2. Performance evaluation of proposed real-time emotion recognition system for CK+ dataset

Emotion	CASE 1			CASE 2			CASE 3			CASE 4			CASE 5		
	KNN	LDA	SVM	KNN	LDA	SVM	KNN	LDA	SVM	KNN	LDA	SVM	KNN	LDA	SVM
Anger	60.00	68.88	73.33	57.77	64.44	73.33	55.56	73.33	77.33	55.56	73.33	77.78	55.55	75.56	77.78
Disgust	75.00	83.33	78.33	43.33	86.67	83.33	50.00	88.33	83.33	50.00	88.33	83.33	50.00	88.33	83.33
Fear	64.00	96.00	76.00	16.00	76.00	68.00	32.00	80.00	68.00	12.00	80.00	68.00	12.00	84.00	68.00
Happy	94.12	98.53	92.65	69.13	98.53	95.06	63.24	95.59	97.10	63.23	95.59	89.70	63.23	97.10	89.71
Sad	14.28	42.86	42.86	3.57	46.43	42.86	3.57	71.43	42.86	3.57	71.43	57.14	3.57	67.88	57.14
Surprise	98.73	98.73	96.20	77.22	97.47	94.94	78.48	96.20	94.94	78.48	96.20	88.61	78.45	98.73	88.61
Accuracy	**76.72**	**85.9**	**81.97**	**54.10**	**84.26**	**82.51**	**55.41**	**87.54**	**83.55**	**53.76**	**87.54**	**81.63**	**53.76**	**88.86**	**81.64**

The proposed system is designed using a Toshiba laptop with following configuration: Quadcore AMD Processor operating at 1.5 GHz with 8 GB RAM and 1 MB L2 cache. Details about the time taken by various blocks in the system are reported in Table 3. Face detection takes 275 ms, but it is performed only once every 2 s. Once a face is detected, only face tracking is performed which takes only 60 ms and this significantly improved the performance of our system. The computation time can be further reduced on using a system with better configuration.

Table 3. Time taken by various steps

SI No.	Step	Time taken in ms
1	Face detection	275
2	Face tracking	60
3	Landmark detection	7.6
4	Geometrical feature extraction	2
5	Pre-processing	0.45
6	LBP feature extraction	45
7	Classifier	0.001

5 Conclusion

Design and implementation of a real-time human emotion recognition system is proposed in this paper. The entire system is designed using freely available open source softwares and libraries only, motivating the readers to design their own system. The system works well with low resolution images as input too, which can be easily obtained from webcam, security cameras etc., which will be useful for IoT based applications, where most of the images captured have lesser resolution. The system designed is real-time, fully automated, doesn't require any prior information about the person in front of the camera, independent of factors like background lighting and needs just one time training. The system designed has potential applications for gaming, marketing, e-learning etc.

References

1. Izard, C.E.: Human Emotions. Springer Science & Business Media, Berlin (2013)
2. Mauro, R., Sato, K., Tucker, J.: The role of appraisal in human emotions: a cross-cultural study. J. Pers. Soc. Psychol. **62**(2), 301 (1992)
3. De la Torre, F., Cohn, J.F.: Visual analysis of humans: looking at people. In: Moeslund, T., Hilton, A., Krüger, V., Sigal, L. (eds.) Facial Expression Analysis, pp. 377–409. Springer, London (2011). https://doi.org/10.1007/978-0-85729-997-0_19
4. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human-computer interaction. Sig. Process. Mag. IEEE **18**(1), 32–80 (2001)
5. Anderson, K., McOwan, P.W.: A real-time automated system for the recognition of human facial expressions. IEEE Trans. Syst. Man Cybern. Part B (Cybern) **36**, 96–105 (2006)
6. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 2, pp. 568–573. June 2005
7. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
8. Saeed, A., Al-Hamadi, A., Niese, R., Elzobi, M.: Frame-based facial expression recognition using geometrical features. Adv. Hum.-Comput. Interact. **2014**, 4 (2014)
9. Thanh Do, T., Hoang Le, T.: Facial feature extraction using geometric feature and independent component analysis. In: Richards, D., Kang, B.-H. (eds.) PKAW 2008. LNCS (LNAI), vol. 5465, pp. 231–241. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01715-5_20
10. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
11. Darwin, C.: The Expression of the Emotions in Man and Animals. John Murray, London (1872)
12. Suwa, M., Sugie, N., Fujimora, K.: A preliminary note on pattern recognition of human emotional expression. In: International Joint Conference on Pattern Recognition, vol. 1978, pp. 408–410. (1978)
13. Pande, S., Shinde, S.: A survey on: emotion recognition with respect to database and various recognition techniques. Int. J. Comput. Appl. **58**(3), 9–12 (2012)
14. Hemalatha, G., Sumathi, C.: A study of techniques for facial detection and expression classification. Int. J. Comput. Sci. Eng. Surv. **5**(2), 27 (2014)
15. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. Pattern Recogn. **36**(1), 259–275 (2003)
16. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 94–101. IEEE (2010)

17. Niese, R., Al-Hamadi, A., Farag, A., Neumann, H., Michaelis, B.: Facial expression recognition based on geometric and optical flow features in colour image sequences. Comput. Vis. IET **6**(2), 79–89 (2012)
18. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, pp. 568–573. IEEE (2005)
19. Yeasin, M., Bullot, B., Sharma, R.: From facial expression to level of interest: a spatio-temporal approach. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 2, pp. II–922. IEEE (2004)
20. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**, 137–154 (2004)
21. Bradski, G.: The OpenCV library. Dr. Dobb's J. Softw. Tools **120**, 122–125 (2000)
22. Freund, Y., Schapire, R.E.: A short introduction to boosting. J. Jpn. Soc. Artif. Intell. **14**, 771–780 (1999)
23. Danelljan, M., Häger, G., Khan, F., Felsberg, M.: Accurate scale estimation for robust visual tracking. In: British Machine Vision Conference, Nottingham, 1–5 September 2014. BMVA Press (2014)
24. King, D.E.: Dlib-ml: a machine learning toolkit. J. Mach. Learn. Res. **10**, 1755–1758 (2009)
25. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874 (2014)
26. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
27. van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., Scikit-Image Contributors: Scikit-image: image processing in Python. PeerJ **2**, e453 (2014)
28. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**(1), 2825–2830 (2011)

Trace and Track: Enhanced Pharma Supply Chain Infrastructure to Prevent Fraud

Archa(✉), Bithin Alangot, and Krishnashree Achuthan

Amrita Center for Cybersecurity Systems and Networks,
Amrita School of Engineering, Amritapuri Campus,
Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore, India
archa91@gmail.com, bithina@am.amrita.edu, krishna@amrita.edu

Abstract. The menace of counterfeit drugs has prompted the regulatory authorities to mandate trace and track systems to verify the provenance of drugs as it travels across the supply chain. Here, we propose a highly scalable and enhanced trace and track system for pharma supply chain. Our novel idea of using an IoT framework known as GDP (Global Data Plane) integrated with blockchain helps in communication and management of data between untrusted parties. The blockchain maintains an immutable record of drugs each party holds and transactions the parties make amongst themselves. This makes it hard to introduce counterfeit drugs into the supply chain.

Keywords: Supply chain · Blockchain · IoT

1 Introduction

The pharma industry provides critical health care service by supplying life saving drugs to people. But the recent surge of counterfeit drug supply in the market has raised issues of mistrust among the people. The statistics show that this issue is especially prevalent in the developing countries. There are many cases of counterfeit drugs reported worldwide, for example, the anti malarial drugs that were bought in Southeast Asia or the usage of toxin diethylene glycol in the manufacture of fake paracetamol syrup [4]. Many of the pharmaceutical companies and government do not publish information regarding counterfeit drugs as it might harm their sales [12]. They also believe in the philosophy that "as much as possible should be done behind the scene and that no great publicity should be sought because it could damage public confidence in medicines" [4].

In order to counter menace of counterfeit drugs, the regulatory authorities (such as the FDA) has mandated to implement trace and track system into the pharma supply chain. The Drug Quality and Security Act (DQSA) details the requirements to build a system to trace some prescription drugs that are distributed in United States. The key recommendations of the system are: (1) to include unique product identifiers by manufacturers and repackagers with lot granularity, (2) support tracing, verification and notification by manufacturers,

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 189–195, 2018.
https://doi.org/10.1007/978-3-319-73423-1_17

wholesale distributors, repackagers and dispensers, (3) include details on wholesale as well as third party logistics licensing [6]. The pharma companies take help from companies such as Tracelink Inc. [16] to implement a trace and track system which follows all the compliance put forth by the regulatory authorities. The companies provide services over the cloud which allow to deploy the system over a global scale. But applications (especially IoT applications) build over the cloud if not properly implemented can lead to lot of drawbacks as mentioned by Zhang et al. The trace and track system can be considered as a large scale IoT infrastructure with barcode or RFID readers which are connected to the cloud over the internet.

In this paper, we propose to build a highly scalable and trusted trace and track system for the pharma industry that trace their drugs across the distribution lifecycle. Here, we model the whole system into a large scale Internet of Things (IoT) infrastructure where IoT devices such as barcode readers, smartphones and so on scan the serial numbers or RFID tags integrated into the drug package. In order to manage the IoT applications and devices across the distributed infrastructure we use an IoT framework called GDP [3]. We also design a controller for the GDP leveraging blockchain technology to track the transaction between different entities. A single entity cannot make any changes to the transaction record without the knowledge of others. This will bring in more accountability into the system and in turn eliminate the possibility of inclusion of counterfeit drugs into the distribution cycle by malicious entities.

Contribution of our paper:

- Propose a scalable trace and track system for pharmaceutical industry using GDP IoT framework.
- Integrate blockchain into the trace and track system to create a trusted environment amongst different entities in the supply to prevent fraud.

The paper is structured as follows. Section 2 overviews the related works. Section 3 describes the background knowledge of the GDP framework and blockchain technology. Section 4 proposes the trace and track system and the paper concludes with Sect. 5.

2 Related Works

For fraud detection many trace and track systems have been developed to be used as part of the supply chain cycle of various products such as with drugs. The Tracelink is a pharma supply chain trace and track system. It introduced Life Sciences cloud on Amazon Web services to provide a scalable solution to trace drugs on a global level if needed [16]. This is further enforced by techniques to neutralize data format and transport preferences allowing partners to integrate under a uniform level. Bosch Packaging Technology [2] is another such system. It provides the mass serialization services in food and pharmaceutical companies. It makes use of an open standard for data format GS1 application identifiers to represent the data. Data is encoded into 2D matrix code and printed on the product in a human readable format. A camera reads this and stores the information in a centralized database unlike our decentralized approach for tracking. Recently IBM and Maersk [9] released a system that uses Hyperledger Fabric to implement a tracking system for shipping industries. Our solution offers more flexibility and trust than the existing systems as we are using the GDP framework that better supports the IoT infrastructure along with the blockchain controller using Tendermint to improve security and privacy.

3 Background

In this section, we discuss two main technologies used to build our trace and track system: GDP and Blockchain. The GDP is a framework which enables us to develop applications on top of IoT infrastructure with ease and Blockchain is a distributed ledger which is maintained in a decentralized manner by a group of untrusted people.

3.1 Global Data Plane

The growth of Internet and the development of communicative devices have led to high interactivity through computing platforms and services leading to the development of Internet of Things (IoT). The recent trend in easy availability of cheap IoT devices and economic model of cloud has accelerated the deployment of highly scalable IoT applications without much effort. Anyone with a sensor can start streaming data to the cloud from which it could be analyzed to get useful information. But connecting these IoT devices directly to the cloud has drawn in issues related to scalability, bandwidth utilization, latency, durability management along with privacy and security concerns as stated by Zhang et al. In order to overcome these issues related to the present IoT infrastructure they have come with a IoT framework called GDP. This framework handles data protection, preservation and distribution amongst the distributed edge servers and backend cloud nodes. GDP follows a data-centric architecture which advocate the use of a append-only log data structure as the fundamental storage abstraction for transferring and managing data [3]. A log abstract sensors, actuator and IoT applications; any read or write to the device is done via a log just like how a file abstract devices in Operating Systems. Logs are durable, lightweight and support multiple readers at the same time and can be migrated as necessary to meet Quality of Service (QoS) requirements.

The GDP is logically divided into different planes: Application Plane, Control Plane and Data Plane. The data plane contains the logs while control plane holds the program which need to communicate with both applications and logs. The controller application such as the one which decide the follow of data across the infrastructure or access controllers are placed in the control plane. The application plane contains the IoT applications which is used to interact with the IoT devices via log. In our trace and track system we are developing both the controller with blockchain backend and IoT application.

3.2 Blockchain

Blockchain was first introduced as a immutable public ledger for tracking transaction of cryptocurrency Bitcoin [15]. Its decentralized approach relies on the distributed network to validate the consensus for the transaction and prevents any single one authority to control it. The presence of a third party may monopolize the transactions, eliminating them reduces transaction costs and time delays. This feature of blockchain further adds reliability to the transaction as it prevents the occurrence of human errors. Also, the records are accessible from anywhere with any device and by anyone. Confidentiality of the data in the blockchain is maintained by encryption and hashing methods.

The research into the various consensus algorithms has led to diverse applications for blockchains [7,14] and to support uses apart from just cryptocurrencies. This is achieved by introduction of private blockchains. While the public blockchain implemented by Bitcoin application can be used by any user of the Internet it has its limitations. In applications which poses constraints on who has the authority to read the data, private blockchain is a more ideal alternative. It still maintains the property of blockchain and does not allow writes onto it based on a single party only. Private blockchain combines the advantages of public blockchain and maintains confidentiality provided by traditional databases.

The difference in the types of blockchains is brought about by the consensus algorithms they use. Consensus algorithm mainly defines a method whereby all the participants can come to an agreement on a matter in a fair manner in the best interest of the parties involved in the transaction. Many projects are developed based on different consensus algorithms like Proof of Work (PoW)[15], Proof of Stake (PoS)[18], Proof of Elasped Time (PoET)[10] and Practical Byzantine Fault Tolerance (PBFT)[13]. Projects like Hyperledger Fabric [8], Intelledger [10] and Tendermint [11] are designed on the basis of these algorithms.

There are several options to choose from with respect to consensus algorithms and blockchain frameworks. In our case study, the design prioritized privacy followed by scalability and also ensures more accountability. The combination of Tendermint blockchain framework and Tendermint Consensus Algorithm was found most suitable to cater to these requirements.

4 Pharma Trace and Track System

In this section, we introduce the trace and track system and its design details. The pharmaceutical supply chain consists of many unit processes that transform the input resources to medicines and medical equipments for distribution in the market. In each of the unit processes (manufacturing, registration, distribution etc.) fraudulent occurrences can lead to the development of substandard or counterfeit drugs. Pharma companies may not adhere to proper manufacturing technique; register products that do not meet quality requirement; hoard drugs to manipulate the market prices and so on. An example is when FDA discovered counterfeit versions of Avastin, a cancer drug, introduced in the U.S.

supply chain in 2012 [5]. Such activities adversely affect people's trust on pharma companies and the regulatory agencies.

The work flow at each stage of supply chain varies, hence, we will have to employ different approaches to tackle the different frauds. We focus on a solution to tackle fraud in the distribution stage as our first step. Currently, there are companies such as Tracelink [16], Axway [1] and VerifyBrand [17] which provide services to prevent or detect counterfeits. The present implementations makes use of distributed databases with central administrator while our blockchain approach help maintain the database among nodes that do not trust each other in a decentralized manner.

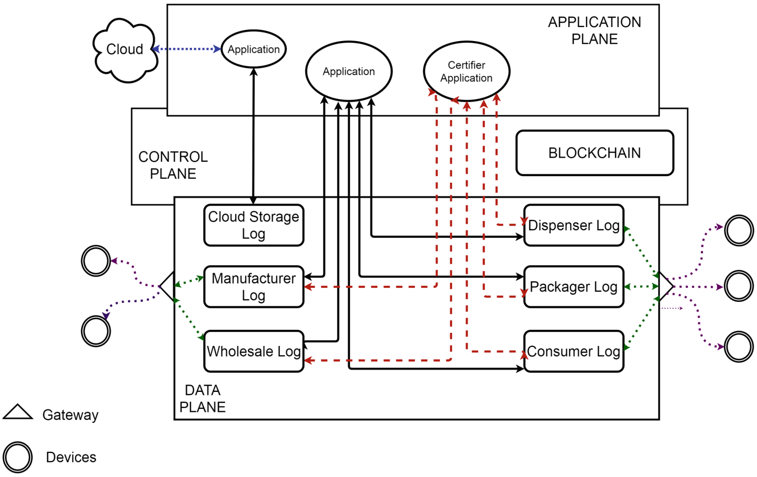


Fig. 1. Pharma supply chain infrastructure

In our implementation, we use GDP IoT framework proposed by Zhang et al. for efficient management and communication of data across the pharma supply chain. The whole supply chain infrastructure consist of sensors (RFID reader, Barcode scanner, Smartphones) that generate time series data that need to be stored and processed in a distributed manner. All data including information regarding the drug as well as transaction details are collected by the devices and stored in log files. The GDP framework is specifically designed to manage time-series data through its log based storage abstraction. The logs are signed with unique cryptographic keys for different devices, hence authenticity, integrity and non-repudiation is also incorporated.

The framework is logically divided into different planes as shown in the Fig. 1, such as the data plane, control plane and application plane. The data plane along with the application plane holds the trace and track application that we are going to build and control plane contains the controller that can detect counterfeit drug. The applications include querying the requested details of a medicine, taking care of data transfer from one log to another simultaneously with transfer of medicine from one participant to another. It also includes transferring relevant data to the cloud for further high end processing. The novelty of our approach is on the design of the GDP controller which uses blockchain to store verified transaction details on drug transfer. The design of controller workflow and role of blockchain in detecting fraud is further elaborated below.

While GDP help in storage and communication amongst different entities in a supply chain, the controller helps in tracing the drug as it travel across the entities of supply chain. When one entity transfers a set of drugs to another, the transaction details are sent to the controller. The controller in turn stores the details onto a private blockchain. For example, a transaction is recorded when a Distributor A transfers 100 units of a drug from manufacturer A to supplier B. The controller with the help of blockchain maintains the exact details of drugs that each entity hold thus making it very difficult to introduce counterfeits. The blockchain is maintained by all the entities that are part of the supply chain. The entities use a consensus algorithm to update the blockchain in order to avoid any inconsistency.

In blockchain, transactions are stored into as blocks which are linked using cryptographic hashes. A block consists of a header, set of transactions (that occurred during a time window) and a set of signatures from the participants to verify the transactions of the previous block. Each transaction will consist of a unique ID to identify the batch number, the source entity, and receiving entities, the quantity of medicines transferred and the timestamp. The header will consist of length of the chain, the last block ID as well as the hash of the previous chain state.

We have started the implementation of our trace and track system using Tendermint. Tendermint is divided into two sub components: a blockchain consensus engine and a generic application interface. Former, helps in the creation of blockchain and the later facilitates the communication between the trace and track controller and the blockchain [11]. The reason why we have chosen Tendermint is that it provide more accountability in case of occurrence of fraud. It is designed to provide the provenance of drugs to prevent introduction of counterfeit drugs into the supply chain while transactions happens between different entities. During distribution each of the product will have a unique hash representing a product state that can be tracked, but cannot be replicated. Due to the addition of transactions at each stage governed by consensus of its participants, the introduction of counterfeit drugs without other's knowledge is impossible.

5 Conclusion

Technologies for Healthcare today is a growing domain witnessing significant leaps in innovative use of science and engineering to enhance patient care. It is also a critical domain directly impacting health and well being. Pharma based healthcare industries serve large populations, and recently they have witnessed increasing instabilities in their supply chains by miscreants trying to introduce counterfeit drugs for additional profit. The increase in counterfeit drugs and the distribution of substandard drugs that can cause disability and loss of life have led to the implementation of trace and track systems over the pharmacy supply chains by the federal agencies around the world. In our work we have described a novel trace and track system that combines the advantages of GDP and blockchain technology. We utilize the GDP framework to build the IoT application that tracks the provenance of drugs. We also put forward the idea of using the blockchain technology in the controller application to ensure the trust among the different participants of the system. The blockchain by design imparts transparency, authentication, auditability to trace the origins of a product. Using the blockchain distributed design reduces the data tampering risks Further providing a unique identity to the blockchain makes it a cheap but error free operation.

References

1. Axway. https://www.axway.com/en/datasheet/Axway-track-trace-gs1-epcglobal-certified
2. Bosch Packaging. http://www.boschpackaging.com/en/pa/services/after-sales-services/modernization/track-and-trace/track-and-trace-4.html
3. Zhang, B., Mor, N., Kolb, J., Chan, D.S., Lutz, K., Allman, E., Wawrzynek, J., Lee, E.A., Kubiatowicz, J.: The cloud is not enough: saving IoT from the cloud. In: HotCloud (2015)
4. Cockburn, R., et al.: The global threat of counterfeit drugs: why industry and governments must communicate the dangers. PLoS Med. **2**(4), e100 (2005)
5. Drug Safety. https://www.fda.gov/drugs/drugsafety/ucm291960.htm
6. FDA. http://www.fda.gov/Drugs/DrugSafety/DrugIntegrityandSupplyChainSecurity/DrugSupplyChainSecurityAct/
7. Fischer, M.J.: The consensus problem in unreliable distributed systems (a brief survey). In: Karpinski, M. (ed.) FCT 1983. LNCS, vol. 158, pp. 127–140. Springer, Heidelberg (1983). https://doi.org/10.1007/3-540-12689-9_99
8. Hyperledger Project. www.hyperledger.org
9. IBM. https://www-03.ibm.com/press/us/en/pressrelease/51712.wss
10. Intelledger. http://intelledger.github.io
11. Kwon, J.: Tendermint: consensus without mining. http://tendermint.com/docs/tendermintv04.pdf (2014)
12. Gibson, L.: Drug regulators study global treaty to tackle counterfeit drugs. BMJ. Br. Med. J. **328**(7438), 486 (2004)
13. Castro, M., Liskov, B., et al.: Practical byzantine fault tolerance. In: OSDI, vol. 99, pp. 173–186 (1999)
14. Crosby, M., Pattanayak, P., Verma, S., Kalyanaraman, V.: Blockchain technology: beyond bitcoin. Appl. Innov. **2**, 6–10 (2016)
15. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system (2008)
16. TraceLink Inc. http://www.tracelink.com/
17. Verifybrand. http://verifybrand.com/pharma-serialization/
18. Vasin, P.: Blackcoins proof-of-stake protocol v2 (2014)

Management of IoT Devices in Home Network via Intelligent Home Gateway Using NETCONF

Savita Vijay(✉) and M. K. Banga

Department of Computer Science and Engineering, Dayananda Sagar University,
Kudlu Gate, Bangalore, India
{savitavijay-cse,chairman-cse}@dsu.edu.in

Abstract. Internet of things (IoT) is surrounded by heterogeneous entities such as sensors, mobile devices and actuators in a constrained environment which are running on very low power and lossy networks. These entities are also of very small memory and can handle small computational overhead. To this end, complete IoT system and different devices which are working in home network management system will be presented in this paper. Applications for home network are considered under different architectures and their designs are discussed. Conventional simple network management protocol (SNMP) is generally applied for network management but it is not optimal due to lack of flexibility in configuring devices and lack of capabilities in managing operations in an IoT network. A design of smart washing machine device using the IoT design methodology is being discussed using network configuration protocol (NETCONF) and yet another next generation (YANG) data modeling language to illustrate how it would be a better alternative for managing home network.

Keywords: Internet of things (IoT)
Intelligent home gateway network management (IHGNM)
Simple network management protocol (SNMP)
Network configuration protocol (NETCONF) · Netopeer

1 Introduction

There are number of heterogeneous devices present in home network which are expected to be connected. Here, devices are based on different hardware platforms, controller services are also of different nature, the software components that enable network access on them are also of different nature for each device. For example, health and lifestyle wearable IoT devices like smartwatch, wristband have different capability in terms of memory usage, power consumption, processing speed as compared to smart home appliances like washing machine.

IoT devices can be broadly classified on the basis of their key characteristics like communication pattern, memory usage, data processing capability and power consumption. For example, devices like smart keys or a smart washing machine doesn't need to be connected always and are switched on to perform certain tasks when required; these devices consume less power for communication. This paper is focused on such low

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 196–207, 2018.
https://doi.org/10.1007/978-3-319-73423-1_18

power or normally off devices in a home network as they need a gateway for effectively managing operations and communication with internet.

Particularly, there are two common approaches for managing devices in home automation system: simple network management [1] and intelligent home gateway network management (IHGNM).

The main difference between them is involvement of the intelligent home gateway for management of the devices as shown in Fig. 1.

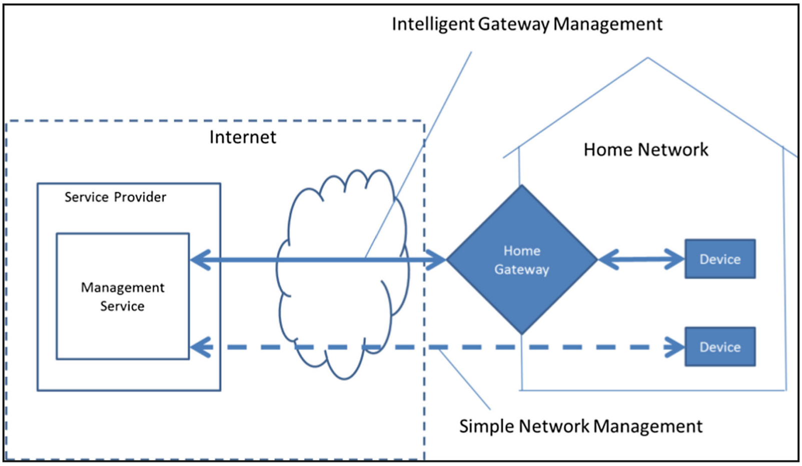


Fig. 1. Various components in home automation services architecture

Each of these approaches is as per the application. If the devices have enough resources that it supports a direct connection to the internet in a secure manner and does not support multiple device classes, then implementation is possible with light weight machine to machine (LWM2M) [2]. LWM2M is a remote device management standard. In this, IoT devices can be directly managed. This is an example of simple network management.

In IHGNM architecture, multiple device classes from low resource constrained to high resource constrained devices are considered. The proposed architecture monitors large number of heterogeneous devices in a home network. Here, IHGNM architecture for a high resource constrained device (i.e. washing machine) in a home network is discussed.

The rest of the paper is organized as follows: Sect. 2 describes IoT system components for home network management system and different devices for the same. Sections 3 and 4 describes intelligent home gateway design approach and study of same using netopeer tools. Section 5 gives conclusion and future research possible.

2 IoT System Components for Home Network Management System

2.1 A Home Network Management IoT System Comprises of Following Components

Device: An IoT device allows identification, remote sensing, actuating and monitoring from remote locations capability. Generally, IoT devices have unique identities, can exchange data with other connected devices and different web or mobile applications (directly or through Intelligent gateways), or collect data from other devices and store the data in local databases and process the data locally. Data can also be processed on cloud based application backend (like Amazon Web Services, Microsoft Azure) or on centralized servers. In different IoT design levels, we can perform some tasks locally and other tasks within IoT infrastructure, based on temporal and space constraints (i.e. memory, communication latencies and speeds, processing capabilities and deadlines).

IoT software: On IoT devices, some software components are required and installed for accessing the information from different sensors running in the home management system. They are also responsible for storing different sensor information or controlling the actuators connected to the devices. To enable network access on the device, resources are required.

Native controller service: This service is the native service that runs on the device. It interacts with the web services. For controlling the devices, it sends data from the device to the web service and receives commands from the application (via web services).

IoT database: Database for storing the collected data can be either local or it can be on cloud.

IoT data security: On local network, it will be done with symmetric key encryption decryption technique. And if it will run on internet then asymmetric key encryption decryption technique will be used.

Web service: Web services work as a connection link between all the IoT system components i.e. IoT device, IoT application, database for storing collected data and analysis components. Web services can be either implemented using hypertext transfer protocol (HTTP), constrained application protocol (CoAP) or representational state transfer (REST) principles or if it's a real time application then WebSocket protocol can also be used.

Analysis component: The Analysis component is responsible for analyzing the IoT data and generating results. Generally, results are published in a format that user can easily understand. Local and cloud, at both places, analysis can be done and then results are kept in local and cloud databases.

Applications: Users can control and monitor various features of IoT system with IoT applications only. To perform any operation on the IoT system or to see its status and to view the processed data, applications are used.

2.2 Device Heterogeneity and Applicable Management Approaches

There are numerous IoT devices across different applications in a home network as illustrated in Table 1. Based on the characteristics and constraints of devices in a home network, classification of devices is being done and appropriate management approach is recommended as highlighted in Tables 2 and 3 below.

Table 1. IoT devices for home network management system [3–5]

IoT device type	Device name
Smart lighting	Internet protocol enabled lights (tubelight, bulb), solid state lighting (LED lights)
Smart appliances	Refrigerator, air conditioner, television, television remote, music system, washer/dryer
Intrusion detection	Security camera, door sensor
Smoke/gas detectors	Smoke detectors (optical detection, ionization or air sampling technique), gas detector
Health and lifestyle	Wearable IoT devices (smartwatches e.g. moto 360 smart watch), smart glasses e.g. Google glass, wristbands (fitbit), fashion electronics (electronics in clothing and accessories, smart shoes), wearable ubiquitous healthcare monitoring system (integrated electrocardiogram (ECG)), accelerometer and oxygen saturation (SpO2) sensors)

Table 2. Classes of constrained IoT devices-Class 0, Class 1, Class 2 [6]

Class	RAM	Flash	Description
Class 0	<1 KB	<100 KB	Use gateway for basic communication need
Class 1	Approx 10 KB	Approx 100 KB	Use protocol stack as per IoT devices using CoAP. Can interact with other devices without the need of gateway
Class 2	Approx 50 KB	Approx 250 KB	These devices support regular IPv4 and IPv6 protocol. They function similar to other network devices

Device heterogeneity could be due to multiple aspects. This study is focusing on heterogeneity in terms of:

(1) Characteristics of the device: In [9], device classification is done.

Depending on (i) memory usage and data processing capabilities and (ii) strategies for power consumption because existing management technologies use different protocol stacks and different protocol stacks consumes different amount of memory and power. Class 0 devices cannot support simple network management because they lack resources require for proper communication and they cannot support any security standards also. However, both Class 1 and Class 2 devices support both simple and intelligent gateway management approaches.

Table 3. Devices and corresponding management approaches

	Simple network management	Intelligent home gateway network management
Suitable devices	IoT devices communicate directly to cloud	Communication between device to device
	Devices which can share background data	Communication between IoT device and gateway
	Class 1 devices, Class 2 devices	Class 0 devices, Class 1 devices
	Devices which are always on	Low power devices which are normally off [7, 8]

Table 4 lists and categorizes general strategies for power usage. Low-power or normally-off devices are not recommended in direct management approach as they cannot maintain the connection with the simple management service.

(2) Communication pattern of devices: In home gateway network system, IoT devices can communicate in between or through gateway.

Table 4. Strategies of using power for communication

Name	Strategy	Ability to communicate
Class 0	Normally off	Reattach when required
Class 1	Low power	Appears connected, perhaps with high latency
Class 2	Always on	Always connected

3 Design Discussion

In the home network, devices are connected and controlled by the home gateway or directly managed by the remote management platform (RMP) running on the internet.

To manage multiple devices within a single system requires enhanced management capabilities because at home we have different IoT devices; which use sensors, different software and data collection, data analysis services and interfaces to interact with users.

3.1 Simple Network Management System

Simple management services manage IoT devices directly without any gateway. The application running remotely on internet can communicate directly with the devices. Performance and the latency introduced because of gateways will be minimized here. The majority of devices that primarily exchange real-time sensory and control data in small but numerous messages, direct management should be preferred in them, due to the aforementioned advantages.

Class 2 of IoT devices directly communicate to central servers for data storage. It supports IPv6 protocol. These classes of devices use powerful processors. They are not constrained by battery power. They also support gateway functionalities wherein they support different types of communication ports such as digital subscriber line (DSL), WiMax, WiFi etc. They support multiple sensor devices.

3.2 Intelligent Home Gateway Network Management

In IoT intelligent home gateway network management system, all the home network devices are connected to gateway and gateway is connected to internet as shown in Fig. 1. In the given diagram, every device that perform sensing and/or actuation, stores the collected data on their datastores, perform analysis and hosts the application on gateway.

3.2.1 Intelligent Home Gateway Network Management with NETCONF and YANG

There are two standard protocols for network management viz. network configuration protocol (NETCONF) [10, 11] and simple network management protocol (SNMP).

3.2.2 Introduction to NETCONF and YANG

In 2002, internet architecture board (IAB) workshop held, in that workshop it was concluded that SNMP is not suitable for configuration management. This is documented in RFC 3535. That was a point to initiate research on NETCONF and YANG.

NETCONF is an internet engineering task force (IETF) network management protocol and recorded in RFC 4741. It is a session based network management protocol. It provides mechanisms to install, manipulate, and delete the configuration of network devices. Its operations are running on top of a simple remote procedure call (RPC) layer. It uses XML based remote procedure calls for framing request and response messages. It works on secure shell transport layer protocol. It also supports block extensible exchange protocol (BEEP) which is a transport layer protocol. Transport layer provides port to port connectivity and reliable delivery of messages. It can also replace command line interpreter (CLI) based programming interfaces like perl running over secure shell (SSH). It uses structured schema driven data and provides error return information also in structured format, which even CLI cannot provide. Figure 2 shows the layered architecture of NETCONF protocol.

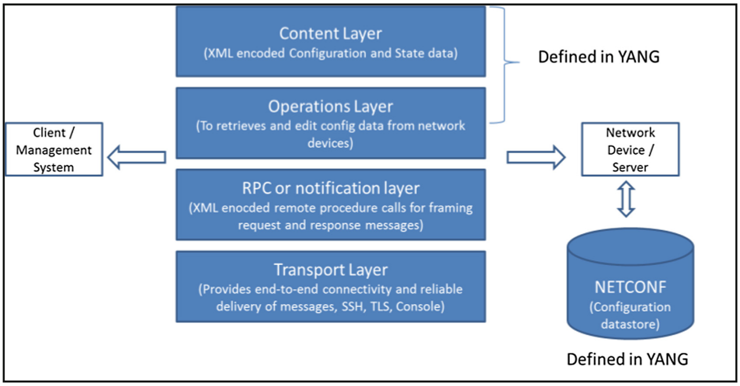

Content Layer
(XML encoded Configuration and State data)
Operations Layer
(To retrieves and edit config data from network devices)
RPC or notification layer
(XML enocded remote procedure calls for framing request and response messages)
Transport Layer
(Provides end-to-end connectivity and reliable delivery of messages, SSH, TLS, Console)
Defined in YANG
Client / Management System
Network Device / Server
NETCONF
(Configuration datastore)
Defined in YANG

Fig. 2. NETCONF protocol layers

The content layer consists of state data and configuration of each device running in home like TV, refrigerator in XML format. YANG is a data modeling language used to model configuration and state data of devices manipulated by NETCONF protocol [8, 9]. The definition of data exchanged between NETCONF client and server i.e. device and management system running on gateway. NETCONF operation <get-config> retrieves the configuration data of devices, and the operation <get> retrieves the state and configuration data of devices. On every device a NETCONF configuration datastore is running to keep configuration data. Here, the client and server maintains the NETCONF session for communication. Client manages the server (device) with 'hello' message exchange to share each other capabilities. Client can then send n number of requests to the server for retrieving and editing the configuration data. NETCONF allows management client to discover the capabilities of the device and also access of its native capabilities.

NETCONF defines on the devices one or more configuration datastores. A configuration store contains all the configuration information to bring the device from its initial state to the operational state. By default a <running> configuration store is present. Additional configuration datastores as per the device need such as <startup> and <candidate> can be defined in the capabilities.

NETCONF is a connection oriented protocol and uses SSH or transport layer security (TLS) transport protocol for providing security features on server like authentication, data integrity and confidentiality. NETCONF connection persists between protocol operations.

3.2.3 YANG

YANG is a data modeling language. It is a standard defined by the IETF in the network modeling (NETMOD) working group. YANG can said to be tree-structured. Configuration data is structured into the tree and data can be of complex type such as lists and unions. It is used to model configuration and state data manipulated by the NETCONF protocol.

YANG modules define configuration data, remote procedure calls and state data that can be issued and it decides the format of notifications also. Whatever data is exchanged between client and server, format of that data is decided by the NETCONF protocol. A YANG module comprises of number of 'leaf' nodes which are specified using the 'leaf' or 'leaf list' constructs. Leaf nodes are organized using 'container' or 'list' constructs. On data nodes constraints and data validation can also be defined. YANG module can use other modules also by introducing their definition in it. 'config' statement is used to model both configuration data and state data.

3.2.4 SNMP and NETCONF

SNMP is also widely used network management protocol which is responsible for monitoring and configuring network devices such as router, printer, scanner, server etc. Table 5 shows the comparison in between SNMP and NETCONF, and shows the suitability of NETCONF over SNMP.

Table 5. Comparison of SNMP and NETCONF protocols [12]

S. no.	SNMP	NETCONF
1	SNMP uses user datagram protocol (UDP). UDP is a transport layer connectionless protocol which makes SNMP unreliable	It uses SSH protocol. It ensures reliable delivery of messages
2	It is stateless in nature. Management application should be intelligent to manage the device	NETCONF is session-based protocol
3	Generally lacks writeable objects without which device configuration is not possible	SNMP can only be used for device monitoring and status polling while NETCONF allows to retrieve state or configuration data of network devices
4	Very difficult to differentiate between configuration and state data	<get-config> retrieve configuration data only. <get> retrieve both state and configuration data
5	It does not support easy retrieval and playback of configurations	NETCONF gives access to the native capabilities of the device
6	Latest version of SNMP providing security support is very complex	NETCONF uses SSH or TLS for security services

4 Study of Intelligent Home Gateway Network Management Through NETCONF and YANG Using Netopeer Tools

We have heard about the smart devices and at home also we have electronic devices like washing machine, dishwasher, refrigerator, etc. Initially Amazon launched Amazon dash button for washing machines. Using this button, washing machine prompts users to order more detergent for itself when it is running low. Afterward, around 40 services were launched in the form of Amazon dash replenishment service also known as DRS buttons. They are linked to the Amazon account and allow anyone in the home to instantly order staples from toothbrush heads, kitchen rolls and washing up liquid to coffee. Internet of things is big technology space and in home network also there are series of interconnected devices that look to automate tasks such as lights that respond to presence or timers or smart thermostats that save energy by only putting the heating on when people are in the house.

A design of smart washing machine device using the IoT design methodology is being discussed. The purpose of the washing machine automation system is to control the machine in a home remotely using a gateway running management system application and connected to the internet and other IoT devices running in home network. Implementation tool considered is netopeer. Netopeer is a set of open source NETCONF tools built on libnetconf library [13].

Figure 3 shows how to manage a washing machine using the netopeer tools. It includes:

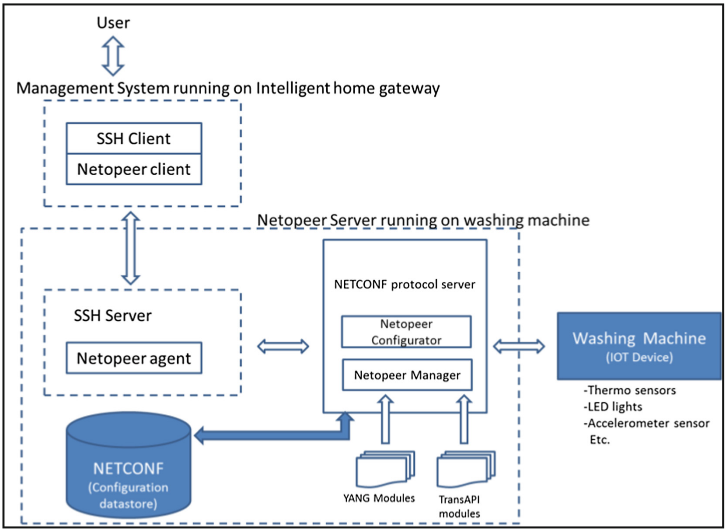


Fig. 3. Washing machine management with NETCONF using netopeer tool

1. **Netopeer-server:** Netopeer-server is a NETCONF protocol server that runs on the washing machine. It provides an environment for configuring the washing machine using NETCONF RPC operations and also retrieving the state data from the washing machine.

   On washing machine, three services are running, it includes:
   (a) Native controller service – runs as a native service on washing machine. Gets the current mode (auto/manual), current state (on/off/wash/spin) and sends to the netopeer-cli.
   (b) Mode service: Every machine includes auto and manual modes.
      i. In auto mode, system measures clothes in machine and switch it on if clothes are there.
      ii. In manual mode, the system provides the option of manually and remotely switching on/off the machine.

(c) There are four states of washing machine we are considering:
   i. On
   ii. Off
   iii. Wash (for washing clothes)
   iv. Spin (for spinning clothes).
2. **Netopeer-agent:** Netopeer-agent is the NETCONF protocol agent running as a secure shell subsystem. It accepts incoming NETCONF connection and passes the NETCONF RPC operations received from the NETCONF client (running on gateway) to the netopeer-server (running on washing machine). It is also responsible for authentication and checking integrity of message (if required) in the request message. It checks that request message is for washing machine only and checks the syntax of message.
3. **Netopeer-cli:** It is a NETCONF client that provides a command line interface for interacting with the device running netopeer-server. The operator can use the netopeer-cli from the gateway management system to send NETCONF RPC operations for configuring the washing machine and retrieving the state information.
4. **Netopeer-manager:** Netopeer-manager allows managing the YANG and libnetconf transaction API (TransAPI) modules on the washing machine. With netopeer-manager modules can be loaded or removed from the washing machine.
5. **Netopeer-configurator:** Netopeer-configurator is a tool that be used to configure the netopeer-server.

4.1 Steps for Managing Washing Machine with NETCONF-YANG

i. Create a YANG module of the washing machine in home management system that defines the configuration and its state data on its hierarchical tree structure [14].

ii. Compile the YANG model with the ‘inctool’. Inctool is part of libnetconf library. Whatever the changes done in the configuration file of washing machine, to reflect those changes in actual washing machine device, TransAPI framework is used. The ‘Inctool’ generates the TransAPI module (callbacks C file) and the YIN file. The callbacks C file contains the functions for making the change on the washing machine. YIN file contains an XML representation of the YANG module.
   - Inctool – model washingmachine.yang convert
   - Inctool – model washingmachine.yang validation
   - Inctool – model washingmachine.yang transapi – paths washingmachine.paths

iii. Fill in the IoT device management code in the transaction API module also known as callbacks C file. This file comprises of configuration callbacks, RPC callbacks and state data callbacks.

iv. Below commands are issued to build the callbacks C file as a result generate (.so) library file.
   - Autoreconf
   - ./configure
   - make

v. Netopeer manager tool loads the YANG module and .so binary generated by TransAPI into the washing machine.
   - Sudo netopeer - manager add – name washingmachine – model washingmachine.yin – datastore/home/ubuntu/washingmachine.xml

vi. The user can now connect from the management system running on gateway to the netopeer server using the netopeer-cli.
   - netopeer-cli
   - netconf> connect
   - Host: localhost
   - Password:

vii. User can issue NETCONF commands from the netopeer client, CLI. Commands can be issued to change the configuration data of washing machine, get operational data or execute an RPC on it.
   - netconf> get
     ..
     ..
     - netconf> get-config running
     - netconf> edit-config running
       ..
       ..
   - netconf> user-rpc.

5 Conclusion and Future Directions

This paper discusses IoT system for intelligent home gateway network management system to handle different IoT devices in the network. Complete design of IoT devices, their software and services running on them are discussed. Here, the study is done by choosing washing machine device from home network system and how it can be managed using the architecture involving NETCONF protocol, YANG data modeling language and netopeer tools. With this architecture, we can manage multiple device classes especially devices which are of limited resource capability like low power, slow processing, and less communication capability. NETCONF protocol manages the operation and configuration of devices in home network in a reliable and efficient manner as it is session based thus reduces the traffic to the home gateway and also easily retrieves states and configuration of devices. As future research, some work can be done around introducing artificial intelligence into the IoT devices of the system.

References

1. Pham, C., Lim, Y., Tan, Y.: Management architecture for heterogeneous IoT devices in home network. IEEE (2016)
2. Rao, S., Chendanda, D., Deshpande, C., Lakkundi, V.: Implementing LWM2M in constrained IoT devices. In: ICWiSe, pp. 52–57 (2015)

3. Caldeira, J.M.L.P., Rodrigues, J.J.P.C., Lorenz, P.: Toward ubiquitous mobility solutions for body sensor networks on healthcare. IEEE Commun. Mag. **50**, 108–115 (2012)
4. Chung, W.Y., Lee, Y.D., Jung, S.J.: A wireless sensor network compatible wearable U-healthcare monitoring system using integrated ECG, accelerometer and SpO2. In: International Conference of the IEEE Engineering in Medicine and Biology Society (2008)
5. Bahga, A., Madisetti, V.: Internet of Things, A Hands on Approach (2015)
6. ITU-T: Overview of the Internet of Things, Y.2062 (2012)
7. Ersue, M., Romascanu, D., Schoenwaelder, J.: Management of networks with constrained devices: problem statement and requirements, RFC 7547 (2015)
8. Bormann, C., Ersue, M., Keranen, A.: Terminology for constrained-node networks, RFC 7228 (2014)
9. Sehgal, A., Perelman, V., Kuryla, S., Schonwalder, J.: Management of resource constrained devices in the internet of things. IEEE Commun. Mag. **50**, 144–149 (2012)
10. Enns, R., Bjorklund, M., Bierman, A.: Network configuration protocol (NETCONF), RFC 6241 (2011)
11. Schönwälder, J., Björklund, M., Shafer, P.: Network configuration management using NETCONF and YANG (2010)
12. Harrington, D., Preshun, R., Wijnen, B.: An architecture for describing simple network management protocol (SNMP) management frameworks, RFC 3411 (2002)
13. libnetconf (2014). https://github.com/CZ-NIC/libnetconf
14. Tschofenig, H., Arkko, J., Thaler, D., McPherson, D.: Architectural considerations in smart object networking, RFC 7452 (2015)

DNA Based Cryptography to Improve Usability of Authenticated Access of Electronic Health Records

C. S. Sreeja[1(✉)], Mohammed Misbahuddin[2],
and B. S. Bindhumadhava[2]

[1] Christ University, Bangalore 560029, Karnataka, India
sreejasukumaran@gmail.com
[2] Centre for Development of Advanced Computing (C-DAC),
Bangalore 560100, Karnataka, India
mdmisbahuddin@gmail.com, bindhu@cdac.in

Abstract. The quality of health care has been drastically improved with the evolution of Internet. Electronic health records play a major role in interoperability and accessibility of patient's data which helps in effective and timely treatment irrespective of the demographic area. The proposed model is to ensure and monitor maternal health during pregnancy and to create awareness alerts (options include messages, voice alerts or flash the system) based on the individual health record. The system aims to prevent maternal death due to medical negligence and helps to make recommendations to prevent future mortality based on medical history and take appropriate action. Authentication is a critical aspect considering the trade-off between usability and security whereas data breach and related cybercrime are major concerns in health care. The proposed model uses DNA based authentication techniques to ensure usability and confidentiality of electronic data, Aadhaar to prevent unauthorized access to patient's data in case of emergency without affecting availability.

Keywords: DNA Cryptography · Confidentiality · Usability · Aadhaar
EHR · Data breach · Authentication · M-health

1 Introduction

India as a developing nation health care is also an important domain to be taken care and it needed special care as it involves confidential and sensitive data of patients. Manipulation of health records is always a matter of life and death. Recently death related to medical negligence has become common news, especially during pregnancy and childbirth. This not only affects the citizen's trust in hospitals but also affects the reputation of doctor's community. As per World Health Report ranking India's healthcare system is at 112 out of 190 countries [1, 2]. So, it is high time to implement a proper health care management system to improve the quality of life of citizens and to reduce infant and maternal mortality rate. A proper and systematic medical data recording system based on individual health record will help in reducing the infant

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 208–219, 2018.
https://doi.org/10.1007/978-3-319-73423-1_19

mortality rate (IMR) and maternal mortality rate (MMR) also helps in minimizing medical malpractices.

The scope of the study includes improving the quality of health care system in India which has key issues like maternal mortality, birth defects etc. by providing an Electronic health record (EHR) and a usable authentication system as the digital data plays a vital role. Digital society is in store for near future for all the countries across the globe and it has a high impact as the online users are increasing day by day. Most of the organizations are globally connected leading to economic growth whereas few domains such as healthcare are not exploiting the online facilities apart from few private hospitals as the data breach and confidentiality issues are major inhibitors.

This paper proposes a simple and usable health care system which concurrently maintains EHR and prevents unauthorized access to it by using DNA based encryption techniques. Aadhaar, 12-digit unique identification number issued by the Unique Identification Authority of India [3] is also incorporated into the proposed EHR and related authentication system as it helps and plays a major role for unique identification of a citizen. The proposed system will be beneficial for m-health based services as the system ensures security without using complex algorithms and the model is Universal and can be implemented in the countries where a unique national ID is used for residents. Linking DNA and EHR has various benefits as the future is of personalized medicine, (e.g. Pharmacogenomics) and real DNA can be used for authentication of EHR. Next section explains the existing EHR systems across the world, merits and demerits, usage of internet connectivity and mobile and its impact on health care.

2 Electronic Health Records Welfare, Challenges and Existing Systems

EHR plays an important role considering the quality of care, patient safety and on time treatment whereas information security issues and data breach related to the health care industry are major concerns. Nir Menachemi and Taleah H Collum gives details on the benefits and drawbacks of EHR systems. In the paper, the authors highlight the potential benefits of EHR and related uses in clinical, organizational and societal and outcomes. EHR systems have many capabilities but main functionalities which improve the quality of health care are Clinical decision support (CDS) tools, Computerized physician order entry (CPOE) systems and Health information exchange (HIE).

All the three criteria mentioned above are of "meaningful use" set forth in the HITECH Act of 2009 [4]. Benefits of using EHR also includes reducing malpractices, increased availability of data which in turn helps in clinical analysis and helps in providing best practices. The major drawbacks for the adoption of EHR are security and privacy concerns whereas financial issues and changes in the workflow are also considered [5].

Most of the countries are shifting towards digitization of data in every field whereas health care is a domain which has more benefits of digitization. Digital health data provides availability and interoperability of health information which is essential during an emergency as it ensures timely treatment and care and helps to reduce mortality. Setting up digital databases will help in HIE at the same time ensuring security and privacy is a major concern. The majority of the private hospitals are using the benefits of digitization whereas government sectors are still inert to tackle the situation.

The majority of the developed countries have well-defined Health Care System and it varies from nation to nation depending on the economic development and available resources. EHR implementation around the world is reviewed by the author in [6] and gives a quick look of adoption of EHR and related policies across various countries - The United States, The United Kingdom, France including the initiatives by India and claims that the EHR system in India does not have adequate security and privacy. Singapore has one of the most successful health care systems in the world-National Electronic Healthcare Record (NEHR) and has been rated as most efficient in the world by Bloomberg in 2014 [7].

A mobile based application - MOTHER (MObile based maTernal HEalth awaReness is a good initiative developed by Centre for Development of Advanced Computing (C-DAC), India [8]. It is a scheduler for sending customized alerts on maternal and child health related information straightway to the mobile phones of the pregnant and lactating women as voice calls in regional languages. The government of India has taken an initiative to define India's Healthcare future by developing an integrated Health Information system – National eHealth Authority (NeHA) proposed by Ministry of Health and Family Welfare. It will also be responsible for enforcing the laws and regulations relating to the privacy and security of the patient's health information and records. Vision of the program is

- To facilitate the integration of multiple health IT systems through health information exchanges.
- To ensure that security, confidentiality, and privacy of patient data.
- To prepare documents relating to architecture, standards, policies and guidelines for e-Health stores, National Health Information Network (NHIN) and HIE [9].

Population from the rural areas will be more benefited by this initiative as it will promote telemedicine and m-health services. But concurrently data security issues persist and there is a need for secure but less complex techniques considering the fact most of the population rely on mobile based systems [10].

Internet usage statistical reports claim India as the second largest online market with 460 million internet users [11] which is a clear demonstration of growing internet usage. Internet penetration rate of India as per 2016 is 34.8% [12]. World Bank measured Indian rural population as 67.25% as per 2015 report [13], which explicitly indicates most of the people lives in rural areas. Digital India campaign was initiated for setting up broadband services in rural India. Considering Internet connectivity and mobile usage in rural area m-health services will be more usable.

Defining a digital health care system which is secure but with less complexity will be efficient and effective considering the usability of the public. Usability and Security always conflicts but in EHR a balance between both are important. Password-based authentication is still prominent and dominant because of easy to use, considering these facts the proposed system is a password-based authentication using DNA cryptography which is easy to use for both patients and doctors. Instead of using complex algorithms, simple but secure algorithms will be more reliable especially if the EHR is accessed using mobile and here comes the importance of DNA encryption techniques, which provides security based on biocomputations. Next section gives a review on DNA cryptography, basics of DNA, structure, DNA computing and the digital representation of DNA.

3 DNA Cryptography

Data breach and security issues are major drawbacks resisting the adoption of electronic data on health care as most of the organizations are concerned about the security of patient's electronic data. Encryption and Authentication techniques can play a major role in securing EHR but usability and data availability must be considered. DNA based encryption techniques are the recent branch in Cryptography and it's getting wide acceptance due to the vast parallelism and computational complexity. Computational properties of DNA were first introduced in 1994 by Dr. Leonard M. Adleman of the University of South California to solve the complex computational problem of Mathematics [14].

3.1 DNA

DNA is the genetic blueprint of all living cells. In 1869 DNA was first identified by Swiss physician Friedrich Miescher. DNA is a double-stranded helix of nucleotides. A DNA sequence consists of four nucleic acid bases A (adenine), C (cytosine), G (guanine), T (thymine), where A and T, G and C are complementary [15]. Figure 1 represents nitrogenous bases of DNA [16]. Figure 2 depicts the pictorial representation of the helical structure of DNA [17].

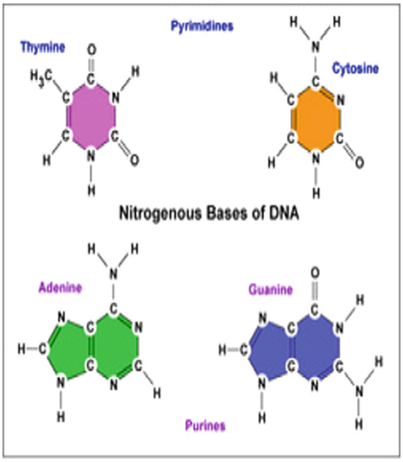


Fig. 1. Nitrogenous bases of DNA

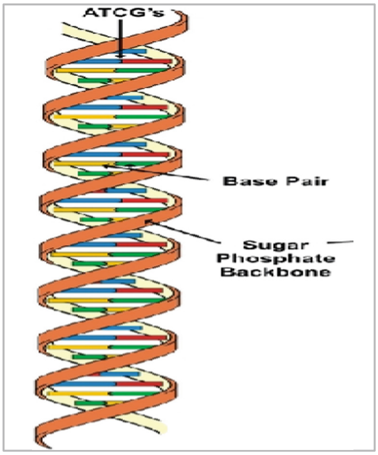


Fig. 2. Helical structure of DNA

3.2 DNA Computing

DNA Computing allows to code A, G, C, T in binary form and this digital representation of the DNA sequences formed the basis of DNA based encryption. Table 1 represents the binary coding of DNA. DNA bases A, G, C and T can be represented as 4! = 24, which adds computational complexity to the sequences if used with biomolecular concepts.

Table 1. Binary representation of DNA bases

A	00
G	01
C	10
T	11

DNA based cryptographic techniques can be used for securing information as it ensures security by means of biocomputations. DNA based encryption methods are widely classified as [18] Symmetric DNA cryptography, Asymmetric DNA cryptography, Pseudo DNA cryptography and DNA Steganography.

Conventional cryptographic techniques used in hybridization with DNA encryption enhances security. DNA based authentication has more applications, especially in EHR. In [18] authors proposed Image and DNA based authentication which ensures usability and security. DNA authentication can be performed using a wet lab or by using digital sequences which are available in the databases in billions. DNA Steganography can be used for transferring confidential medical data especially using Image based Steganography which plays a vital role if medical images are used.

4 Proposed Methodology

The proposed model is to develop an EHR which is a pregnancy registry and tracking system for pregnant women by linking Aadhaar. The purpose of using Aadhaar card is to provide authorized access to patient's data for doctors. The system will have patient's portal where the patient can log in through two-step verification. It also has doctor's portal where registered doctor can login into the database using hospital ID and his own user ID and view patient's health records by using patients Aadhaar. Aadhaar has a QR code which has details of patient embedded in it and can be read by a simple barcode reader or by entering Aadhaar number.

The system aims to protect data as each patient would like to maintain their information privacy. The mother can sign up once pregnancy and estimated due date is confirmed. Based on registered user's EHR and demographic dividend options such as message alerts, voice alerts or flash the system can be used to connect with mothers and share the information related to infant and pregnancy care. The alerts are evoked based on individual's health data and pregnancy stage, emergency numbers can also be shared with expecting moms. Users can rate or raise complaints through the system after each medical consultation. The information registered will be secured in the centralized database using DNA-based authentication techniques. The centralized database system impacts on reducing IMR and MMR and the mortality due to medical negligence as every act and events are updated in the database and these records also help patients to sue for the medical malpractice.

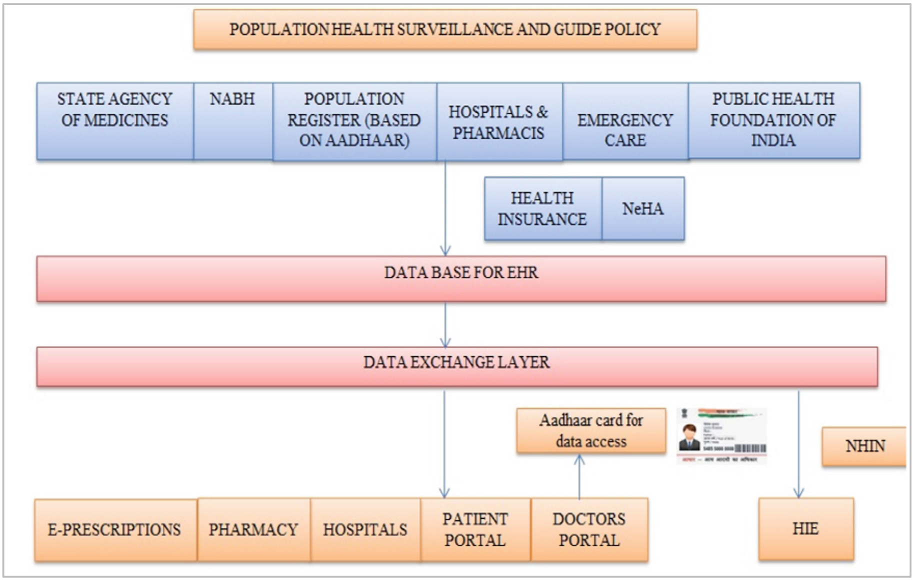

POPULATION HEALTH SURVEILLANCE AND GUIDE POLICY
STATE AGENCY OF MEDICINES
NABH
POPULATION REGISTER (BASED ON AADHAAR)
HOSPITALS & PHARMACIS
EMERGENCY CARE
PUBLIC HEALTH FOUNDATION OF INDIA
HEALTH INSURANCE
NeHA
DATA BASE FOR EHR
DATA EXCHANGE LAYER
Aadhaar card for data access
NHIN
E-PRESCRIPTIONS
PHARMACY
HOSPITALS
PATIENT PORTAL
DOCTORS PORTAL
HIE

Fig. 3. Proposed architecture for health care system in India

Figure 3 is the architecture proposed for Health Care System in India by including the fields which are essential for the functioning of the centralised database system. It includes National health information network (NHIN), National Accreditation Board for Hospitals & Healthcare Providers (NABH), National health information network (NHIN) and Public health foundation of India. The system can be used for maternal health care and this can be extended for general healthcare in India (Table 2).

Table 2. Notations used in the algorithm

U	User
ID_U	User ID
AID_U	Aadhaar number of ID_U
pwd_{U1}	User selected password during registration
pwd_{U2}	User selected password for security question during registration
AS	Authentication Server
UDS	Unique DNA sequence selected for the user by AS based on AID_U
Pwd_N	New password selected during password change phase

For registration process the patients can take support from Primary health care (PHC) agent or officer and once the registration process is completed the user can view her health record using Aadhaar and OTP received or the details can be shared with the PHC agent to view the health records. If the user wants to view or edit the records the proposed authentication methodology can be used to login and edit the details. Doctors can log in to the EHR portal using his ID and hospital ID and view the patient details using patient Aadhaar number, these will help in case of emergency to access the EHR. The proposed authentication model has three phases – Registration phase, Authentication Phase and password change phase.

4.1 Registration Phase

R1: In the interface or EHR portal two options will be available for Registration.

a. Register as a Patient Login. b. Register as a Doctor's Login
User, Mother can register as a patient by entering essential credentials including $AID_{U.}$

R2: AS allows the user to select a valid ID_U and pwd_U.
R3: AS allows the user to select a security question and user can set an answer of her choice.
R4: AS picks a DNA sequence (UDS) depending on user credentials and $AID_{U.}$ This act as a key factor in DNA Cryptography during user authentication phase.
R5: Registration Phase completed Successfully. An EHR is created for the user, U. Figure 4 depicts the registration process.

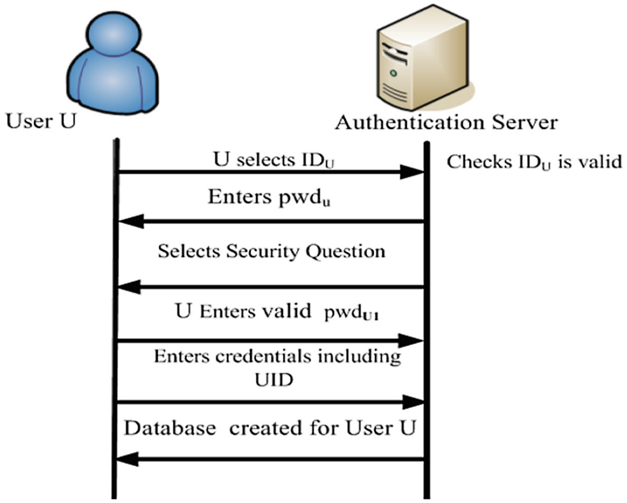

User U
Authentication Server
U selects ID_U
Checks ID_U is valid
Enters pwd_u
Selects Security Question
U Enters valid pwd_{U1}
Enters credentials including UID
Database created for User U

Fig. 4. Registration phase

4.2 Authentication Phase/Login Phase

The proposed authentication phase is a simple, secure and usable technique. This method is a two-step verification phase, but only password based verification is included to ensure usability.

A1: Interface has two options.

a. Patient Login b. Doctor's Login, User enters ID → $\mathrm{ID_U}$.

A2: U enters password → $\mathrm{pwd_{U1}}$.

A3: AS displays the security question.

A4: U enters $\mathrm{pwd_{U2.}}$

A5: AS picks the unique DNA sequence (UDS) and computes a key value based on $\mathrm{ID_U}$, $\mathrm{pwd_{U1}}$, $\mathrm{pwd_{U2}}$ and UDS.

A6: Key computation is based on DNA coding and DNA encryption and generates a hash. During authentication AS computes the key value and if it matches the key value computed during registration the server authenticates the user to the corresponding EHR system.

$$\text{Computation of key value} = \mathrm{pwd_{U1\rightarrow}}\ \text{binary} \tag{1}$$

$$= \mathrm{pwd_{U2\rightarrow}}\ \text{binary} \tag{2}$$

$$= (\mathrm{pwd_{U1}}) \oplus (\mathrm{pwd_{U2}}) \tag{3}$$

$$\text{Conversion of Eq. (3) to DNA code based on Table 1} \tag{4}$$

$$\text{DNA encryption by Eq. (4) and UDS} \tag{5}$$

Hash of cipher text generated from Eq. (5) is the key value.

A7: AS checks the key value computed for the user, U during the registration phase, if the key value matches, authentication to the system is granted else denied.

Figure 5 represents authentication phase.

4.3 Password Change Phase

In this phase user, can change the passwords and set the new passwords.

P1: User can enter the details on the interface.

P2: After Authentication phase user, can change the password, security question or both can be changed. The users who prefer additional security can also opt for one-time password (OTP) instead of the security question or as an additive layer depending on the usability of the end-user.

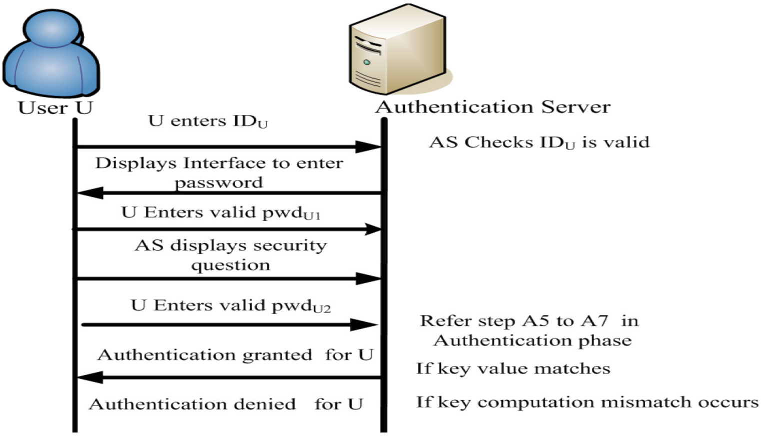

User U
Authentication Server
U enters ID$_U$
AS Checks ID$_U$ is valid
Displays Interface to enter password
U Enters valid pwd$_{U1}$
AS displays security question
U Enters valid pwd$_{U2}$
Refer step A5 to A7 in Authentication phase
Authentication granted for U
If key value matches
Authentication denied for U
If key computation mismatch occurs

Fig. 5. Authentication phase

5 Security Analysis

The proposed system is simple and usable and can be used by common people without any technical background, considering this fact there is no two-factor verification involved but a two-step password based verification is included for usability. But information security of the system must be maintained for this DNA encryption and hash computations are performed at the server side which ensures security. This section gives security analysis of the proposed system.

5.1 Attack on Password File

The proposed scheme resists password file compromise attack as the password file saved in the database is a computed key value based on hash algorithm as a final output using inputs as:

Binary $((\mathrm{pwd}_{U1}) \oplus (\mathrm{pwd}_{U2})) \rightarrow$ Eq. (1)
DNA Sequence Conversion (Eq. (1)) → Eq. (2)
DNA encryption (Eq. (2) and UDS) → Eq. (3)

Hash Eq. (3) → Eq. (4) → output. If the password file is compromised it will be difficult to extract the passwords and enter the system.

5.2 Attack on DNA Sequence

DNA encryption is based on DNA coding, which can be represented as 4! = 24, DNA coding in Table 1 represents one of the digital representation of DNA. The sequence selected by the server will be difficult to guess considering the availability of millions of DNA sequences in the digital databases and the selection criteria used for DNA reference sequence. It is also possible to generate random sequences, say n which resists the possibility of reference sequence attack.

5.3 Attack on Unique ID

Availability of the Aadhaar or a unique ID of a user to an unauthorized person will not provide access to the EHR. To view the EHR details the knowledge of both passwords are vital. During the registration process users also registers their mobile number for receiving message alerts, voice alerts and to the same number the user receives alert message or notification whenever her EHR is accessed and alerts the valid user about unauthorized attempt to access the data and user can take appropriate action to prevent data access. An option to add OTP ensures additive security and in that scenario, the unauthorized person must have access to all three values.

6 Merits of Aadhaar Usage

Use of Aadhaar in the protocol has potential benefits especially during emergency cases where including the availability of health records and time plays a critical role. If the doctor has the knowledge of patient's Aadhaar he can retrieve the users EHR using his login and hospital login credentials which help in providing effective and timely treatment. Aadhaar will also help in HIE as the unique ID is valid and unique across India. Implementation time of the proposed system can be reduced by using Aadhaar card rather than introducing a new health card system or a smart card for 1.324 billion people, India's huge population.

7 Proof of Concept

This section demonstrates the working of proposed protocol. This section only describes and discusses on the key computation of authentication server for the user U.

Step 1: The user after selecting her valid Id selects the pwd_{U1} as "myehr".

Step 2: AS displays list of security questions, sample question "who is your best friend?" The user enters (pwd_{U2}) as "bob".

Step 3: AS selects a random DNA sequence based on user credentials. DNA sequence selection can be done based on digital sequence, random sequence or from user data. In this study, a sample random sequence of size 300 base pairs (bp) [19] has been generated. Random DNA Sequence generated is:

CTGGTACATTATGTGAACAATGTTCTGAAGAAAATTTGTGAAAGAAGG
ACGGGTCATCGCCTACTATTAGCAACAACGGTCGGCCACACCTTCCATTGT
CGTGGCCACGCTCGGATTACACGGCAGAGGTGCTTGTGTTCCGACAGGCTA
GCATATTATCCTAAGGCGTTACCCCAATCGTTTACCGTCGGATTTGCTATA
GCCCTGAACGCTACATGTACGAAACCATGTTATGTATGCACTAGGTCAACA
ATAGGACATAGCCTTGTAGTTAACACGTAGCCCGGTCGTATAAGTAC

Step 4: Key value computation mainly includes 3 stages

(a) Converting passwords to binary and performing XOR the value generated 110110101111001011001010110100000010000001100010011 01000 → (1), converting (1) into DNA coding based on Table 1.

(b) The value generated is: TGCCTTACTACCTGAAACAATAGATGAA → (2). A parity bit is added to correct the conversion.

(c) DNA Encryption: The encrypted password which is in DNA form is camouflaged into Random DNA sequence to generate:

CTGGTACATTATGTGAACAATGTTCTGAAGAAAATTTGTGAAA
GAAGGACGGGTCATCGCCTACTATTAGCAACAACGGTCGGCC
ACACCTTCCATTGTCGTGGCCACGCTCGGATTACACGGCAGAG
GTGCTTGTGTTTGCCTTACTACCTGAAACAATAGATGAACCGA
CAGGCTAGCATATTATCCTAAGGCGTTACCCCAATCGTTTACC
GTCGGATTTGCTATAGCCCCTGAACGCTACATGTACGAAACCA
TGTTATGTATGCACTAGGTCAACAATAGGACATAGCCTTGTAG
TTAACACGTAGCCCGGTCGTATAAGTAC →(3),Cipher sequence.

(d) Computing Hash: Hash of the encrypted sequence is computed using SHA-2. 843a81d0051ede0a3a4b1affc7e1a9f3589d91db29048a7059e 74f0a57f11022 → (4)

8 Conclusion and Future Work

In this paper, a methodology has been proposed to reduce the IMR and MMR rate in India which also reduces the medical malpractice related to childbirth and pregnancy as EHR helps to track every records and event. The healthcare system can be integrated for general health care to ensure better treatment. The proposed health care architecture can be integrated for general healthcare system by combining with regulatory agencies for better performance. The work also focuses on the usable authentication of the system by a novel DNA based authentication which uses DNA encryption and unique ID- Aadhaar considering patients usability and data security.

Future work is to propose a single sign-on mechanism for securing electronic health records using Aadhaar by integrating maternity and infant care, health insurance policies, national immunization registry, HIE etc. into a single platform and connect to a centralized database. DNA based authentication techniques can be used with conventional techniques for a secure environment.

References

1. Jayaraman, V.R.: 5 Things to know about India's Healthcare System. http://forbesindia.com/blog/health/5-things-to-know-about-the-indias-healthcare-system/#ixzz3S3WIt74N. Accessed 11 Sept 2014
2. Srinivisan, R.: Health Care in India-Vision 2020, vol. 1. Government of India, Planning Commission of India, New Delhi (2010)
3. What is Aadhar Card – Its Uses, Benefits and Why You Should Have it! http://www.aadharcardkendra.org.in/what-is-aadhar-card-benefits-uses-1424/
4. Blumenthal, D., Tavenner, M.: The meaningful use regulation for electronic health records. N. Engl. J. Med. **363**(6), 501–504 (2010)
5. Menachemi, N., Collum, T.H.: Benefits and drawbacks of electronic health record systems. Risk Manag. Healthc. Policy **4**, 47–55 (2011)
6. Stone, C.P.: A Glimpse at EHR Implementation Around the World: The Lessons the US Can Learn. The Health Institute for E-Health Policy, May 2014
7. Wee, Y.H., Zhou, Y., Tayi, G.K.: IT-enabled healthcare integration: the case of National Electronic Health Records in Singapore. In: PACIS (2015)
8. eIndia 2012 Award to Mother Project. https://cdac.in/index.aspx?id=aboutus_mother_award
9. Concept Note- National eHealth Authority (NeHA). https://www.mygov.in/sites/default/files/master_image/NeHA%20Concept%20Note%20Eng.pdf. Accessed 16 Mar 2015
10. Statistics and facts on Internet Usage in India. https://www.statista.com/topics/2157/internet-usage-in-india/
11. Rural population (% of total population) in India. http://www.tradingeconomics.com/india/rural-population-percent-of-total-population-wb-data.html
12. India internet users. http://www.internetlivestats.com/internet-users/india/
13. The World Bank, Rural population (% of total population). http://data.worldbank.org/indicator/SP.RUR.TOTL.ZS
14. Adleman, L.M.: Molecular computation of solutions to combinatorial problems. Science **266** (5187), 1021–1023 (1994). AAAS-Weekly Paper Edition
15. Sreeja, C.S., Misbahuddin, M., Mohammed Hashim, N.P.: DNA for information security: a survey on DNA computing and a pseudo DNA method based on central dogma of molecular biology. In: International Conference on Computer and Communications Technologies (ICCCT), pp. 1–6, 11–13 December 2014. https://doi.org/10.1109/iccct2.2014.7066757
16. Structure of DNA. http://geneticsk8vaneckv.weebly.com/structures-of-dna.html
17. DNA: Definition, Structure & Discovery. http://www.livescience.com/37247-dna.html
18. Misbahuddin, M., Sreeja, C.S.: A secure image-based authentication scheme employing DNA crypto and steganography. In: Proceedings of the Third International Symposium on Women in Computing and Informatics. ACM (2015). https://doi.org/10.1145/2791405.2791503
19. Random DNA Sequence Genenartor. http://www.faculty.ucr.edu/~mmaduro/random.htm

Design and Safety Verification for Vehicle Networks

Debasis Das(✉) and Harsha Vasudev

Department of CS&IS, BITS Pilani K.K. Birla Goa Campus,
Zuarinagar 403726, Goa, India
{debasisd,p2016411}@goa.bits-pilani.ac.in

Abstract. There is a serious mismatch between the growing traffic volume and the availability of resources to support the traffic. Some of the important reasons for this mismatch are the rapid development of our economy, increased affordability of our society, multiple vehicles per family, and so on. We believe that the mismatch will continue to grow and adversely affect our traffic infrastructure unless efficient traffic management solutions that include system integration, design, prediction, safety verification, validation, and security are developed and deployed. Security has appeared as an important issue for Intelligent Transportation Systems (ITS). Some security threats become more challenging task with the emergence of Vehicle-to-Vehicle (V2V) communication and Vehicle-to-Roadside (V2R) communication in vehicular networks. Addressing the security issues in smart vehicular communication systems requires new effective and efficient algorithm that encompass considerations of new security techniques, safety things, communication related resource limitations, and other related new performance metrics. In this paper, we recommend a unified framework and new metrics that combines integrated modelling, system integration and optimization, official certification and validation, and automatic synthesis approaches for analysing the security and safety of ITS and booming out design space investigation of both in-vehicle electronic control systems and vehicle-to-vehicle communications. This integrated framework will facilitated the system integration and optimization and enable validation of various design the new metrics for vehicular networks such as timing, direction, reliability, speed, security and performance.

Keywords: Vehicular networks · V2V · V2I · Safety verification

1 Introduction

The Intelligent Transportation Systems (ITS) [1] for smart cities holds promise for a sustainable, effective, cost effective and a balanced solution for Vehicles at road, highway and intersection [2]. ITS is basically the use of computer and communications technologies coming in support of the transportation problems. They enable assembling of data or intelligence and then give timely response

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 220–232, 2018.
https://doi.org/10.1007/978-3-319-73423-1_20

to traffic supervisors and road-users. The objective of vehicular networks [3,4] is to provide an efficient, effective, fast and safe interchange of safety related information such as obstacle warnings or lane change notices or the communication with road side units for the purpose of traffic information and infotainment applications is based on the smart transportation.

As vehicles can easily be equipped with positioning capabilities and wireless transceivers [5,6], they provide a suitable platform for geographic routing protocols [7,8]. Primary application areas of these technologies are vehicular networks, society application and military communications. A real-world implementation of an efficient communication and routing protocol was used for inter-vehicle communication in Vehicular Ad hoc Networks (VANETs) [17,18,20]. In recent years, the improvement of intelligent transportation for next generation smart city communication has achieved great leaps in the field of information and communication technology. Various applications have been developed to improve vehicle safety (e.g., collision avoidance, emergency warning message, weather condition warning, traffic condition warning and road condition warning) and traffic efficiency (e.g., speed management and cooperative navigation). Most of all applications are constructed through vehicular communication networks [18–20], where vehicles transmit and receive security acute data such as speed, acceleration and distance, as well as general information such as traffic and weather, so that drivers or vehicles themselves may respond to various situations in a timely manner.

The architecture of the smart vehicle for smart city communication has became more heterogeneous. The different types of sensors such as Light Detection And Ranging Light (LIDAR), radar, Global Positioning System (GPS), smart phone and cameras are used to collect the environmental information and fast and efficient computation units includes the high speed processors, Graphics Processing Units (GPUs), and Field Programmable Gate Arrays (FPGAs) are used to process by the sensors (i.e., Internal Sensors and External Sensors) and make driving decisions based on the collected data from real environment [3]. The sensors, high speed computation units (i.e., processors) and actuators are connected to a heterogeneous bus systems where the multiple buses are such as Controller Area Network (CAN), FlexRay, Local Interconnect Network (LIN) and Ethernet connected through the gateways [13–15]. However, a large amount of highly complex data generated by the intelligent sensors needs to be processed at real time, and the workload could vary significantly due to the changing environment while driving because vehicular networks means highly dynamic nature of mobile nodes and constantly changing the environments.

A hybrid framework that integrates the hybrid modelling, system integration and optimization, official certification, validation and automatic production techniques for analysing the security and safety of smart transportation and booming out the design space investigation of both intra vehicle electronic control systems and inter vehicle communications for smart city communications. Some examples of intelligent transportation for next generation smart communication include innovative traffic management systems, innovative traveller information systems, innovative vehicle control systems, innovative electronic toll collection systems, innovative public transportation systems, etc.

The rest of the paper is organized as follows. Section 2 describes the summary of the Related Works. Section 3 explains the proposed Model for Smart Communications for Smart Cities. Section 4 illustrates the Solutions for Next Generation Automotive/Semi Automotive Systems. Section 5 gives the New Parameter for Autonomous Driving. Section 6 gives the Proposed Framework for Next Generation Automotive/Semi Automotive Systems. Section 7 illustrates the Limitations and Possible Solutions. Finally we have presented the Conclusion in the Sect. 8.

2 Related Work

In the literature review, several existing automated design space exploration techniques have been proposed to address the system integration challenge for autonomous, semi-autonomous and human driven vehicle for smart cities. Zheng [1] proposed an architecture modelling and exploration outline for assessing various software and hardware architecture possibilities. The proposed system framework allows system integration and optimization. They have also mentioned that this system allows the authentication of different design parameters such as security, reliability, timing etc. The main challenges mentioned are prediction, verification, validation, system integration etc. Zheng [3] proposed a unified framework that combines hybrid modelling, formal verification, and automated creation methods for analysing the security and safety of transportation systems. They have mentioned that all modern vehicles, more precisely the most safety acute components are attacked now a days, so in order to address these issues some broad approaches like, resource constraints, safety properties, security mechanisms and related some system metrics should be considered.

In [5], Schatz et al. proposed a technique of using constraint-based formulation for automotive optimal software and hardware design, under the requirements of ISO26262 standard for automotive safety. In [6], Oetjens discuss the advantages and research challenges of virtual prototypes that can virtually mix motorized software and hardware and do design space examination and system certification. In [7], Eberl further combine firmware related functionalities like diagnostic tests into a holistic design space examination structure for motorized electrical/electronic (E/E) design for automotive networks. In [8], the Yu propose a model-based formal structure for motorised control software in terms of challenges interoperability. In recent work, the authors in [9] propose an early model-based design and verification framework for automotive control system software. In [21], Petrenko introduces the model based testing technique and envision the future test generation tools with "tester-in-the- loop" support. In [22], Krishna introduces a formal model for analysis the automotive feature production lines that is capable of capturing variability and real-time behaviour. In the literature, there are a lot of work based on enhancing automotive software control reliability [12]. In [1], the authors proposed a structure to check and verify temporary errors for motorised security acute applications. In [3], the authors study and discuss the faults during the start up and operation of a FlexRay network, and propose a bus guardian.

3 Proposed Model for Smart Communications for Smart Cities

Autonomous/semi-autonomous/human driven vehicles have the ability to percept the environment and make driving decisions without human intervention or partially or fully human intervention for smart city communication. The basic methods of a typical autonomous/semi-autonomous/human driven vehicle include Awareness (i.e., perception), efficient route planning, innovative behavioral executive and innovative speed control. The awareness method is to collect and fuse internal and external information from various sensors (i.e., internal and external sensors) and perform works such as efficient obstacle detection, road/highway/intersection shape estimation and localization (locate the relative position of the vehicle to the road using GPS). The efficient route planning method is to generate high level route to fulfill the travel mission while considering travel time, distance, speed, direction and traffic condition at road/highway/intersection. The behavioral executive methods is to decide driving behaviors such as lane change at simple highway, intersection behavior means direction changing and parking lot behavior, based on the traffic information and internal vehicle information from the awareness method, while following the navigation from the efficient route planning method. The speed/velocity control method is to physically execute the behavior generated from the behavioral executive method by controlling the actuators such as steering, acceleration and braking. Besides the basic functions, some systematic services may be needed including efficient jobs management, efficient and effective communication services for smart city communication, system configuration for smart vehicles, and fault/error handling for vehicular networks, etc. With the emerging vehicular network (e.g., the dedicated short range communications technology), Autonomous/semi-autonomous/human driven vehicles are able to communicate with each other through the V2V or V2R and exchange important information with the roadside units or other vehicle they are with in the range of the sender vehicles. This can further help the planning and control methods to make driving decisions and efficient communication for smart cities.

Traditional buses like CAN are reaching their boundaries as the increasing volume of data processed by Autonomous/semi-autonomous/human driven vehicles requires increasing bandwidth and scalability of the systems. Fully switched Ethernet is presented in the Autonomous/semi-autonomous/human driven domain to be a good candidate for Autonomous/semi-autonomous/human driven driving technologies.

4 Solutions for Next Generation Automotive/Semi Automotive Systems

The new movements of autonomous/semi-autonomous vehicles post challenges to the design of next generation fully automotive or semi automotive system integration, design, prediction, verification and validation to analysis and optimization of various design metrics such as reliability, security and performance (i.e., throughput, end-to-end delay, and fault tolerance). The fully automotive or semi automotive systems will be equipped large number of heterogeneous components such as sensors, GPS, actuators, buses, and computation units. The fully automotive or semi automotive system integration have to process a high traffic volume of data using intelligent algorithm for smart transportation.

4.1 Hybrid Systems for Autonomous/Semi-autonomous Vehicle

The Next Generation autonomous/semi-autonomous driving applications require large number of different useful units to run concurrently, while ensuring several timing and vehicular resource limitations to be met. In our propose hybrid architecture, each job inside the dissimilar functional component needs to finish execution within its deadline, and the end-to-end latency from sensor to actuator must not exceed certain threshold value.

4.2 Safety Verification and Validation for Smart Vehicle

Safety is considered as particularly significant in Next Generation Automotive/Semi Automotive System [16,17] design. To guarantee safety, the hybrid system should be unified and verified through a holistic structure that confirms proper operation and execution of the complex software units on a heterogeneous platform.

4.3 Analysis and Optimization of New Design Metrics

Analysis and optimization of new design metrics (e.g., timing, control performance, reliability, security, fault tolerance, throughput, end to end delay, and energy consumption) also post great challenges to fully automotive or semi automotive design process.

Reliability. Due to the non-stop scaling and lower power of integrated circuits, and the radiation from the environment, next-generation fully automotive or semi automotive systems are more prone to soft errors.

Security. Security Different types of sensors, GPS, actuators, the heterogeneous bus system and the incorporation of vehicle-to-vehicle (V2V) and vehicle to infrastructure (V2I) communication provide attackers with a large number of attack surfaces. In this work, we proposed a unified framework that combines hybrid modelling, formal verification, and automated synthesis techniques for analyzing the security and safety of transportation systems for smart cities.

Energy Consumption. In Vehicular Networks energy consumption is an important parameter for smart cities. Here, we developed an efficient and effective method to reduce the cost associated with automotive idling and some new technique to reduce an energy consumption of controllers through model-based design for automotive control systems and also reduce the stand-by power consumption.

5 New Parameter for Autonomous Driving

5.1 High Volume of Data

Autonomous/semi-autonomous vehicles need to accumulate and process large amount of data at real time. It is reported [1] in that Google's experimental autonomous vehicle generates 750 MB sensing data per second to be transferred through internal CAN buses and Design and Safety Verification for Vehicle Networks processed by various components. In the proposed system, it takes extra time for the insight unit to detect objects in the complex down-town area than on simple rural roads.

5.2 Dynamic Data Generation and Transmission

In the case of autonomous/semi-autonomous vehicle they dynamically generate and transmit the large amount of traffic data. The autonomous/semi-autonomous vehicles need to collect and process large amount of data at real time and these data are dynamic in nature.

5.3 Heterogeneous Design

Traditional autonomous/semi-autonomous vehicles system approves a united architecture, that is to allocate each function to one ECU and attach numerous ECUs through buses like CAN and FlexRay. This strategy style leads to 30–80 ECUs controlling dozens of complex physical processes. In our proposed approaches we are trying to reducing the number of ECUs due to the increasing complexity of autonomous/semiautonomous vehicles driving applications. This pointers to a replacement of united architecture with integrated design, where software tasks are assigned to a heterogeneous platform with single core or multi-core processors and possibly accelerators such as GPUs and FPGAs for computationally-intensive applications. The embedded hybrid architecture design for such platforms will become meaningfully more difficult and challenging.

6 Proposed Framework

6.1 Modelling of Software

The propose software model can be captured by a synchronous reactive task graph shown in Fig. 1. Synchronous software models are prevalent practice for modelling control centric cyber-physical systems in fully automotive or semi automotive and avionics domain, and used in popular tools such as Simulink. Synchronous software system contains a fixed set of synchronized communicating processes, as shown in the software model in Fig. 2. For timing analysis, many timing-related parameters need to be abstracted from the synchronous reactive model. For tasks, the most important parameters are the execution time in worst-case is ET_{Ti} for task Ti on certain ECU, and the period activation is T_{Ti}. For messages, the most import parameters are total length message L_{Mi} for message Mi, the message period T_{mi}, and the source and destination tasks of the message mi.

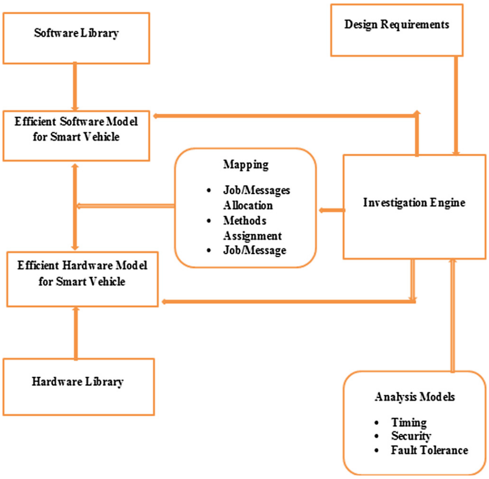


Fig. 1. Framework for proposed design architecture

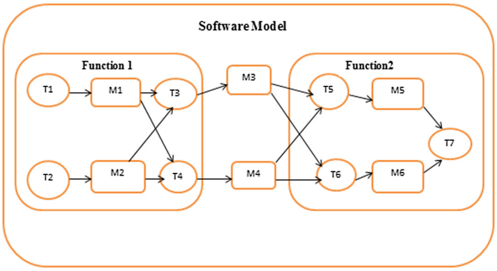


Fig. 2. Example of proposed software model

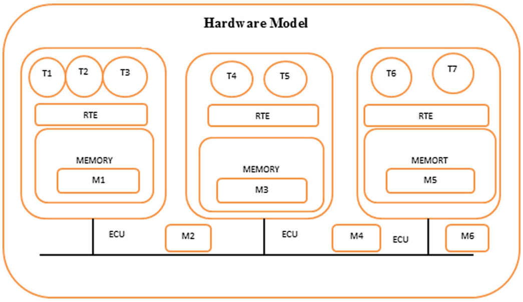


Fig. 3. Example of proposed hardware model

6.2 Modelling of Hardware

The propose hardware model can be taken through hybrid architecture explanation languages, i.e., architecture modeling and description language (AADL). AADL language captures the system through components, and each component is characterized by its identity, large number of interfaces, some properties and subcomponents. The proposed hardware architecture for autonomous/semi-autonomous vehicle is shown in Fig. 3 where different type of buses such as Ethernet, CAN and FlexRay are connected through the gateway, and high speed computation units which include the ECU, FPGA and GPU are connected to the bus system.

6.3 Traffic and Environmental Data Analysis Modelling

The Traffic and Environmental Data analysis models refer to the advanced mathematical models used to quantitatively study various new metrics of automotive/semi automotive system. The various new metrics are timing, reliability, security and fault tolerance.

Task Model. In our propose model, every task is required to finish its execution before a specific deadline (typically set as its activation time slots). The task response time R_{Ti} in worst case is represents the large time it may take to finish task Ti. If we consider static priority preemptive scheduling (where maximum priority tasks can pre-empt minimum priority tasks), the task response time in worst-case for tasks on the same computation vehicle (i.e. node) can be presented in Eq. (1). The task response time in worst-case includes the worst-case execution time ET_{Ti}. $MaxP(Ti)$ denotes the set of maximum priority tasks on the same core.

$$R_{Ti} = ET_{Ti} + \sum_{T_k \epsilon p_T i} (\lceil \frac{R_{Ti}}{T_{Tk}} \rceil) ET_{Ti} \tag{1}$$

Message Model. As a distributed system, different type of buses such as Ethernet, CAN and FlexRay are connected through the gateway, and high speed computation units which includes the ECU, FPGA and GPU are connected to the bus system. The messages can be transmitted on CAN bus or through memory. The message access latency for the messages in memory is assumed to be a small constant in our propose model. However, the latency for messages transmitted on CAN bus should be carefully studied in the proposed model. For instance, for the prevalent CAN bus protocol that uses non-preemptive scheduling algorithm, the message response time in worst-case R_{Mi} for message Mi can be represent as Eq. (2). T_{Mi} represents the transmission time in worst-case for transmuted and generated message Mi, and BT_{max} represent the maximum blocking time (approximated as the highest transmission time of any generated and transmitted message in the system). Similarly, $MaxP(Mi)$ represents the set of highest priority messages on the same bus in the buses.

$$R_{mi} = ET_{mi} + BT_{max} + \sum_{m_j \epsilon p_m i} (\lceil \frac{R_{Mi}}{T_{Mk}} \rceil) ET_{Mj} \tag{2}$$

End to End Delay Model. In automotive/semi-automotive system, the deadlines can also be set on functional paths from source to destination. The path delay or end to end delay from the action of pressing brake to the action of the corresponding actuator should be bounded within a present value to ensure safety. The worst end-to-end delay for path p can be calculated based on the Eq. (3). Because of the asynchronous nature of the communication, task and message periods may need to be added. The details of calculating path latency using a formula (3).

$$L_p = \sum_{T_i \epsilon p} (R_{Ti} + T_{Ti}) + \sum_{M_i \epsilon p} (R_{Mi} + T_{Mi}) \tag{3}$$

Proposed Optimization Model. A new quantitative mathematical for fully autonomous or semi-autonomous vehicle for optimization consists of a new objective function and a set of limitations in the form of a system of equations or inequalities. Propose optimization prototypes are used comprehensively in almost all areas of decision-making, such as vehicular architecture design, verification, safety, validation, fault tolerance and security. This paper presents a motivated and organized process for new optimization problem formulation, design of optimal strategy, and quality-control tools that include safety, validation, verification, security, and post-solution activities. By applying the quantitative or mathematical models, the investigation can be done by solving the optimization problem or simply finding a feasible solution to the problem. Besides the timing properties, we can set constraints such as response time in worst case R_{Ti} must be lower than a execution time, ET_{Ti} in worst case and worst case response time for message R_{Mi} must be lower than message period, T_{Mi}. We can also set the time constraints such as path delay, L_p must be lower than D_p and security level *Sec* must be higher than a value S_0. We can also set constraints such as reliability level *Rel* must be higher than a preset value REL0 and throughput and fault tolerance level must be higher than TPT_0 and FLT_0.

Optimize Objective Function for Design

$$s.t. R_{Ti} \leq ET_{Ti}(Timing) \tag{4}$$

$$R_{Mi} \leq ET_{Mi}(Timing) \tag{5}$$

$$L_p \leq D_p(Timing) \tag{6}$$

$$Rel \geq REL_0(Reliability) \tag{7}$$

$$Sec \geq S_0(Security) \tag{8}$$

$$Tpt \geq TPT_0(Throughput) \tag{9}$$

$$Fault \leq FLT_0(FaultTolerance) \tag{10}$$

7 Limitations and Possible Solutions

Given below are a few of the known limitations and possible solutions:

7.1 Limitations

- For Authentication and Confidentiality, it isn't mentioned which cryptographic technique should be used.
- If the common bus is failed, either LIN, MOST or CAN then whole system fails or partially affects the system performance.
- If any one of the ECU failed, it should not affect the overall braking system.
- If number of ECU increased, then framework will be more difficult.
- It is mentioned that CACC, Cooperative Adaptive Cruise Control directly make use of the leading vehicles velocity and acceleration, through the information from DSRC. If the obtained information is false, then there is a chance for collision between two vehicles.
- Once the attacker gained access to in-vehicle, they directly access all devices.

7.2 Possible Solutions

1. For ensuring Authentication and Confidentiality these mechanisms can be used.
    - User Ownership: A driver owns some unique identity (e.g.: identity card, driving licence etc.).
    - Human Knowledge: A user knows some unique things (e.g.: passwords, human responses through secret questions etc).
    - Biometric Clarifications: These include the signature, thump expression, face and voice.

2. A thumb expression and password mechanism can be added to increase the security of the car, whenever an attacker gain access (in this case only the car owner can drive).
3. In CACC, an additional mechanism can be added that is whenever the values of acceleration and velocity comes from the leading vehicle, instead of using directly, if it is authenticated by the actual forwarder, then a higher level of security can be achieved.

8 Conclusion

The design of new hybrid architecture for autonomous or semi- autonomous or human driving vehicles is great challenges, from the increasing the generation and transmission of data volume from the environment, the usage of real life heterogeneous architecture, and the necessity to address multiple aspect and sometimes contradictory, the new design metrics for smart city communication such as consistency, safety and performance. We propose a novel hybrid model-based architecture for automotive/semi-automotive/human driven vehicle modelling and exploration for smart cities for handling these problems. By formally modelling software and hardware architecture with crucial abstracted some new properties; design space investigation is conducted through quantitative analysis. By resolving numerous optimization problems in case of smart city communications, we can achieve decisions such as task to electronic control unit (ECU) allocation, task scheduling, message allocation, message scheduling, security methods assignment and fault tolerance procedures assignment for smart communication.

Acknowledgement. The author gratefully acknowledge the support from the SERB-DST Early Research Award $ECR \setminus 2015 \setminus 0256$ and the BITS RIG award 2016.

References

1. Zheng, B., Liang, H., Zhu, Q., Yu, H., Lin, C.W.: Next generation automotive architecture modeling and exploration for autonomous driving. In: IEEE Computer Society Annual Symposium on VLSI, pp. 53–58 (2016)
2. Pedroza, G., Apvrille, L., Pacalet, R.: A formal security model for verification of automotive embedded applications. In: SAFA, pp. 1–4 (2010). https://doi.org/10.13140/RG.2.1.4890.1609
3. Zheng, B., Li, W., Deng, P., Gérard, L., Zhu, Q., Shankar, N.: Design and verification for transportation system security. In: Proceedings of the 52nd Annual Design Automation Conference, pp. 1–6 (2015). Article No. 96
4. Lin, C. W., Yu, H.: Invited - cooperation or competition?: coexistence of safety and security in next-generation ethernet-based automotive networks. In: 53rd Annual Design Automation Conference, pp. 1–6 (2016). Article No. 52
5. Schatz, B., Voss, S., Zverlov, S.: Automating design-space exploration: optimal deployment of automotive SW-components in an ISO26262 context. In: Proceedings of the 52nd Annual Design Automation Conference (DAC), pp. 1–6 (2015)

6. Oetjens, J.H., Bannow, N., Becker, M., Bringmann, O., Burger, A., Chaari, M., Chakraborty, S., Drechsler, R., Ecker, W., Gruttner, K.: Safety evaluation of automotive electronics using virtual prototypes: state of the art and research challenges. In: Proceedings of the 51th Annual Design Automation Conference (DAC), pp. 1–6 (2014)
7. Eberl, M., Gla, M., Teich, J., Abelein, U.: Considering diagnosis functionality during automatic system-level design of automotive networks. In: Proceedings of the 49th Annual Design Automation Conference, pp. 205–213. ACM (2012)
8. Yu, H., Joshi, P., Talpin, J.P., Shukla, S., Shiraishi, S.: The challenge of interoperability: model-based integration for automotive control software. In: Proceedings of the 52nd Annual Design Automation Conference, pp. 51–58 (2015)
9. Shahbakhti, M., Amini, M.R., Li, J., Asami, S., Hedrick, J.K.: Early model-based design and verification of automotive control system software implementations. J. Dyn. Syst. Meas. Control **137**(2), 021006 (2015)
10. Zhu, Q., Zeng, H., Zheng, W., Natale, M.D., Sangiovanni-Vincentelli, A.: Optimization of task allocation and priority assignment in hard real-time distributed systems. ACM Trans. Embed. Comput. Syst. (TECS) **11**(4), 1–30 (2012)
11. Zheng, B., Deng, P., Anguluri, R., Zhu, Q., Pasqualetti, F.: Crosslayer codesign for secure cyber-physical systems. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **35**(5), 699–711 (2016)
12. Jo, K., Kim, J., Kim, D., Jang, C., Sunwoo, M.: Development of autonomous car part I: distributed system architecture and development process. IEEE Trans. Ind. Electron. **61**(12), 7131–7140 (2014)
13. Kordes, A., Vermeulen, B., Deb, A., Wahl, M.G.: Startup error detection and containment to improve the robustness of hybrid FlexRay networks. In: IEEE Design, Automation and Test in Europe Conference and Exhibition (DATE), pp. 1–6 (2014)
14. Zhu, Q., Zeng, H., Zheng, W., Natale, M.D., Sangiovanni-Vincentelli, A.: Optimization of task allocation and priority assignment in hard real-time distributed systems. ACM Trans. Embed. Comput. Syst. (TECS) **11**(4), 85–95 (2012)
15. Zheng, B., Deng, P., Anguluri, R., Zhu, Q., Pasqualetti, F.: Crosslayer codesign for secure cyber-physical systems. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **35**(5), 699–711 (2015)
16. Davare, A., Zhu, Q., Di Natale, M., Pinello, C., Kanajan, S., Sangiovanni-Vincentelli, A.: Period optimization for hard real-time distributed automotive systems. In: ACM Proceedings of the 44th Annual Design Automation Conference, DAC, pp. 278–283 (2007)
17. Zheng, B., Lin, C.W., Yu, H., Liang, H., Zhu, Q.: CONVINCE: a cross-layer modeling, exploration and validation framework for next-generation connected vehicles. In: ICCAD, pp. 1–8 (2016)
18. Das, D., Misra, R.: Parallel processing concept based vehicular bridge traffic problem. In: Kumar Kundu, M., Mohapatra, D.P., Konar, A., Chakraborty, A. (eds.) Advanced Computing, Networking and Informatics- Volume 2. SIST, vol. 28, pp. 1–9. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07350-7_1
19. Das, D., Misra, R., Raj, A.: Approximating geographic routing using coverage tree heuristics for wireless network. Wirel. Netw. (WINE) **21**(4), 1109–1118 (2015). Springer US
20. Das, D., Misra, R.: Improvised k-hop neighborhood knowledge based routing in wireless sensor networks. In: IEEE International Conference on Advanced Computing, Networking and Security (ADCONS), pp. 128–134 (2013)

21. Petrenko, A., Timo, O.N., Ramesh, S.: Model-based testing of automotive software: some challenges and solutions. In: 52nd ACM/EDAC/IEEE Design Automation Conference (DAC), pp. 1–6 (2015)
22. Krishna, S.N., Narwane, G., Ramesh, S., Trivedi, S.: Compositional modeling and analysis of automotive feature product lines. In: Proceedings of the 52nd Annual Design Automation Conference, pp. 1–6 (2015). Article no. 57

Outdoor Millimeter-Wave Channel Modeling for Uniform Coverage Without Beam Steering

M. Sheeba Kumari[1](✉), Sudarshan A. Rao[2], and Navin Kumar[3]

[1] New Horizon College of Engineering, Bangalore, India
sheeba.bnm@gmail.com
[2] BigSolv Labs Pvt Ltd., Bangalore, India
dr.sudarshan.rao@ieee.org
[3] Department of ECE, Amrita School of Engineering, Amrita Vishwa Vidyapeetham University, Bangalore, India
navinkumar@ieee.org

Abstract. Diverse performance requirements of the emerging 5G cellular systems and their deployment challenges motivate researchers to explore high frequency millimetre wave (mmWave) spectrum as a potential solution. The allocation and utilization of mmWave spectrum in cellular communication is a new frontier. In this paper, we investigate a directional mmWave small cell outdoor propagation channel by exploiting its deterministic nature; using ray tracing method. Estimates on specific attenuation measurements and free space propagation parameters reveal that directional transmission is inevitable and would result in a channel model divergent from that of an omnidirectional propagation model. In addition, we examine the effect of antenna tilting in the access links to establish highly directional adaptive link. Capacity of the sparsely faded channel is also analyzed. Results exemplify that, beyond 50 m propagation range, the received signal strength in mmWave small cells employing base stations of height 5 m as opposed to macro cells with base station height of 20 m is independent of beam steering. A simplified geometry devoid of the complexity of adaptive beam steering is hence proposed to provide uniform signal strength in an outdoor small cell channel to affirm low latency.

Keywords: 5G · mmWave · Ray tracing · Channel model · Beam steering

1 Introduction

Upcoming 5G cellular communication system with its claim for higher data rate as well as lower latency motivates the use of mmWave frequency spectrum to meet the desired performance specifications [1, 2]. The wide bandwidth available at these high frequencies eliminates spectrum shortage, a major concern in currently deployed cellular systems. Contrary to the fragmented spectrum available in long term evolution (LTE) standards, large contiguous spectrum of approximately 200 GHz enhances the acceptance of mmWave spectrum in cellular communication. The short wavelength of mmWaves (1 to 10 mm) grants integration of many antennas into the same physical dimension, enabling massive multiple input multiple output (MIMO) technique.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 233–244, 2018.
https://doi.org/10.1007/978-3-319-73423-1_21

However, mmWave signals even with the promised high data rate are confined to short communication range due to high free space path loss [3]. The transmission is additionally affected by atmospheric variations such as rain, fog and other moisture content present in the atmosphere. Unlike the traditionally evolved pre-5G technology, mmWave based 5G introduce significant engineering challenges; starting with an accurate analysis of mmWave propagation channel. The channel characteristics vary for diverse propagation scenarios leading to varying system performance. Hence, massive research is driven to assess this technology in 5G context essentially with directional antennas to offer significant extension of link distance than would omnidirectional antennas [4–9]. Initial estimates revealed that signal outages become prominent beyond 200 m. This serves to realize small cell structures with ease contributing to the network densification paradigm.

Ongoing investigations explore diverse methods to model mmWave channel [5–9], improve coverage and capacity [7], exploit beam forming techniques [3] and design network architecture [10]. Path loss data for distinct mmWave frequencies of 28 GHz and 73 GHz are presented in [4] for the perusal of researchers to analyse mmWaves. However, a major work is based on modelling the channel statistically or empirically which lacks accuracy or generality respectively. Computationally complex deterministic modelling can be simplified for the mmWave case due to relatively low dense multipath components [7–9]. Additionally, they assist beam forming techniques through real time estimation of link's directional parameters eliminating the need for extensive beam switching search techniques. Though beamforming technique enhances system gain, its practical implementation involves acquiring comprehensive channel state information. Also, 5G networks due to their smaller cell size have increased handoff probability. The observation, that implementing complex adaptive beam steering technique frequently would affect the latency of the network, is the motivation for this paper. Ray tracing for an mmWave indoor model is demonstrated in [8] and outdoor channel modelling with six ray model is presented in [7, 9]. Yet the models were not fitted in with beam steering technique. Channel characteristics at 60 GHz were examined in [7] whereas [9] explored the entire mmWave spectrum. However, the model geometry in either case was limited to the backhaul transmission.

In this paper, a ray tracing channel model is used to analyse mmWave propagation characteristics for two different use cases namely, wireless backhaul and cellular access. The outdoor base stations implemented with backhaul provisioning, have predictable point to point channel geometry. This, as well as a sparsely populated access channel scenario is modelled using simple deterministic radio channel model. We assumed highly directional horn antenna at both transmitting and receiving sides to assess the received signal strength. The increased outage probability in access channel for regions close to the transmitting antenna is further corrected using beam steering. Beam tilt angle synthesized from the cell geometry is successively updated in the model. Though the probability of deep fade reduces with beam steering, the mmWave link quality excepting beamsteering was observed to be significant at receiver locations beyond 50 m. Hence, we propose a simple layout of 2×2 MIMO for a low latency moderate data rate mmWave system, with one pair of antenna, aligned along the boresight and another pair oriented at an angle. Hence a minimal of two orientations in the elevation plane and the beam swept over the entire azimuth plane suffices to provide uniform coverage in the cell area. This is possible under the assumption of a small cell network with non-blockage, employing base stations/access points at a nominal height of 5 m–8 m. The suggestion is analytically validated by evaluating antenna tilt at all locations within the cell range in distinct steps. Furthermore, the channel capacity of the fading channel is analysed and compared for spatial diversity.

Rest of this paper is organized as follows. Section 2 discusses the propagation characteristics of mmWave channel. mmWave outdoor propagation model is elucidated in Sect. 3 highlighting the antenna design and beam steering strategy. All the results together with significant discussions are provided in Sect. 4. The conclusive statements are included in Sect. 5.

2 Propagation Characteristics of mmWave Channel

2.1 Propagation Characteristics

The propagation characteristics of mmWaves vary from that of existing microwaves and may appear adverse for cellular communication; given the increased propagation loss affiliated with such high frequencies. However, the increase in path loss is supportive for small cell design which increases cellular capacity. Also, the increased pathloss can be compensated by increasing the directivity of antenna, if desired; leading to high effective isotropic radiated power for the same transmitted power levels. To have realistic estimation of channel propagation, it is essential to consider atmospheric absorption loss of mmWaves while evaluating the received signal strength. The empirical model developed by Liebe [11] and the approximate model proposed by ITU-R [12] are two popular techniques. The complex Liebe's model calls for empirical parameters obtained from laboratory measurements to estimate the absorption coefficient whereas the approximate model is based on a rather straight forward approach and is used where accuracy is compromised for simplicity. The modified received power is given by [15]:

$$P_{r,dBm}(min) = P_{t,dBm} + G_{t,dBi} + G_{r,dBi} - L_{dB} \tag{1}$$

where, P_t and P_r represent the transmit and receive power, G_t and G_r the transmitter and receiver gain factor respectively in dBi and L represents the path loss in dB. The total path loss is given as:

$$L_{dB} = L_{FSL,dB} + L_{abs,dB} + L_{Margin,dB} \tag{2}$$

where, L_{FSL} is free space path loss, L_{abs} is the absorption loss and L_{Margin} is the link margin.

2.2 mmWave Link Budget

The simulation is conducted for a 2 Gbps QPSK link, maximum 200 m link range, having a minimum Eb/N0 of 12 dB and a channel bandwidth of 1.5 GHz. The choice of link distance is made to budget for a distance which, even in worst case would result in a tolerable path loss. Using the above parameters, to obtain an un-coded bit error rate (BER) of less than 10^{-7}, the required signal-to-noise ratio (SNR) is 14 dB. For a nominal receiver noise figure of 3 dB, the receiver sensitivity is −65.5 dBm. This value is used as the free space benchmark in our simulations and is the threshold for examining outage probability. The directional antenna is designed for 21.6 dBi directivity. Applying oxygen absorption loss of 16 dB/km, the loss for 200 m link is 3.2 dB. For an additional link margin of 5 dB, we establish that the transmit power is 13.5 dBm for 60 GHz transmission. The maximum allowable path loss is thus budgeted as 114.3 dB using the equation [13]:

where, *EIRP* represents the effective isotropic radiated power obtained as 35.14 dBm,

$$PL_{dB}(max) = EIRP_{dBm} - P_{r,dBm}(min) - L_{dB} + G_{dBi} \tag{3}$$

$P_{r,dBm(min)}$ represents the receiver sensitivity, L_{dB} represents the absorption loss together with the link margin in dB and *G* represents the receiver gain factor in dBi.

3 Proposed Channel Model

mmWave system performance is well analyzed with a suitable propagation model that captures major propagation challenges including higher carrier frequency, wider bandwidth, larger antenna array elements and directional transmission. In this work, we used a deterministic ray tracing model employing highly directional horn antenna to generate narrow beam signals resulting in sparser multipath components (MPCs).

3.1 Antenna Design

This work assumes a highly directional horn antenna with side lobe level 14 dB lower than the main lobe peak and having a beam width of 13° and 15° at half power in the elevation and azimuth planes, respectively.

$$G(\phi, \theta) = G_0\left[sinc^2(a.\sin(\phi))\cos^2(\phi)\right].\left[sinc^2(b.\sin(\theta))cos^2(\theta)\right] \tag{4}$$

Equation (4) avoids analytical solving of complex double integration [14] wherein $G(\phi, \theta)$ is the antenna gain at azimuth and elevation angles respectively, G_0 represents the peak gain at antenna boresight which occurs for a value of $\phi = \theta = 0$. The constants *a* and *b* were estimated using azimuth and elevation HPBWs for normalized half power values.

3.2 Six Ray mmWave Channel Model

The directional mmWave outdoor channel with reduced MPCs has been modeled as a six ray street canyon (SC) model. The direct ray (LOS), ground reflection, first order reflections from either side of the walls and additional second order reflections from both the walls forms the six multipath components. The specific backhaul geometry chosen is an outdoor link with transmitting and receiving nodes deployed on lamp-posts separated by 200 m as in Fig. 1. (Ground reflection is analyzed, but not shown in the illustration due to limitation on dimensionality.) The assumption of closely placed multi-storey buildings [7] helps to treat the buildings on either side of the street as a single reflecting wall resulting in worst case reflection scenario leading to maximum signal attenuation. The effect of diffraction from building edges, due to its insignificance in mmWave links, is excluded in the analysis. The model parameters assumed in channel simulation are listed in Table 1.

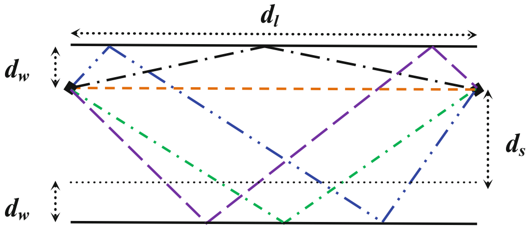


Fig. 1. Top-down view of street canyon six ray propagation geometry.

Table 1. Model parameters for ray tracing channel simulation.

Simulation parameter	Unit	Value
Link range (d_l)	m	200
Transmitting frequency	GHz	60
Width of the street (d_s)	m	12
Transmitter to wall distance (d_w)	m	4
Receiver to wall distance (d_w)	m	4
Antenna gain	dBi	21.6

For six ray channel model, the MPCs contributing to the total response are:

$$h = h_{Los} + h_g + \sum_{i=1}^{M} h_i \tag{5}$$

where h_{Los}, h_g and h_i are the contributions of the LOS, ground reflected and first/second order wall reflected paths respectively. M represents the total number of wall reflections, having a value of 4 in the six ray model accounting for 2 single and double reflections each. The directional mmWave channel impulse response is hence expressed as:

$$h(f, \theta, \varphi) = G_{Los}(\theta, \varphi)H_{Los} + G_g(\theta, \varphi)H_g + \sum_{i=1}^{M} G_i(\theta, \varphi)H_i \tag{6}$$

where $G_{Los}(\theta, \varphi)$, $G_g(\theta, \varphi)$ and $G_i(\theta, \varphi)$ are the products of transmit and receive antenna power patterns for LOS, ground reflected and i^{th} wall reflected rays at appropriate elevation and azimuth angles evaluated from (4) and H_{Los}, H_g and H_i are their channel impulse response components,

$$
\begin{aligned}
H_{Los} &= \left[\frac{\lambda}{4\pi p_{Los}} e^{(-K_p p_{Los}/2)}\right] e^{\left(-j\frac{2\pi}{\lambda}p_{Los}\right)} \\
H_g &= \left[\Gamma_g \frac{\lambda}{4\pi p_g} e^{(-K_p p_g/2)}\right] e^{\left(-j\frac{2\pi}{\lambda}p_g\right)} \\
H_i &= \left[\Gamma_i \frac{\lambda}{4\pi p_i} e^{(-K_p p_i/2)}\right] e^{\left(-j\frac{2\pi}{\lambda}p_i\right)}
\end{aligned}
\tag{7}
$$

where λ is the wavelength, p_{Los}, p_g and p_i are the path lengths of LOS, ground reflected and first/second order wall reflected paths respectively, Γ_g and Γ_i are the reflection coefficient for ground and wall reflections, K_p is the coefficient for exponential absorption and $2\pi/\lambda$ is the wave number.

The mmWave sparsely faded channel characterization can be obtained as follows:

(1) Generate mmWave SISO channel model from Eq. (8) for the given outdoor street canyon geometry using the path lengths of MPCs and their related AoAs and AoDs taking antenna directionality into account. Analyze the received power for wireless backhaul with antenna height fixed at 5 m.

(2) Adapt the model to accommodate receiver antenna height of 1.5 m to explore access link. To perform antenna beam steering, change the orientation of antenna radiation pattern through ongoing tilt angle calculation, using Eq. (9).

(3) Compute the CIR contributions $h_{m,n}$ from the nth transmit antenna to the mth receive antenna to determine the MIMO channel matrix H. The multipaths with AoAs and AoDs within the antenna beam solid angle are used in evaluating the response.

(4) Obtain the MIMO channel capacity relative to the free space signal to noise ratio ($\mathrm{SNR_{LOS}}$):

$$
C = log_2\left|I_{N_r} + \frac{SNR_{Los}}{N_t} HH^H\right| b/s/Hz \tag{8}
$$

where N_r and N_t are the number of receive and transmit antennas, I_{Nr} refers to $N_r \times N_t$ identity matrix, SNR_{Los} refers to signal-to-noise ratio relative to LOS and HH^H refers to the matrix product of MIMO channel matrix and its Hermitian.

3.3 Beamsteering vs Non-beamsteering

In backhaul analysis, the antenna placed on either lamp post at a height of 5 m is oriented along the boresight. However, for wireless access channel characterization, the receiver is assumed at a height of 1–2 m which for optimal performance suggests the transmitter

(AP) antenna radiation pattern to be tilted down by an angle $\Delta\theta$ with respect to horizontal axis. The channel model performance with beamsteering is hence investigated by tilting the main beam of BS and user equipment (UE), leading to main lobes aligned for maximum received signal strength. Thus, by varying the antenna tilt angle for Tx-Rx pair in the model, a direct link with maximum antenna gain is created from the transmitter to the receiver for every receiver locations. The tilt angle is obtained from the vertical angle between BS antenna and UE antenna given by:

$$\Delta\theta = \arctan\left(h_t - h_r/d\right) \tag{9}$$

where h_t and h_r are the heights of transmitting and receiving antenna respectively and d is the distance of separation between them. The concept of SISO hitherto explained can be extended to simulate transmit/receive diversity or MIMO to enhance performance.

Though beamforming in MIMO enhances channel performance in terms of extended coverage as well as reduced deep fades, it requires smart signal processing algorithms to estimate spatial signature like direction of arrival (DoA) of the signal increasing the system's complexity. For small cells, with cell radius typically below 200 m and access points/BS mounted at relatively low heights as in the modeled channel geometry the tilt angle variation is trivial. Alternately, we propose a simplified geometry, in Fig. 2, wherein the antenna array with two elements at the lampposts is oriented differently leading to uniform coverage with interference mitigation traded off. A pair of array elements on either lamppost is oriented horizontally to serve the small cell backhaul link as well as the UE far from the transmitter say, at the cell edge. The UE position in the proximity of lamp post, outside the antenna beam width, receives signal from the second antenna element oriented at an angle of 15° relative to boresight.

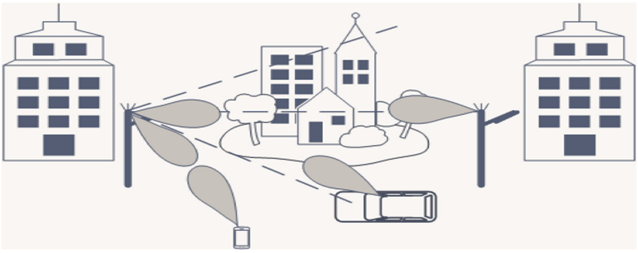


Fig. 2. Proposed small cell layout geometry devoid of beamsteering for uniform coverage.

4 Results and Discussion

First, we compare the absorption loss simulated for mmWave E-band frequencies using two standard modeling methodologies listed in Sect. 2 [11, 12]. As illustrated in Fig. 3, we observe that variations in attenuation due to oxygen absorption for the models are less than 0.3 dB/km for frequencies in the upper E-band. Even at the frequency of interest, 60 GHz, the observed variation is not more than 0.62 dB/km leading to our choice of approximate model. The high attenuation value of 16 dB/km obtained justifies the suitability of 60 GHz mmWave band in limited communication range small cell network design. The normalized antenna radiation pattern for a horn antenna used in the channel model simulation is shown in Fig. 4. This reference antenna model forms the basis for evaluating the MPCs' strength at each pointing angle (AoAs and AoDs) which in turn is obtained through ray optic analysis of street canyon geometry. The antenna being vertically polarized, E-plane and H-plane determine the antenna gain factor for ground and wall reflections respectively. Note that 0° implies horizontal orientation of beam along the reference. The constants a and b are evaluated as 3.35 and 3.88 by equating Eq. (4) to 1/2, solving for ϕ and θ respectively. The narrow beam width of receiving antenna guarantees that a reflected path greater than second order is either received outside the beam area or attenuated significantly relative to LOS ray. The simulated results demonstrate that six ray channel modeling suffices directional mmWave channel characterization for the chosen geometrical layout (refer Fig. 8).

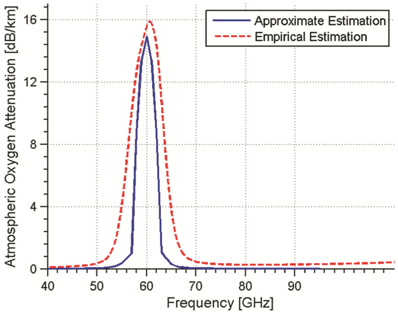


Fig. 3. Specific attenuation simulated using Leibe's model and approximation model.

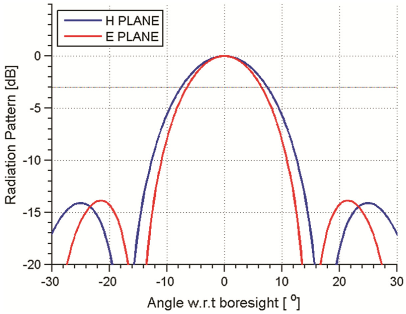
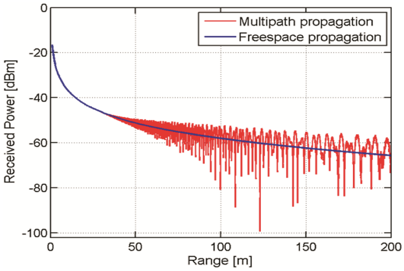


Fig. 4. Normalized power pattern for a horn antenna, HPBWϕ = 15° and HPBWθ = 13°.

The backhaul channel is analyzed by placing in the frequency dependent absorption loss and angle dependent antenna gain factor. Figure 5 depicts received signal variations observed in backhaul ecosystem wherein transmitting and receiving antennas are mounted on lampposts at 5 m from the ground and 200 m apart. As illustrated, the variation in faded signal strength with respect to free space benchmark is more for link distances exceeding 100 m with the deepest fade of −99.09 dBm occurring at 123 m. This is due to the existence of all the higher order reflections. The mmWave link up to a range of 30 m is predominantly LOS with zero contribution from reflected rays as their AoAs fall outside the beamwidth leading to a non-faded channel. The significant variation of received power owes to the small wavelength of mmWave carrier signal. Another relevant aspect is the reduced deviation of the statistical distribution from its free space value below 100 m.



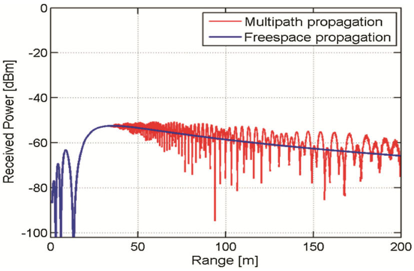
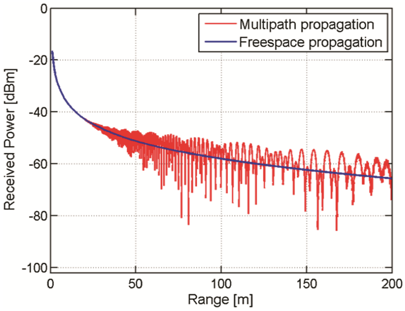


Fig. 5. With antenna aligned at the boresight for backhaul ecosystem.

Fig. 6. Antenna aligned at the boresight for cellular access.

The signal strength in access network analysed for a receiving antenna height of 1.5 m with antenna radiation pattern aligned horizontally is shown in Fig. 6. It is analogous to the backhaul case for Tx-Rx separation distance greater than 36 m. For the chosen antenna beamwidth and the small cell geometry, the pattern spans such that even a horizontally aligned beam ensures signal reception. It is also observed that at distance closer to the lamp post signal outage occurs suggesting the need for beam steering. The beam is hence adaptively tilted by an angle generated by Eq. (6) to correct the probability of signal outage. The resulting received signal strength variations can be easily compared from Figs. 6 and 7. Hence, we observe that the probability of fade at the close-in distance of transmitting antenna is corrected with beam steering. The comparison between simulated six and eight ray models, in terms of relative received power level in dB, for varying cases of backhaul and access transmissions is presented in Fig. 8. When transmitting and receiving antennas were assumed to be of same height, introducing two additional reflections at the receiver had negligible impact. This illustrates the triviality of the number of reflecting bodies in a directional transmission scenario. The statistical distribution of received power in six ray model for the chosen geometry is observed to be identical to eight ray model with 89% of the mmWave links. For the access case, variation between two models is more apparent indicative of choosing higher order ray tracing model. However, as the difference is less than 10 dB for 90% of the links, we reckon six ray models as a fair approximation for mmWave channel propagation characterization.



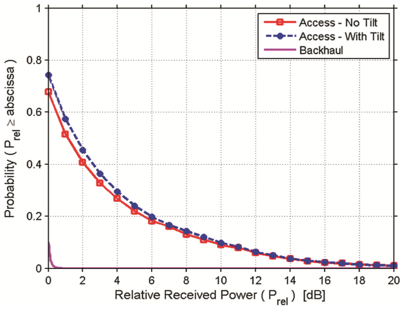


Fig. 7. Rx signal strength with appropriate antenna alignment using beam steering.

Fig. 8. Adequacy of six ray model to characterize mmWave link is depicted.

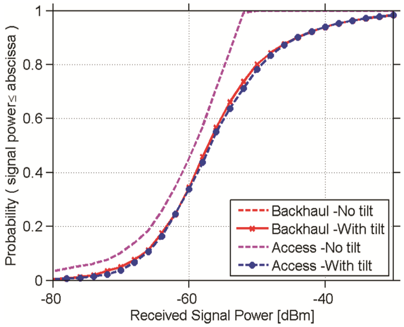


Fig. 9. Probability of signal strength required analyzed for diverse cases.

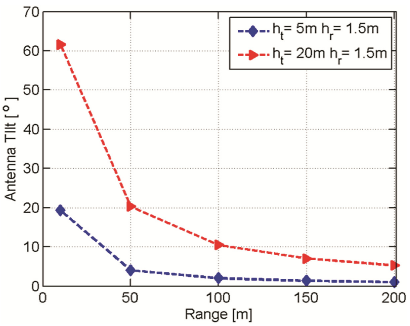


Fig. 10. Antenna tilt angle to maintain the beam orientation.

The cumulative distribution function (cdf) plot (Fig. 9) shows that the probability of received signal fitting in a precise range, for backhaul and access case, is approximately same even with adequate inclination of antenna beam to route energy in the desired direction. It is depicted that without antenna tilt, signal strength for a given transmission link within the cell coverage area is always less than −52 dBm. Yet, this value is higher than the free space signal benchmarked at −65.5 dBm. The antenna tilt technique adapted offers a direct LOS link aligned with the strongest signal. It may be noted that the model, for convenience, in analysis assumes an un-obstructed LOS link. The tilt angle evaluated for two different cases, first being with the proposed height of 5 m and second with the traditional BS height of 20 m, is provided in Fig. 10. For a small cell of size 200 m, employing base station on lamp posts or like structures, a tilt is required primarily in the serving area close to the base station. The results show that in addition to the antenna beam oriented along the horizontal direction, it may be sufficient to have a single additional beam tilted by an angle in the range [5°–15°] with respect to horizontal to provide uniform signal strength at any UE location within the cell area of radius 200 m.

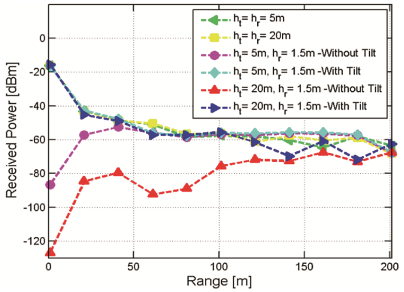


Fig. 11. SISO received power compared to emphasize proposed antenna tilt strategy.

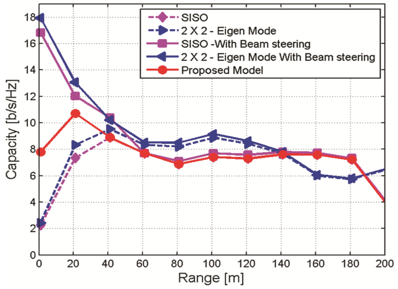


Fig. 12. Capacity analysis of access channel.

The received signal power at distinct receiver locations within the selected link range are presented in Fig. 11. We learn that significant reduction in received power is observed only when the access points are placed high above ground level, say 20 m, for receiver located anyplace in the cell range. Else, except for the deep fades at certain receiver locations, the statistical distribution of received signal strength is approximately uniform even in the absence of beam steering. The proposed method hence demonstrates a simple and viable solution for effective transmission in a cell layout where LOS condition prevails. The capacity of access channel is analysed in the presence and absence of beam steering and compared with the proposed model in which beam is oriented in two fixed directions. The analysis is conducted with a static beam oriented at a tilt angle of 15° with respect to horizontal. It is observed that though the model architecture yields a uniform capacity over the entire range even though it fails to acquire the peak rates offered by steerable antenna at low link ranges (Fig. 12).

5 Conclusion and Future Work

mmwave exhibits the prospective for use as small cell backhaul and access in upcoming 5G cellular networks. Antenna patterns and specific attenuation due to atmospheric absorption are to be considered to derive an accurate directional mmWave channel model. Steering the beams toward major paths and/or the nulls toward interfering paths will minimize interference to an extent where even the MPC that contributes destructively can be phased out. We examine a six ray model with directional transmission incorporating antenna steering to evaluate the received signal strength for cellular backhaul and access networks. The probability of fading for backhaul channels, irrespective of antenna steering is observed to be approximately 9%. An equivalent fading probability is realized for access channels incorporating antenna steering. However, an increased outage probability of 17% is obtained for access channel without antenna tilting. Our mmWave signal strength study indicates that incorporating beamsteering in an outdoor small cell street canyon channel hence offers an 8% enhancement in the channel performance.

This study on antenna angle-dependent mmWave channel model assimilates the directionality and steerability of mmWaves. Hence, for the chosen channel specification, significant performance is assured in a small cell network with simple antenna geometry of two guided beams at the transmitting base station trading off the aforementioned 8% performance enhancement. Yet, in a deployment scenario of highly populated urban area with large number of buildings and moving vehicles, such an approximation may not work and have to be analyzed using a site generic probabilistic model to capture the channel characteristics effectively. Hence, with small cell size and reduced Tx antenna height, we demonstrate that adequate performance is guaranteed with two different vertical orientations of the transmitting antenna. The model can be extended to analyze higher order multipath components.

References

1. Sadri, A.: mmWave technology evolution: from WiGig to backhaul and access for 5G. In: Proceedings of International Workshop on Cloud Cooperated Heterogeneous Networks, Osaka (2013)
2. Warren, D., Dewar, C.: Understanding 5G: perspectives on future technological advancements in mobile. White paper, GSMA Intelligence (2014)
3. Rappaport, T.S., et al.: Millimeter wave mobile communications for 5G cellular: it will work! IEEE Access **1**, 335–349 (2013)
4. Maccartney, G.R., Rappaport, T.S., Samimi, M.K., Sun, S.: Millimeter-wave omnidirectional path loss data for small cell 5G channel modeling. IEEE Access **3**, 1573–1580 (2015). ISSN 2169-3536
5. Sun, S., et al.: Synthesizing omnidirectional antenna patterns, received power and path loss from directional antennas for 5G millimeter-wave communications. In: Proceedings of IEEE Global Communications Conference, GlobeCom 2015, pp. 1–7. IEEE Press (2015)
6. Rappaport, T.S., MacCartney Jr., G.R., Samimi, M.K., Sun, S.: Wideband millimeter wave propagation measurements and channel models for future wireless communication system design. IEEE Trans. Commun. **63**(9), 3029–3056 (2015)
7. Zhang H., Venkateswaran, S., Madhow, U.: Channel modeling and MIMO capacity for outdoor millimeter wave links. In: Proceedings of IEEE Wireless Communications and Networking Conference, WCNC, pp. 1–6. IEEE Press (2010)
8. Steinmetzer, D., Classen, J., Hollick, M.: mmTrace: modeling millimeter-wave indoor propagation with image-based ray-tracing. In: Proceedings of Millimeter wave Networking Workshop, mmNet 2016 (2016)
9. Kumari, M.S., Rao, S.A., Kumar, N.: Characterization of mmWave link for outdoor communications in 5G networks. In: Proceedings of IEEE International Conference on Advances in Computing, Communications and Informatics, ICACCI, pp. 44–49. IEEE Press (2015)
10. Gotsis, A., Stefanatos, S., Alexiou, A.: UltraDense networks: the new wireless frontier for enabling 5G access. IEEE Veh. Technol. Mag. **11**(2), 71–78 (2016)
11. Liebe, H.J.: MPM-An atmospheric millimeter-wave propagation model. Int. J. Infrared Millim. Waves **10**, 631–650 (1989)
12. ITU-R Recommendation: Attenuation by atmospheric gases. ITU-R P.676-11 (2016)
13. Madhow, U.: Introduction to Communication Systems. Cambridge University Press, Cambridge (2014)
14. Far field radiation from electric current. http://www.thefouriertransform.com/applications/radiation.php
15. Goldsmith, A.: Wireless Communications. Cambridge University Press, Cambridge (2005)

Dimensional Modification Induced Band Gap Tuning in 2D-Photonic Crystal for Advanced Communication and Other Application

R. R. Sathya Narayanan[1], T. Srinivasulu[1], Chitrank Kaul[1], Arvind Narendran[1], Ashit Sharma[1], Jhilick Ghosh[2], Nabanita Acharjee[2], and Kaustav Bhowmick[1](✉)

[1] Photonics Research Lab, Department of Electronics and Communication Engineering,
Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham,
Amrita University, Bengaluru, India
k_bhowmick@blr.amrita.edu

[2] Netaji Subhash Engineering College, Kolkata, West Bengal, India

Abstract. We present a systematic simulation study of dimension-induced photonic band-gap tuning in two-dimensional (2-D), hexagonal lattice photonic crystals, consisting of air-holes in dielectric slabs. Photonic crystals are interesting candidates for application in various fields e.g. communication ranging from optical to THz regime, sensing, spectroscopy, imaging etc., using their property to trap and harness light and to produce high-Q resonances by the principle of localization and photonic bandgap formation. The insensitivity towards launched light wavelength shown herein by the bandgap response of a given 2-D planar photonic crystal is promising for enabling cheaper visible or NIR light sources to produce desired response in Mid-IR wavelengths with ease. The structures and material studied lie within the range of popular fabrication methodology. The results show that silicon photonic crystals, operated at 1.55 μm, can produce sharp resonances and large band transmission in Mid-IR wavelengths (3–5 μm) as well.

Keywords: Photonic crystal · Photonic band gap · Visible · NIR · MIR

1 Introduction

Interaction and manipulation of light and matter in photonic crystals (PhC) have mobilized the scientific community of photonics towards their exploration since their first appearance [1]. Despite an acquired understanding of the photonic bandgap (PBG) in such photonic structures [1, 2], further exploration of their optical properties continues [2]. The established wavelength-scalability in optical waveguides have been demonstrated in PhCs by Yablonovitch et al. [3], for a 3-D PhC in a microwave dielectric with refractive-index (r.i.) value similar to Si, to produce PBG-like forbidden bands in microwave ranges. Also, a Si 2.5-D PhC was shown [4] with PBG around 0.1 THz with 75 GHz (λ ≈ 3.99 mm) to 110 GHz (λ ≈ 2.72 mm) input.

Wavelength tuning and PBG tuning of various types have been reported for various application, e.g., Temperature-induced tuning of PBG-span [5] in 2012 and interlayers

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 245–255, 2018.
https://doi.org/10.1007/978-3-319-73423-1_22

in the holes in 2.5-D PhCs shifting the band-edges of the minimum PBG 1.5–3% [6]. However, fabrication of such structures may be challenging in optical domain. In 1-D PhCs [1], varying N or number of layers and external stimulus like magnetic fields, can vary the PBG [7] although, refractive-index contrast between layers (Δn) remains a greater controlling factor [1]. Hence, availability of suitable material dictates the fabrication and usage. Other factors influencing the band-edge e.g. free-carrier injection around 1.9 μm [8], structural parameter tuning of band-gap in 2-D square lattice [9], effect on modal reflectivity of geometry tuning in photonic crystal [10], and, effect of various positions on grown oxide layer in Si on the band-gap [11] were also presented in the recent past.

Herein, we present a systematic simulation study of the effect of dimensional parameters in a hexagonal lattice 2-D PhC. As light red-shifts from visible to MIR, the cost of source laser increases manifold [12]. It is envisaged that using the dimensional band-gap tuning features in 2-D PhCs, a lower wavelength 1550 nm pulsed laser may be used to get desired operation in more expensive MIR ranges. Hexagonal lattice structures, which were found to be the best for fabrication of bent defects and better optical operation [1] were chosen. Analyses have been done based on two materials, namely, $Al_xGa_{1-x}As$ (x = 1) [13] for visible-to-NIR PBG shift study and Si [14–16] for NIR-to-MIR PBG shift study. The effective-indices for the PhC-structures were calculated using the effective-index method [17] and, the PBG features using plane-wave expansion method by MPB [1] and by FDTD-method using commercial software RSoft-FullWAVE [18]. The most important result hence found is a broadband response in MIR (3-5 μm) using Si PhC excited at 1550 nm laser pulse. Till date, applications based on Si for MIR are rare except for some studies on MIR high-Q resonance [19], using an expensive MIR tunable CW-laser at ~4.5 μm. The aforesaid result is important from the viewpoint of making 1550 nm communication devices using visible lasers and MIR (3–5 μm) spectroscopic devices on Si.

2 Simulation Work and Results

The material consideration, simulation work and Results with Discussion are presented in the following sub-sections.

2.1 Material Choice and Refractive-Index

The material(s) chosen for the present study were $Al_xGa_{(1-x)}As$ (x = 1) [13], for visible to NIR wavelengths, and Si [19] for NIR to MIR wavelengths. The principal criterion for the choices was to obtain a high enough absolute-index value at the chosen launch wavelength of the bulk absolute indices of $Al_xGa_{(1-x)}As$, the values for x = 1.0 (i.e., AlAs) are suitably high between λ = 650 to 700 nm in the visible range (with maximum current price of USD 702 [12]). The lower fractions of x provide high absolute indices, however, their UV-band-edges show successive red-shifts. Practically, AlAs being hygroscopic, it may need to be housed within some other material for usage. Similarly, data for Si absolute-index was obtained from three sources (for NIR excitation) [14–16]. Further, it was recognized that full 3-D FDTD simulation of PhC could be highly time-consuming with not much of advantage [18]. Hence, the effective-index method [17] was adopted to find the effective-index of the PhCs. Figure 1 shows the schematic of a 2.5-D PhC and the direction of incident light along which the effective-indices were calculated for different 2.5-D PhCs studied, thus removing the slab- thickness and making 2-D simulation of the structure possible, for lower computational time. The correctness of the method was tested with that of a reported PhC structure [20] wherein the effective-index reported being ~2.797 at 1.55 μm.For the same work, we used the effective-index method and generated ~2.798 at 1.55 μm as the effective-index, which was close to the experimental measurement of refractive-indices. The convention shown for coordinate axes in Fig. 1(a), is at par with the convention used in the FDTD-solver of RSoft [18], and Fig. 1(b) shows the essential parameters of a 2.5D PhC.

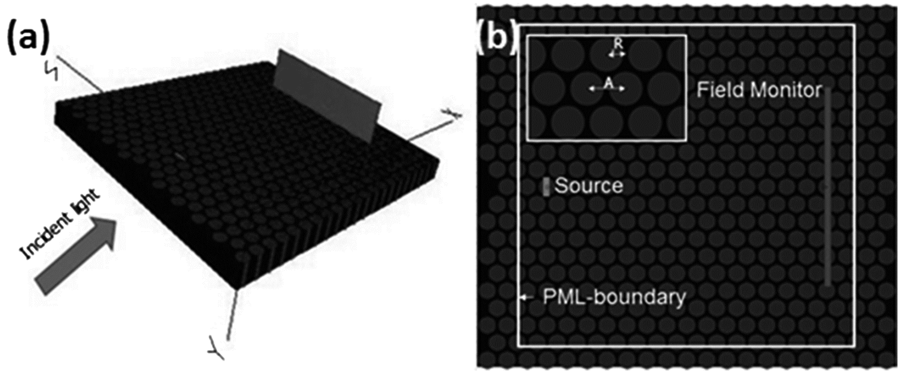


Fig. 1. (a) Schematic of typical 2.5-D PhC showing spatial coordinates used and light launching direction in which effective-index calculations have been performed. (b) A typical PhC (from Rsoft-FullWAVE) used in the present work showing the source, the field monitor and PML-boundary location. The inset shows the hole radius ‘R’ and period ‘A’.

2.2 Numerical Verification of Simulation Mesh

Following the choice of materials and adopting the effective-index method, the simulation environment was setup using the ab-initio MPB software and the FDTD-based FullWAVE software. For both the cases first the verification of the mesh dependence of bandgap results were performed. For the said study, a 2.5-D PhC (shown in Fig. 1(b)) with hole-radii ‘R’ of 0.267 μm and period ‘A’ of 0.710 μm (both parameters as illustrated in Fig. 1(b) inset) were considered. The thickness of the 2.5-D structure was kept equal to the period ‘A’. Maintaining the same structural parameters, different mesh size effects were studied in AlAs and in Si, with relevant effective-indices and launch-wavelengths (for FDTD). Invariably, throughout the present work, the launch light source have been considered to have a Gaussian distribution with the centre-wavelength as will be mentioned and with a 40 nm spread about the centre-wavelength, which is at the resolution limit of FullWAVE. Also, a PML [1] boundary was chosen and was position-optimized in the order of λ/2 near the source and the monitor, both of which are shown in Fig. 1(b).

A schematic of the PML boundary is shown in Fig. 1(b), drawn along the true simulation boundary, to make the location visible. Actual PML thickness is chosen as per optimization of the output till the achievement of mesh-independence. TE-mode excitation was done in each case. The band-gaps in the same structure were found using MPB, and was studied for different grid-sizes (which in MPB is in terms of number of k- points [1]), namely, 4, 25, 81, 100, in succession, chosen for random increase of fine-ness. The two cases studied had effective-indices ~3.018 at 700 nm (Al-As) and ~3.343 at 1532 nm (Si). However, the band gap obtained (in MPB) had very low percentage variation (order of $10^{-3\%}$), with mean-square error calculated as $\sqrt{\left(PBG\_mesh_{j-1} - PBG\_mesh_j\right)^2 / \left(PBG\_mesh_j\right)^2}$, with no appreciable pattern.

However, in the case of FullWAVE, some appreciable mesh-to-mesh errors were obtained, using the same error formula, which is shown for AlAs and Si, in Fig. 2. The RSoft mesh study was done with grid-sizes 0.071 μm, 0.0473 μm, 0.035 μm and 0.01 μm, in the order of increasing fine-ness, in two ways, namely, error % between each progressively finer mesh-pair, and, error % between the coarsest mesh 1 paired with progressively finer meshes. The mesh error analyses shows that other than the coarsest first mesh in FDTD, the finer meshes quickly fall below 5% error which is statistically acceptable. Also, a plateau-ing trend can be observed, which is confirmed by Fig. 2(b), depicting increasing error with the coarsest mesh 1. Hence, the fineness of meshes 2 to 4 were maintained, as appropriate for the subsequent simulations. All dimensions have been chosen bearing the minimum dimension that can be fabricated by e-beam lithography, i.e. 0.2 μm.

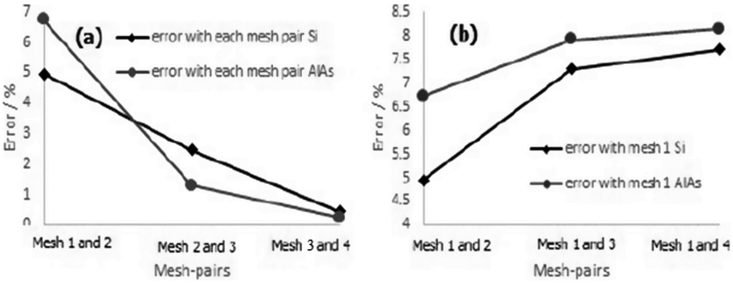


Fig. 2. Mesh error analyses: (a) error between mesh pairs, coarse to fine; (b) error between coarsest mesh 1 paired with each progressively finer mesh, in terms of PBG.

2.3 Simulation of Al-As PhC

For the Al-As PhC mentioned in Sect. 2.2, the results of PBG-analyses are presented here. The result of MPB simulation is presented in Fig. 3(a), with the corresponding band-edges in wavelength marked as (a) 1724 nm and (b) 2573 nm (see the gray band). Only the TE-like PBG was sought and not the complete band-gap [1]. Similarly, using FDTD solver, the same structure was simulated with Gaussian light centered at 680 nm (Fig. 3(b)) effective-index ~3.030, and at 700 nm (Fig. 3(c)) with effective-index ~3.018. In the FDTD results, the y-axis depicts transmitted modal intensity arbitrary units (a.u) and the PBG is where the transmission is at zero-level. In both the cases, it can be seen that the band-gap and band-edges are approximately constant and in agreement with the MPB solution showing that visible light in a suitable PhC can produce PBG in NIR to MIR region. Similar studies were carried out for different R/A ratios and the trend curve is presented in Sect. 2.5.

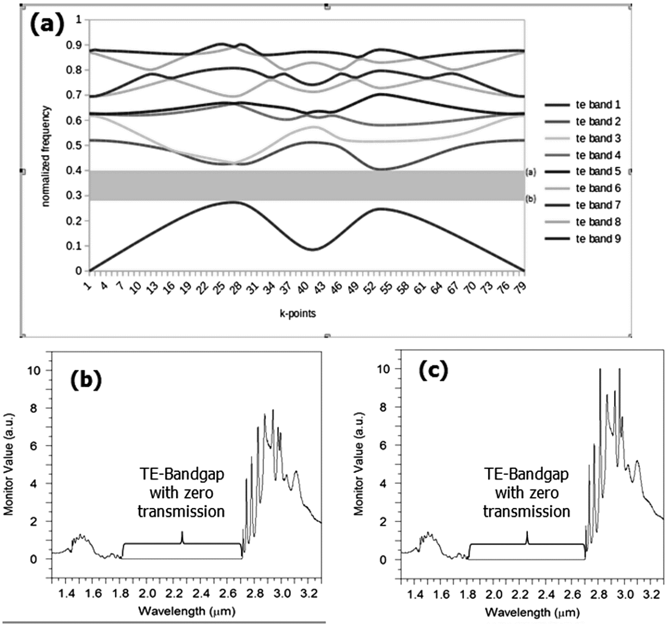


Fig. 3. Al-As PhC simulation for R = 0.267 μm, A = 0.71 μm excited in TE-mode; (a) MPB results showing PBG between as 1724 nm and 2573 nm; (b) FDTD simulation for 680 nm showing similar PBG as in (a); (c) FDTD simulation for 700 nm showing similar PBG as in (a).

2.4 Simulation of Si PhC

Following the PBG obtained from NIR to MIR in Al-As PhC, it was tested whether the established Si technology could produce PBG deeper into the MIR, to cover the 3–5 μm region, where the absorptive losses are the least. Thus, a PhC with the same R/A ratio was considered but, made in Si, with an effective-index ~3.343 at 1532 nm. The results of the same are presented as simulated in MPB (Fig. 4(a)) and FDTD (Fig. 4(b)). It can be seen that the PBG obtained in the present case extends from ~2.0 μm to 3.0 μm in the MIR. The results for various values of R/A are presented in Sect. 2.5.

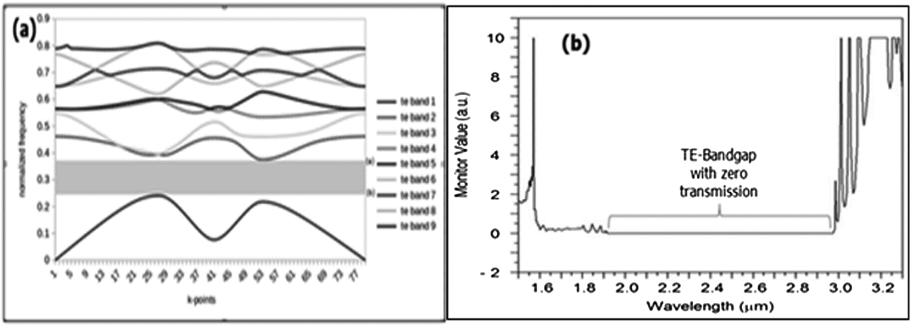


Fig. 4. Si PhC simulation for R = 0.267 μm, A = 0.71 μm excited in TE-mode; (a) MPB results showing PBG between as 1900 nm and 2900 nm (gray shaded band); (b) FDTD simulation for 1532 nm showing similar PBG as in (a).

2.5 Trend Curves in PBG and R/A Ratio

As mentioned in Sects. 2.3 and 2.4, the exercises were repeated for different values of R/A for both Al-As and Si PhCs, the trends of which are shown here. The R/A ratio has been identified in the theory of PhCs as the major structural parameter [1]. So, for the present study, certain major variations of the ratio were identified and the corresponding variations of the PBG were recorded. The choice of values for R and A was based on obtaining the maximum PBG and the input lights were taken as 850 nm for Al-As and 1532 nm for Si, with appropriately calculated effective-indices. Firstly, the value of

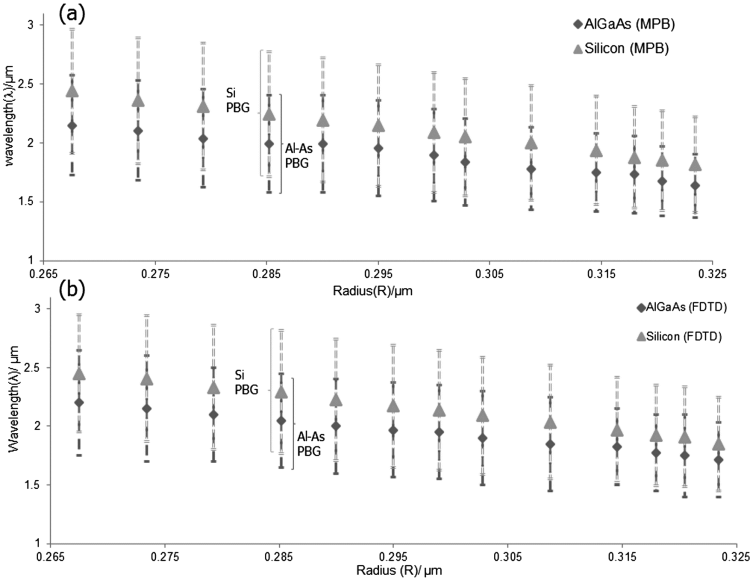


Fig. 5. Hole radius 'R' variation at period 'A' = 0.71 μm: (a) MPB; (b) FDTD.

period ‘A’ was fixed at 0.71 µm, and the hole-radius ‘R’ was increased. The result for the same is shown in Fig. 5(a) and (b) for MPB and FDTD, respectively. Figure 5 is oriented such that each data point shows the center-wavelength of PBG for either Al-As or Si, with the error-bars depicting the PBG span on either side of the centre-wavelength. Both simulation tools show that by increasing ‘R’ for a given ‘A’ the center-wavelengths of the PBGs show a blue-shift, accompanied by a simultaneous slight and progressive decrease in the PBG-span.

Thereafter, the hole-radius ‘R’ was fixed at 0.3 µm and increasing the period ‘A’. The input lights were chosen same as in the previous. The results of the study show a trend opposite to that of varying ‘R’ with a fixed ‘A’ (see Fig. 6). The MPB and FDTD results are again found to have a good agreement (Fig. 6(a) and (b), respectively).

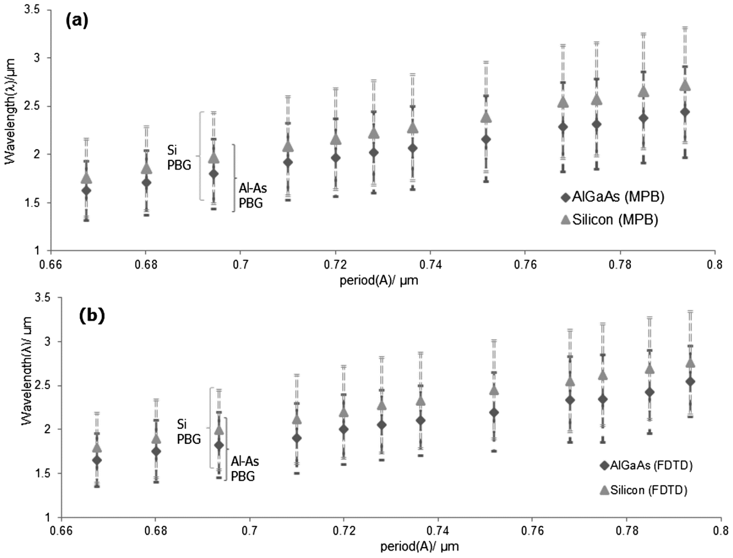


Fig. 6. Period’A’ variation at hole radius ‘R’ = 0.3 µm simulated by: (a) MPB; (b) FDTD.

Finally, R/A ratio was taken with R = 0.3 µm and A = 0.65 µm, wherein ‘R’ and ‘A’ were simultaneously increased and decreased by factors representing ±2.4%, ±5% etc. The results of study are presented in Fig. 7. Again, Fig. 8(a) shows the results from MPB, while Fig. 7(b) shows the results from FDTD. While the R/A is always maintained, the PBG spread can be seen to remain more or less the same, with little variation. However, the simultaneous decrease in R and A shows a blue-shift in the center-wavelength of PBG and a simultaneous increase in R and A shows a red-shift.

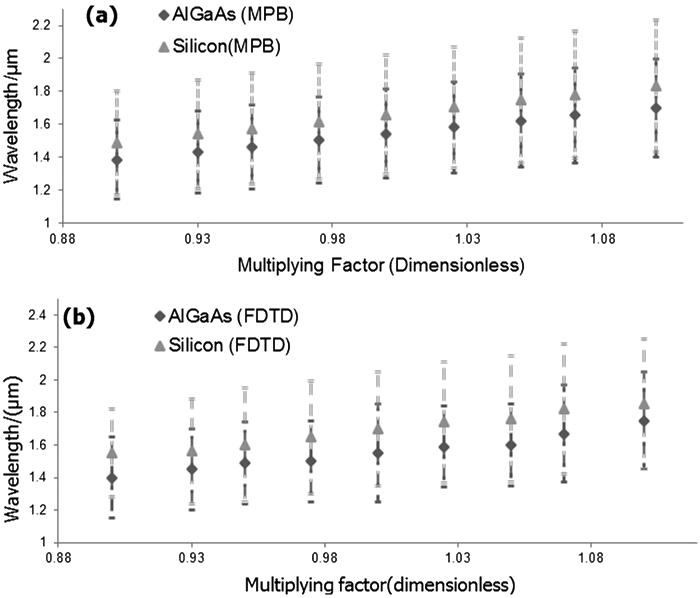


Fig. 7. Effect of period' A' and hole-radius 'R' simultaneously varying by percentages, keeping the R/A ratio constant, about the central values of R = 0.3 µm and A = 0.65 µm.

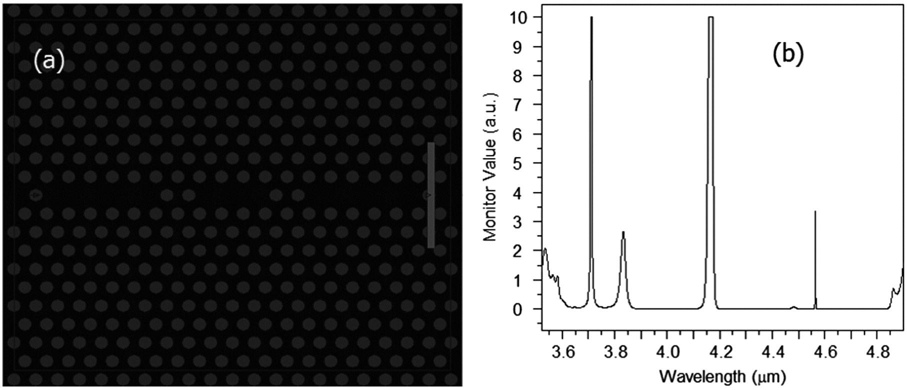


Fig. 8. Si MIR operation study: (a) Line defect(s) in Si 2-D photonic crystal at 1532 nm light source; (b) high intensity resonant peak and narrow high intensity band in 3–5 µm MIR range.

Overall, the study presented in Sect. 2.5 shows that by the interplay of dimensional parameters 'R' and 'A' with chosen wavelength, hence, refractive-index of the structure at the chosen wavelength, the PBG can either be blue-shifted or red-shifted. As can be seen from Figs. 5, 6 and 7, even a visible transparent material like Al-As can be illuminated with suitable coherent light, and the PBG may be obtained in MIR. The same can be achieved with the Si-technology. Thus, cheaper material and laser options may be utilized to obtain PBG in MIR and THz, where optical-sources and transparent material are otherwise expensive, and may help to develop long-wavelength devices, using defect light-localization [1], as discussed in Sect. 3.

3 Suggestive Devices Possible for MIR

Some preliminary MIR devices were simulated based on the concept presented in Sect. 2. Figure 8(a) shows a Si Photonic crystal with R = 0.304 μm and A = 1.012 μm, ordained by a series of line-defects created, which was simulated by FDTD with an input light of 1532 nm. It can be seen that a sharp high intensity resonant peak and a narrow, high intensity band is obtained in the 3-5 μm band (see Fig. 8(b)). The peaks are similar to resonant peaks possible to obtain with defect modes in PhC [1], but in the low-absorption MIR-band with light source at 1532 nm. Potential application of such spectral response lies in the resonance dependent application [20], but at a cheaper cost of material and source.

Further, a waveguide without any discontinuity can be formed in the same PhC, as shown in Fig. 9(a). It can be seen that a high-intensity band is obtained, between 3.6 μm to 4.8 μm wavelengths (see Fig. 9(b)), which can be used for spectroscopic application, for which 3–5 μm wavelength region is mostly sought. Further study maybe performed on the same for sensing and free-space communication.

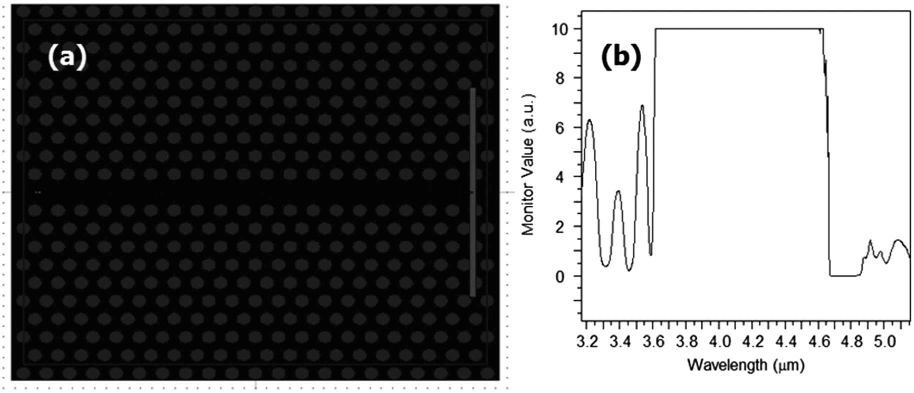


Fig. 9. Si MIR broadband operation study: (a) Waveguide line defect in Si 2-D photonic crystal at 1532 nm light source; (b) high intensity broadband response in 3–5 μm MIR range.

A final study was done in the present work, to find out how the aforesaid wavelength band between 3.6 μm to 4.8 μm might change with varying thickness of the PhC slab, which in effect would vary the effective-index of the slab. The result of the same is presented in Fig. 10(a). It can be seen within the range of thicknesses studied, the band remains unchanged, demonstrating a high stability over thickness. However, with higher thickness, slightly lower intensity was obtained as shown in Fig. 10(b), compared to that shown in Fig. 9(b).

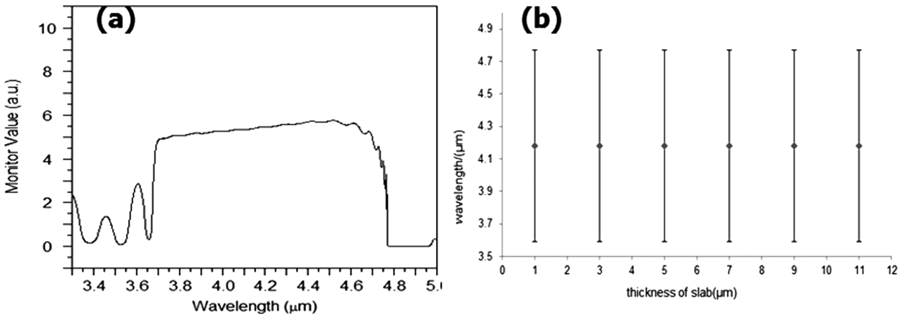


Fig. 10. Si MIR broadband operation study: (a) High intensity broadband response in 3–5 μm MIR range for 2-D PhC shown in Fig. 10 (thickness = period), but with thickness 11 μm, showing lower yet high, intensity; (b) unchanged bandwidth in 3–5 μm with increasing thickness.

Thus, with all the studies presented in the current work, the structural parameters ‘R’ and ‘A’ play the most important part in the optical outputs from the structures. Following the parameters mentioned, wavelength of source has some effect on the span of the PBG. However, the effect of the thickness is largely nil as it is compensated by the effective refractive-index at a given wavelength.

4 Conclusion

A detailed study of the effect of structural parameters of 2-D hexagonal-lattice photonic crystals (PhC) have been presented. The study brings out the fact that in 2-D PhCs, the hole-radii and period between holes are the most important parameters controlling the spectral response of such crystals, which have been hinted earlier. The new fact that came to light from the present work is that it is possible to manipulate the span and location of the photonic band-gap to a great extent. Such tuning capability have been presented as capable of producing long wavelength devices for communication and many other application, using 2-D PhCs made of cheaper material, using cheaper and shorter wavelength sources. The resulting trends show the promise of developing cheap broadband MIR devices as well for spectroscopy and free-space communication, which will be further explored in our future works.

References

1. Joannopoulos, J.D., Johnson, S.G., Winn, J.N., Meade, R.D.: Photonic Crystals: Molding the Flow of Light, 2nd edn. Princeton University Press, Princeton (2008)
2. Kitzerow, H.: Tunable photonic crystals. Liq. Cryst. Today **11**(4), 3–7 (2002)
3. Yablonovitch, E., Gmitter, T.J., Leung, K.M.: Photonic band structure: the face-centered-cubic case employing nonspherical atoms. Phys. Rev. Lett. **67**(17), 2295–2298 (1991)
4. Kim, J.I., Jeon, S.G., Kim, G.J., et al.: Two-dimensional terahertz photonic crystals fabricated by wet chemical etching of silicon. J. Infrared Milli Terahz Waves **33**, 206–211 (2012)

5. Suthar, B., Kumar, V., Kumar, A., Singh, K., Bhargava, A.: Thermal expansion of photonic band gap for one dimensional photonic crystal. Prog. Electromagnet. Res. Lett. **32**, 81–90 (2012)
6. Glushko, A., Karachevtseva, L.: PBG properties of three-component 2D photonic crystals. In: International Conference on Photonics and Nanostructures - Fundamentals and Applications, vol. 4, no. 3, November 2008
7. Chun-Zhen, F., Jun-Qiao, W., Jin-Na, H., Pei, D., Er-Jun, L.: Theoretical study on the photonic band gap in one-dimensional photonic crystals with graded multilayer structure. Chin. Phys. B **22**(7), 074211-1–074211-5 (2013)
8. Leonard, S.W., van Driel, H.M., Schilling, J., Wehrspohn, R.B.: Ultrafast band-edge tuning of a two-dimensional silicon photonic crystal via free-carrier injection. Phy. Rev. B **66**, 161102-1–161102-4 (2002)
9. Muhammad, H.M., Jamil, N.Y., Abdullah, A.I.: Photonic bandgap tuning of photonic crystals by air filling fraction. Raf. J. Sci. **24**(5), 96–102 (2013)
10. Sauvan, C., Lecamp, G., Lalanne, P., Hugonin, J.P.: Modal-reflectivity enhancement by geometry tuning in photonic crystal microcavities. Opt. Exp. **13**(1), 245–255 (2005)
11. Thitsa, M., Albin, S.: Band gap tuning of macro-porous si photonic crystals by thermally grown SiO2 interfacial layer. ECS Trans. **11**(17), 1–9 (2008)
12. [online resource]. www.thorlabs.com. Accessed 30 Mar 2017
13. Pikhtin, A.N., Yaskov, A.D.: Sov. Phys. Semicond. **14**(4), 389–392 (1980)
14. Aspnes, D.E., Theeten, J.B.: Spectroscopic analysis of the interface between si and its thermally grown oxide. J. Electrochem. Soc. **127**(6), 1359–1365 (1980)
15. Frey, B.J., Leviton, D.B., Madison, T.J.: Temperature dependent refractive index of silicon and germanium. In: International Conference on Proceedings of SPIE 6273 (Orlando) (2007)
16. Green, M.A., Keevers, M.J.: Optical properties of intrinsic silicon at 300 K. Prog. Photovoltaics Res. Appl. **3**, 189–192 (1995)
17. Coldren, L.A., Corzine, S.W., Masanovic, M.L.: Introduction to optical waveguiding in simple double-heterostructures. In: Diode Lasers and Photonic integrated circuits, 2nd edn., New Jersey, Ap. 3, sec. 3, pp. 551–554 (2012)
18. [online resource]. https://optics.synopsys.com/rsoft/rsoft-passive-device-FullWAVE.html. Accessed 30 Mar 2017
19. Shankar, R.: Mid-infrared photonics in silicon. Ph.D. thesis, Harvard University, April 2013
20. Lee, C., Thillaigovindan, J.: Optical nanomechanical sensor using a silicon photonic crystal cantilever embedded with a nanocavity resonator. Appl. Opt. **48**(10), 1797–1803 (2009)

Power Aware Network on Chip Test Scheduling with Variable Test Clock Frequency

Harikrishna Parmar(✉) and Usha Mehta

EC Department, Nirma University, Ahmedabad, Gujarat, India
hk_parmar@yahoo.co.in, usha.mehta@nirmauni.ac.in

Abstract. For a stated core, the test time changes in a staircase pattern with the width of Test Access Mechanism (TAM). The core test time cannot decrease all time with increase in TAM width. However, the test time can always be diminished with increasing the test clock speed but clock speed cannot be increased beyond power limits. Here, a new method is proposed to reduce the Network on Chip (NoC) test time, by differing the test clock frequency such that it doesn't cross the predefined power limit. The power dissipation, test clock frequency and overall test time is the three trade off. In the proposed method, the clock frequency is optimized to minimize the total test application time (TAT) considering the power limits. Experimental results show an reduction of 48% over existing solution for the benchmark system on chip (SoC) D695, P93791 and P22810.

Keywords: Power · NoC · TAM
Test clock frequency · Overall test application time (TAT)

1 Introduction

Now a days, huge number of transistors are integrated on a wafer, which shows the certainty of Gordan Moore prediction, who had stated that the number of transistor on a wafer will be doubled on every 18 months. State-of-the-art technologies for manufacturing integrated circuits allow integrating a huge number of transistors on a single chip and reuse of IP cores for the sake of settling the time-to-market issues [1]. However, as the SoC is becoming more and more complex, typical bus based TAM architecture for SoC is subjected to scalable global synchronous clock, performance issues and communication bandwidths [2]. To avoid the limitation of SoC bus based architecture, Network on chip (NoC) system is introduced. Routers, channels, IP Cores and packet switching interconnections make NoC systems ideal to overcome limitations of SoC. [3]. The NoC system dispenses multiple benefits over long-established bus based architecture for its superior parallelism.

NoC is an emerging design paradigm deliberated to cope with future systems-on-chips (SoCs) comprising numerous built-in cores. Since NoC have some excellent distinctive attribute like scalability, design complexity, power dissipation, timing and so on, extensive interest is probable to grow towards NoC. The test strategy is the main

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 256–264, 2018.
https://doi.org/10.1007/978-3-319-73423-1_23

aspect in the feasibility and practicability of the NoC based SoC. In SoC, TAM architecture is designed to fetch the test data from the automatic test equipment (ATE) to the core and to transfer test response from core to sink. Among the existing test objective for NoC based SoC, test scheduling and TAM architecture peculiarly influence the overall test performance of NoC [4].

Since, minimization of test time with NoC as TAM is an intractable problem, requiring the co-optimization of the core assignment to TAM for test data transportation, effective exploitation of the channel bandwidth and the number and location of the test interface. One or more of these features were ignored in the past.

In this paper, a new method is proposed to minimize the test time in NoC by differing the test clock frequency for each test session. Here, the test power and the test time is formulated as a function of test clock frequency, and hence this method gets the test time reduction for the predefined power limit. In the proposed method dynamic clock control based on power dissipation of test session is adopted.

The paper is organized as follow: Sect. 2 covers the prior work similar to the NoC testing whereas Sect. 3 covers the proposed algorithm based on variable test clock rate. Outcomes are discussed in Sect. 4 and then Sect. 5 concludes the paper.

2 Prior Work

Prior work shows the test time as a function of the TAM width and assignment of core to the TAM width to minimize the test time [5]. TAM architecture is the mediator to transfer the test data from automatic test equipment to the core and core to the sink. In [6], it is shown that the test time of the core varies in staircase pattern with TAM width. [5, 7–9] shows the optimal assignment of TAM width to the core under test to reduce test time significantly.

In NoC, NoC fabric can be used as TAM. So, the requirement of dedicated TAM can be avoided here. Since no extra hardware is required to build TAM, it reduces the cost of NoC testing [10, 11]. Number of scheduling algorithm is designed to minimize NoC test time with different constraints are proposed in [12–20].

The fundamental of reusing NoC as TAM are first introduced in [4]. Here, the core having longer test time is given higher priority in scheduling to reduce testing time. This method was further developed in [10] with power constraint and increased test parallelism. The Time division multiplexing (TDM) approach was discussed in [18] to have high speed test data transportation over the network and low speed test execution of NoC core. In [16, 17], Poweraware test scheduling is shown by effectively utilizing on chip network. Here the on chip clocking is used in a smart way such that the faster clock is assigned to some cores and slower to remaining to limit the overall power consumption. In short, clock rate distribution is effectively designed in this methodology to have lower test time. Test scheduling using rectangle packing solution and use of multiple test clocks for NoC test was proposed in [21]. Test scheduling with different topology of network was described in [12]. It also gives the idea of fast wiring test time minimization blueprint for different test structure. In [22], a unicast based multicast complication of NoC core testing is explained, where different techniques like Test data compression, power constraint scheduling, vector compactions are used to minimize test time. Power and thermal aware NoC test scheduling with multiple clock rate is proposed in [23]. The algorithm is designed based on Integer linear programming and simulated annealing technique. Co-optimization of pin assignment to access point and NoC core test scheduling was proposed in [24]. Minimization of test time with given pin count is well described here. In [25], test delivery optimization of many core system is proposed. Here, NoC partitioning difficulty is formulated with dynamic programming. It also emphasize the optimization of the access point location, distribution of automatic test equipment (ATE) to access point and assignment of core to access point. In [26], hybrid test data transportation system for advance NoC based SoC is described. As the scheduler is affected by the location of the access point and the position of the embedded core, a new technique is developed here for concurrently testing several diverse cores.

3 Proposed Method

Assume that there are 'n' numbers of cores $C_1 \ldots C_n$ in an NoC. Individual core is initialized with its test time t_i and power consumed P_i. Maximum power limit of the NoC is P_{max}. Assume that core can be scheduled individually or in a group called sessions. Each session can have more than one core. The length of each session can be defined as

$$L_{Sj} = \max(t_i | \text{for all } t_i \in S_j) \tag{1}$$

And the power dissipated in that session can be defined as

$$P_{Sj} = \sum (P_i), \text{for all } t_i \in S_j \tag{2}$$

Since, routine test scheduling algorithm doesn't give any information regarding test clock frequency, here assume that all test time and power are evaluated on nominal clock frequency f_{nom}. Since, frequency is inversely proportional to the overall test time along with directly proportional to the power, increase in the clock frequency decreases the test time but increase power. Keeping this fundamental in mind, a new idiom is introduced called Frequency factor 'F', which will decide, for how many fold - frequencies should be increased or decreased to have optimum test time of an NoC. To understand this fundamental, two cases are evaluated here.

Case 1: If each core is scheduled individually.

$$\begin{aligned} \text{Frequency factor} \quad = F_{core1} &= P_{max}/P_{core1} \\ F_{core2} &= P_{max}/P_{core2} \\ &. \\ &. \\ &. \\ F_{coren} &= P_{max}/P_{coren} \end{aligned}$$

Case 2: If each core is scheduled in a group called session.

e.g. If core 1, 2 … m of n cores are scheduled in a group, then

$$\text{Frequency factor} = F_{session1} = P_{max}/(P_{core1} + P_{core2} \ldots P_{corem})$$

In both the case, if Frequency factor F is greater than 1 then the test clock frequency will be increased by frequency factor times and if Frequency factor F is less than 1 then the test clock frequency will be decreased by frequency factor times. If F = 1 then test clock frequency will remain unchanged which indicates that all the sessions are executed at the nominal frequency f_{nom}. Now, the test scheduling of sessions can be framed as

Objective: Min $\sum \left(L_{Sj}/F_j\right) * x_j$ for $j = 1$ to k

where $x_j = 1$, if S_j is Scheduled
0, otherwise

Constraints: (1) $P_{Sj} * F_j * x_j \leq P_{max}$, where P_{max} is the power limit for the NoC. (2) Each core, Ci, $i \in \{1, 2, \ldots, n\}$ is made performed at least once.

Here, the first constraint indicates that frequency factor F cannot be increased more than (P_{max}/P_{si}), so that power constraint will not be violated. Power of discrete core can be intensify up to P_{max} but not beyond it. The power limit and test time of SoC D695 is depicted in Table 1.

If each core is scheduled individually, then the lower bound of the test time is set, as the test-clock-frequency is increased till the power consumed for each core is the same as the power limit P_{max}. The results are represented in Table 2 for case II and case I results are shown in Tables 3 and 4 for SoC D695.

3.1 Test Time Calculation

The entire test time is set by

$$T = \max(1 \leq j \leq B) \sum (T_{testi}) \text{ for } i = 1 \text{ to } n \tag{3}$$

where, B is number of test session
T_{testi} is the test−time of all cores on TAMj
n is the overall number of core.

Here, T_{testi} is the combination of two entities 1. Transmit time T_{trai} 2. Test time of core T_{corei}

$$T_{testi} = T_{trai} + T_{corei} \tag{4}$$

Since, transmit time depends on number of channels and routers used in NoC. So, it can be given as,

$$T_{trai} = nb_{chan} * T_{chan} = nb_{ro} * T_{ro} \tag{5}$$

where T_{ro} is the time consumed in router

T_{chan} is the time consumed in NoC channels

nb_{chan} is number of channels

nb_{ro} is number of routers.

Core test time depends on the TAM width selection and arrangement of scan chain with Best Fit Decreasing algorithm [21]. So total core test time is given as

$$T_{corei} = (1 + \max(S_i, S_o)) * p + \min(S_i, S_o) \tag{6}$$

where S_i = Wrapper−scan−in chain

S_o = Wrapper−scan−out chain

P = Test pattern count of the core

In NoC, Testing time of core is considerably higher than the transmit time. So here, transmit time is neglected as in contrast to core test time. Here, S_i and S_o is basically flip flops and it works on the edge of clocks, so the test time measured here is in number of clock cycle it used.

4 Results and Discussions

Here, the proposed algorithm is implemented on three SoC D695, P93791 and P22810 form ITC 2002 benchmark [27]. Since power consumption of each core is not mentioned in ITC 2002 benchmark database, it is taken from [28] where power consumption is calculated from the number of input, output and scan chain. For the proposed algorithm, here it is assumed that NoC has similar configuration as given in [17, 25, 29] like network topology, core placement etc. In all SoC, each core has different combination of scan chain, input, output and circuit structure, so the power consumption varies from core to core. Since power $P = fCV^2$, we are keeping capacitance - C and voltage - V constant and analyzing the effect of changing the frequency on power.

Here, for power constraint test scheduling, maximum power limit is set as the percentage of the gross of total power consumption of sole core i.e. 30% power edge means 30% of summation of total power consumed by each core.

Test database for SoC D695 is shown in Table 1. Column 1 list the core numeral, Column 2 catalogue the test time in clock cycles when TAM width is equal to 32. Core test time is evaluated from Eq. 6. Column 3 lists the power consumption of each individual core.

Here, the simulation is done on MATLAB 14, and LPSOLVE. The results of the proposed algorithm is compared with [17, 25, 29] and shown in Table 5 with SoC D695, P22810 and P93791 respectively with different maximum power limit and different Input output. Column 2–3 shows the results generated from [17] with 2 cases (1) 50% power limits (2) 30% power limit. Results shown in Tables 3 and 4 are the smallest test time achieved ever. Results are compared with [29] and it shows that for 50% power limit, average test time reduces by 48% and for 30% power limit; test time reduces by 24%.

Table 1. Test database for SoC D695

Core	Test clock cycles	Power
Core-1	25	600
Core-2	584	602
Core-3	2475	823
Core-4	5775	275
Core-5	5843	690
Core-6	9828	354
Core-7	3325	530
Core-8	4559	753
Core-9	834	640
Core-10	3859	1144

Table 2. Results with SoC D695 with 50% power limit (core scheduled in sessions)

Core	Power	Time	Frequency factor	Test time
Core-1, 2, 3, 4, 5	2990	5843	1.07	5460
Core-6, 7, 8, 9, 10	3421	9828	0.93	10567
			Total	16027

Table 3. Results with SoC D695 with 50% power limit (core scheduled individually)

Core	Power	Time	Frequency factor	Test time
Core-1	600	25	6.35	4
Core-2	602	584	5.33	110
Core-3	823	2475	3.9	635
Core-4	275	5775	11.067	522
Core-5	690	5843	4.65	1257
Core-6	354	9828	9.06	1085
Core-7	530	3325	6.05	550
Core-8	753	4559	4.26	1070
`Core-9	640	834	5.0	167
Core-10	1144	3859	2.80	1378
			Total	6778

Table 4. Results with Soc D695 with 30% power limit (core scheduled individually)

Core	Power	Time	Frequency factor	Test time
Core-1	600	25	3.2	8
Core-2	602	584	3.19	183
Core-3	823	2475	2.33	353
Core-4	275	5775	7.0	825
Core-5	690	5843	2.78	2101
Core-6	354	9828	5.43	1810
Core-7	530	3325	3.62	919
Core-8	753	4559	2.55	1788
Core-9	640	834	3.00	278
Core-10	1144	3859	1.68	2297
			Total	10562

Table 5. Results with SoC D695, P93791, P22810 with I/O = 2/2

SoC	From ref [29]		Proposed			
	50%	30%	50%	Reduction	30%	Reduction
D695	11927	14250	6778	44%	10562	25%
P93791	443548	444350	203117.3	54%	338534.6	23%
P22810	165302	165302	8447	48%	12389	24%
			Average reduction	48%	Average reduction	24%

5 Conclusion

Here, it is proved that significant test time minimization is achieved by managing the test clock frequency of the test sessions. For the given assumption that the clock frequency confined through the power limit of the NoC, optimization attained is much better. It's also shown that, if the cores are scheduled individually, then the optimum test time is achieved, thereby setting the lower bound of the test time in NoC. Experimental results present an enhancement of 48% on to existing solution for the benchmark SoC D695, P93791 and P22810.

References

1. Moreno, E., Webber, T., Marcon, C., Moraes, F., Calazans, N.: NoC: a monitored network on chip with path adaptation mechanism. Syst. Archit. **60**, 783–795 (2014)
2. Ansari, A., Song, J., Kim, M., Park, S.: Parallel test method for NoC-based SoCs. In: Proceedings IEEE International SoC Design Conference (ISOCC), pp. 116–119 (2009)
3. Touzene, A.: On all-to-all broadcast in dense Gaussian network on-chip. IEEE Trans. Parallel Distrib. Syst. **26**, 1085–1095 (2015)

4. Cota, E.: The impact of NoC reuse on the testing of core-based systems. In: Proceedings of the IEEE VLSI Test Symp, pp. 128–133 (2003)
5. Larsson, E.: Introduction to Advanced System-on-Chip Test Design and Optimization. Springer, Heidelberg (2005). https://doi.org/10.1007/b135763
6. Iyengar, V., Chakrabarty, K., Marinissen, E.: Test wrapper and test access mechanism co-optimization for system-on-chip. J. Electron. Test. Theory Appl. **18**, 213–230 (2002)
7. Iyengar, V., Chakrabarty, K., Marinissen, E.J.: On using rectangle packing for SOC wrapper/TAM co-optimization. In: Proceedings of the 20th IEEE VLSI Test Symposium (2002)
8. Zhao, D., Upadhyaya, S.: Power constrained test scheduling with dynamically varied TAM. In: Proceedings of the 21st IEEE VLSI Test Symposium (2003)
9. Larsson, E., Fujiwara, H.: Power constrained preemptive TAM scheduling. In: Proceedings of the Seventh IEEE European Test Workshop (2002)
10. Cota, E., Carro, L., Lubaszewski, M.: Reusing an on-chip network for the test of core-based systems. ACM Trans. Des. Autom. Electron. Syst. **9**, 471–499 (2004)
11. Cota, E., Liu, C.: Constraint-driven test scheduling for NoC based systems. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **25**, 2465–2478 (2006)
12. Amory, A., Lazzari, C., Lubaszewski, S., Moraes, S.: A new test scheduling algorithm based on networks-on-chip as test access mechanisms. J. Parallel Distrib. Comput. **71**, 675–686 (2011)
13. Chakrabarty, K.: Test scheduling for core based systems using mixed-integer linear programming. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **19**, 1163–1174 (2000)
14. Chattopadhyay, S., Reddy, K.: Genetic algorithm based test scheduling and test access mechanism design for system-on-chips. In: Proceedings of the 16th International Conference on VLSI Design, pp. 341–346 (2003)
15. Iyengar, V., Chakrabarty, K.: System on-a-chip test scheduling with precedence relationships, preemption, and power constraints. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **21**, 1088–1094 (2002)
16. Liu, C., Iyengar, V.: Test scheduling with thermal optimization for network-on-chip systems using variable-rate on-chip clocking. In: Proceedings Design, Automation and Test in Europe Conference and Exhibition (DATE), pp. 650–655 (2006)
17. Liu, C., Shi, J., Cota, E., Iyengar, V.: Power-aware test scheduling in network-on-chip using variable-rate on-chip clocking. In: Proceedings of the 23rd IEEE VLSI Test Symposium (VTS), pp. 349–354 (2005)
18. Nolen, M., Mahapatra, R.: TDM test scheduling method for network-on-chip systems. In: Proceedings of the Sixth International Workshop on Microprocessor Test and Verification (MTV), pp. 90–98 (2005)
19. Su, C., Wu, C.: A graph-based approach to power-constrained SOC test scheduling. J. Electron. Test. Theory Appl. **20**, 45–60 (2004)
20. Zou, W., Reddy, M., Pomeranz, I., Huang, Y.: SOC test scheduling using simulated annealing. In: Proceedings of the 21st IEEE VLSI Test Symposium (VTS), pp. 325–330 (2003)
21. Ahn, J., Sungho, K.: Test scheduling of NoC-based SoCs using multiple test clocks. ETRI J. **28**, 475–485 (2006)
22. Xiang, D., Zhang, Y.: Cost-effective power-aware core testing in NoCs based on a New unicast-based multicast scheme. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **30**, 135–147 (2011)
23. Aktouf, A.: Complete strategy for testing an on-chip multiprocessor architecture. IEEE Des. Test Comput. **19**, 18–28 (2002)

24. Richter, M., Chakrabarty, K.: Optimization of test pin- count, test scheduling, and test access for NoC-based multicore SoCs. IEEE Trans. Comput. **63**, 691–702 (2014)
25. Agrawal, M., Richter, M., Chakrabarty, K.: Test-delivery optimization in manycore SOC. IEEE Trans. Comput.-Aided Des. Integr. Circuits. Syst. 33(7) (2014)
26. Ansari, A., Kim, D., Jung, J., Park, S.: Hybrid test data transportation scheme for advanced NoC-based SoCs. J. Semicond. Technol. Sci. **15**, 85–95 (2015)
27. Marinissen, E., Iyengar, V., Chakrabarty, K.: A set of benchmarks for modular testing of SOCs. In: Proceedings International Test Conference (ITC), pp. 519–528 (2002)
28. Pouget, J., Larsson, E., Peng, Z.: SOC test time minimization under multiple constraints. In: Proceedings of the 12th Asian Test Symposium (ATS), pp. 312–317 (2003)
29. Hu, C., Li, Z., Lu, C., Jia, M.: Test scheduling for network-on-chip using XY-direction connected subgraph partition and multiple test clocks. J. Electron. Test. **32**, 31–42 (2016)

Author Index