Mahdi H. Miraz
Peter Excell
Andrew Ware
Safeeullah Soomro
Maaruf Ali (Eds.)

LNICST

200

LNICST

Emerging Technologies in Computing

First International Conference, iCETiC 2018
London, UK, August 23–24, 2018
Proceedings

EAI

Springer

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 200

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

Mahdi H. Miraz · Peter Excell
Andrew Ware · Safeeullah Soomro
Maaruf Ali (Eds.)

Emerging Technologies in Computing

First International Conference, iCETiC 2018
London, UK, August 23–24, 2018
Proceedings

Springer

Editors
Mahdi H. Miraz
University of Hong Kong
Hong Kong
Hong Kong

Peter Excell
Wrexham Glyndwr University
Bradford
UK

Andrew Ware
University of South Wales
Pontypridd, Mid Glamorgan
UK

Safeeullah Soomro
AMA International University
Salmabad
Bahrain

Maaruf Ali
International Association of Educators
and Researchers (IAER)
London
UK

ISSN 1867-8211 ISSN 1867-822X (electronic)
Lecture Notes of the Institute for Computer Sciences, Social Informatics
and Telecommunications Engineering
ISBN 978-3-319-95449-3 ISBN 978-3-319-95450-9 (eBook)
https://doi.org/10.1007/978-3-319-95450-9

Library of Congress Control Number: 2018947459

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

It is our great pleasure to introduce the Proceedings of the First International Conference on Emerging Technologies in Computing (iCETiC 2018), held during 23rd – 24th August, 2018, at London Metropolitan University, London, UK. This conference drew together researchers and developers from both academia and industry - especially in the domains of computing, networking and communications engineering.

iCETiC 2018 was organised by the International Association for Educators and Researchers (IAER) and technically cosponsored by the IEEE ComSoc Bahrain Chapter as well as the Chester and North Wales Branch of the British Computer Society (BCS).

The technical program of iCETiC 2018 consisted of 26 full papers in oral presentation sessions in the main conference tracks. The conference tracks were:

- Track 1 - Cloud, IoT and Distributed Computing
- Track 2 - Software Engineering
- Track 3 - Communications Engineering and Vehicular Technology
- Track 4 - AI, Expert Systems and Big Data Analytics
- Track 5 - Web Information Systems and Applications
- Track 6 - Security
- Track 7 - Database System and Application
- Track 8 - Economics and Business Engineering
- Track 9 - mLearning and eLearning

Apart from the high quality technical paper presentations, the technical program also featured two keynote speeches and one invited talk. The two keynote speakers were Professor Andrew Ware, Professor of Computing, Faculty of Computing, Engineering and Science, University of South Wales, UK and Professor Andrew Jones, Director of the Cyber Security Centre and Professor of Cyber Security, School of Computer Science, University of Hertfordshire, Hatfield, Hertfordshire, UK.

It was also a great pleasure to work with such an excellent Organising Committee team for their arduous work in arranging and supporting the conference. In particular, we thank the Technical Program Committee, who have completed the peer-review process of technical papers and made a high quality technical programme.

We strongly believe that the iCETiC 2018 Conference provides a good forum for all researchers, developers and practitioners to discuss recent advancements in computing, networking and communications engineering. We also expect that future iCETiC conferences will be as successful and stimulating, as indicated by the contributions presented in this volume.

Thank you.

Yours cordially,

August 2018

Mahdi H. Miraz
Peter S. Excell
Andrew Ware
Safeeullah Soomro
Maaruf Ali

Organization

Steering Committee Co-chairs

Maaruf Ali	International Association of Educators and Researchers (IAER), London, UK
Safeeullah Soomro	AMA International University Bahrain (AMAIUB), Bahrain
Mahdi H. Miraz	The Chinese University of Hong Kong, Hong Kong

Organizing Committee

General Co-chairs

Safeeullah Soomro	AMA International University Bahrain (AMAIUB), Bahrain
Ali Kashif Bashir	University of the Faroe Islands, Denmark

Technical Program Committee Chair

Mahdi H. Miraz	The Chinese University of Hong Kong, Hong Kong

Web, Publicity, and Social Media Chair

Shayma K. Miraz	International Association of Educators and Researchers (IAER), London, UK

Publications Chair

Mahdi H. Miraz	The Chinese University of Hong Kong, Hong Kong

Local Chairs

Anowarul Karim	International Association of Educators and Researchers (IAER), London, UK
Emily Thomas	Wrexham Glyndwr University, UK

Track Chairs

Cloud, IoT, and Distributed Computing Track Chair

Virginia N. L. Franqueira	University of Derby, UK

Software Engineering Track Chair

M. Abdullah-Al-Wadud	King Saud University, Saudi Arabia

Communications Engineering and Vehicular Technology Track Chairs

Bhawani Shankar Chowdhry	Mehran University of Engineering and Technology, Pakistan
Mohab A. Mangoud	University of Bahrain, Bahrain

AI, Expert Systems, and Big Data Analytics Track Chair

Christian Esposito	Università degli Studi di Salerno, Italy

Web Information Systems and Applications Track Chair

Seifedine Kadry	Beirut Arab University (UMB), Beirut, Lebanon

Security Track Chair

Aniello Castiglione	University of Salerno, Italy

Database System and Application Track Chair

Basit Shahzad	King Saud University, Saudi Arabia

Economics and Business Engineering Track Chair

Olga Angelopoulou	University of Hertfordshire, Hatfield, UK

mLearning and eLearning Track Chair

Garfield Southall	University of Chester, UK

General Track Chair

Andrew Jones	University of Hertfordshire, Hatfield, UK

Technical Program Committee

Ajith Abraham	Monash University, Australia
Renaud Lambiotte	University of Oxford, UK
Ajay K. Gupta	Western Michigan University, USA
Ljiljana Trajkovic	Simon Fraser University, Canada
Been-Chian Chien	National University of Tainan, Taiwan
Victor Preciado	University of Pennsylvania, USA
Lin Liu	Tsinghua University, China
Guanghui Wen	Southeast University, China
Nowshad Amin	Solar Energy Research Institute (SERI), Malaysia
Yalin Zheng	University of Liverpool, UK
AbdelRahman H. Hussein	Al-Ahliyya Amman University, Al-salt, Jordan
Ali Kashif Bashir	University of the Faroe Islands, Faroe Islands, Denmark
Rabie Ramadan	University of Ha'il, Saudi Arabia
Vincenza Carchiolo	Università di Catania, Italy

Yousuf M. Islam	Daffodil International University, Bangladesh
Prabhat K. Mahanti	University of New Brunswick, Canada
Massimo Ficco	Università degli Studi della Campania Luigi Vanvitelli, Italy
Rosa María Benito Zafrilla	Universidad Politécnica de Madrid, Spain
Syed Faiz Ahmed	Universiti Kuala Lumpur, Malaysia
Touhid Bhuiyan	Daffodil International University, Bangladesh
Anurag Singh	National Institute of Technology (NIT) Delhi, India
Mohammad Siraj	College of Engineering, King Saud University, Saudi Arabia
Anthony Chukwuemeka Ijeh	American University in the Emirates, UAE
José Javier Ramasco	Insitute for Cross-Disciplinary Physics and Complex Systems (IFISC), Spain
Zi-Ke Zhang	Hangzhou Normal University, China
I-Hsien Ting	National University of Kaohsiung, Taiwan
Francisco Rodrigues	University of São Paulo, Brazil
Amir Rubin	Ben-Gurion University, Israel
Ahmed N. AL Masri	American University in the Emirates, UAE
Ahmed Bin Touq	United Arab Emirates University, UAE
Daniel Onah	University College London, UK
Oussama Hamid	University of Kurdistan Hewlêr, Iraq
Souvik Pal	Elitte College of Engineering, Kolkata, India
Matteo Zignani	Università degli Studi di Milano, Italy
Stephen Uzzo	New York Hall of Science, NY, USA
Ali Hessami	IEEE UKRI Section and Innovation Director at Vega Systems Ltd., UK
Ezendu Ariwa	University of Bedfordshire, UK
Amirrudin Bin Kamsin	University of Malaya, Malaysia
Mohab A. Mangoud	University of Bahrain, Bahrain
Thamer Al-meshhadany	IT Consultant, Canada
Umair Ahmed	Gulf University, Bahrain
Hafiz Abid Mahmood Malik	AMA International University Bahrain, Bahrain
Hafeez Siddiqui	London South Bank University, UK
Valentina E. Balas	Aurel Vlaicu University of Arad, Romania
Mahmood Shah	Coventry University, UK
Aamir Zeb Shaikh	NED University of Engineering and Technology, Pakistan
Farhat Naureen Memon	University of Sindh, Pakistan
Ghulam Ali Mallah	Shah Abdul Latif University, Pakistan
Muhammad Mansoor Alam	University Kuala Lumpur (UniKL), Malaysia
Naveed Ahmed Shaikh	Management Sciences, Szabist, UAE
Fida Hussain Chandio	University of Sindh, UAE
Muhammad Yaqoob Koondhar	Sindh Agriculture University, Pakistan

Madad Ali Shah	Sukkur IBA University, Pakistan
Riaz Ahmed Shaikh	Shah Abdul Latif University, Pakistan
Muniba Memon	Najran University, Saudi Arabia
P. Vijaya	Birla Institute of Technology-Muscat Centre, Waljat College of Applied Sciences, Oman
Mansoor Hyder Depar	Sindh Agriculture University (SAU), Pakistan
Jawdat Alshaer	AMA International University Bahrain, Bahrain

Contents

Cloud, IoT and Distributed Computing

Checkpoints and Requirements Based Cloud Service Ranking

Mohammad Riyaz Belgaum[1(✉)], Shahrulniza Musa[2], Muhammad Alam[2], Mazliham Mohd Su'ud[2], Safeeullah Soomro[1], and Zainab Alansari[1,3]

[1] AMA International University, Salmabad, Kingdom of Bahrain
bmdriyaz@amaiu.edu.bh
[2] Universiti Kuala Lumpur (UniKL), Kuala Lumpur, Malaysia
[3] University of Malaya, Kuala Lumpur, Malaysia

Abstract. Cloud computing and the services offered by cloud computing in the field of Information and Communication Technology has impacted the enterprises and is stimulating at a more prominent pace in the recent years. Various studies have been conducted in the field to meet the client's requirements and raise the quality of services offered. Based on the client's requests and integrating it with load balancing as one of the challenges of cloud computing to be addressed the cloud services are ranked. In order to offer better services to the clients and to maintain the trust, load balancing as one of the criteria in real time scheduling is adopted. In order to attain ranking, different services are required to be invoked in the cloud, the requirement factors and the ranking criteria for each factor has been considered to rank them using entropy analysis of Shannon Furthermore, a framework has been proposed to rank them based on the weights attained by the requirements and the ranking criteria.

Keywords: Checkpoints · Service ranking · Load balancing · Scheduling

1 Introduction

The technological advancements in the field of Information and Communications Technology has brought tremendous changes and has changed the daily life of the customers by enabling him to store, manage and access data from different machines connected through internet and is termed as a cloud. Different models of cloud like private, public and hybrid offer services based on the named functionality and are charged based on the policy pay-as-you use. The Cloud Service Provider (CSP) are responsible for the management of the resources based on the types of services offered by the respective service model like Software as a Service, Platform as a Service and Infrastructure as a Service. Resources could be processing devices, storage devices, specialized tools to perform tasks etc.

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 3–15, 2018.
https://doi.org/10.1007/978-3-319-95450-9_1

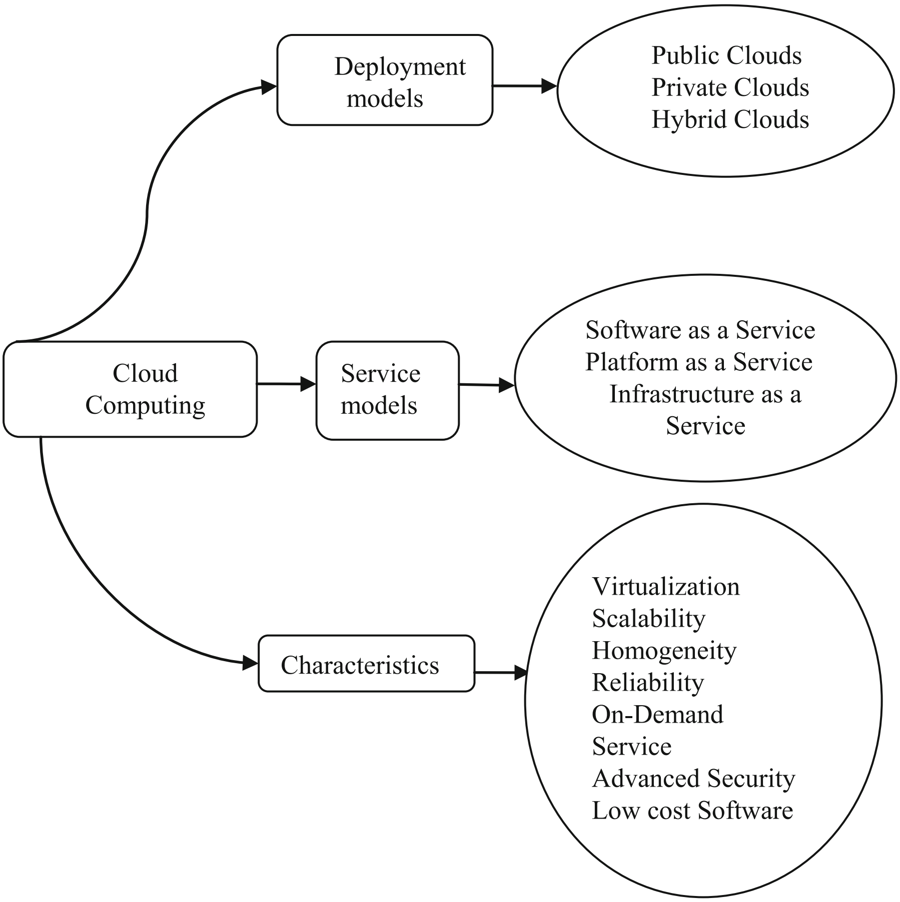


Fig. 1. Cloud computing overview [15].

Figure 1 gives an overview of cloud computing, showing the various deployment models, services offered and the characteristics of cloud computing. There are different regular qualities of Cloud like the virtual processing, extensibility of the machines, though heterogeneous systems are connected showing the homogeneous nature, dependability, wide accepted on demand service, high level security, affordable cost etc. Unmistakable of those are Service level understanding, Sharing of load, safety and security, Quality of service in cloud administrations, costs, and so forth [14].

Figure 2 shows how the tasks are assigned to each Virtual Machine. The point where the system is stable and the data is consistent is called a checkpoint. Point in time (PIT) permits the VMs to roll back to a stable state to meet the client's demand. There are two types of checkpoints as programmed and manual [5]. Programmed checkpoints are taken consistently at regular intervals or as right on time as would be prudent, which are crash reliable and are helpful amid recuperation. Manual checkpoints can be explicitly set at a specific time like at a non-peak hour so as not to effect the user's trust [10]. Therefore, the client has an alternative to do PIT recuperation utilizing manual checkpoints routinely. Executing the checkpoints makes the framework be in a reliable state. Checkpoints are advantageous in an environment where there is non pre-emption along with implementing load balancing.

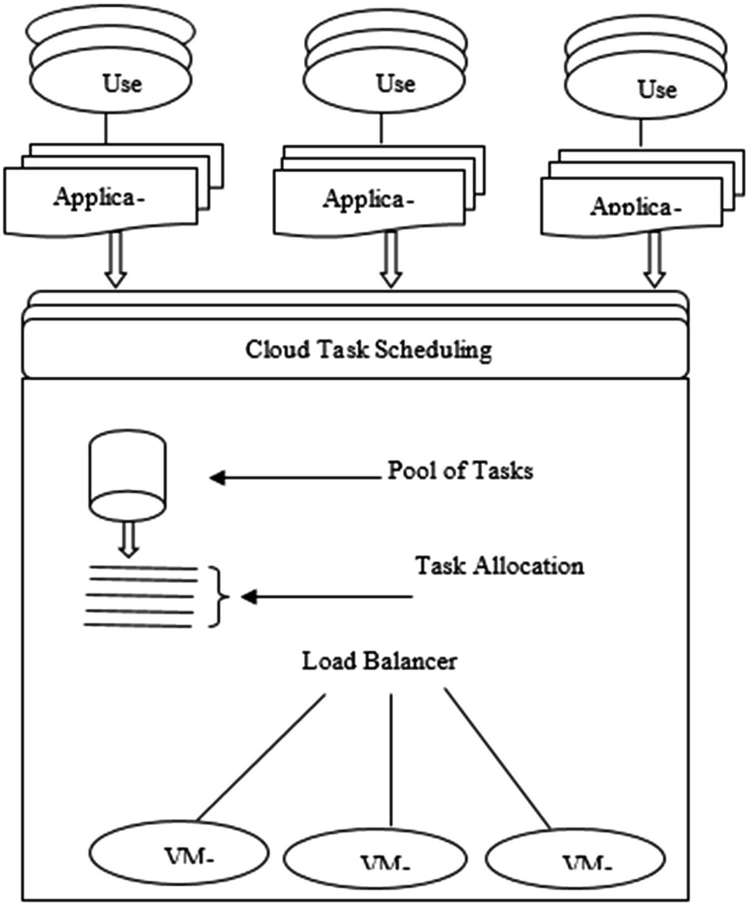


Fig. 2. Cloud architecture for load balancing.

2 Related Work

Various authors have ranked the cloud services using different methodologies. Some have taken the user requirements as a base to rank the cloud service, and some others have made the criteria to rank it [8].

The cloud services being offered by various cloud service providers, there is no standard way to evaluate and select an appropriate service provider. A framework and a mechanism were proposed in order to measure the quality and prioritize the service needed showing its impact on to satisfy the SLA and improve the QoS [6].

In [18], the authors have taken after multi specialist framework (MAS) in cloud computing, keeping in mind the goal to adjust the heap. This framework is an inexactly combined system with the product specialists in order to address the issues which couldn't be settled utilizing singular limit [1]. Creators utilized diverse parameters for load adjusting which are dependability, reconfigurability and capacity to alter [2].

Agreeable Checkpointing talked about in was contrasted and occasional checkpointing with a trial examination which demonstrates the helpful checkpointing helps the application or a procedure to continue promote under a few gathering of disappointment appropriations. Likewise, the agreeable checkpointing is utilized to enhance the unwavering quality procedures like QoS, adaptation to non-critical failure in this manner making the framework strong and increment the execution [12].

The authors proposed a calculation in [20] for decreasing the time taken for migration of the tasks between virtual machines as the faster this process happens so will be the processing. Taking in to consideration the factors like throughput, benefit, misfortune a reproduction is done to demonstrate the non-preemptive ongoing planning checkpointing diminishes the execution time [7].

For keeping the security dangers in distributed computing, the authors in [9] proposed a plan of inside movement checkpoint demonstrate. The proposed model considered three segments which are utilized to distinguish and avoid dangers to make the cloud assets secure. The load balancing techniques alongside the measurements are talked about, addressed the problems in categorizing different types of load calculations are ordered in light of framework load and framework topology [4].

The authors in [11] have proposed another ongoing booking calculation for distributed computing whose point is to have a most extreme utility of the assets utilizing the time utility capacity. Two-time utility capacities in particular benefit and punishment have been utilized as a part of their work. The punishment was utilized to rebuff the assignments that have missed due dates and the benefit was utilized to compensate the errands which have met the due dates.

A preemptive cloud booking calculation was utilized as a part of with a settled need appointed to each assignment keeping in mind the end goal to enhance the QoS [25]. Two variations of preemptive booking calculations were numerically ended up being fit for administration situated errands. In the evaluation process, a dispatcher assumes the part of appropriating the low need errand when a high demand assignment lands with less overhead and keeping up optimality to accomplish QoS [22].

The authors in [13] considered load is adjusting as the primary test alongside accomplishing green processing with the different studies. Given the exponential increment of the distributed computing, the requirement for the server farms expanded which thus is bringing about the abundance of carbon outflows polluting the earth. Different measurements to assess the heap adjusting calculations alongside Carbon discharge and vitality utilization measurements were utilized to show which calculation is productive [17].

Stack Balancing in conjunction with accessibility is examined in alongside with a Hospital Data Management framework. All things considered review the information of a patient should be gotten to by various specialists, nurture internationally from various frameworks when the data of the patient is accessible [24]. An asset director is in charge of the entire operations like observing, accessibility and execution.

The privacy issue and an important challenge in cloud computing is the protection of data. Data being a basic and important element of any organization need to be protected and kept secured. And its security is more important and complicated when it is kept in the cloud as security plays an important role in cloud computing when compared with the traditional way of storing the data [3]. Though encrypted information is a solution to keep the data protected still it is not free from other vulnerable attacks for the reason that it is transmitted over web.

3 Methodology

This study employed a practical descriptive survey and the required data for determining the sample size provided by the decision team. We prepared a list of twelve experts who somehow deals with cloud service and provided four different questionnaires to each to pursue the following:

- Questionnaire 1: was prepared to confirm the proposed structure.
- Questionnaire 2: was developed based on a nine-level scale of Saaty [19] and its aim was the comparison of each two criteria and determining the preference among them.
- Questionnaire 3: was designed as an open-ended which required the respondents to establish the weight of each criterion to its pair from the same group.
- Questionnaire 4: was prepared to implement the interview and its design was based on the computing need and related literature. The level of each criterion graded and identified in a ten-point ranking scale.

In the paired comparison method, as each factor should estimate by the other factors, it assures the consideration of all. Consequently, questionnaires have somehow a logical content validity that is linked with a method used. In the paired comparison method, all factors should be evaluated to each other which remove all the possibilities of not being considered for each criterion. Also, in another used questionnaire, all criterions were assessed and reviewed. Therefore, they also reject the possibility of not considered for being measured. Moreover, the validity and content of the questionnaires were confirmed by the experts and decision teams. Hence it can be said that the used data collection tools in this study have been proved of the content validity.

The theory of Dempster-Shafer is recognized as the most used methods for uncertainty reasoning, modeling, and efficiency of intelligent systems with unstructured data. Dempster founded it, and then Shaffer introduced it as a theory [23]. Besides, uncertainty measurement of a particular situation is one of the most important roles of entropy as a primary concept of big data.

The reliability measure in this study doesn't benefit from the quantitative methods however the reliable estimation of evaluators considers as a factor for analysis of reliability. Nonetheless, the compatibility rating can evaluate the reliability for questionnaire's paired comparison that used Saaty's system.

3.1 Research Process

First, by extensive research in library and literature reviews, the cloud service ranking based on requirements and criteria in a load balancing environment were recognized. Then we used them to create our new conceptual framework which consists of eight requirements factor which is in a cycle with eight ranking criteria. The structure of the proposed new conceptual framework is displayed in Fig. 3. Then we ranked each factor using entropy analysis of Shannon [21].

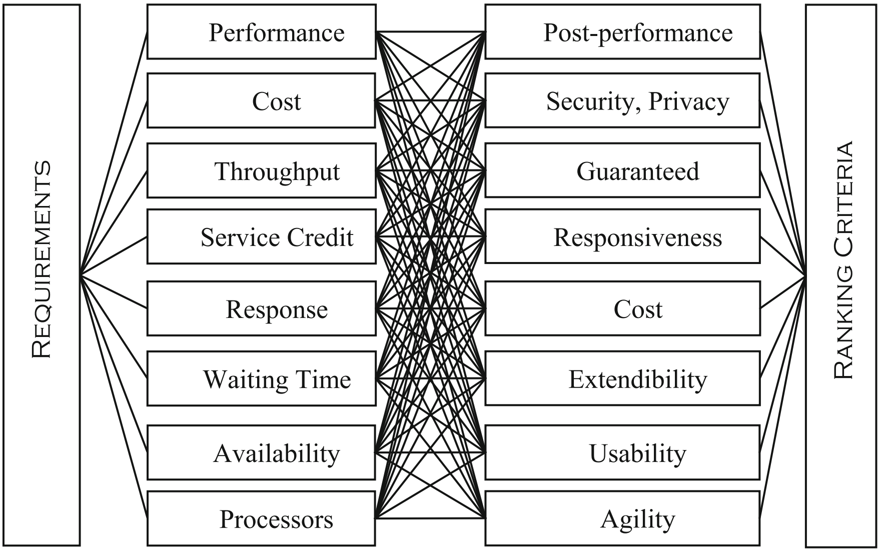


Fig. 3. Cloud service ranking evaluation mapping in load balancing environment.

By developing specific matrices to each questionnaire, we could accomplish a valid result. Due to the formation of multiple matrices, to obtain the final matrix we had to use the geometric mean for the variable of each matrix and calculated the compatibility rank which is used as an input for entropy Shannon and computed the final weight of each variable.

The ranking of cloud service should be evaluated based on requirements and criteria in load balancing environment. By collecting the acknowledged data from the fourth questionnaire to be used as input in fuzzy functions, the next level starts which is data analysis and evaluation of final decision. Remarking that the method used at this stage is proposed by [26]. Uncertainty is a primary motivation for the fuzzy sets. By achieving the rank of each factor, it is time to combine factors in the same rank. Therefore, we reduce the factors to increase the certainty given to each. Dempster-Shafer theory revised the combination conflict. Given that this study contains some conflicts, the averaging method of Murphy [16] is used to overcome the conflicts. Murphy recommended that if all shreds of evidence are available concurrently, calculate the weight average and determine the final weight by joining the averaged values several times. Therefore, using this method, we prevent the over-dependence to a section of conflicting evidence.

3.2 Proposed Algorithm

The following section explains computation of the cloud service ranking using checkpoint based load balancing and then considering the various requirement factors and the criteria to rank the services. The following formulae are used to evaluate the node correspondence value, value for the preferred node and the priority value for a service at a node. The correspondence value of node is evaluated by Eq. (1):

$$\mathrm{CV}(\chi, \mathrm{y}) = \frac{\mathrm{a} - \mathrm{b}}{\mathrm{n}(\mathrm{n}-1)/2} \tag{1}$$

Where n is total services, a is a total number of consistent pairs and b is a total number of variant pairs among two lists, $n(n-1)/2$ are the total number of pairs in the cloud with n services. Preferred nodes among the correspondent nodes are selected by subtracting the ranks of services.

$$\mathrm{P}(\chi, \mathrm{y}) = \mathrm{S}_\chi - \mathrm{S}_\mathrm{y}. \tag{2}$$

Where $P(\chi, y)$ = prefer value among node x and y, S_χ = rank of node x's service, S_y = rank of node y's service. The greater prefer value indicates that the service is more reliable than the other service.

$$\mathrm{PV} = \sum\nolimits_{\mathrm{y} \in \mathrm{S}} \mathrm{P}(\chi, \mathrm{y}) \tag{3}$$

Where PV = priority value of service x. The system then arranges the list having services with greater priority values higher in the list. To improve the accuracy of rank prediction of services the system prefers the higher priority values of implicit services which the user has already accessed.

Algorithm 1. Proposed Algorithm
Input: A set of service S, x is a cloud service and π stacks in the ranking.
Output: ranked service list x
Step 1: for each service from 1 to n
Step 2: calculate correspondence value of each service based on user's requirements using eq. (1)
Step 3: end for
Step 4: for each service from 1 to n
Step5: calculate prefer value of each service using eq. (2)
Step 6: end for
Step 7: R=S;
Step 8: for each service from 1 to n
Step 9: rank each service by checkpoints and the load balancing, present on the cloud,
x = max rank in S,
π(x) = S-R+1;
R=R-x;
Step 10: end for
Step 11: for each service from 1 to n
Step 12: Calculate the priority value of each service using eq. (3)
Step 13: rank the services by their priority values,
R=μ(i)
a= max rank in priority value set, μ(i),
π(x) = μ(i)-R+1,
R=R-x,
Step 14: prioritize the implicit services with greater rank.
Step 15: update the service set S with the ranked services and save it in ranked service list x
Step 16: end for

4 Analysis and Results

In this section, we brought just some sample results due to the high volume of outputs. According to the decision matrix, we calculated the weight of each requirement factors as shown in Table 1:

Table 1. The weight of requirement's factors.

Factors	Entropy value	Uncertainty value	Factor's weight	Intellectual weight	Adjusted weight
A1	0.533	0.145	0.143	0.132	0.113
A2	0.422	0.137	0.063	0.086	0.136
A3	0.539	0.135	0.106	0.193	0.105
A4	0.544	0.118	0.139	0.16	0.127
A5	0.441	0.123	0.161	0.066	0.091
A6	0.432	0.093	0.091	0.103	0.145
A7	0.59	0.144	0.173	0.118	0.155
A8	0.499	0.105	0.124	0.142	0.128

Similarly, we gained the weight of each ranking criteria as shown in Table 2:

Table 2. The weight of ranking criteria

Factors	Entropy value	Uncertainty value	Factor's weight	Intellectual weight	Adjusted weight
B1	0.542	0.118	0.063	0.174	0.113
B2	0.414	0.091	0.107	0.071	0.136
B3	0.517	0.128	0.129	0.124	0.105
B4	0.538	0.154	0.116	0.179	0.127
B5	0.523	0.113	0.175	0.063	0.091
B6	0.556	0.152	0.153	0.154	0.145
B7	0.436	0.103	0.169	0.091	0.155
B8	0.474	0.141	0.088	0.144	0.128

Based on Tables 1 and 2 and after calculating the mean of entropy, uncertainty values, factor's weight with intellectual and adjusted weight, we found that almost all the sixteen factors have near rating and similarities which are shown in Fig. 4.

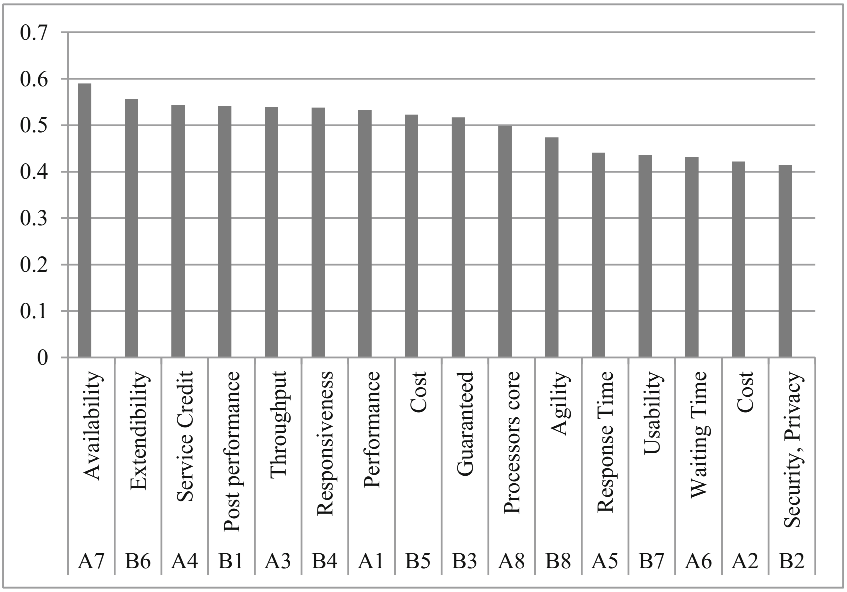


Fig. 4. Cloud service ranking based on requirements and criteria in load balancing environment.

4.1 Factor's Evaluation

After calculating each factors' weight, the ranking of each factor has been evaluated based on the conducted interviews with experts. Table 3 shows the results:

Table 3. Factor's ranking

Factors	Description	Factor's ranking
A1	Performance	H
A2	Cost	VL
A3	Throughput	H
A4	Service credit	VH
A5	Response time	L
A6	Waiting time	VL
A7	Availability	VH
A8	Processors core	M
B1	Post-performance	VH
B2	Security, privacy	VL
B3	Guaranteed	M
B4	Responsiveness	H
B5	Cost	M
B6	Extendibility	VH
B7	Usability	L
B8	Agility	L

By determining the factor's score, the next step is to combine the factors of each group. For this purpose, the following five diagnosis hypotheses were considered:

$$\theta = \{(\mathrm{VL})\ \text{Very Low},\ (\mathrm{L})\ \text{Low},\ (\mathrm{M})\ \text{Medium},\ (\mathrm{H})\ \text{High},\ (\mathrm{VH})\ \text{Very High}\}$$

Each one of these is intimating the cloud service ranking based on requirements and criteria in load balancing environment and applied as input in Dempster-Shafer Theory. Remarking, these evidence are preliminary and uncertain for the combination. Therefore, they need to be reduced first. By synthesizing evidence, almost 100% assurance allocated to an individual factor which several ways have been offered for facing such conflicts by other researchers. In this study, we used Murphy's proposed idea. The results are shown in Table 4:

Table 4. Overall evaluation of cloud service ranking based on requirements and criteria in load balancing environment

Combination of evidence	VL	L	M	H	VH	VL, L	L, M	M, H	H, VH
A1, A2, A3, A4	0.17	0.00	0.00	0.43	0.22	0.08	0.00	0.21	0.32
B1, B2, B3, B4	0.17	0.00	0.21	0.22	0.22	0.08	0.10	0.21	0.22
A5, A6, A7, A8	0.17	0.18	0.20	0.00	0.24	0.17	0.19	0.10	0.12
B5, B6, B7, B8	0.00	0.36	0.21	0.00	0.22	0.18	0.29	0.10	0.11
Average	0.13	0.14	0.15	0.16	0.22	0.13	0.14	0.16	0.19

5 Conclusion

This methodology is an extended version of the article entitled “Cloud Service ranking using Checkpoint based Load balancing in real time scheduling of Cloud Computing.” A new framework is proposed to evaluate the cloud service ranking based on the requirements and criteria. In the previous work, a mathematical evaluation was proposed for evaluation and the factors affecting them were not considered. The study, by the use of interviews, expert opinions and review of previously related researches provided a new framework model which includes the requirements and criteria to evaluate the cloud service provided by CSP. Each of these considered eight sub contents in them for evaluation to prioritize a cloud service. This research prioritizes aspects of the conceptual model using modified Fuzzy Analytic Hierarchy Process.

Undoubtedly, this research can be the basis for selecting the cloud service rendered by a specific Cloud Service Provider. As all the CSPs are providing the same kinds of service by balancing the load at their end, the customer is confused in selecting a specific CSP for the service intended. Therefore this model can be a decision making tool in selection of a service with CSP. On the other hand, it helps the Cloud Service Providers to enhance themselves in the competitive market for reaching the targets and improving their businesses. Then Cloud Services ranking enables the customers to adopt the service with high raking to satisfy their requirements. The Quality of Service and optimization of resources at the data center should be improved to meet the increasing demands. Whenever the CSPs fail to meet the criteria, the enterprises always have an option to migrate to other service providers resulting in perishing the business.

6 Future Studies

Although the Cloud Computing has made the tasks easier for the enterprises, still the trust and security is a big challenge. The Service level Agreements are in its place to guarantee the enterprises regarding the performance yet still there lacks reliability and efficiency. The lack of support from the management of the enterprises also can have adverse effects on performance and usage. Furthermore, Cloud Computing needs standards to be benchmarked and provide the service to the customers. With the increasing demand of adoption of cloud, the need of protecting these resources and data against cyber attacks also increase giving scope for the researchers to propose solutions for protection.

References

1. Agar, J., Smith, C. (eds.): Making Space for Science: Territorial Themes in the Shaping of Knowledge. Springer, London (2016). https://doi.org/10.1007/978-1-349-26324-0
2. Ahmed, T., Dubois, E., Dupé, J.B., Ferrús, R., Gélard, P., Kuhn, N.: Software defined satellite cloud RAN. Int. J. Satell. Commun. Netw. **36**(1), 108–133 (2017)
3. Ali, M., Miraz, M.H.: Recent Advances in cloud computing applications and services. Int. J. Cloud Comput. (IJCC) **1**(1), 1–12 (2014)

4. Bala, K., Kumar, A.: A hybrid approach for load balancing: using random forest and PSO approach (RFPSO). Int. J. **8**(5), 1554–1559 (2017)
5. Breytenbach, A.: Comparative accuracy evaluation of fine-scale global and local digital surface models: the Tshwane case study I. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **4**(2), 211–222 (2016)
6. Calheiros, R.N., Masoumi, E., Ranjan, R., Buyya, R.: Workload prediction using ARIMA model and its impact on cloud applications' QoS. IEEE Trans. Cloud Comput. **3**(4), 449–458 (2015)
7. Clark, T.R.: Leading with Character and Competence: Moving Beyond Title, Position, and Authority. Berrett-Koehler Publishers, Oakland (2016)
8. Coutinho, E.F., de Carvalho Sousa, F.R., Rego, P.A.L., Gomes, D.G., de Souza, J.N.: Elasticity in cloud computing: a survey. Ann. Telecommun. annales des télécommunications **70**(7–8), 289–309 (2015)
9. Gonzales, D., Kaplan, J., Saltzman, E., Winkelman, Z., Woods, D.: Cloud-trust-a security assessment model for infrastructure as a service (IaaS) clouds. IEEE Trans. Cloud Comput. **5**, 523–536 (2015)
10. Gupta, G.: Semantically ordered parallel execution of multiprocessor programs Doctoral dissertation, The University of Wisconsin-Madison (2015)
11. Jain, N., Menache, I., Naor, J.S., Yaniv, J.: Near-optimal scheduling mechanisms for deadline-sensitive jobs in large computing clusters. ACM Trans. Parallel Comput. **2**(1), 3 (2015)
12. Jain, S.: Security and trust in mobile ad-hoc networks. Doctoral dissertation, University of Maryland, College Park (2015)
13. Kansal, N.J., Chan, I.: Cloud load balancing techniques: a step towards green computing. Int. J. Comput. Sci. Issues **9**(1), 238–246 (2012)
14. Kumar, B., Boaddh, J., Mahawar, L.: A hybrid security approach based on AES and RSA for cloud data. Int. J. Adv. Technol. Eng. Explor. **3**(17), 43 (2016)
15. Belgaum, M.R., Soomro, S., Alansari, Z., Alam, M.: Cloud service ranking using checkpoint-based load balancing in real-time scheduling of cloud computing. In: Saeed, K., Chaki, N., Pati, B., Bakshi, S., Mohapatra, D.P. (eds.) Progress in Advanced Computing and Intelligent Engineering. AISC, vol. 563, pp. 667–676. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6872-0_64
16. Murphy, C.K.: Combining belief functions when evidence conflicts. Decis. Support Syst. **29**(1), 1–9 (2000)
17. Pal, J., Dixit, S., Pathik, B., Sahu, S.K.: A review of load balancing algorithm based on swarm intelligence in heterogeneous cloud environment. Imp. J. Interdisc. Res. **3**(1), 414–419 (2016)
18. Pandey, R., Ranjan, R.M.S.: Distributed load balancing in cloud computing. In: International Conference on Computer Science and Information Technology, pp. 32–36 (2013)
19. Saaty, T.L.: What is the analytic hierarchy process? In: Mitra, G., Greenberg, H.J., Lootsma, F.A., Rijkaert, M.J., Zimmermann, H.J. (eds.) Mathematical Models for Decision Support, pp. 109–121. Springer, Heidelberg (1988). https://doi.org/10.1007/978-3-642-83555-1_5
20. Santosh, R., Ravichandran, T.: Non preemptive realtime scheduling using checkpointing algorithm for cloud computing. Int. J. Comput. Appl. **80**(9), 1–4 (2013)
21. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. **5**(1), 3–55 (2001)
22. Singh, S., Chana, I.: QRSF: QoS-aware resource scheduling framework in cloud computing. J. Supercomput. **71**(1), 241–292 (2015)

23. Yoshimura, A.: An autoethnography of Kin-aesthetics: retrieving family Folklore through the wearing of used Kimonos. Doctoral dissertation, The University of Wisconsin-Madison (2015)
24. Alansari, Z., Soomro, S., Belgaum, M.R., Shamshirband, S.: The rise of Internet of Things (IoT) in big healthcare data: review and open research issues. In: Saeed, K., Chaki, N., Pati, B., Bakshi, S., Mohapatra, D.P. (eds.) Progress in Advanced Computing and Intelligent Engineering. AISC, vol. 564, pp. 675–685. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6875-1_66
25. Zeyad M.A., Mohammad R.B.: An enhanced multipath strategy in mobile ad hoc routing protocols. In: 2017 9th IEEE-GCC Conference and Exhibition (GCCCE), pp. 1088–1093 (2017)
26. Zhang, Y., Deng, X., Wei, D., Deng, Y.: Assessment of E-Commerce security using AHP and evidential reasoning. Expert Syst. Appl. **39**(3), 3611–3623 (2012)

Risks in Adopting Cloud Computing: A Proposed Conceptual Framework

Ali Al-Badi[1], Ali Tarhini[1(✉)], and Nabeel Al-Qirim[2]

[1] Information Systems Department, Sultan Qaboos University, Muscat, Oman
{aalbadi,alitarhini}@squ.edu.om
[2] Department of Information Systems and Security, UAE University, Al Ain, UAE
Nalqirim@uaeu.ac.ae

Abstract. Cloud computing has become highly strategic and necessary technology in the IT industry. Cloud computing provide many benefits to organizations but there are risks associated with it which hamper its adoption. Therefore, it is important in this research identify the risks which negatively affect cloud computing adoption decision in order to accelerate its adoption. The research has reviewed relevant literature and accordingly selected fifty research papers from leading information systems journals and conferences. Accordingly, several critical factors were identified. The most important critical factors are grouped into three different categories namely, legal (data privacy, compliance and regulations), technical (bandwidth, data integration, security, vendor lock-in) and operational (loss of control over the services, lack of equipment and knowledge, business continuity and disaster recovery) risks. Finally, a conceptual framework on cloud computing adoption risks is proposed based on those identified factors. This research is of great importance to researchers, cloud computing professionals and policymakers. It will also help in formulating strategies to encourage the adoption and acceptance of cloud computing services, where cloud computing is still considered a risky endeavor and outcomes are seen as uncertain.

Keywords: Cloud computing · Adoption · Risk management
Cloud computing inhibitors · Cloud computing barriers · Deployment models
Service models · Framework

1 Introduction

Cloud computing has gained popularity in recent years. The benefits of the cloud computing include scalability, availability, and cost saving. The cloud computing adoption enable organization to gain competitive advantage in the market (Bisong and Rahman 2011). Several statistics shows the important growth and future of cloud computing. Globally 80% of enterprises will be using cloud computing service (IaaS) and private clouds by 2016 (Illsley 2014). Spending in public cloud will double to $127.5 by 2018 (Leopold 2014). Cloud computing has reshaped the way of managing and acquiring IT resources in more efficient and profitable ways (Phaphoom et al. 2015).

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 16–37, 2018.
https://doi.org/10.1007/978-3-319-95450-9_2

There is a transformation of delivering IT services. The new cloud-computing paradigm helps enterprises concentrate more on their core businesses and hence, increasing their productivity. Cloud computing is based on existing technologies like virtualization and grid computing (Oliveira et al. 2014). Virtualization refers to the creation of virtual version of the computing resources like storage device, operating system and hardware platform that hides the physical characteristics of such resources. Grid computing refers to the aggregation of distributed systems available in different geographical locations allowing users to use the system from almost anywhere.

However, there are some risks associated with this technology. Enterprise Management Associates (EMA) study covered 400 IT professionals around the world. The findings showed that the major risks included difficulties in management of the cloud service, services downtime, cloud provider support, performance and pricing model (EMA 2014). Security, cost, compliance, cloud service reliability, and limited configuration are considered global challenges in cloud computing (Internap 2014). KPMG's international Global cloud survey found several adoption challenges such as data loss, lack of visibility on future costs, legal and regularity compliance and interoperability (KPMG 2013). Other researchers pointed to other challenges that potential clouds computing users should avoid (Dave 2013). According to Research in Action CIO revealed that hidden costs relating to cloud computing adoption an important factor as well.

Cloud computing enables organizations adapt to changing business environments (Armbrust et al. 2010). Around 94% of the CIO have accepted importance of cloud computing (Dave 2013). Nowadays many cloud services are available for free and used by many users, e.g., Dropbox, Google Doc and Google app., and social networking sites like Facebook (Bernard 2011).

Cloud computing offers many benefits to enterprise (Al-Qirim 2011; Phaphoom et al. 2015). The important advantages of cloud computing include on-demand self-service, ubiquitous network access, location-independent resource pooling, rapid elasticity, and measured service (Takabi et al. 2010). The basic characteristics of cloud computing are accessibility, ease of scale-up and -down of resources, and charging for the service based on consumption. The pay as you use model is an attractive method pushing the organization to adopt cloud computing. The cost saved by using cloud computing could be three to five times less than existing infrastructure (Han 2010). Cloud computing helps to reduce the up-front cost needed to purchase the hardware and software (Armbrust et al. 2010). The other advantage of cloud computing is data integration across the networks (Ali et al. 2016). While the adoption of cloud computing keeps growing, there are still some skepticisms about its advantages. Existing research highlighted factor that are hindering cloud computing adoption such as data security, loose of control over the IT service, regulatory and compliance issues, lack of experience and knowledge about it from business and IT managers, compromised accounts' details and business continuity and disaster recovery challenges (Priyadarshinee et al. 2017). The same authors found perceived IT security risk (PITR), risk analysis (RA), technology innovation (TI), management style (MS) and trust (T) as main challenges in cloud computing adoption. Ackermann (2012) reported that IT security risks as important factors in cloud computing adoption. There is a relationship between perceived IT security risk and cloud computing adoption (Haile and Altmann 2016).

The risks associated with cloud computing from customers' point of view consisted of unavailability of data, unauthorized access to data (damage, modification, disclosure of data), non-provision of services (outages, defects, not ensuring continuity and restoration of services) (Fortinová 2013). In Saudi Arabia privacy, trust, and security are reported as important determinants of cloud computing adoption in private sector (Alkhater et al. 2018). Another kind of risk is the high amount of capital needed to invest in cloud computing (Ali et al. 2016). Cloud computing usage amongst government agencies is less than the private sector as they are more concerned about data security and risks of exposing their sensitive data to untrusted entities (Ali et al. 2016).

Given the gloom surrounding cloud computing research and its results, it is hoped here to focus on deterrents only as a proxy here to detect for cloud computing adoption. Therefore, by focusing on deterrents, the research will identify the factors that affects negatively cloud computing adoption decision. The problem that this research is attempting to resolve here is to shed more light on the relationship between such impediments and cloud computing adoption (Khajeh-Hosseini et al. 2010). The literature reported that IT managers fail to take decisions about adopting new technologies (Bisong and Rahman 2011). It is hypothesized here that by capitalizing on deterrents alone, this will provide sufficient help and guidelines for IT managers to mitigate risks and improve their adoption decisions.

The research aims to answer the following two questions:

1. What are the major risk-factors that influence the decision to adopt cloud computing?
2. What precautions are required to mitigate such risks associated with cloud computing adoption?

Cloud computing has been investigated from different perspectives: organizational level, individual level and technical level relating to cloud services specifics (IaaS, PaaS, SaaS). This study is limited to the adoption of cloud computing at the organization level only. The objective of this research is to identify the different cloud computing risks that influence cloud computing adoption decisions in organizations. The research will examine the influencers of cloud computing adoption (independent variables) from four different aspects technical, security, financial and legal and organization.

2 Literature Review

This section describes the cloud computing concepts, cloud-computing trends, cloud computing risks and, proposed conceptual framework.

2.1 Cloud Computing Concepts

Cloud computing has attained greater popularity in recent years. Cloud computing is defined by NIST as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models" (Mell and Grance 2011). There are four deployment methods:

1. *Public Cloud*: The ownership of the infrastructure is dedicated to the cloud provider. Amazon, Microsoft Azure and Rackspace are example of public cloud providers.
2. *Private Cloud*: The services are offered to specific client through private network. The offered e.g., hardware service is dedicated to one client in a single-tenant environment. Normally this method is used with the application that contain sensitive data especially for government and legal organizations. Therefore, it ensures the full control in the hand of the organization. The infrastructure is owned by the organization itself.
3. *Hybrid Cloud*: It is a combination of two or more private, public or community clouds. Normally, the organization will keep the critical application as private. Here the organization can control the cost and keep the security at an acceptable level based on the organization requirements.
4. *Community Cloud*: The cloud infrastructure is shared by many organizations, which have similar requirements like policy and compliance standards. The ownership of the infrastructure can be on one or more inside the community.

There are three main service models of cloud: Software as Service (SaaS), Platform as Service (PaaS), and Infrastructure as Service (IaaS).

1. *Software as Service (SaaS)*: Applications or software's are hosted somewhere and delivered via web browser like Google Doc, Salesforce CRM and Microsoft Exchange. Usually the cloud provider offer the software based on the user demand through license model. It eliminates the upfront investment by charging the user per usage and demand. Also, clients don't spend effort and time in the software support, maintenance and patch since it is provider's responsibility. Normally, the user cannot change the basic configuration of the software since it is provider's accountability. For example, web content management, social networking and video conferencing.
2. *Platform as Service (PaaS)*: It provides a platform that allows developing, running and deploying the application without maintaining the low-level infrastructure. The PaaS consists of a set of a development environment (servers used by programmers to continuously develop applications), testing and/or Q&A environment (set of servers is used by testers) (Fortinová 2013). Google app Engine and Microsoft Azure are example of the PaaS. These allow the developers or customers to manage database, application server and other middleware remotely. Therefore, the cloud client can develop the application without installing the software or purchasing any hardware. The cloud providers provide all required development toolkits, which help to eliminate any upfront cost for buying hardware or software and saves time. The potential risk of this service model is that each provider has its own platform and development language and this creates difficulties to move to another vendor.
3. *Infrastructure as Service (IaaS)*: It provides full infrastructure services such as server, storage, virtualized resources over the internet like virtual machine and

operating system instead of buying the physical hardware and software in the house. These allow organizations to respond quickly to the change through scaling up and scaling down easily. Amazon Elastic Compute Cloud (EC2) and Rackspace are example of IaaS. The implementation of IaaS depends on virtualization technologies, as well as grids or clusters (Fortinová 2013).

The importance of the cloud is derived from its essential characteristics. The basic characteristics of the cloud computing according to NIST definition (Mell and Grance 2011) are as follows:

- *Self-Service* (Client can provision and de-provision cloud services without human interaction. Client can order and manage cloud services through web page).
- *Broad Network Access* (Cloud services can be accessed through network through internet) *Resource Pooling* (Cloud computing resources like storage, virtual machine, memory and processing are shared among different clients)
- *Elasticity* (Client can scale up and down computing resources)
- *Measured Service* (Client can control and optimize the usage of the cloud services. Pay as you use model used to calculate the client service usage)

2.2 Cloud Computing Trends

Several statistic data shows positive trend towards cloud computing adoption around the world. More than 60% of the organizations will have half of the infrastructure on the cloud-based platform by 2018 (McNee 2014). Software as a service (SaaS) is expected to continue at 20% rise each year throughout 2018 (Comella-Dorda et al. 2015).

Cloud computing is highly beneficial to clients, reduced investment cost in infrastructure, offer increased scalability, improve accessibility from anywhere and anytime, and reduced the number of needed skilled employees (Erl et al. 2013, Siefker and Lucas 2013). Cost reduction is considered as the most important benefit (KPMG 2013). Pay as you use in cloud computing allows charging the clients as resources used on hourly basis. Thus saving costs as compared to the high cost of owning the hardware or software (Armbrust et al. 2010). The other type of benefits include time saving to implement new solutions and avoid over provisioning of the resources (Archer 2015). Cloud computing adoption is influenced by flexibility, accessibility, and low investments at the beginning of the project (Wease et al. 2018, Nayar and Kumar 2018). But moving from traditional to cloud computing system is still seen as a challenge (Al-Shamsi and Al-Qirim 2016; Fahmideh and Beydoun 2018). Security and privacy are the main issues (Ahmad and Jolly 2018, Garrison et al. 2018) such as security challenges in the integration of Internet of Things (IoT) and cloud computing (Stergiou et al. 2018).

Cloud based services can eliminate the up-front cost of the hardware and software. It enables the organizations focus more on their business. Moreover, cloud providers help focus on applying the security procedures in more effective and efficient manner, since it is a centralized environment. System availability in multiple locations is ensured in the cloud.

2.3 Cloud Computing Risk

This section provides an overview of previously published research in cloud computing risks. Those are slit into two parts. Firstly, the articles published by IT companies on the inhibitors and risks of adopting cloud computing. Secondly, this part discusses the research that explored the risks associated with cloud computing adopting decisions. In addition, cloud computing adoption risks are presented in a frequency table (below) to understand the common risks.

Most of the past research studies have not illustrated risks as a challenge to cloud computing adoption (Priyadarshinee et al. 2017). According to a research conducted by “Research in Action”, 79% of the CIOs were suspecting hidden costs as hindering cloud computing adoption. The CIOs were also concerned about the performance that would negatively affects user experience (Siefker and Lucas 2013). Low performance is one of the reported obstacles in cloud computing adoption (Armbrust et al. 2009). A survey conducted by Symantec concluded that potential users should avoid risks before adopting cloud computing (Dave 2013).

It is difficult to restore the backup data from the cloud and confidential information are often exposed (Dave 2013). Vendor lock or interoperability is another concern for the client (Ghanam et al. 2012, Marinescu 2013). Lack of standards among the different vendors force the client to tie with one vendor (Marinescu 2013). Enterprise Management Associates surveyed 400 IT professionals around the world and concluded that the major risks faced by clients were difficulties in management of the cloud service, services downtime, cloud provider’s support, performance and pricing models (EMA 2014).

Internap Network Services Corporation released findings of its global survey showing challenges and risks faced by cloud clients which included security, compliance, cost increase, compliance, cloud service reliability and limited configuration (Internap 2014). KPMG international Global cloud survey addressed the adoption challenges, including data loss, lack of visibility of the future cost, legal and regularity compliance and interoperability (KPMG 2013). In 2014, there was a downtime and service outage in the cloud services e.g., Microsoft Azure service storage went down for 10.89 h, Google cloud computing platform went down for 4 h and Amazon storage services went down for fewer than 5 h (Coldewey 2014). Table 1 highlights important cloud computing risks as discussed by articles published by the IT industry.

Risks in Moving to Cloud Computing from an Industrial Viewpoint

Evaluating the business perspective in the cloud is another study where the report listed cloud trends, factors deriving transformation, workforce mobility, cloud and data analytical, cloud challenges and tips to raise the success of cloud transformation (KPMG 2014). The cloud security alliance noted that data security as the top barrier to cloud adoption (Coles et al. 2015). Enterprise Cloud Computing conducted a study in 2014 on cloud trends and challenges faced by decision-makers in planning and adoption phase, adoption drivers from business perspective and effects of cloud adoption on business strategy (IDGEnterprise 2014). The study of the future of cloud computing conducted by North Bridge and partners explored the key drivers, inhibitors and trends of cloud computing (NorthBridge 2014). Cloud Connect and Everest Group conducted a survey

Table 1. Industrial papers

S. no	Organization	Year	Method	Num	Geography	Cloud Computing Risks and their percentage
1	Cloud Security Alliance (CSA)	2015	Online questionnaire	212	Americas, Asia-Pacific (APAC), Europe-Middle-East-Africa (EMEA)	• Data Security (73%) • Loose of control over the IT services (38%) • Regulatory and compliance (38%) • Lack of experience and knowledge of business managers and IT (34%) • Compromised accounts (30%) • Business continuity and disaster recovery (28%)
2	KPMG	2014	Online questionnaire	539	United states, Asia-Pacific (APAC), Europe-Middle-East-Africa (EMEA), Latin and south America, Canada and Mexico	• Privacy and data loss (53%) • Risk of intellectual property theft (50%) • Impact on IT organization (49%) • Measuring the ROI (48%) • High cost of implementation (48%) • Legal and regulatory compliance (46%) • Integrity with existing architecture (46%) • Lack of clarity of total cost ownership (46%)

(*continued*)

Table 1. (*continued*)

S. no	Organization	Year	Method	Num	Geography	Cloud Computing Risks and their percentage
3	IDG	2014	Online questionnaire	1,600	Global	• Security (61%) • Integration challenge (46%) • Information governance (35%) • Measuring ROI (30%) • Meeting industry standard (27%) • Lack of vendor strategy on implementing the solution (20%) • Business leader not receptive the cloud solution (11%) • Employees are not receptive to the cloud (10%)
4	North bridge Venture Partners	2014	n-a	1358	Global	• Security (49%) • Regulatory and compliance (34%) • Vendor lock-in interoperability (29%) • Interoperability (17%) • Privacy (31%) • Reliability (22%) • Network bandwidth (25%) • Complexity (15%) • Expense (12%)

(*continued*)

Table 1. (*continued*)

S. no	Organization	Year	Method	Num	Geography	Cloud Computing Risks and their percentage
5	Cloud Connect and Everest Group	2013	Online questionnaire	302	EU	• Security (30%) • Integration of the cloud solution (27%) • Lack of budget for new initiatives • Fear of vendor lock (27%) • Lack of suitable cloud solution (25%) • Lack of management buy-in (24%) • Lack in house capability to evaluate cloud solution (24%) • Lack of attractive business cases for cloud adoption (22%)

on 2013. The objective of the survey was to identify the cloud adoption patterns, cloud adopting barriers and making decision patterns for cloud adoption (EversetGroup 2013). International Data Corporation (IDC) explored the status of cloud adoption and barriers such as security, loose of control over IT asset, and regulatory in the financial services and government entities (IDC 2015). Table 2 consolidates the cloud computing adoption risks as reported by different journal publications.

Risks in Moving to Cloud Computing from an Academic Viewpoint
Phaphoom et al. (2015) explored the major technical and security-related inhibitors to organizational adoption decisions of the cloud. The study consisted of 352 participants from different organizations. The study compared non-adopters' and adopters' perceptions. The study concluded that the major inhibitors were security, data privacy and portability. Another study identified the critical factors affecting cloud computing adoption decision by CIOs in hospitals in Taiwan (Lian et al. 2014). The study revealed data security, cost, top manager support, complexity and perceived technical competence as critical factors. Therefore, technology dimension is the most influential amongst the other dimensions (organization, environment and human) pertaining to cloud computing adoption (Liana et al. 2014). Oliveira et al. (2014) explored the factors that influenced cloud computing adoption in service and manufacturing sectors in Portugal using a questionnaire distributed to 369 firms. The theories of TOE

Table 2. Research papers

Dependent variable	Independent variables	Theory	Methods	Num	Geography	Key findings	Author
Adoption decision of cloud computing	Availability, portability, integration with current enterprise system, migration complexity, data privacy and security	–	Logistic regression	352	Europe, North America, Asia, Africa, Australia and south America	Security, data privacy and portability are the inhibitors factors for cloud adoption	(Phaphoom et al. 2015)
Adoption decision of cloud computing	Technological factor (Security, compatibility, complexity & cost) Human factor (CIO innovativeness, technical competence) organizational factor (management support, relative advantage, resources & benefits) environmental factors (government policy & industry pressure)	TOE and HOT-FIT	ANOVA and Mean Value	60	Taiwan	Most critical factors data security, cost, top manager support, complexity and perceived technical competence	(Jiunn-Woei Liana et al. 2014)
Cloud computing adoption	Technology (Technology readiness), organization (Top management support & firm size), environmental (competitive pressure and regulatory support), innovation (relative advantage, complexity & compatibility), security concern & cost saving	Diffusion of innovation (DOI) and TOE	Structural model	369	Portugal	Cost saving, relative advantage, complexity, technology readiness, top management support and firm size are significant factors to cloud adoption	(Oliveira et al. 2014)

(*continued*)

Table 2. (*continued*)

Dependent variable	Independent variables	Theory	Methods	Num	Geography	Key findings	Author
Cloud computing adoption	Technology (complexity, compatibility & relative advantage), Organization (technology readiness, top management support & firm size) and Environment (trading partner pressure, competitive pressure)	TOE	Logistic regression and principal component analysis	257	UK	Competitive pressure, complexity, technology readiness and trading partner pressure	(Anabel Gutierrez et al. 2015)

(Technology-organization-environment) and DOI (diffusion of innovation) are used to organize the different factors. The factors used in the TOE theory were technology readiness (Technology context), top management support and firm size (organizational context) and competitive pressure and regulatory support (environmental context). The factors used in the DOI were relative advantage, complexity and compatibility. Security concern and cost saving used as the factors affect relative advantage. The study revealed cost saving, relative advantage, complexity, technology readiness, top management support and firm size as significant determinants of cloud adoption (Oliveira et al. 2014). Gutierrez et al. (2015) investigated the factors that influence decision makers adopt cloud computing in UK by using TOE framework. The data collected through questionnaire from 257 decision makers and IT professional in UK. The factors used for analysis were complexity, compatibility, technology readiness, trading partner pressure, competitive pressure, relative advantage, top management support and firm size. Among the eight factors, competitive pressure, complexity, technology readiness and trading partner pressure were found to be significant. Table 3 shows the frequency of each risks reported by industrial and research papers

Conceptual Framework and Hypotheses Construction

Adoption of new technology is one of the complex issues in information systems research. It plays a significant role in the organization's ability to gain a competitive advantage. The decision to adopt new technology require clear understanding of its impact on the organization either positively or negatively. The adoption decision criteria need to be as much comprehensive as possible. Adoption of new technology such as cloud computing is complicated due to the numerous uncertainties in terms of captured value and risk-susceptibilities.

There are different theories used as a framework to determine the factors influencing the adoption decision. Many researches have used theories like diffusion of innovations theory (DOI) and the technological, organizational and environmental

Table 3. Frequency table

Risk name	Data security	Loose of control over IT service	Lack of experience and knowledge	Regulatory compliance	Compromised account	Privacy	Lack of clarity of total cost ownership in terms of ROI	High cost of implementation	Integrating with existing architecture	Meeting industry standards	Bandwidth	Vendor lock	Measuring the ROI	Reliability	Business continuity and recovery
Cloud security Alliance (CSA)	*	*	*	*	*									*	*
IDC	*	*		*											
KMPG	*			*		*	*	*					*		
IDG	*						*		*				*		
North bridge venture partner	*			*		*				*	*	*			
Cloud connect and event group	*		*						*			*			
Phapoom et al. (2015)	*					*						*			
Lian et al. (2014)	*								*						
Total	7	2	2	4	1	3	2	1	3	1	1	3	2	1	1

theory (TOE) as a comprehensive framework to identify the factors that affecting adoption decision. The DOI theory measures adoption by five factors: relative advantage, complexity, compatibility, observability and trial ability while the TOE framework is classified into three main categories: technical, organizational and environmental context. Both theories aid IT managers and decision makers understand the business values and risks of implementing new technologies. However, as this research focuses on the risks side of adopting cloud computing, none of these theories have been used in this research.

There are many studies on cloud computing risks and corresponding categories. There are 39 cloud computing risks and they are classified into four categories i.e., organizational, technical, legal and operational risks (Dutta et al. 2013). The studies conducted by (Alzadjali et al. 2015, Sharma et al. 2016, AlKharusi and Al-Badi 2016, Al-Harthy and Al-Badi 2014, Al-Musawi et al. 2015) highlight the challenges faced by decision makers in accepting cloud computing in Oman. The success of cloud computing depends upon providing effective and efficient solutions to risks associated with cloud computing security and privacy (Takabi et al. 2010). There are organizational, legal, security, technical, and financial risk. In this research, as per the risks identified in Table 3, the research framework in this research comprises of the following three components:

Organizational Risks: It assesses the impact of cloud adoption from an organizational aspect including IT employees and non-IT employees, IT planning, business and IT operations and IT governance and management.

Technical Risks: The complicated cloud infrastructure and inherent IT deficiencies that exists inside the organization can raise a set of technical risks during cloud computing adoption.

Legal Risk: The features of the cloud computing may lead to different legal issues relating to regulations and compliance policies, contracts, data privacy and intellectual property (Fig. 1).

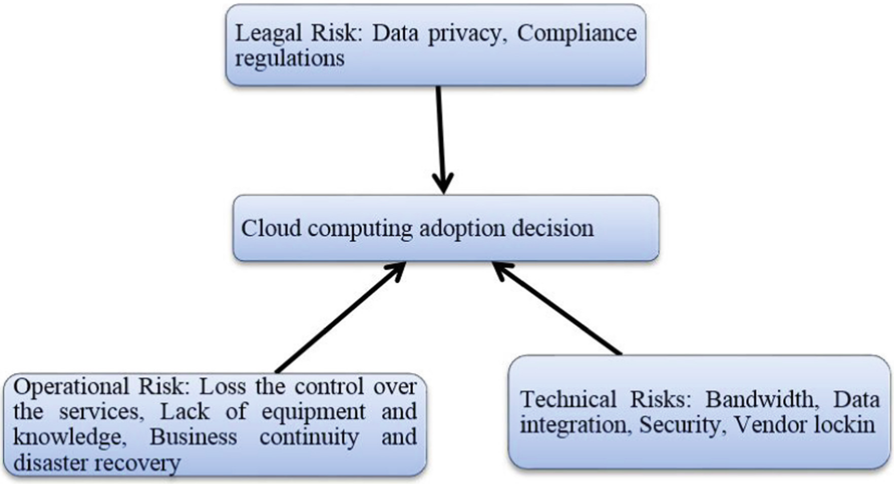


Fig. 1. Conceptual framework

The identified risks in the conceptual framework is presented as hypothesis such as H1, H2, H3, H4, H5, H6, H7, H8, H9, and H10:

Security: Cloud Security Alliance (CSA) in a survey distributed across 17 countries found that security of data was the greatest challenge for cloud adaptors (Coles et al. 2015). In general, there is unwillingness to allow confidential data be hosted outside the organization's firewall (Gnanasambandam. et al. 2014). One of the major adoption barriers in European countries is security (Porter 2015). In Tunisia, risk plays a mediating factor in cloud computing adoption (Hachicha and Mezghani 2018). Usually it is cited as the number one concern and resistance for cloud computing adoption (Armbrust et al. 2010, Akande et al. 2013). Accordingly, the following hypothesis is posited:

H1: Security risk negatively influences the cloud computing adoption decision.

Data Privacy: Data privacy was cited as one of the cloud computing adoption risks (Phaphoom et al. 2015, KPMG 2014, Donnelly 2015, Dutta et al. 2013). Privacy is directly related to security (Ali et al. 2016). Security was raised as the most important roadblock in the way of cloud computing adoption (Grobauer et al. 2011). Security and hence, trust needs to be shared between the different clients and service providers (Yeager and Morin 2018). Cloud providers should be responsible for the data once it is moved to the cloud (Phaphoom et al. 2015). The laws of protecting data vary from one country to another country which represents a challenge here. It creates inconsistencies in data protection between countries that generate the data and the country which store the data (Dutta et al. 2013). O'Donohue speaking at the Data Cloud Europe event claimed that the diversity of data protection laws is one of the legislative challenges that may prevent some EU members from accessing cloud services (Donnelly 2015). The cloud client should read the terms and conditions before moving to the cloud to avoid any legal issue. Accordingly, the following hypothesis is posited:

H2: Data privacy risk negatively influences the cloud computing adoption decision.

Vendor Lock-In or Portability: It refers to the difficulty of the cloud client to move from one vendor to another. It is considered as one of the risks that affect cloud computing adoption (Phaphoom et al. 2015, NorthBridge 2014, EverestGroup 2013). The decision of switching between cloud providers could be expensive and time taking (Siefker and Lucas 2013). The effort spent in customizing the solution with one cloud provider may requires redoing it again with another (Erl et al. 2013). Bringing back the service inhouse including data and applications is not an easy mission (Leavitt 2009). For example, there is no standard API (application programming interface) between the different vendors. The closed and different architecture and unique applications of the cloud services of the different vendors make switching between them a very difficult task for adopters (Gordon 2010, Phaphoom et al. 2015). Accordingly, the following hypothesis is posited:

H3: Vendor lock-in risks negatively influence the cloud computing adoption decision.

Bandwidth is defined as the amount of data transferred in a given time from one point to another. Basically, cloud computing is an internet based services and having adequate bandwidth is therefore very essential. Adopting cloud is cost effective for hardware, software, maintenance and other services but could increase as the bandwidth increases (Leavitt 2009). Applications are becoming data-intensive which means that data going back and forth from a client to a cloud provider will require more bandwidth and hence, the cost will increase (Armbrust et al. 2010). Bandwidth remains an issue especially in developing countries (NorthBridge 2014). Accordingly, the following hypothesis is posited:

H4: Bandwidth risk negatively influences the cloud computing adoption decision.

Data Integration: Moving some systems to the cloud and keeping others in house creates data integration issues (Neske 2015). Integration with existing systems is considered one of the cloud computing risks (KPMG 2014, Phaphoom et al. 2015, EverestGroup 2013). Some organizations tend to adopt hybrid clouds, integrating private and public cloud that might address integration complexities (Kim et al. 2009, Phaphooma et al. 2015). Integration between partners require constant connectivity and standardization of data (Phaphooma et al. 2015). Accordingly, the following hypothesis is posited:

H5: Data integration risk negatively influences the cloud computing adoption decision.

Compliance and Regulations: Compliance is considered one of the major obstacles in adopting the cloud in North America and Asia Pacific (Gordon 2010). Compliance and regulations remain an inhibitor to moving to the cloud (Coles et al. 2015, KPMG 2014, NorthBridge 2014, IDC 2015). Organizations are subject to some regulations and compliance issues that must be met. The compliance is concerned with privacy, secure storage and disclosure of data. According to the cloud security alliance, clients face many cloud computing adoption risks including compliance. Once data moves to the cloud, organization needs to ensure that cloud providers apply the same compliance standards (Coles et al. 2015). Compliance and regulations are considered major barrier, especially for some organizations that have sensitive data like the financial sector and the government that need to keep data safe (IDC 2015). Accordingly, the following hypothesis is posited:

H7: Compliance and regulations risk negatively influence the cloud computing adoption decision.

Loose Control over IT Service: The on-premise allows the organization to have full control over the IT assets including the hardware and software. Moreover, the organization has the accountability for data security and access control as well. Moving to the cloud transfer these privileges to cloud providers (Morgan and Conboy 2013). The cloud provider is responsible for physical hardware, location, security, accessibility and other services (Sheppard 2014). Therefore, many organizations still deem this issue as an inhibitor to adoption (Coles et al. 2015, KPMG 2014). Accordingly, the following hypothesis is posited:

H8: Loose control over IT service risk negatively influence the cloud computing adoption decision.

Business Continuity and Disaster Recovery: According to the cloud security alliance, 28% of the respondents highlighted business continuity and disaster recovery as a barrier to cloud adoption (Coles et al. 2015). Business continuity refers to the procedure and polices that are put in place to ensures the essential functioning and processing of the organization's operations before and after a disaster. This includes disaster recovery processes that help the organization recover and restore their operations (Techopedia 2017). Accordingly, the following hypothesis is posited:

H9: Business continuity and disaster recovery risk negatively influence the cloud computing adoption decision.

Lack of Experience and Knowledge: Cloud based services require some training before acceptance. The lack of understanding of cloud, its benefits, and how to deal with it play a role in adopting cloud computing (Coles et al. 2015, IDC 2015). Many studies found that prior experience as important factor in technology adoption decisions. The IT staff's familiarity with cloud computing technologies like clustering and virtualization can influence the adoption decision (Alshamaila et al. 2013). Introducing cloud services in the organization may result in mustering employee's resistance if there is a lack of knowledge (Morgan and Conboy 2013). Accordingly, the following hypothesis is posited:

H10: Lack of experience and knowledge risk negatively influences the cloud computing adoption decision.

3 Research Methodology

The objective of this research is to identify as to what extent the different types of risks influences cloud computing adoption decisions and accordingly, proposes a conceptual adoption framework. Online databases such as Science direct, Scopus, Google scholar, Springer, and IEEE explore were searched to get relevant studies for this research. This study reviewed 50 research papers from leading information systems journals and conferences. The articles were searched and ordered based on the categories of risks in adopting cloud-computing services such as technical (data integration, security, vendor lock-in and bandwidth), legal (data privacy and compliance and regulation) and operational risks (lack of knowledge and experience, loose of control over IT service, and business continuity and disaster recovery).

4 Discussion

It has already been established that cloud computing numerous advantages. On the other hand, there are many researchers that raise alarming concerns about security and privacy issues. Therefore, providing effective and efficient solutions to risks related to cloud computing adoption could contribute to its success. The risks associated with cloud computing can be reduced by the following guidelines.

4.1 Avoiding Cost and Vendor Locks

The project and corresponding processes need to be studied thoroughly before implementing a cloud computing project. Clients often develop negative attitude towards cloud computing adoption due to recurring costs and costs related to moving to another cloud. Most of the recurring costs are related to the services and corresponding infrastructure. Therefore, top management must select the right cloud-service providers.

4.2 Clear and Defined Technical Contract and Service Level Agreement (SLAs)

The contracts between the cloud provider and clients must be detailed specifying exact roles and responsibilities. Legal consultations could be sought to craft a legally binding agreement. It should serve as a legally binding technical document for both parties to avoid litigations in the future.

4.3 Outsourcing and Authorization of Data

What part of the data belongs to the provider and what parts belong to the client must be documented to avoid conflicting data ownership and privacy issues. Only authorized persons are expected to access the data stored in the cloud. It is important to develop both technical and non-technical frameworks regarding data specifications, relationships and interrelationships. In case of data theft and leakage, proper legal frameworks must be available to protect the interests of clients.

4.4 Trust Building, Security, and Usage of Software and Programs

Trust is required between the different stakeholders of cloud computing. Cloud computing has a large amount of data uploaded every day and hence, cloud computing providers must ensure that the data does not go in the wrong hand e.g., a third party, without the owner's consent. The specific program and software running on the cloud that is used by a client cannot be transferred to another without the consent of that client. The security and privacy relating to the usage of software and hardware need to be shared between client and service providers only. For example, public clouds service providers are more responsible as compared to private ones in the case of SaaS. In PaaS, clients are more responsible for running and usage of programs on the platforms (Aleem and Al-Qirim 2012). In IaaS, the client is expected to ensure secure operating systems, applications, and content.

4.5 Management Decisions and Lack of Knowledge

Drafting proposer policies by policymakers influence cloud computing adoption decisions. The limited knowledge and the lack of understanding of cloud computing technology could lead to taking wrong adoption decisions. Such decision makers like top management are often have no IT background and hence, have limited awareness of the new technology. This further aggravate the risks involved in cloud computing adoption decisions.

4.6 Data Integration, Bandwidth, and Disaster Recovery

Both clients and service providers share data integration and limited bandwidth risks. The client must ensure that the data is regularly updated from their sides while the service providers should ensure the consistency and visibility of the data irrespective of networks failure through e.g., backup plans. Also, it is the responsibility of client to have enough bandwidth for uninterrupted access to the cloud.

4.7 Compliance of Agreement and Regulations

It must be ensured that both clients and cloud service providers comply with the "legally" signed contracts and agreements relating to cloud computing services. The top management must study carefully review the past records of the service provider with respect to fulfillment and compliance with the agreement's roles and responsibilities. The security-risks must be covered in the SLAs between the client and the service provider (Rojas et al. 2018). Also, proper channels need to be established to guide pursuing and reporting violations (Hussain et al. 2018).

5 Conclusion

There is no doubt that cloud computing has gained significant attention from IT industries. Cloud computing provides more flexibility in the usage of IT services, help in cost reduction, and increase accessibility. Nevertheless, there are challenges which still hinders the adoption of cloud computing. This study suggests a conceptual framework for the smooth adoption of cloud computing. The framework provides different guidelines relating to cost-reduction, avoiding vendor's locking, having clear and defined contracts and agreements, outsourcing and authorization of data, trust building, security, and usage of software and programs, management decisions and lack of IT knowledge, data integration, bandwidth, and disaster recovery, compliance of agreement and regulations. The present study is theoretical in nature and hence, future empirical study will be conducted to assess the significance of the proposed model in detecting cloud computing adoption barriers.

References

Ackermann, T.: IT Security Risk Management: Perceived IT Security Risks in the Context of Cloud Computing. Springer, Wiesbaden (2012). https://doi.org/10.1007/978-3-658-01115-4

Ahmad, S.J., Jolly, R.K.: Immense implementation of Cloud Computing on distinct pilot projects as a specimen of the delineation of cost effectiveness to manifest as Cloud Computing democracy to be or not to be. GSJ **6**, 191 (2018)

Akande, A.O., April, N.A., Belle, J.-P.V.: Management issues with Cloud Computing. In: Proceedings of the Second International Conference on Innovative Computing and Cloud Computing. ACM (2013)
Aleem, S., Al-Qirim, N.: Cloud applications: a perspective on PaaS. In: The Proceedings of the 2012 International Conference on Business and Information (BAI2012), 3–5 July 2012, Sapporo, Japan (2012)
Al-Qirim, N.: A roadmap for success in the clouds. In: IEEE Explore and in the Proceedings of the 7th International Research Conference on Innovations in Information Technology (Innovations 2011), 25–27 April, Abu Dhabi, UAE (2011)
Al-Harthy, L.Y., Al-Badi, A.H.: To cloudify or not to cloudify. World Acad. Sci. Eng. Technol. Int. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng. **8**, 2453–2464 (2014)
Al-Musawi, F., Al-Badi, A.H., Ali, S.: A road map to risk management framework for successful implementation of Cloud Computing in Oman. In: 2015 International Conference on Intelligent Networking and Collaborative Systems (INCOS), pp. 417–422. IEEE (2015)
Al-Shamsi, A., Al-Qirim, N.: Cloud healthcare-records manager. In: The Proceedings of the WorldCist 2016 - 4th World Conference on Information Systems and Technologies, 22–24 March 2016, Recife, PE, Brazil (2016)
Ali, O., Soar, J., Yong, J.: An investigation of the challenges and issues influencing the adoption of cloud computing in Australian regional municipal governments. J. Inf. Secur. Appl. **27**, 19–34 (2016)
Alkharusi, M.H., Al-Badi, A.H.: IT personnel perspective of the slow adoption of cloud computing in public sector: case study in Oman. In: 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), pp. 1–8. IEEE (2016)
Alkhater, N., Walters, R., Wills, G.: An empirical study of factors influencing cloud adoption among private sector organisations. Telematics Inform. **35**, 38–54 (2018)
Alshamaila, Y., Papagiannidis, S., Li, F.: Cloud computing adoption by SMEs in the northeast of England: a multi-perspective framework. J. Enterp. Inf. Manage. **26**, 250–275 (2013)
Alzadjali, A.M., Al-Badi, A.H., Ali, S.: An analysis of the security threats and vulnerabilities of Cloud Computing in Oman. In: 2015 International Conference on Intelligent Networking and Collaborative Systems (INCOS), pp. 423–428. IEEE (2015)
Gutierrez, A., Boukrami, E., Lumsden, R.: Technological, organisational and environmental factors influencing managers' decision to adopt cloud computing in the UK. J. Enterprise Inf. Manage. **28**, 788–807 (2015)
Archer, G.: Government CIOs see expected cloud cost savings evaporate (2015). https://www.cio.com.au/article/572878/government-cios-see-expected-cloud-cost-savings-evaporate/. Accessed 30 June 2017
Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I.: A view of Cloud Computing. Commun. ACM **53**, 50–58 (2010)
Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: Above the clouds: a berkeley view of Cloud Computing (2009)
Dutta, A., Peng, G.C.A., Choudhary, A.: Risks in enterprise cloud computing: the perspective of IT experts. J. Comput. Inf. Syst. **53**(4), 39–48 (2013)
Bernard, L.: A risk assessment framework for evaluating Software-as-a-Service (SaaS) cloud services before adoption, University of Maryland University College (2011)
Bisong, A., Rahman, M.: An overview of the security concerns in enterprise cloud computing (2011). arXiv preprint arXiv:1101.5613
Coldewey, D.: Google services suffer outage (2014). http://www.nbcnews.com/tech/internet/google-services-suffer-outage-n34986. Accessed 12 Apr 2015

Coles, C., Yeoh, J., Guanco, F., Mishra, E.: Cloud adoption practices & priorities survey report. Cloud Security Alliance, 9 (2015)
Comella-Dorda, S., Gnanasambandam, C., Shah, B., Strålin, T.: From box to cloud: An approach for software development executives. McKinsey & Company (2015). http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/from-box-to-cloud. Accessed 30 June 2017
Dave, D.: Avoiding the hidden costs of the cloud (2013). http://www.symantec.com/connect/blogs/avoiding-hidden-costs-cloud. Accessed 30 June 2017
Donnelly, C.: European Commission calls for cloud data protection law reforms (2015). http://www.computerweekly.com/news/4500247442/European-Commission-calls-for-cloud-data-protection-law-reforms. Accessed 30 June 2017
Dutta, A., Peng, G.C.A., Choudhary, A.: Risks in enterprise cloud computing: the perspective of IT experts. J. Comput. Inf. Syst. **53**, 39–48 (2013)
EMA. Casualties of Cloud Wars: customers are paying the price (2014). http://www.iland.com/wp-content/uploads/2014/06/EMA-iland_VMware_CloudWars-0614-WP.pdf. Accessed 12 Apr 2015
Erl, T., Puttini, R., Mahmood, Z.: Cloud Computing: Concepts, Technology & Architecture. Prentice Hall, Upper Saddle River (2013)
Everestgroup. Enterprise Cloud Adoption Survey 2013: Summary of Results (2013). http://www.everestgrp.com/wp-content/uploads/2013/03/2013-Enterprise-Cloud-Adoption-Survey.pdf. Accessed 30 June 2017
Eversetgroup Enterprise Cloud Adoption Survey 2013: Summary of Result (2013)
Fahmideh, M., Beydoun, G.: Reusing empirical knowledge during cloud computing adoption. J. Syst. Softw. **138**, 124–157 (2018)
Fortinová, J.: Risks of cloud computing. Systémová Integrace, 20 (2013)
Garrison, G., Rebman Jr., C.M., Kim, S.H.: An identification of factors motivating individuals' use of cloud-based services. J. Comput. Inf. Syst. **58**, 19–29 (2018)
Ghanam, Y., Ferreira, J., Maurer, F.: Emerging issues & challenges in cloud computing-a hybrid approach. J. Softw. Eng. Appl. **5**, 923 (2012)
Gnanasambandam, C., Sprague, K., Huskins, M.: Next-Generation IT Infrastructure. McKinsey & Company (2014)
Gordon, J.: Exploring the future of cloud computing: riding the next wave of technology-driven transformation. In: World Economic Forum, White Paper (2010)
Grobauer, B., Walloschek, T., Stocker, E.: Understanding cloud computing vulnerabilities. IEEE Secur. Priv. **9**, 50–57 (2011)
Gutierrez, A., Boukrami, E., Lumsden, R.: Technological, organisational and environmental factors influencing managers' decision to adopt cloud computing in the UK. J. Enterp. Inf. Manage. **28**, 788–807 (2015)
Hachicha, Z.S., Mezghani, K.: Understanding intentions to switch toward cloud computing at firms' level: a multiple case study in Tunisia. J. Global Inf. Manage. (JGIM) **26**, 136–165 (2018)
Haile, N., Altmann, J.: Structural analysis of value creation in software service platforms. Electron. Mark. **26**, 129–142 (2016)
Han, Y.: On the clouds: a new way of Computing. Inf. Technol. Libraries **29**, 87 (2010)
Hussain, W., Hussain, F.K., Hussain, O., Bagia, R., Chang, E.: Risk-based framework for SLA violation abatement from the cloud service provider's perspective. Comput. J. (2018)
IDC. Public Cloud Spending to Reach Nearly $70 Billion Worldwide in 2015 (2015). According to IDC. http://www.businesswire.com/news/home/20150721005355/en/Public-Cloud-Spending-Reach-70-Billion-Worldwide. Accessed 30 June 2017
IDG Enterprise 2014. IDG Enterprise Cloud Computing Study (2014)

Illsley, R.: The role of cloud in IT modernisation: the DevOps challenge. http://docplayer.net/10458737-The-role-of-cloud-in-it-modernisation-the-devops-challenge-ovum.html. Accessed 30 June 2017

Internap. Intrernap public cloud survey reveals performance as top challenge (2014). http://cloudtweaks.com/2014/01/internap-public-cloud-survey-reveals-performance-as-top-challenge/. Accessed 12 Apr 2015

Liana, J.W., Yenb, D.C., Wang, Y.-T.: An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital. Int. J. Inf. Manage. **34**, 28–36 (2014)

Khajeh-Hosseini, A., Sommerville, I., Sriram, I.: Research challenges for enterprise cloud computing (2010). arXiv preprint arXiv:1001.3257

Kim, W., Kim, S.D., Lee, E., Lee, S.: Adoption issues for cloud computing. In: Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia, pp. 2–5. ACM (2009)

KPMG. The cloud take shapes (2013). http://www.kpmg-institutes.com/institutes/advisory-institute/articles/2013/02/cloud-takes-shape.html. Accessed 30 June 2017

KPMG. Elevating business in the cloud (2014). http://www.kpmginfo.com/EnablingBusinessInTheCloud/downloads/2014%20KPMG%20Cloud%20Survey%20Report%20-%20Final%2012-10-14.pdf. Accessed 30 June 2017

Leavitt, N.: Is cloud computing really ready for prime time? Computer **42**, 15–20 (2009)

Leopold, G.: Forecasts call for cloud burst through 2018 (2014). http://www.enterprisetech.com/2014/11/03/forecasts-call-cloud-burst-2018/. Accessed 19 Sept 2015

Lian, J.-W., Yen, D.C., Wang, Y.-T.: An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital. Int. J. Inf. Manage. **34**, 28–36 (2014)

Marinescu, D.C.: Cloud Computing: Theory and Practice. Morgan Kaufmann, Boston (2013)

Mcnee, B.: Digital Business Rethinking Fundamentals. NYC: Saugatuck Technology (2014)

Mell, P., Grance, T.: The NIST definition of cloud computing. U.S. Department of Commerce (2011)

Morgan, L., Conboy, K.: Factors affecting the adoption of cloud computing: an exploratory study. In: 21st European Conference on Information Systems 2013 (ECIS), 5–8 June 2013. Utrecht University (2013)

Phaphooma, N., Wanga, X., Samuelb, S., Helmera, S., Abrahamssona, P.: A survey study on major technical barriers affecting the decision to adopt cloud services. J. Syst. Softw. **103**, 167–181 (2015)

Nayar, K.B., Kumar, V.: Cost benefit analysis of cloud computing in education. Int. J. Bus. Inf. Syst. **27**, 205–221 (2018)

Neske, D.: Hidden costs of the cloud: here is your umbrella (2015). http://services.ricoh.com/knowledge-center/blog/2015-02-25-hidden-costs-of-the-cloud-here-is-your-umbrella. Accessed 30 June 2017

Northbridge. The Future of the Cloud Computing (2014). https://www.slideshare.net/mjskok/2014-future-of-cloud-computing-4th-annual-survey-results. Accessed 30 June 2017

Oliveira, T., Thomas, M., Espadanal, M.: Assessing the determinants of cloud computing adoption: an analysis of the manufacturing and services sectors. Inf. Manage. **51**, 497–510 (2014)

Phaphoom, N., Wang, X., Samuel, S., Helmer, S., Abrahamsson, P.: A survey study on major technical barriers affecting the decision to adopt cloud services. J. Syst. Softw. **103**, 167–181 (2015)

Porter, S.: The state of cloud computing in Europe (2015). http://www.thoughtsoncloud.com/. Accessed 19 Sept 2015

Priyadarshinee, P., Raut, R.D., Jha, M.K., Gardas, B.B.: Understanding and predicting the determinants of cloud computing adoption: a two staged hybrid SEM - Neural Networks Approach. Comput. Hum. Behav. **76**, 341–362 (2017)

Rojas, M.A.T., et al.: Managing the lifecycle of security SLA requirements in cloud computing. In: Rocha, Á., Reis, L.P. (eds.) Developments and Advances in Intelligent Systems and Applications. SCI, vol. 718, pp. 119–140. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-58965-7_9

Sharma, S.K., Al-Badi, A.H., Govindaluri, S.M., Al-Kharusi, M.H.: Predicting motivators of cloud computing adoption: A developing country perspective. Comput. Hum. Behav. **62**, 61–69 (2016)

Sheppard, D.: Is loss of control the biggest hurdle to cloud computing? (2014). http://www.itworldcanada.com/blog/is-loss-of-control-the-biggest-hurdle-to-cloud-computing/95131. Accessed 30 June 2017

Siefker, K., Lucas, J.: Hidden cost of cloud computing (2013). http://investor.compuware.com/releasedetail.cfm?ReleaseID=777461. Accessed 12 Apr 2015

Stergiou, C., Psannis, K.E., Kim, B.-G., Gupta, B.: Secure integration of IoT and cloud computing. Future Gener. Comput. Syst. **78**, 964–975 (2018)

Takabi, H., Joshi, J.B., Ahn, G.-J.: Security and privacy challenges in cloud computing environments. IEEE Secur. Priv. **8**, 24–31 (2010)

Techopedia. Business continuity and disaster recovery (BCDR) (2017). https://www.techopedia.com/definition/13767/business-continuity-and-disaster-recovery-bcdr. Accessed 30 June 2017

Wease, G., Boateng, K., Yu, C.-J., Chan, L., Barham, H.: Technology assessment: cloud service adoption decision. In: Daim, T.U., Chan, L., Estep, J. (eds.) Infrastructure and Technology Management. ITKM, pp. 447–471. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-68987-6_16

Yeager, W.J., Morin, J.-H.: Introduction to the minitrack on security and critical infrastructure for Cloud, IOT and Decentralized trust. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018)

Blockchain Enabled Enhanced IoT Ecosystem Security

Mahdi H. Miraz[1(✉)] and Maaruf Ali[2]

[1] The Chinese University of Hong Kong, Shatin, NT, Hong Kong
m.miraz@cuhk.edu.hk
[2] International Association of Educators and Researchers (IAER), London, UK
maaruf@ieee.org

Abstract. Blockchain (BC), the technology behind the Bitcoin crypto-currency system - is starting to be adopted for ensuring enhanced security and privacy in the Internet of Things (IoT) ecosystem. Fervent research is currently being focused in both academia and industry in this domain. Proof-of-Work (PoW), a cryptographic puzzle, plays a vital rôle in ensuring BC security by maintaining a digital ledger of transactions, which are considered to be incorruptible. Furthermore, BC uses a changeable Public Key (PK) to record the identity of users – thus providing an extra layer of privacy. Not only in crypto-currency has the successful adoption of the BC been implemented, but also in multifaceted non-monetary systems, such as in: distributed storage systems, proof-of-location and healthcare. Recent research articles and projects/applications were surveyed to assess the implementation of the BC for IoT Security and identify associated challenges and propose solutions for BC enabled enhanced security for the IoT ecosystem.

Keywords: Blockchain · Proof-of-Work (PoW) · Internet of Things (IoT) Security

1 Introduction

The primary aim of this article is to investigate the research question, "To what extent can the Blockchain be used in enhancing the overall security of the Internet of Things (IoT) ecosystems?" and to draw appropriate conclusions. Considering the fact that the Blockchain is comparatively an avant-garde technology, this paper presents a representative sample of research conducted in the last ten years, commencing with the early work in this domain. Although, identifying how the Blockchain can further enhance the security paradigm of IoT is the main focus of the paper, to do so various other usages of the Blockchain and similar digital ledger technologies were explored along with their applications, impediments, privacy and security concerns.

Like many other domains of computing, security and privacy issues are the major concerns of the Internet of Things (IoT) eco-system. To fortify the backbone for improved security and privacy of IoT, the Blockchain is considered to be able to play a vital rôle. In fact, Blockchain research has become truly multifaceted as researchers from both industry as well as academia are applying the Blockchain in new dimensions

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 38–46, 2018.
https://doi.org/10.1007/978-3-319-95450-9_3

on a regular basis. In the Proof-of-Work (PoW) concept, as shown in Fig. 1, an algorithm based on mainly solving a mathematical challenge, is the major method to assure the security aspects of the BC by recording and maintaining a complete digital ledger of all the completed transactions. These transactions are thus unalterable.

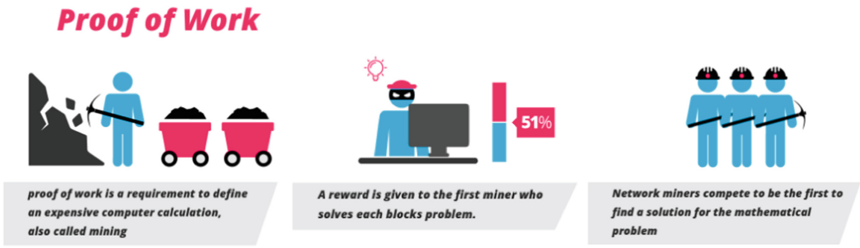


Fig. 1. The proof of work concept [1].

A high-level system block diagram of how the BC technology works is shown in Fig. 2.

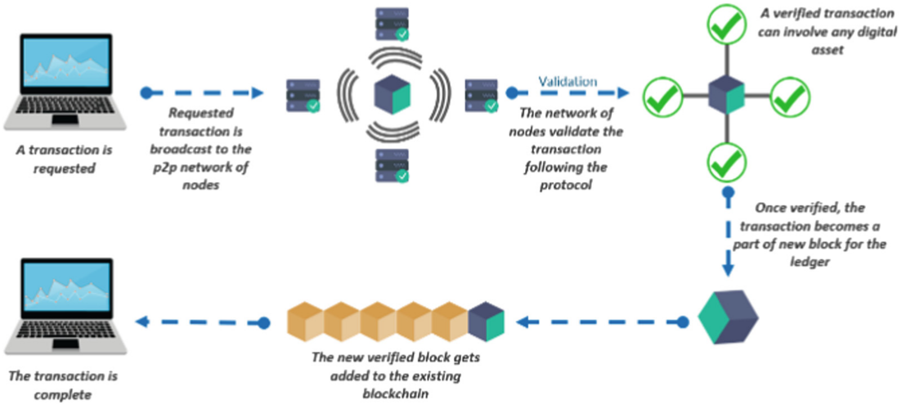


Fig. 2. A high level view of the Blockchain technology [2].

In addition to this, the BC also takes advantage of the Public Key, as shown in Fig. 3, which is purposely made chaotic in nature for ensuring the highest level of security, in order to register the identity of the users. Thus, an extra layer of privacy is ensured automatically. As evident by many research and project reports, the adoption of the Blockchain technology has been found to be successful in many non-monetary domains such as in the supply chain, healthcare systems, online/electronic voting, proof of location, distributed cloud storage, even in human resource management and recruitment [3].

The authors of this paper not only surveyed research articles but also considered relevant projects/applications to ascertain the applicability of Blockchain technology

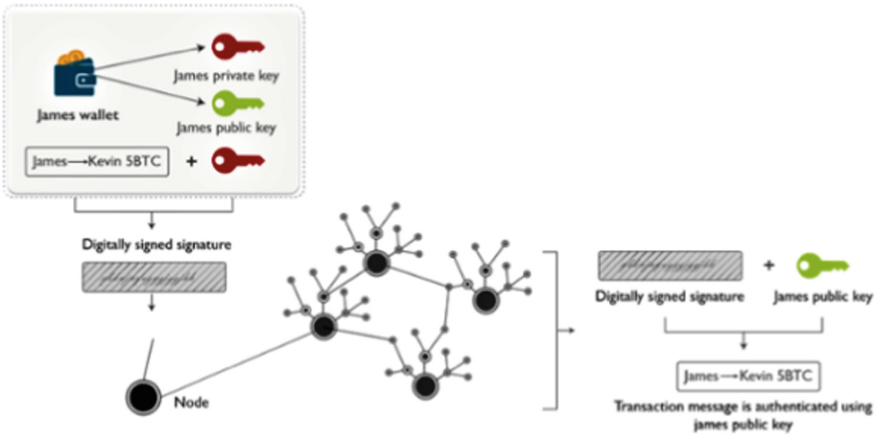


Fig. 3. The verification of signatures by the miners in the blockchain [2].

for augmented IoT security and to distinguish the challenges associated with such application of the BC and thence to put forward probable solutions for BC enabled enhanced IoT security systems.

The knowledge domain of the research is in the realm of the Internet of Things (IoT), Internet of Everything (IoE), Wireless Sensor Network (WSN), digital ledger, specifically, in Blockchain and crypto-currency.

2 Blockchain Fundamentals

To understand how Blockchain can be applied for enhancing IoT security, it is very important to understand how these two technologies are put into function. In this section, the basic technological fundamentals of Blockchain have been briefly described while the next session introduces the Internet of Things (IoT) ecosystem.

A Blockchain mainly consists of two separate but interrelated integrant. These are as follows:

1. **Transaction:** In a digital ledger system such as the Blockchain, a transaction is basically the action triggered by the participant.
2. **Block:** A block, in a Blockchain system, is a collection or pool of data which records the transaction and other relevant details such as the correct sequence, timestamp of creation, *et cetera*.

Based on the scope of how a Blockchain is going to be used, it can be of two types: private or public. In a public Blockchain, usually all the users have both read and write permissions. One example of such a public Blockchain application is in recording the generation and financial flow of the Bitcoin cryptocurrency. However, there are also some public Blockchains where access is limited to either write or read rights, depending on the rôle of the user in the system. The aim of a private Blockchain, on the contrary, is to conceal the details of the users. To ensure that, access is limited to some trusted participants or members of a single organization. A Blockchain that is controlled by a consortium is known as a consortium blockchain. This is particularly pertinent amongst governmental institutions and allied sister concerns or their subsidies thereof.

The implementation of the Blockchain technology being public puts the BC well ahead of other technologies, especially in terms of security aspects. Since each participating nodes possesses its own copy of the complete blockchain i.e. whole blocks of updated records and transactions, the data thus remain unaltered. Any unauthorised or unexpected changes will thus be publicly verifiable. However, the data recorded in such publicly available blocks are hashed and encrypted (by the private key) to ensure security and anonymity. Because the private key is used to encrypt the data, it cannot be publicly interpreted, as shown in Fig. 3.

Although a centralised implementation of the Blockchain technology is possible, it is mostly decentralized in nature, which is considered to be another one of its major advantages. It is decentralized in the sense that:

- The data, comprising the transactions and associated blocks, are distributed among the participating nodes of the Blockchain network, rather than storing them in a single piece of node or storage device.
- The transactions are approved by a set of specific rules or algorithms, thus eliminating the influence of being biased by one single authority involving substantial trust in order to reach a consensus.
- The Blockchain systems only allows new verified blocks be appended to the old chain. As the previously added blocks are already public and distributed, they are openly verifiable and hence cannot be altered or revised. Thus, the overall security of a Blockchain ecosystem is another advantage over other technologies.

Once a transaction is triggered by a participant, it is not added straightaway to the chain of blocks i.e. the blockchain. In order for a newly initiated transaction to be appended with the existing chain, the transaction has to go through the validation and verification processes. The participating nodes of the Blockchain networks must apply a set of predefined rules or specific algorithms for this purpose. The set of rules or algorithms basically defines what is perceived as "valid" by the respective Blockchain system and may vary from one to another. Rather than adding one single transaction in a block, usually a number of such transactions are put together in order to construct a new block. This newly prepared block is then sent to all other participating nodes of the Blockchain network so that they can be appended to their copy of the existing chain of blocks. Each succeeding block of the chain comprises a hash, a unique digital fingerprint, of the preceding one.

The Blockchain not only verifies and validates all the newly triggered transactions but also maintains an irreversible lifelong record of them, while assuring that all the identification related information of the users or the participants are kept incognito. Thus all the personal information of the users is sequestered while substantiating all the transactions. This is achieved by reconciling mass collaboration by cumulating all the transactions in a computer code based digital ledger. Thus, in a Blockchain system, instead of trusting each other or an intermediary, the participants need to trust the decentralized network system itself. Thus the Blockchain itself has become the ideal "Trust Machine" [4, 5] paradigm.

Although the Bitcoin cryptocurrency first used the Blockchain, it is considered to be just an exemplary use of the BC. Blockchain technology is a relatively novel technology in the domain of computing that is enabling illimitable applications, such as in and not just limited to: healthcare systems, human resource management, recruitment, storing and verifying legal documents including deeds and various certificates, IoT and the Cloud. In fact, Tapscott [6] has perfectly connoted Blockchain to be the "World Wide Ledger", facilitating many novel applications beyond just the simple verifying of transactions such as in: recording smart deeds, decentralized and/or autonomous organizations/government services *et cetera*. Figure 4, shows the typical and diverse applications of the blockchain technology.

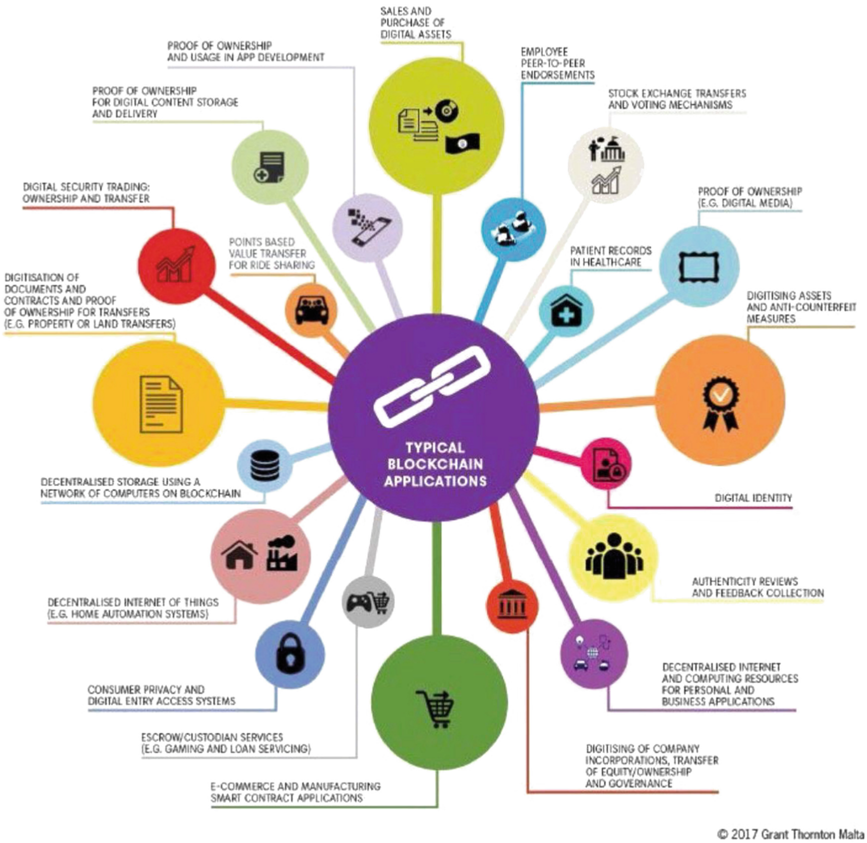


Fig. 4. Typical application of the Blockchain technology [7].

3 Internet of Things (IoT)

The term 'Internet of Objects' or 'Internet of Things' (more commonly referred to as 'IoT') - denotes the electronic or electrical devices of many different sizes and capabilities connected to the Internet. This connection is mainly by using wireless sensors, but excluding those primarily involved in communications with human beings, i.e. the traditional Internet. New IoT devices are being marketed on a regular basis and thus the scope of the connections is ever broadening beyond just basic machine-to-machine communication (M2M) [8].

There are many types of IoT devices employing a wide range of applications, protocols, and network domains [9]. The growing preponderance of IoT technology is enabled by the physical objects being connected to the Internet by various types of short-range wireless technologies such as sensor networks, RFID, ZigBee and through location-based technologies [10].

The emergence of IoT as a distinctive entity was reached (according to the Internet Business Solutions Group (IBSG)) when more inanimate objects were directly connected to the Internet bypassing human users [11]. This accelerating process has been gaining momentum ever since the rollout of CISCO's 'Planetary Skin', the Smart Grid and intelligent vehicles [11]. IoT is already on the verge of making the Internet truly pervasive, with devices already embedded into consumer white goods, including personal and intimate devices in our daily lives. IoT devices are only standardised in their use of the Internet networking protocols and not how they interface to the Internet or with each other. This immediate potential inhibiting factor needs to be addressed.

IoT may be deployed with added privacy, security and management features to link, for example, vehicle electronics, home environmental management systems, telephone networks and control of domestic utility services. The broadening scope of IoT and how it can link with heterogeneous networks is shown in Fig. 5 [11], below.

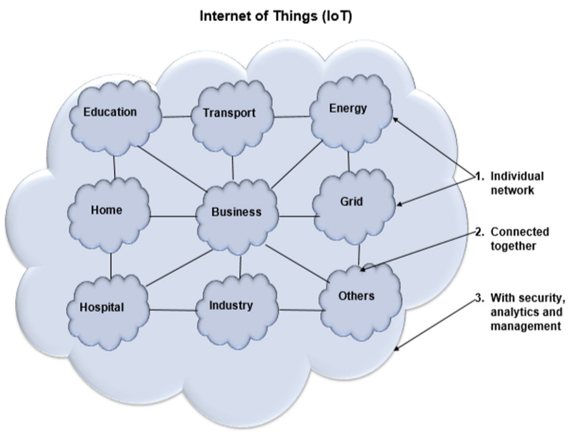


Fig. 5. IoT viewed as a network of networks [11].

A standard IoT ecosystem typically comprises of the following five components:

1. Sensors: sensors are mainly responsible to collect and transduce the required data;
2. Computing Node: such nodes containing the central processing unit (CPU), are required for processing the data and information received from a sensor;
3. Receiver: which is actually a transceiver, facilitates the collection of the message sent by the local and remote computing nodes or other associated devices;
4. Actuator: which could be electro-mechanical in nature, works on the basis of the decision taken by the Computing Node, processing the information received from the sensor and/or from the Internet, then triggering the associated device to perform a function; and
5. Device: to perform the desired task as and when triggered.

4 BC Enabled Enhanced IoT Security

In an IoT ecosystem [12, 13], most of the communication is in the form of Machine-to-Machine (M2M) interactions, that is, without any human intervention whatsoever. Thus establishing trust among the participating machines is a big challenge that IoT technology still has not met extensively. However, Blockchain may act as a catalyst in this regard, by enabling enhanced scalability, security, reliability and privacy [4, 5]. This can be achieved by deploying Blockchain technology to track billions of devices connected to the IoT ecosystems and then used to enable and/or coordinate transaction processing. In fact, a specific search engine exists, called “Shodan”, that describes itself as “the world’s first search engine for Internet-connected devices” [14]. The use of this search engine by anyone will also expose any insecure IoT devices and hence their need for rectification. Application of the Blockchain in any IoT ecosystem will further enhance the reliability by completely eliminating any Single Point of Failure (SPF). In Blockchain, data is encrypted using cryptographic algorithms as well as the hashing techniques. Thus, the application of Blockchain in an IoT ecosystem can offer better security services. However, to perform the hashing techniques and implement the cryptographic algorithms, the systems shall obviously demand more processing power, which IoT devices currently lack. Thus further research is required to overcome this present limitation, including extending longevity of the powering source.

Underwood [15] considers the application of Blockchain technology to completely overhaul the digital economy. Ensuring and maintaining trust is both the primary and initial concern of the application of the blockchain. BC can also be used to gather chronological and sequence information of transactions, as it may be seen as an enormous networked time-stamping system. For example, NASDAQ is using its ‘Linq blockchain’ to record its private securities transactions. Meanwhile the Depository Trust & Clearing Corporation (DTCC, USA) is working with Axoni in implementing financial settlement services such as post-trade matters and swaps. Regulators are also interested in BC’s ability to offer secure, private, traceable real-time monitoring of transactions.

Securing operational technology is also of paramount importance. Thus the Blockchain can help to prevent tampering and spoofing of data by managing and securing industrial IoT and operational technology (OT) devices. So once a sensor, device or controller has been deployed and is working, it cannot be touched. Since any compromised devices will be recorded in the BC. Thus as Pindar propounds:

5 Concluding Discussions

To answer the research question “To what extent can the Blockchain be used in enhancing the overall security of the Internet of Things (IoT) ecosystems”, this paper first introduced how these two emerging technologies works. The current security issues related to IoT systems were also discussed. The authors of the article then investigated how the application of the Blockchain can eliminate these security concerns inherent in the IoT ecosystem and improve its overall security.

Both the Blockchain and Internet of Things (IoT) are two relatively new but promising technologies being successfully used in multifaceted applications. The way the application of the Blockchain has widened beyond its initial use for Bitcoin generation and dealing has conclusively shown its relevance and versality in general networked secure transactions. IoT also proved itself to be capable of doing far more things than being a simple wireless sensor network. In fact, Blockchain and its variants combined offers many security aspects such as enhanced privacy, stronger security, full traceability, inherent detailed data provenance and accurate time-stamping which other technologies still could not offer as standalone features. Thus BC has seen its adoption beyond its initial application areas and is now used to secure any type of transactions, whether: human-to-human, machine-to-machine or human-to-machine communications. The adoption of Blockchain appears to be secure, especially allied with the world emergence of the Internet-of-Things (IoT). Its decentralized application across the already established global Internet is also very appealing, in terms of ensuring network redundancy, data redundancy through distribution and hence survivability.

References

1. Rosic, A.: Proof of Work vs Proof of Stake: Basic Mining Guide. Blog 2017. https://blockgeeks.com/guides/proof-of-work-vs-proof-of-stake
2. Shashank: Blockchain Technology – Unfolding the Technology behind Bitcoins. Blog 2017. https://www.edureka.co/blog/blockchain-technology/
3. Onik, M.M.H., Miraz, M.H., Kim, C.-S.: A recruitment and human resource management technique using Blockchain technology for Industry 4.0. In: Proceeding of Smart Cities Symposium (SCS 2018), Manama, Bahrain, pp. 11–16 (2018)
4. Miraz, M.H., Ali, M.: Applications of Blockchain technology beyond cryptocurrency. Ann. Emerg. Technol. Comput. (AETiC) **2**(1), 1–6 (2018). http://aetic.theiaer.org/archive/v2n1/p1.pdf
5. Miraz, M.H.: Blockchain: technology fundamentals of the trust machine. Machine Lawyering, December 2017. https://doi.org/10.13140/RG.2.2.22541.64480/2
6. Tapscott, D., Tapscott, A., Revolution, B.: How the Technology Behind Bitcoin Is Changing Money, Business, and the World, 1st edn. Penguin Publishing Group, New York (2016)

7. Sutcliffe, M.: An overview of Blockchain applications — this is just the beginning! Blog 2017. https://twitter.com/MikeSutcliff/status/912382978680082433
8. Benattia, A., Ali, M.: Convergence of Technologies in the M2M Space. In: IEEE International Conference on Applied Electronics, Pilsen, Czech Republic, pp. 9–12 (2008)
9. Höller, J., et al.: From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence, 1st edn. Academic Press Ltd., London (2014)
10. Feki, M.A., Kawsar, F., Boussard, M., Trappeniers, L.: The Internet of Things: the next technological revolution. Computer **46**(2), 24–25 (2013). http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6457383
11. Evans, D.: The Internet of Things: how the next evolution of the internet is changing everything. In: White Paper, Cisco Internet Business Solutions Group (IBSG), Cisco Systems, Inc., San Jose (2011). http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
12. Miraz, M.H., Ali, M., Excell, P., Rich, P.: A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT). In: The Proceedings of the Fifth International IEEE Conference on Internet Technologies and Applications (ITA 2015), Wrexham, UK, pp. 219–224 (2015). http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7317398
13. Miraz, M.H., Ali, M., Excell, P.S., Picking, R.: Internet of nano-things, things and everything: future growth trends. Future Internet (2018)
14. Shodan: Shodan, May 2018. https://www.shodan.io/
15. Underwood, S.: Blockchain beyond bitcoin. Commun. ACM **59**(11), 15–17 (2016). https://doi.org/10.1145/2994581

Challenges of Internet of Things and Big Data Integration

Zainab Alansari[1,2](✉), Nor Badrul Anuar[1], Amirrudin Kamsin[1], Safeeullah Soomro[2], Mohammad Riyaz Belgaum[2], Mahdi H. Miraz[2,3], and Jawdat Alshaer[4]

[1] University of Malaya, Kuala Lumpur, Malaysia
z.alansari@siswa.um.edu.my, {badrul,amir}@um.edu.my
[2] AMA International University, Salmabad, Kingdom of Bahrain
{zeinab,s.soomro,bmdriyaz}@amaiu.edu.bh
[3] Wrexham Glyndŵr University, Wrexham, UK
m.miraz@gmail.com
[4] Al-Balqa Applied University, Salt, Jordan
jawdat_alshaer@bau.edu.jo

Abstract. The Internet of Things anticipates the conjunction of physical gadgets to the Internet and their access to wireless sensor data which makes it expedient to restrain the physical world. Big Data convergence has put multifarious new opportunities ahead of business ventures to get into a new market or enhance their operations in the current market. considering the existing techniques and technologies, it is probably safe to say that the best solution is to use big data tools to provide an analytical solution to the Internet of Things. Based on the current technology deployment and adoption trends, it is envisioned that the Internet of Things is the technology of the future; while to-day's real-world devices can provide real and valuable analytics, and people in the real world use many IoT devices. Despite all the advertisements that companies offer in connection with the Internet of Things, you as a liable consumer, have the right to be suspicious about IoT advertisements. The primary question is: What is the promise of the Internet of things concerning reality and what are the prospects for the future.

Keywords: Internet of Things · Big data · Cloud computing

1 Introduction

Convergence between wireless communications, Digital electronic devices, and Micro-electro-mechanical systems (MEMS) technologies led to the rise of the Internet of Things. Internet-connected objects like computers, smartphones, tablets and Wi-Fi devices, sensors, wearable devices and household appliances are all the objects of the IoT components [1] and considered as "Things". The Internet of Things means the production of tremendous amount of data and a collection of substantial different data bulk that has not seen so far. Big data management and generating smart data are the research interests of the companies which produces these data. Without the application

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 47–55, 2018.
https://doi.org/10.1007/978-3-319-95450-9_4

of big data analytics, the vast volume of data generated by the Internet of Things adds to the overhead of any organization and considered to be one of the most significant obstacles towards the deployment of this technology [2]. In other hand, organizations must know what to do with the massive amount of data that collected. The explosive growth of Internet use, along with smartphone and social programs and machine-to-machine (M2M) communication, has revolutionized big data [3].

The primary challenge is the design of a model to analyze big data. In other words, we need to change our point of view about the blocks created by the Internet of Things [4]. Instead of looking at data as a data warehouse, we should look at the supply chain. Since the tools are enabling the extraction of numerous unstructured and new data sources, the lack of sufficient data issue will reveal soon. Therefore, must overcome the following two fundamental problems [5]. Firstly, not to miss the data that truly needed and secondly, make sure not to spend time on unnecessary data. Despite the supply chain, organizations can fill in existing gaps in any way they need [6]. To this end, companies can take advantage of these three strategies:

- Design interfaces for applications that created before.
- Request help from partners or third parties who can provide the required data.
- Generate data with the commission of the physical environment around the business.

Apparently, getting the accurate data is not the only problem with organizations. The other challenge is to acquire the necessary skills in the field of analytical analysis to deal with the big data. Traditional skills in the domain of data analysis on the Internet of Things have not been practical. Companies need people who know analytics, as well as understanding the new meaning of the initial data for a specific industry. Along with the growing trend in data generation and analysis required for these big data, organizations are forced to prepare devices that connect customers and objects at any time and any place [7].

One of the critical infrastructures required by active companies in the Internet of Things is having the culture of data-driven decision-making. The Internet of Things, in essence, provides a flow of accurate data derived from the real world. These data must go beyond the process of transforming data into information, then knowledge and awareness, and ultimately wisdom, using traditional analytical skills in this area to be meaningful. For example, in the field of agriculture, an expert scientist needs to know how much irrigation required to produce a product under different climatic conditions [8].

In intermittent IoT, the ability to collect weather information, field, and product information is automatic and accurate. However, when the data collected, the actions to be taken on data depends on the expert opinion. Coupled with the growing amount of data and analysis needed, companies need to be ready for a range of devices that connect consumers and objects at any point and any time. Those who accept the philosophy of data supply chain will go through these waves of information without dwelling in detail. Apart from the cases mentioned, the investment required in the field of sensors, analytical capabilities, and data security and support are among the other obstacles faced by the technology of the Internet of Things [9].

Hitherto, considering the existing techniques and technologies, it is probably safe to say that the best solution is to use big data tools to provide an analytical solution to the

Internet of Things. Based on the current technology deployment and adoption trends, it is envisioned that the Internet of Things is the technology of the future; while today's real-world devices can provide real and valuable analytics, and people in the real world use many IoT devices. Despite all the advertisements that companies offer in connection with the Internet of Things, you as a liable consumer, have the right to be suspicious about IoT advertisements. The primary question is: What is the promise of the Internet of things concerning reality and what are the prospects for the future?

The core value of an IoT system is the ability to analyze the data needed and achieve practical and useful insight without making any mistake. Hence for two reasons, creating a communication medium is not easy. Firstly, it must have the ability of scalable analysis. Secondly, it must make comprehensive usage of this ability regarding the volume and speed of the IoT devices that generate their data [10]. In this research, we discussed different solutions to help everyone to stay away from a series of issues on how to develop an ideal analysis of big data.

2 Internet of Things and Its Impact on Big Data

Nowadays, with the help of the Internet of Things technology, the ability to connect each object to the network is provided. The Internet of Things offers a chain of connected people, objects, applications, and data over the Internet for remote control, interactive, services integration and management. Hence this network is overgrowing; we need a platform that can collect and store the data generated by IoT devices. Some of the advanced Internet of Things services require a mechanism to collect, analyze and process raw data from sensors to be used as operational control information. Some types of sensor data may have very high volumes because of the significant number of sensors in the Internet of Things ecosystems. Possibly, we would see a data flood coming from these devices. Accordingly, we need new technologies or architectural patterns in the area of data collection, storage, processing, and data retrieval [11].

Databases designed and implemented to work with the Internet of Things have their specific conditions and characteristics. The proliferation of NoSQL technologies can be considered as an indication that the management of the Internet of Things requires the use of novel approaches in administering and utilizing databases. The provision of cloud computing platforms based on the internet of things eases the opportunities to enter this arena and take advantage of its achievements and services for many businesses of various dimensions [12]. Moreover, there will be a need for IoT data analysts of an acquaintance and entry into the fascinating world of big data. The Internet of things affects people, processes, data, and things.

- People: More objects can be monitored and controlled, and subsequently increased individual's abilities.
- Processes: Users and more machines will be able to interact with each other in real time. Therefore, very complex tasks can finish in less time as the percentage of engagement and participation in doing a job is far more significant.
- Data: The ability to collect data at a higher frequency and reliability provided which can lead to a correct decision making.

- Things: the ability to control things more accurately. Therefore, the value of objects such as mobile devices will be more and can help with much more than the current situation.

Big data convergence, super-efficient networks, social media, low-cost sensors, and a new generation of advanced analytics have provided countless new opportunities for business enterprises to use them to either enter a new market or strengthen their operations in the current market. The Internet of Things is one of these new markets that can provide countless opportunities for businesses in different fields. Significant changes happen by a slight difference, and the Internet of Things can be the source of millions of changes in different areas over the next few years. Consider the Internet of Things as one of the causes for generating data which its impact on IT infrastructure and the use of advanced methods in data analysis are among its exceptional and vital opportunities in this regard [13].

A collection, preparation, and analysis of large volume of data will not be an easy task. Firstly, the amount of data can be doubled in several months and secondly, the gendered nature of this kind of data has its particular complexities. The variety in the template or the format of this type of data is extensive and often includes hundreds of pseudo-structured forms or unstructured formations. Most importantly, to achieve a broad view of the sensor data, it should be possible to analyze and manage every structured and unstructured data. An analysis based on a specific data format can significantly limit the created potential insight. Data analysis, regardless of its composition, is centralized and side by side, which provides a comprehensive analytical perspective to decision makers of business enterprises.

It indicates the consideration of the limitations and stock of traditional enterprise data and current business intelligence software and design. Organizational data warehouses are not able to focus on unstructured data. Accordingly, we need to look for solutions that enable unstructured data storage and analysis [14]. If we want to convert the unstructured data into a specific structure by defining a particular structure and using relational database tables, we will lose time. Consequently, that will surely be possible with the condition of not having the technical limitations. The use of any technology to create an analytical infrastructure that has some limitations can reduce the ability to analyze and, in practice, minimizes the potential for possible value creation.

Analysis of big data with the help of related technologies can be one of the leading actors in this field. Analyzing big data in some cases can help us:

- Combine, integrate and analyze all structured, semi-structured and unstructured data regardless of source, type, size, and format.
- A quick and cost-effective analysis of the high volume data to create an appropriate insight into a decision making process.

The Internet of Things has promised to influence various industries from insurance companies and banks to telecom and other business enterprises. Organizations need to modify data analysis methods so that they can collect, clean, prepare and analyze RFID sensor and tag data in the shortest time possible. Deciding on the actual data and in the shortest reasonable time is the confidentiality of a business firm in today's highly competitive environment. With a proper big data management and the creation of an appropriate atmosphere for their analysis, an organization's vision for proper decisions making increases [15].

3 Benefits of IoT Based on Big Data

In literature, various structures for big data analysis and IoT proposed, which can manage the challenges of storage and analysis of high volume data from intelligent buildings. The first presented structure consists of three main components which are big data management, IoT sensor, and data analysis. These analyzes use are in the real-time management of oxygen level, dangerous gases/soot and the amount of ambient light in smart buildings. In addition to smart building management, IoT devices and sensors for receiving traffic information can be used in real time traffic management with low cost and examine the strengths and weaknesses of existing traffic systems.

In smart city management, the big data used in the analysis of data which obtained from different sensors such as water sensors, transportation network sensors, monitoring devices, smart home sensors and smart car park sensors. These data are generated and processed in a multi-stage model and ultimately reached a decision-making stage. These steps are data production, data collection, data integration, data categorization, data processing and decision making [16].

Sometimes it is essential to pay attention to the concepts of web technology in particular proposed framework to investigate the analytical results obtained from the big data in the Internet of Things. In the literature, this topic has devised, and a conceptual framework has been proposed consisting of 5 layers:

- Data Collection layer - collected data from various sources, the input layer is the proposed framework.
- Extract-transform-load (ETL) layer - provides the ability to change the format of information received from different types of sensors into a defined format.
- The semantic reasoning rules layer - an inference engine that acts on the information received from the ETL layer.
- Learning layer - From the data tailored to the existing extraction data, extract the various specifications and attributes, and finally, Machine learning-based models provided.
- Action layer - executes a set of predefined actions by the outputs of the learning layer.

Other applications of IoT help with geographic information analysis, cloud computing flow processing, big data analysis, and storage, cloud computing security, clustering mechanisms, health, privacy security, performance evaluation of monitoring algorithms, manufacturing systems, and energy development [17].

4 Big Data Collection and Storage

Numerous protocols make it possible to receive events from IoT devices, especially at lower levels. It does not matter if the device connected to a Bluetooth network, cellular network or Wi-Fi, or it communicates through a hardware connection, it is enough to send a message from a broker using a defined protocol. One of the most popular protocols for large IoT applications is MQ Telemetry Transport (MQTT). MQTT refers to the transmission of messages through remote sensing and queuing which is an M2 M IoT connection. This protocol designed as a very lightweight request/response (point-to-point) messaging transfer. MQTT is practical and useful for connecting to distant locations that require low memory or low network bandwidth; For example, this protocol used in sensor communication through a satellite link with a broker in dial-up connections with healthcare providers at different times and in a range of home automation and small devices. Its design principles are to minimize network bandwidth and resource requirements, and at the same time, it also guarantees trust and confidence in the message delivery [18].

There are also other alternatives, such as the limited application protocol, XMPP, and other protocols. Constrained Application Protocol (COAP) is a software protocol that used in straightforward electronic devices which provides communication through the Internet interactively. Constrained RESTful Environments (CoRE) group along with Internet Engineering Force (IETF) did the main standardization work of this protocol.

Extensible Messaging and Presence Protocol (XMPP) is a communication protocol for the Extensible Markup Language (XML). The XMPP enables exchanging close to real time between structured data but expandable between either any two or more network entities.

We recommend starting with MQTT due to its the availability and extensive coverage and the availability of a large number of client applications and open source brokers unless having a convincing reason to choose another protocol. Mosquitto is one of the MQTTs most widely used open source and will be the definitive choice for applications; if this concept should be proved based on limited budgets and want to avoid the cost of dedicated devices, the fact that Mosquitto is open-source brokerage is precious. Regardless of what protocol to choose, it will eventually receive messages that represent events and observations from devices connected to the Internet.

As long as the message received by a broker (Mosquitto), you can send it to the analytics system. The best way is to store source data before any transfer or making any changes to them; this is, of course, worthwhile when debugging problems occurs at the conversion stage. There are several ways to store IoT data which many uses Hadoop and Hive in their projects. Recently researchers are and successfully working with NoSQL databases such as Couchbase. Couchbase provides a right combination of high performance and low latency indicators. Couchbase is a Document-Oriented Database which lacks a layout that covers a significant amount of data and also can add a variety of new events [19] quickly.

Direct data writing on HDFS is another good solution, especially if using Hadoop and batch analysis is considered as part of the analytics workflow. For writing source data in a permanent storage location, the direct code can be added manually to the message broker at the level of the IoT protocol (for example, if you use MQTT, Mosquitto Broker). The other way is to attach messages to a medium-sized middleware broker like Apache Kafka and use different Kafka users to transfer messages to different parts of the system [20].

One of the established patterns is to place messages in Kafka in two user groups based on the subject, where one of the groups writes raw data to their permanent storage, while another, transmits data into a real-time processor engine like Apache Storm. If Storm is used instead of Kafka, a Bolt processor can install in topology, which does nothing except sending messages to a permanent storage location. If MQTT and Mosquitto are used, sending direct messages to the Apache Storm topology via the MQTT's spout source is an easy way to link things together.

5 Conclusion

The development of IoT devices, smartphones, and social media provides decision makers with opportunities to extract valuable data about users, anticipate future trends and fraud detection. With the creation of transparent and usable data, big data can create the organizations' values, make the changes clear and expand their performance. The use of data generated from the IoT and the analytical tools creates many opportunities for organizations. These tools use predictive modeling technologies, clustering, classification to provide data mining solutions.

IoT improves the decision-making habits of decision-makers. The emergence of IoT and related technologies, such as cloud computing, provides the ability to remove data sources in different domains. Typically, any data is considered useful in the domain itself, and data on shared domains can be used to provide different strategies. Machine learning, deep learning, and artificial intelligence are key technologies that are used to provide value-added applications along with IoT and big data in addition to being used in a stand-alone mode. Before the advent of IoT and cloud computing, the use of these technologies was not possible due to the high amount of data and required computational power. Different data analysis platforms, Business intelligence platforms and analytical applications are emerging platforms that have been introduced to help industries and organizations in transforming processes, improving productivity, and the ability to detect and increase agility.

It is anticipated that the speed of technological progress in the next ten years, will be equal to the past thirty years. Therefore, we have to use all our efforts to update our lives to the Internet of Things technology regarding hardware and software.

Acknowledgment. This work is supported by CTRG Research Group of the College of Computer Studies, AMA International University, Kingdom of Bahrain.

References

1. Belgaum, M.R., Soomro, S.U., Alansari, Z., Alam, M.: Treatment of reactive routing protocols using second chance based on malicious behavior of nodes in MANETS. Pak. J. Eng. Technol. Sci. **6**(2) (2018)
2. Alansari, Z., Soomro, S., Belgaum, M.R., Shamshirband, S.: The rise of Internet of Things (IoT) in big healthcare data: review and open research issues. In: Saeed, K., Chaki, N., Pati, B., Bakshi, S., Mohapatra, D.P. (eds.) Progress in Advanced Computing and Intelligent Engineering. AISC, vol. 564, pp. 675–685. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6875-1_66
3. Miraz, M.H., Ali, M., Excell, P.S., Picking, R.: A review on Internet of Things (IoT), Internet of everything (IoE) and Internet of nano things (IoNT). In: Internet Technologies and Applications (ITA 2015), pp. 219–224. IEEE, September 2015
4. Belgaum, M.R., Soomro, S., Alansari, Z., Alam, M., Musa, S., Su'Ud, M.: Challenges: bridge between cloud and IoT. In: 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), Bahrain. IEEE (2017)
5. Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U.: The rise of "big data" on cloud computing: review and open research issues. Inf. Syst. **47**, 98–115 (2015)
6. Ali, M., Miraz, M.H.: Cloud computing applications. In: Proceedings of the International Conference on Cloud Computing and eGovernance, p. 1 (2013)
7. Riggins, F.J., Wamba, S.F.: Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In: 2015 48th Hawaii International Conference on System Sciences (HICSS), pp. 1531–1540. IEEE, January 2015
8. Sun, Y., Song, H., Jara, A.J., Bie, R.: Internet of things and big data analytics for smart and connected communities. IEEE Access **4**, 766–773 (2016)
9. Alansari, Z., Anuar, N.B., Kamsin, A., Soomro, S., Belgaum, M.R.: Evaluation of IoT-based computational intelligence tools for DNA sequence analysis in bioinformatics. In: International Conference on Advanced Computing and Intelligent Engineering 2017, India. Springer (2017)
10. Turner, V., Gantz, J.F., Reinsel, D., Minton, S.: The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, p. 5. IDC Analyze the Future, Boston (2014)
11. Alansari, Z., Anuar, N. B., Kamsin, A., Soomro, S., Belgaum, M.R.: Computational intelligence tools and databases in bioinformatics. In: 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), Bahrain. IEEE (2017)
12. Belgaum, M.R., Soomro, S., Alansari, Z., Alam, M.: Cloud service ranking using checkpoint-based load balancing in real-time scheduling of cloud computing. In: Saeed, K., Chaki, N., Pati, B., Bakshi, S., Mohapatra, D.P. (eds.) Progress in Advanced Computing and Intelligent Engineering. AISC, vol. 563, pp. 667–676. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6872-0_64
13. Perera, C., Ranjan, R., Wang, L., Khan, S.U., Zomaya, A.Y.: Big data privacy in the internet of things era. IT Prof. **17**(3), 32–39 (2015)
14. Alansari, Z., Soomro, S., Belgaum, M.R., Shahaboddin, S.: A new conceptual model for BYOD organizational adoption. Asian J. Sci. Res. **10**, 400–405 (2016). https://doi.org/10.3923/ajsr.2017.400.405
15. Da Xu, L., He, W., Li, S.: Internet of things in industries: a survey. IEEE Trans. Industr. Inf. **10**(4), 2233–2243 (2014)

16. Alansari, Z., Anuar, N.B., Kamsin, A., Soomro, S., Belgaum, M.R.: The internet of things adoption in healthcare applications. In: The IEEE 3rd International Conference on Engineering, Technologies and Social Sciences 2017 (ICETSS 2017) (2017)
17. Belgaum, M.R., Musa, S., Alam, M., Su'ud, M., Soomro, S., Alansari, Z.: Load balancing with preemptive and non-preemptive task scheduling in cloud computing. Int. J. Comput. Sci. Eng. (IJCSE), Inderscience. Special Issue on Recent Innovations in Cloud Computing and Big Data (2017)
18. Suciu, G., Suciu, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., Fratu, O.: Big data, internet of things and cloud convergence–an architecture for secure e-health applications. J. Med. Syst. **39**(11), 141 (2015)
19. Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V., Rong, X.: Data mining for the internet of things: literature review and challenges. Int. J. Distrib. Sens. Netw. **11**(8), 431047 (2015)
20. Miraz, M.H., Ali, M., Excell, P.: Multilingual website usability analysis based on an International user survey. arXiv preprint arXiv:1708.05085 (2017)

Software Engineering

Model-Based Metrics to Estimate Maintainability

Nada Almasri(✉) and Luay Tahat

Gulf University for Science and Technology, West Mishref, Kuwait
{Almasri.n, tahat.l}@gust.edu.kw

Abstract. Software maintenance is becoming more challenging with the increased complexity of software and frequent applied changes to accommodate the rapidly changing technologies and user requirements. In this paper we provide model-based metrics to estimate the maintainability of state-based systems. The purpose of the metrics is to provide a tool that can be used by the system maintenance team to identify critical artifacts of the underlying system and to allow for better planning of the change process. The provided metrics is based on Extended Finite State Machine models (EFSM), and it provides two measures to identify critical transitions. The experimental study shows that the metrics is highly effective in spotting transitions that can cause severe propagation of a change when they are being changed, as well as transitions that are highly sensitive to changes applied to an EFSM model.

Keywords: Maintainability · EFSM · Critical transitions · Sensitive transitions

1 Introduction

The demand for large and complex software systems has been steadily increasing over time. The development and maintenance of these systems are difficult and costly due to the increased complexity of the systems. A major challenge during software maintenance is determining the consequences of applying a requested change to the system. This change may be due to a request to add a functionality, remove a functionality, or fix a bug in the system. Within this context, the system developers would want to estimate (1) if a modification is applied on one component of the system, will other components be affected by this modification? What are these affected components? What percentage of the system do they make? (2) for a stable system component which is not touched by the requested modification, what is the possibility that the modification will propagate to the component? The first set of questions focus on estimating the severity of the requested modification in terms of the number of components affected directly or indirectly by that modification, while the second question focus on estimating the sensitivity of certain components of the system to the modifications applied somewhere else in the model. Estimating the ***severity*** of a modification and the ***sensitivity*** of the system components to modifications can greatly enhance the maintainability of the system as it allows the development team to forecast the scope of the change in order to effectively plan its implementation.

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 59–69, 2018.
https://doi.org/10.1007/978-3-319-95450-9_5

One way to manage the complexity of system development process is to use system models in order to reduce ambiguity, misunderstanding, and misinterpretation of system specifications [1–3]. Furthermore, models can be used for test generation [4, 5], test suite reduction [6, 7], and test case prioritization [8–13]. In this paper we use Extended Finite State Machine models which are used to model state-based systems, and we extract the maintainability metrics from these models instead of dealing with their complex underlying system.

In the context of EFSM models, the severity measure predicts the extent to which a change applied to one EFSM transition will propagate to other transitions in the model. The sensitivity measure predicts how often a particular transition under consideration will be affected by a modification applied somewhere else in the model. A transition is identified as a critical transition when a modification applied to it severely propagates to other transitions, or when it has high probability to be affected by a modification applied elsewhere in the model.

The rest of the paper is organized as follows: Sect. 2 provides an overview of state based modeling. Section 3 introduces the two measures used to identify critical transitions. In Sect. 4 presents the empirical study, while Sect. 5 outlines the related work. Finally, in Sect. 6 the conclusion and the future research are discussed.

2 Related Work

Failure mode, effects, and criticality analysis (FEMCA) is a familiar analytical technique in engineering, and particularly in fields such as aviation and automotive [14]. The technique is usually used during the design and the production of new products to estimate the safety risks and hazards. Within the safety context, the technique is mainly based on brain storming the possible failures and the expected consequences of theses failures from human safety perspective.

In the context of software engineering, how critical a modification applied to the software is, is referred to as "impact analysis". Bohner and Arnold [19] define impact analysis as "identifying the potential consequences of a change, or estimating what needs to be modified to accomplish a change". Several research papers presented code-based impact analysis techniques [15], while only a few targeted model-based impact analyses [16, 17]. Almasri [17] proposed an approach to measure the impact of a change at the model level. Their work focused on measuring the change impact for a change applied to EFSM models using model dependencies.

Generally, impact analysis techniques are used to measure the impact of a given modification. The metrics we are suggesting, however, in this paper attempts to estimate how critical EFSM transitions of a specific EFSM model are in general, for any potential change in the future.

3 State-Based Modeling with EFSM

An EFSM model M can be formally expressed as: M = (Σ, Q, Start, Exit, V, O, R) where:

Σ is the set of events, Q is the set of states, Start ∈ Q is the start state, Exit ∈ Q is the exit state, V is a finite set of variables, O is the set of actions, R is the set of transitions, where each transition T is represented by the tuple: $T = (E, C, A, Sb, Se)$ where: E ∈ Σ is an event, C is an enabling condition defined over V, A is a sequence of actions, $A = \ <a1, a2, \ldots.., aj>$, where ai ∈ O. The action may manipulate variables, read input or produce output. Sb ∈ Q is the transition's originating state, Se ∈ Q is the transition's terminating state.

A transition T in R is triggered when the system is in the originating state Sb(T), the event E(T) occurs, and the enabling condition C(T) is evaluated to TRUE. When transition T is triggered, the A(T) sequence of actions is performed and the system is transferred to the terminating state Se(T). EFSM models may be depicted as graphs where states are represented by nodes and transitions by directed edges between states.

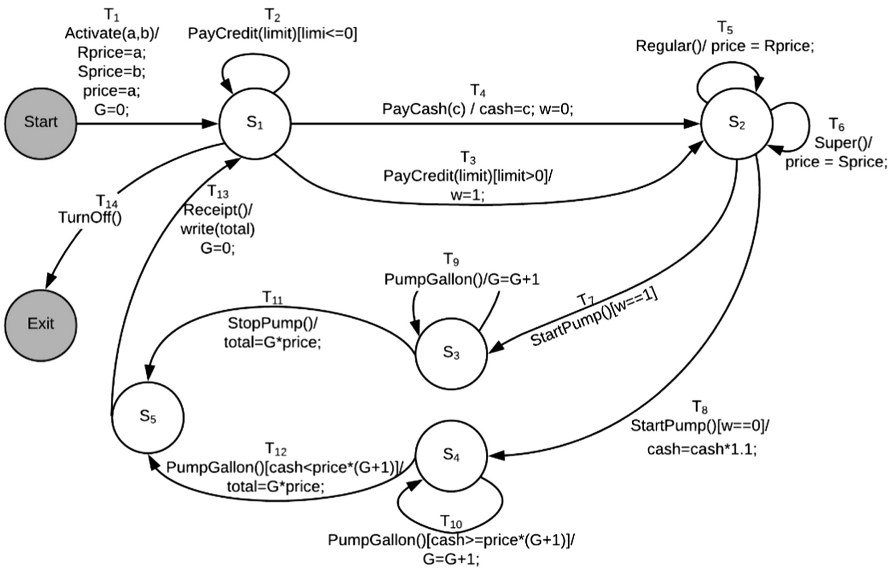


Fig. 1. Fuel pump EFSM model

Figure 1 shows an example of an EFSM model for a Fuel Pump system. According to this model, when the pump is activated, the prices for regular fuel and super fuel are initialized with a default price set as regular price. A person using the pump has the choice to pay by credit or cash. If credit payment is chosen, the credit card is validated based on the available limit. After making the payment choice, the customer gets to choose the type of fuel to pump, and the price to be paid by the user is initialized accordingly. If cash payment is chosen, the customer is rewarded with an extra 10%

once the pumping is started. When the person starts pumping fuel, the amount of gas pumped in is tracked. Finally, as the pump stops pumping gas, the total price is calculated, and a receipt is printed out. At this stage, the pump can be used by the next customer, or it can be turned off.

4 Model-Based Metrics Using Model Dependence

The purpose of the metrics is to identify critical EFSM transitions assuming that such transitions require greater attention from the development team during the maintenance and testing phases of the system development lifecycle.

For a given transition Ti in an EFSM model, if a change is requested, the transition Ti can be subject to change, or the change can be applied somewhere else in the model.

If the change is applied to Ti and it propagates to a large number of other transitions in the model, then Ti is considered as a critical transition since its change affects a large portion of the model. In this case, we call this measure *change-severity* of Ti, and it is denoted as Sv(Ti).

If, on the other hand, the change is applied somewhere else other than Ti, then Ti can still be considered critical if it has a high probability to be impacted by that change. In this case, we call this measure Ti's *sensitivity* to change, and it is denoted as Sn(Ti).

In order to quantify the propagation of a change from Ti to other transitions in the model or vice versa, we use model dependencies which exist between EFSM transitions.

4.1 Model Dependence

The metrics presented in this paper is based on data and control dependence which exist between transitions in EFSM models [13, 17]. These dependencies capture the notion of potential "interactions" between transitions in the model.

Data dependence captures the notion that one transition defines a value to a variable and another transition may potentially use this value. There exists data dependence between transitions Ti and Tk if transition Ti modifies value of variable v, transition Tk uses v, and there exists a path (transition sequence) in the model from Ti to Tk along which v is not modified. For example, there exists data dependence between transitions T1 and T6 in the model of Fig. 1 because transition T1 assigns a value to variable Rprice, transition T6 uses Rprice, and there exists a path (T1, T4, T6) from T1 to T6 along which Rprice is not modified.

Control dependence was originally defined for program's Control Flow Graph (CFG) [18]. Control dependence captures the notion that one node in the control graph may affect the execution of another node. In [1], the concept of program control dependence was extended to EFSM models. Control dependence in an EFSM exists between transitions and it captures the notion that one transition may affect traversal of another transition.

For example, transition T5 is control dependent on T4 in the model of Fig. 1 because (1) Sb(T4) does not post dominate Sb(T5) (condition 1 of control dependence definition is true) and (2) state Sb(T5) post dominates transition T4 (condition 2 is

TRUE). Note that Sb(T4) is S1 and Sb(T5) is S2. The issue of control dependence in EFSMs is discussed in more details in [2, 6, 10, 13, 17].

Data and control dependence in the model can be graphically represented by a directed graph where nodes represent model transitions and directed edges represent model data and control dependencies.

More formally, let M = (Σ, Q, Start, Exit, V, O, R) be an EFSM model and let $G = (R, E)$ be a model dependence graph of model M where:

R is a set of nodes (set of transitions).

E is a binary relation on R, $E \subseteq R \times R$, referred to a set of directed edges where: edge (Ti, Tk) $\in$ E, if there exists data or control dependence between transitions Ti and Tk.

Alternatively, the dependency between transitions can be represented as a matrix where the D, C, or B labels are used to represent data dependency, control dependency, and both data and control dependency between two transitions. Table 1 shows dependence matrix for the Fuel Pump example in Fig. 1. From the matrix we can see that transition T1, T2, T3, T4, and T14 don't depend on any other transition on the model. Other transitions have a mix of dependencies on other transitions. For example, Transition T5 has data dependency on transition T1 with respect to variable price, and it has to control dependencies on T3 and T4.

Table 1. Dependency matrix for fuel pump model

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14
T1														
T2														
T3														
T4														
T5	D		C	C										
T6	D		C	C										
T7			B	B										
T8			B	B				D						
T9	D						C		D				D	
T10	D				D	D		B		D			D	
T11	D				D	D	C		D				D	
T12	D				D	D		B		D			D	
T13			C	C							D	D		
T14														

4.2 Transition's Change Severity

The impact of a change applied to an EFSM model can be measured using the approach presented in [17]. However, in this paper, our purpose is to estimate the expected severity of a change if a particular transition undergoes a change, without actually specifying what type of change the transition may experience. Knowing this information allows identifying critical transitions beforehand prior to applying any changes.

To measure Ti's severity of a change, denoted as Sv(Ti), all transitions that are control or data dependent on Ti are identified, and recursively, their dependent transitions are also identified. The set of dependent transition in this case is called the set of Affected transitions with respect to Ti. The larger this set is, the more severe the change of Ti is considered.

To formally define the set of affecting transitions, we define the relationship "affects" as follows:

Let $G = (R, E)$ be the dependence graph of the model M. A transition T in R "affects" another transition T′ in R *if and only if* there is a non-null path from T to T′ in G.

It is worth mentioning that although control and data dependence relationship is not transitive, the "effects" relationship represents the transitive closure of the dependence relationship [3]. For example, if transition T1 depends on transition T2, and transition T2 depends on transition T3, then T3 "affects" T1.

Below, is the formal definition of the set of transitions affected by a particular transition Ti.

Let $G = (R, E)$ be the dependence graph of the model M. The set of affected transitions for a transition Ti in G, denoted as AD(Ti), is the set of all transitions Tj, where Ti "affects" Tj. Formally, we define this set as:

$$AD(T_i) = \{T_j | T_j \in R, \text{ and } T_i \text{ “affects” } T_j \text{ in } G\} \tag{1}$$

Having identified the set of transitions affected by a given transitions Ti, the percent of transitions affected by Ti out of all transitions in the model represents the severity of the change applied to the transition Ti and denoted as $Sv(T_i)$. The number of the transitions in AD(Ti) is denoted as |AD(Ti)|, and the number of the transitions in the model M is denoted as |R|. More formally, the severity of a change applied to transition Ti is estimated using the following formula:

$$Sv(T_i) = |AD(Ti)|/|R| \tag{2}$$

4.3 Transition's Sensitivity to Change

To measure the sensitivity of transition Ti to a potential change applied to the model, all transitions on which Ti is either data or control dependent on are identified in a recursive manner. These set of identified transition are called Ti's *Affecting* transitions. The larger the set of affecting transitions is, the more sensitive to change the transition Ti is considered. The larger this set is, the more sensitive Ti is considered.

Below, is the formal definition of the set of transitions affecting a particular transition Ti.

Let $G = (R, E)$ be the dependence graph of the model M. The set of affecting transitions for a transition Ti in G, denoted as AG(Ti), is the set of all transitions Tj that "affects" the transition Ti. More formally:

$$AG(T_i) = \{T_j | T_j \in R, \text{ and } T_j \text{ "affects" } T_i \text{ in } G\} \quad (3)$$

Having identified the set of transitions affecting a given transtions Ti, the transition's sensitivity to change, denoted as $Sn(T_i)$, is the percent of transitions affecting Ti out of all the transitions in the EFSM model. The number of the transitions in AG(Ti) is denoted as |AG(Ti)|, and the number of the transitions in the model M is denoted as |R|. More formally, the sensitivity of a given transition Ti in an EFSM model can be calculated using the following formula:

$$Sn(Ti) = |AG(Ti)|/|R| \quad (4)$$

Table 2, shows the "affects" relationship between transitions in the fuel pump model, and for each of the transition in the model, the table displays the size of the set of affected transitions AD(T) and the size of the set of affecting transitions AG(T). Each row in the table shows what transitions are affected by a given transition Ti. For example, the first row shows that T1 is affects T5, T6, T10, T11, T12, and T13.

Table 2. "Affects" relationship in fuel pump model

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	|AG(T)|
T1															0
T2															0
T3															0
T4															0
T5	D		C	C											3
T6	D		C	C											3
T7			B	B											2
T8			B	B				D							3
T9			B	B			B		D						4
T10	D		B	B	D	D		B		D					7
T11	D		B	B	D	D	B		D						7
T12	D		B	B	D	D		B		D					7
T13	D		B	B	D	D	B	B	D	D	D	D			11
T14															0
|AD(T)|	6	0	9	9	4	4	3	4	3	3	1	1	0	0	

Table 3, demonstrates the severity Sv(T) and the sensitivity Sn(T) of each transition. From both tables we can see that transition T13 is the most sensitive to change in the fuel pump model since it is affected by 11 out of the 14 transitions in the model (Sn (T13) = 0.79). Indeed, the value 0.79 could be interpreted as 79% of the transitions in the model affects transition T13. On the other hand, T13 does not affect other

Table 3. Severity and sensitivity measures for transitions in fuel pump model

Transition	Sv(T)	Sn(T)
T1	0.43	0.00
T2	0.00	0.00
T3	0.64	0.00
T4	0.64	0.00
T5	0.29	0.21
T6	0.29	0.21
T7	0.21	0.14
T8	0.29	0.21
T9	0.21	0.29
T10	0.21	0.50
T11	0.07	0.50
T12	0.07	0.50
T13	0.00	0.79
T14	0.00	0.00

transitions in the model (Sv(T13) = 0), so its change is not expected to propagate to any other transitions (assuming that the change doesn't involve setting the value of a variable which was not previously defined at T13).

5 Exploratory Study

In this section we investigate the effectiveness of the two measures in identifying critical transitions in the fuel pump model. To do so, we write a tool which randomly generates a hundred arbitrary changes on the fuel pump model. Then for each transition, we check how many times the transition was touched by the 100 changes, and how many times it touched other transitions.

Finally, we track how many times each transition in the model was touched by the 100 changes. In addition, for each transition Ti, we track how many other transitions where touched by the change of Ti.

The results obtained after running the tool to apply 100 changes on the model are provided in Table 4. The first column of the table which is labeled as "Changed" shows how many times a change was applied on a particular transition. The second column "Touched", shows how many times a transition was touched by a change applied elsewhere in the model, and the third column labeled as "Touching" shows how many times a transition was touched by a change applied to the transition with interest. For example, for transition T1 we can see that it is changed 7 times out of the 100 changes applied to the model. For all of the 100 changes, it was never touched by a change applied to any of the other transitions in the model, while its change touched other transitions 42 times.

The results of the experiment show that the transition that was most frequently touched by a change is T13 which was touched by a change for 80 times. The transition

Table 4. Results of the exploratory study

Transition	Changed	Touched	Touching
T1	7	0	42
T2	6	0	0
T3	8	0	72
T4	9	0	81
T5	8	24	32
T6	2	24	8
T7	4	17	12
T8	6	23	24
T9	11	32	33
T10	6	46	18
T11	7	49	7
T12	12	46	12
T13	3	80	0
T14	11	0	0
SUM	100	341	341

whose change touched a large number of other transitions in the EFSM model was T4 which touched other transitions for 81 times.

The transitions that were not frequently touched by a change were T1, T2, T3, and T4. While the transitions that didn't touch other transitions in the model were: T2, T13, and T14.

These results were consistent with the severity and sensitivity measures estimated for each transition in the model. Indeed, the transitions that have high severity values, touched other transitions more frequently than transitions with lower severity values. For example, T1 (Sv = 0.43), T3 (Sv = 0.64), T4(Sv = 0.64) touched other transitions for 41, 72, and 81 times respectively. While transitions having the severity value of zero (namely T2, T13, and T14) didn't touch any other transition in the model.

Similarly, transitions that have high severity value were touched by a change more frequently than transitions with lower severity values. For example, T1, T2, T3, and T4 have a sensitivity value of zero, and during the experiment they were not touched by any change applied to other transitions in the model. While transitions T10 (Sn = 0.50), T11 (Sn = 0.50), T12(Sn = 0.50), and T13 (Sn = 0.79), were touched by a change for 46, 49, 46, and 80 times respectively.

6 Threats to Validity, Limitations, and Future Work

The major threat to validity for the presented study is the use of a single model (Fuel Pump Model) to test the effectiveness of the two measures. To handle this limitation, the study considered a large number of random changes to be applied to the model. Additionally, it is worth mentioning that the purpose of the current study is simply to illustrate the potential effectiveness of the two measures, while an extended study is planned in the future to cover a larger number of models with different sizes and different characteristics.

Another limitation of the proposed approach is the assumption that the probability of applying a change to any single transition in the model is the same for all transitions in the model. This assumption considers that all transitions in the model have approximately, comparable complexity. While this assumption can be true for some models, other probability metric should be considered for models that don't satisfy this assumption. For example, one can assume that a transition that has a complex condition composed of several sub-conditions joined with logical OR has higher probability to undergo a change compared to a transition that doesn't have any condition associated to it. Consequently, this assumption should be taken into consideration when the metrics are applied. Joining, the results obtained from the metrics with a human expert who can confirm the criticality of a transition given its complexity would generate more reliable conclusions.

7 Conclusion

In this paper we presented two model-based measures that can be very useful during the software maintenance. The severity of an EFSM transition estimates how severe a change applied to the transition can be. The scope of the severity of the change is measured in terms of the number of transitions to which the change may propagate. The propagation of the change is measured using data and control dependencies between transitions in the EFSM model. The sensitivity of an EFSM transition to a change applied to a model is also measured using model dependencies. However, when looking at the sensitivity, we investigate how often a change applied to other transitions in the model will propagate to the transition under consideration.

System development teams can use these two measures as a way to better estimate the severity of a change applied to the model, and to identify the transitions that will more frequently be affected by a change.

In future research, we will apply the measures to a larger set of models, and we will experiment with actual changes instead of random changes.

Acknowledgement. This research is sponsored by Kuwait Foundation for the Advancement of Science (KFAS) as part of project P116-18QA-01.

References

1. Korel, B., Singh, I., Tahat, L., Vaysburg, B.: Slicing of state -based models. In: IEEE International Conference on Software Maintenance, pp. 34–43 (2003)
2. Korel, B., Tahat, L.: Understanding modification in state-based system. In: Proceedings of the 12th IEEE International Conference on Program Comprehension (IWPC 2004), London, UK, pp. pp. 246–250, September 2004
3. Tahat, L., Almasri, N.: Identifying the effect of model modifications in State-Based models and systems. J. Adv. Comput. Sci. Technol. **2**(1), 9 (2013)

4. Cheng, K., Krishnakumar, A.: Automatic functional test generation using the extended finite state machine model. In: The 30th ACM/IEEE Design Automation Conference, pp. 86–91 (1993)
5. Vaysburg, B., Tahat, L., Korel, B., Bader, A.: Automating test case generation from SDL specifications. In: Proceedings of 18th International Conference on Testing Computer Software, pp. 130–139 (2001)
6. Korel, B., Tahat, L., Vaysburg, B.: Model based regression test reduction using dependence analysis. In: Proceedings of the International IEEE Conference on Software Maintenance, pp. 214–223 (2002)
7. Vaysburg, B., Tahat, L., Korel, B.: Dependence analysis in reduction of requirement based test suites. In: Proceedings of the ACM International Symposium on Software Testing and Analysis, pp. 107–111 (2002)
8. Korel, B., Tahat, L., Harman, M.: Test prioritization using system models. In: Proceedings of IEEE International Conference on Software Maintenance, Budapest, Hungary, pp. 559–568 (2005)
9. Tahat, L., Korel, B., Harman, M., Ural, H.: Regression test suite prioritization using system models. J. Softw. Testing Verification Reliab. **22**(17), 481–506 (2012)
10. Tahat, L., Korel, B., Hartman, M., Ural, H.: Regression test suite prioritization using system models. Softw. Test. Verification Reliab. J. (STRV) **27**(17), 481–506 (2012). Wiley Inter science, special edition on Model-Based Testing
11. Yoo, S., Harman, M.: Regression Testing Minimisation, Selection and Prioritisation - A Survey. Department of Computer Science, King's College London, Technical report: TR-09-09, October 2009
12. Thomas, S.W., Hemmati, H., Hassan, A.E., Blostein, D.: Static test case prioritization using topic models. Empirical Softw. Eng. **19**(1), 182–212 (2014)
13. Tahat, L., Korel, B., Koutsogiannakis, G., Almasri, N.: State-based models in regression test suite prioritization. Softw. Qual. J. **25**(3), 703–742 (2016)
14. Stamatis, D.H.: Failure mode and effect analysis: FMEA from theory to execution. ASQ Quality Press, Milwaukee (2003)
15. Li, B., Sun, X., Leung, H., Zhang, S.: A survey of code-based change impact analysis techniques. Softw. Test. Verification Reliab. **23**(8), 613–646 (2013)
16. Briand, L., Labiche, Y., O'Sullivan, L., Sowka, M.: Automated impact analysis of UML models. J. Syst. Softw. **79**, 339–352 (2006)
17. Almasri, N., Tahat, L., Korel, B.: Toward automatically quantifying the impact of a change in systems. Softw. Qual. J. **25**(3), 601–640 (2017)
18. Ferrante, K., Ottenstein, K., Warren, J.: The program dependence graph and its use in optimization. ACM Trans. Program. Lang. Syst. **9**(5), 319–349 (1987)
19. Bohner, S.A., Arnold, R.S.: Software Change Impact Analysis. IEEE Computer Soc. Press, Los Alamitos (1996)

An Analysis of a Methodology that Transforms the Entity-Relationship Model into a Conceptual Model for a Graph Database

Fernán Villa(✉), Francisco Moreno, and Jaime Guzmán

Departamento de Ciencias de la Computación y de la Decisión, Universidad Nacional de Colombia, Medellín, Colombia
{favillao,fjmoreno,jaguzman}@unal.edu.co

Abstract. The graph databases (GDB) have gained a lot of importance in the last years; this is due to the necessity to store and manage very large volumes of data whose natural structure is a graph. However, nowadays there do not exist conceptual models widely accepted to represent a GDB. This fact implies that the analysts are guided considering their experience and best practices. There have been proposed different conceptual models for GDB; in this paper, we analyze a methodology that generates a conceptual model for a GDB from the entity-relationship (E-R) model. We explore several limitations of this methodology and offer some ideas for solving them.

Keywords: Graph databases · Entity-relationship model · Conceptual models
Model transformation

1 Introduction

The basic element of a graph database (GDB) [1] is a graph. A graph is composed of nodes and edges, which show and set up the relationships between the nodes, e.g., the friendship between two users, the distance between two cities. A GDB is appropriate for managing network applications such as social networks, biological networks [2], transport networks, genealogical networks, and citation networks, among others.

In this paper, we analyze a methodology [3] that generates a conceptual model for a GDB from the entity-relationship (E-R) model. We present several limitations of this methodology and offer some ideas for solving them. The remainder of the paper is organized as follows: in Sect. 2, we present the property graph model. In Sect. 3, we explore the Model-Driven Design of Graph Databases methodology. In Sect. 4, we study the methodology limitations and offer some ideas to solve them. Finally, we present the conclusions and future work.

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 70–83, 2018.
https://doi.org/10.1007/978-3-319-95450-9_6

2 Property Graph Model

Today Neo4j [4] is a popular GDBMS (graph database management system). It was launched in 2007 and have been used by organizations such as NASA, Walmart, eBay, among others. This GDBMS uses a property graph model (PGM) to represent the domain of an application, it is a set of nodes related by directed edges. Nodes and relationships have properties (attributes). This model is used as well by other GDBMS such as TinkerPop [5] and Titan [6] and it is the base of GraphX (an API for managing graphs in Spark), among others.

The main elements for modeling are:

- **Nodes:** they are the basic model elements, represent the objects of interest for the application (entities in the real world), i.e., the objects that the analysts are interested to store information.
- **Relationship:** they represent the connections between two nodes. They must have a direction (a source node and a target node) and a type, which describes the nature of the relationship between the nodes, e.g., of friendship, possession, contract, among others. A relationship could have the same source and target node (recursive relationships). Between a couple of nodes, there could be several relationships (multigraph).
- **Properties:** They represent the attributes of nodes and relationships. In a node or in a relationship, a property is associated with a value (a property with its value is called key-value pair [7]). A node or relationship can have zero, one, or many properties.
- **Labels:** The labels allow the analyst to classify the nodes according to its role in the application. A node can have zero, one, or many labels, each with its corresponding name. In Fig. 1, we show three nodes with their labels and two relationships with their properties.

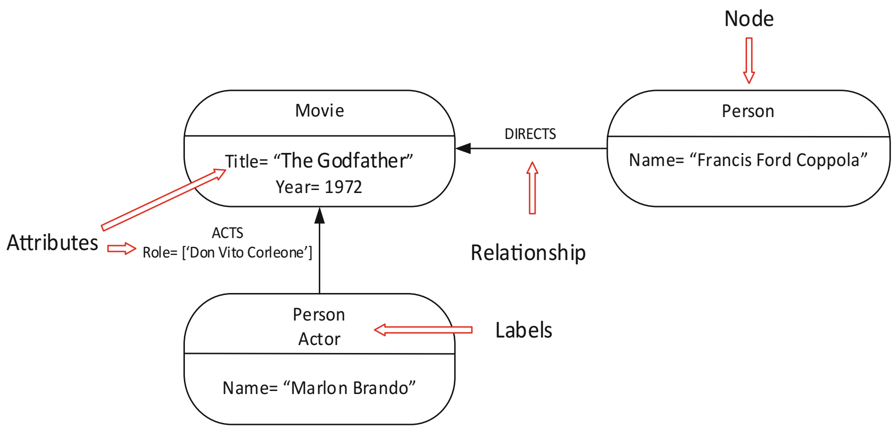


Fig. 1. Example of a PGM.

In addition to the PGM, several GDBMS are based in a hypergraph model. A hypergraph is a group of nodes and edges, but unlike a property graph it allows that a relationship connects more than two nodes or relationships. Thus, a hypergraph is a generalization of a property graph. Hypergraph DB [8] and Trinity [9] support hypergraphs.

3 Model-Driven Design of Graph Databases Methodology

In [3] it is proposed a methodology for modeling a GDB from the E-R model. Considering the relationships between the entities in the E-R model, it is obtained a model similar to the PGM.

To explain this methodology, we consider the E-R model example from Fig. 2. In this notation a rectangle represents an entity, a rhombus a relationship, a black circle a unique identifier, a white circle an attribute, and parentheses represent the cardinality (N stands for many).

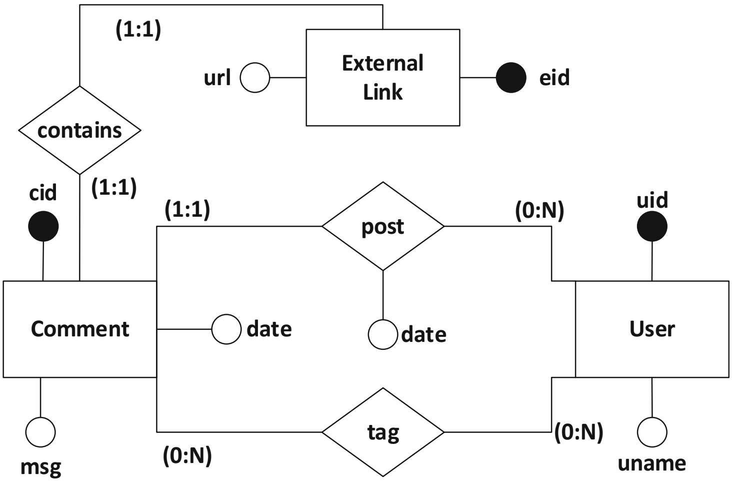


Fig. 2. E-R model.

The methodology has three steps:

3.1 Step 1. Apply Transformation Rules

The E-R model is transformed into an OE-R diagram (Oriented Entity-Relationship Diagram), i.e., a directed graph with labels and weights. The rules for transforming an E-R model into an OE-R diagram are:

(a) A one-to-one relationship is transformed into a bidirectional relationship. It is assigned a weight equal to zero, see Fig. 3a.

(b) A one-to-many relationship is transformed into a relationship that goes from the entity with cardinality one to the other entity. It is assigned a weight equal to one, see Fig. 3b.
(c) A many-to-many relationship is transformed into a bidirectional relationship. It is assigned a weight equal to two, see Fig. 3c.

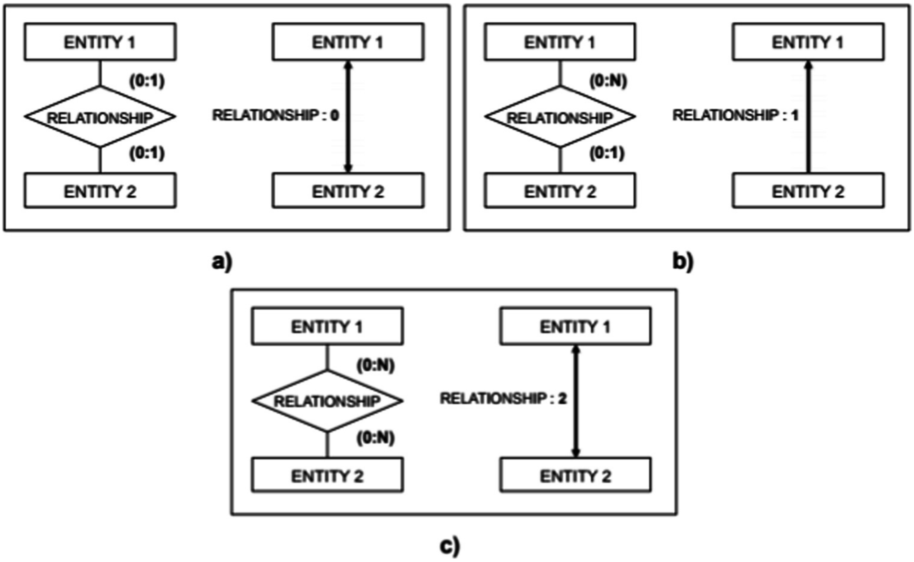


Fig. 3. Rules for transforming relationships: (a) One-to-one relationships, (b) One-to-many relationships, and (c) Many-to-many relationships. Source [3].

After applying these rules to the E-R model of Fig. 2, we obtain the O E-R diagram of Fig. 4

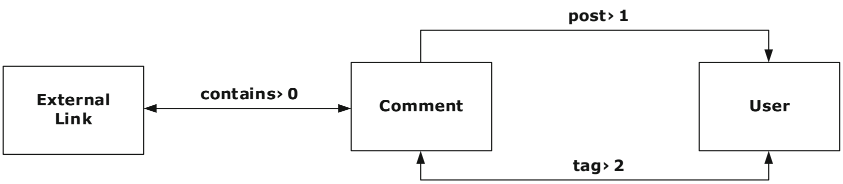


Fig. 4. O E-R diagram for the model of Fig. 2. Source [3].

3.2 Step 2. Merge Entities

This step is intended to merge entities whose instances use to appear together in the queries. To do this, the O E-R diagram is partitioned into groups of entities. To partition it, the authors define the functions W^+ and W^- for an entity n as follows:

$$W^{+}(n) = \sum\nolimits_{e \in out(n)} weight(e). \tag{1}$$

$$W^{-}(n) = \sum\nolimits_{e \in in(n)} weight(e). \tag{2}$$

Where *out*(*n*) is the set of the outgoing relationships of *n* and *in*(*n*) is the set of the incoming relationships of *n*. Thus, W^{+} and W^{-} calculate; respectively, the weights of the relationships that go out and come in of an entity *n*.

For instance, for the *Comment* entity of the O E-R diagram of Fig. 4 we obtain W^{+} (*Comment*) = 1 + 2 = 3 and W^{-} (*Comment*) = 0 + 2 = 2.

The partitions are formed in accordance to the following rules:

- Rule 1: An entity that is isolated, i.e., without relationships, forms a group by itself.
- Rule 2: If for an entity *n* is met that $W^{-}(n) > 1$ and $W^{+}(n) > 1$, then *n* will form a group (however, *n* could merge with some other entity, see Rule 3).
- Rule 3: If for an entity *n* is met that $W^{-}(n) \leq 1$ and $W^{+}(n) \leq 1$, then *n* is merged with other entity *m*, as long as between *m* and *n* there exist a relationship.

From the rules we conclude that:

(a) The merge of entities is done only when a node meets rule 3.
(b) The entities that participate in a many-to-many relationship do not merge, because for each of these entities W^{-} and W^{+} will be greater or equal than 2 (because the weight of this type of relationship is 2) and; therefore, it does not meet Rule 3.
(c) The rules merge the entities that participate in a one-to-one relationship. However, the merge of the entities that participate in a one-to-one requires a more detailed analysis. For a discussion, see [10].
(d) With regard to the one-to-many relationships, these are not necessarily merged; indeed, as we saw in Sect. 4, there are cases in which none of these rules are met and *the methodology does not explain what must be done in such cases*.

After applying ese rules to the O E-R diagram of Fig. 4 we obtain the partition of Fig. 5.

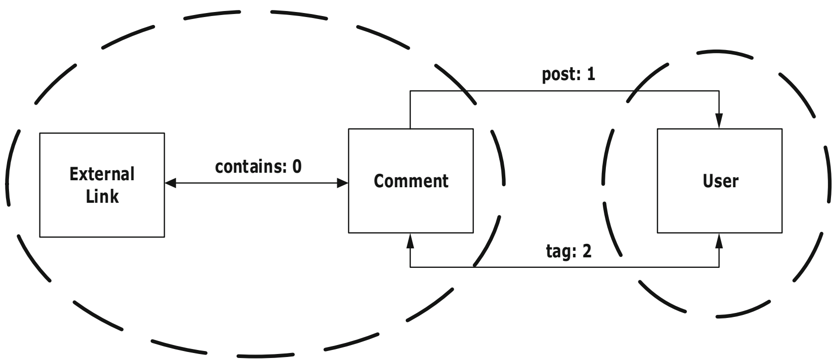


Fig. 5. Partitioned O E-R diagram of Fig. 4. Source [3].

3.3 Step 3. Conceptual Model

Finally, a conceptual model is defined for the GDB. This model is considered as a template for the creation of the instances of the GDB. The template includes all the attributes in this way: entityName.attributeName. See Fig. 6.

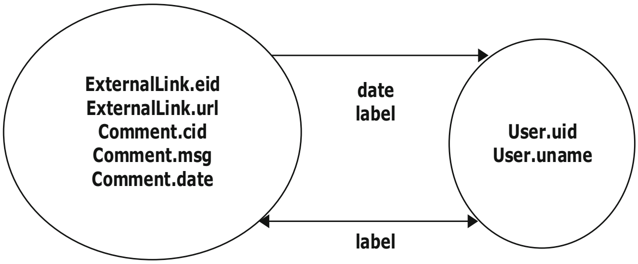


Fig. 6. Resultant template for the model of Fig. 5. Source [3].

Note that in the template appears the word label, which is an attribute that represents the name of the relationship. From the template of Fig. 6 it is possible to generate instances, as it is shown in Fig. 7.

4 Methodology Limitations

This methodology is a first step for the conceptual modelling of a GDB. The methodology generates a model with an abstraction level greater than other proposals, such as the PGM which models the GDB using instances. However, the methodology has some problems and disadvantages.

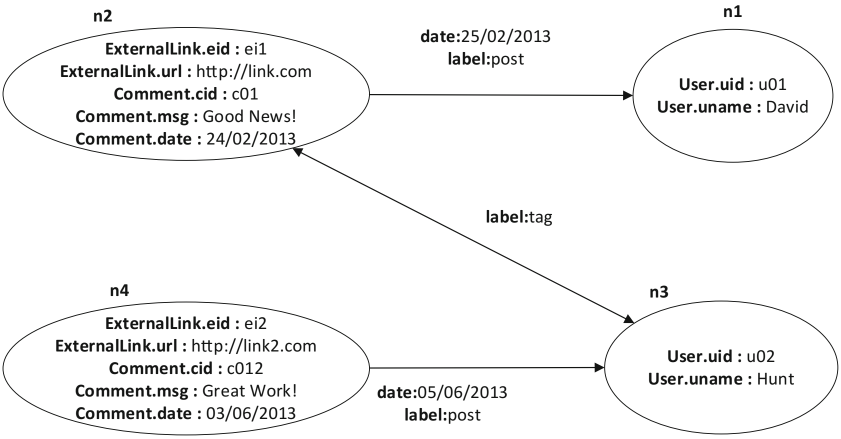


Fig. 7. Instances from the template of Fig. 6. Source [3].

4.1 Insufficient Rules

The proposed rules do not consider all the cases. In particular, it is not indicated how to proceed in these two cases:

- Case 1: if for an entity n is met that $W^{-}(n) > 1$ and $W^{+}(n) < 1$.
- Case 2: if for an entity n is met that $W^{-}(n) \leq 1$ and $W^{+}(n) > 1$.

For example, consider the E-R model of Fig. 8, where Entity1 = Employee, Entity2 = Company, Entity3 = Vehicle, Rel1 = Works for, Rel2 = Works for, Rel3 = Paints, and Rel4 = Drives. We show its corresponding O E-R diagram in Fig. 9.

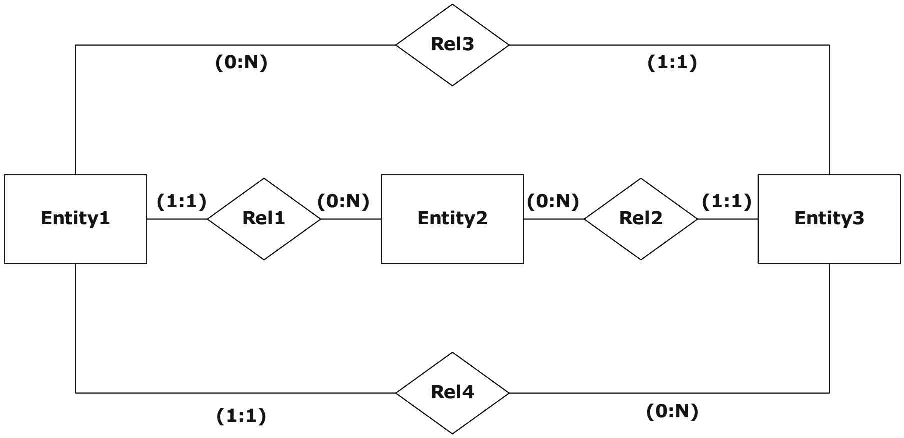


Fig. 8. E-R Model to exemplify insufficient rules.

In Table 1 we show the calculations of W^{+} and W^{-} for each entity.

Because none of the entities meets the rules of the methodology, it is not possible to obtain a GDB template. A possible solution could be: given that all the relationships are of one-to-many type, then we define a node with all the attributes of the participating entities (this seems to be the intention of the methodology as suggested by the previous examples). Another alternative is to define three nodes as follows: (1) to merge Company, Employee, and Vehicle, (2) to merge Employee and Vehicle (relationship Paints), and (3) to merge Employee and Vehicle (relationship Drives). However, the appropriate solution will depend largely on factors not considered by the methodology (e.g., analysis of the most frequent queries in the database), see also Sect. 4.3.

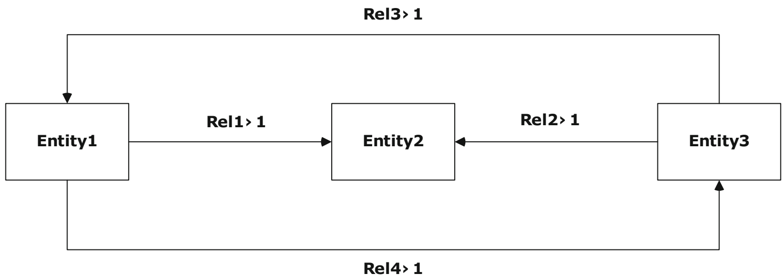


Fig. 9. O E-R diagram for the E-R model of Fig. 8.

Table 1. Calculation of W^+ and W^-.

Entity	W^-	W^+	Rule
Entity 1	1	1 + 1 = 2	None (Case 2)
Entity 2	1 + 1 = 2	0	None (Case 1)
Entity 3	1	1 + 1 = 2	None (Case 2)

4.2 Loss of Semantic Relationships

As entities merge, there is not information on relationships present among them in the original E-R model. This can lead to confusions or inconsistencies in the resultant template. Consider the E-R model of Fig. 10. We show its corresponding O E-R diagram in Fig. 11.

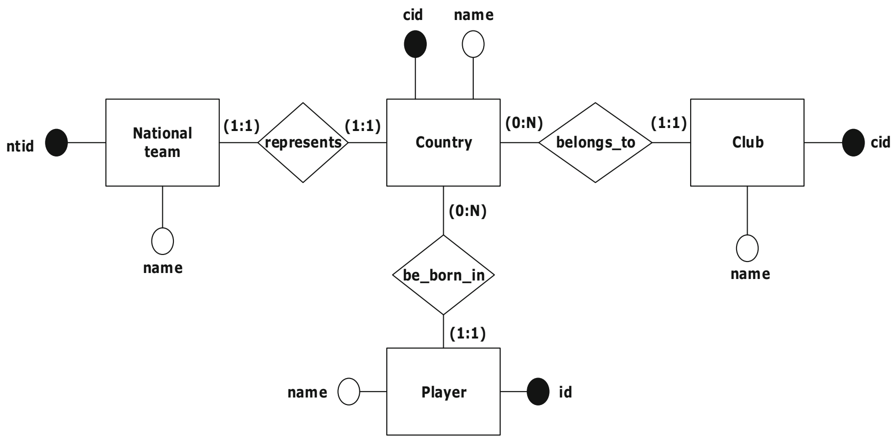


Fig. 10. E-R model with relationship between Country and Player.

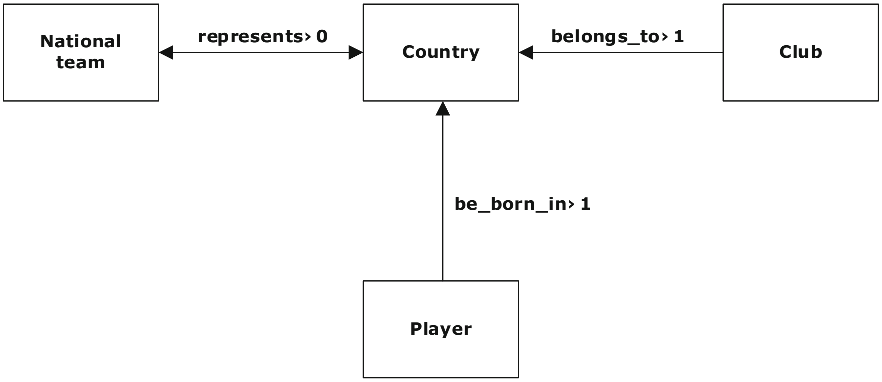


Fig. 11. O E-R diagram for the model of Fig. 10.

Table 2. Calculation of W^- and W^+.

Entity	W^-	W^+	Rule
National team	0	0	3
Country	1 + 1 = 2	0	None (Case 1)
Club	0	1	3
Player	0	1	3

In Table 2, we show the calculations of W^- and W^+ for each entity.

Although the Country entity does not meet any of the rules of the methodology, the other entities meet rule 3; therefore, they must be merged with an entity that they have at least a relationship. We show the resultant template in Fig. 12.

Fig. 12. Resultant template for the model of Fig. 11.

In generating instances with the resultant template some problems arise. For example, if we want to store the data of a player and his country of origin, it is not clear which values should be put in the attributes of the Club entity. The methodology does not indicate how to proceed in these cases.

Note how in the original E-R model, a player is related to a country and not with a club, the resultant template gives the impression that a player is also related with a club, which changes the semantics of the model. For example, in Fig. 13 we show an instance of the template of Fig. 12. The instance, gives the impression that a player belongs to the Boca Juniors club, something that does not correspond with the semantics of the original E-R model.

Fig. 13. Instance of template of Fig. 12.

In addition, if we apply the methodology to the models of Figs. 14 and 15, we obtain *the same template* to the model of Fig. 10.

That is, the methodology generates *the same template for three E-R models with different semantics*. In the resultant template it is not possible to determinate if the relationship is between a player and a club, or if it is between a player and a team, or if it is between a player and a country. The problem is that in the template there is not information about the relationships that existed between the entities in the original E-R model.

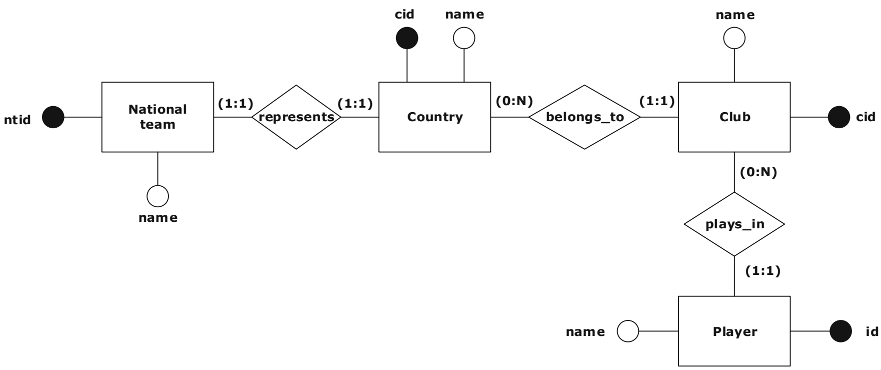


Fig. 14. E-R model with relationship between Club and Player.

Another example where there is a loss of semantic relationships is when between two entities, there is more than one relationship. For instance, consider the E-R model of Fig. 16, where we show a pair of entities with two relationships. After applying the methodology, we obtain the template shown in Fig. 17(a).

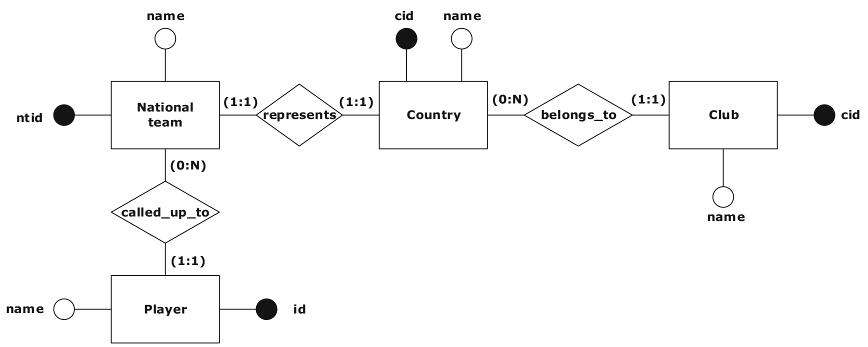


Fig. 15. E-R Model with relationship between National team and Player.

If we use the template of Fig. 16(a) for creating an instance, it is not possible to find if the relationship that exists between the person and the company is of type "Represents" or "Works for", because the template does not include information about the relationship that there was between the two entities. Yet, as in the E-R model a company is related to two kinds of persons (employees and representatives), we change the template as it is shown in Fig. 17(b). Note that in the new template, we include the name of the relationship in the attributes of the entity Person, to distinguish if the person is an employee or a representative. In addition, the relationship ¨works for¨ has an attribute "start_date" but after merging the entities this relationship disappears (it should be included in the resultant template). *These aspects are not considered in the methodology*.

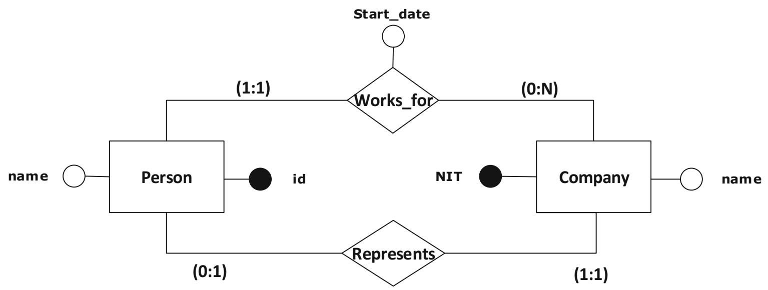


Fig. 16. E-R model with two relationships between two entities.

4.3 Other Limitations

(a) There is a lacking specification for the optionality of the attributes. The methodology does not offer tools that indicate which attributes are mandatory or optional, neither which are unique identifiers. For example, in the template of Fig. 17(a) it is not indicated that "Company.id" and "Person.id" correspond to the unique identifiers of their corresponding entities in the original model.

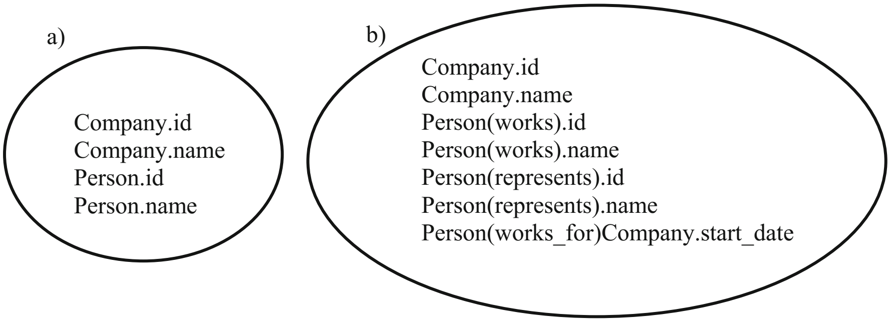


Fig. 17. Template for the model of Fig. 16: (a) generated by the methodology and (b) template proposal.

(b) There are no rules for generalization and exclusive relationships. In the methodology are not considered generalization (inheritance) nor exclusive relationships.

(c) Lack of analysis about the expediency of using a GDB [11]. The methodology is based on a database conceptual model, the E-R model and *mechanically* generates a template (model) for a GDB. However, it is not analyzed if a GDB is appropriate for an application. Although the decision for transforming an E-R model into a conceptual model for a GDB is, in a great part, a responsibility of the analyst team, the methodology could be enriched with elements (e.g., with the most

Table 3. Hints for improving the methodology.

Corresponding limitation	Solution or hint
4.1	Extend or change the current rules to consider all the cases
4.2	Include information about the relationships that disappear after merging entities, e.g., for each group of entities in the resultant template, we could specify in an annex (metadata) the relationships that existed between the corresponding entities in the original model. The template must also include the attributes of such relationship, e.g., the attributes could be named like this: relationshipName_attributeName (see example in Fig. 17)
4.3a	It must be included symbols to represent the mandatory attributes and those corresponding to unique identifiers. For example, it could be used a notation like the proposed in [13]
4.3b	Propose elements to represent in the resultant template generalizations and exclusive relationships. For example, it could be used a notation similar to the proposed in [13, 14]
4.3c	It must be considered the most frequent queries, the data volumes, and the database schema to decide if it is convenient to use a GDBMS
4.3d	It must be considered the most frequent queries to decide the convenience to merge some entities

frequent queries, see next item) which help to determine how convenient is to do the transformation or *transform only a part of the original model*.

(d) There is a lacking analysis on convenience to merge determined entities [12]. This aspect is related with the previous one. For instance, suppose the methodology merge the entities A, B, and C and that in the application the most frequent queries only require data from A and B but not from C. Considering this aspect, it is not convenient to merge C with A and B.

4.4 Some Recommendations

In Table 3 we show some hints to be developed in future works, which could help to improve the methodology.

5 Conclusions

In this paper, we analyzed the methodology "Model-Driven Design of Graph Databases" that transform an E-R model into a conceptual model for a GDB (Template Graph). This methodology presents several disadvantages and problems. We analyzed these aspects and offer some ideas for solving them. Perhaps the main problem is the omission of the relationships in the resulting template; this can lead to semantic confusions (with regard to the original model, i.e., the E-R model) when the template is generated, as we showed in Sect. 4.2.

In future works, in addition to those which can be derived from the identified problems in Table 3, it could be developed a similar methodology to model other non-relational types of databases [15]. Finally, as one of the referees suggested, concepts such as weak entity, strong entity, associative entity should be considered in the transformation from the E-R model to the GDB model. Constraints are also missing in the methodology.

Acknowledgments. This paper has been supported by the research group "SintelWeb, Sistemas Inteligentes Web".

References

1. Patil, S., Vaswani, G., Bhatia, A.: Graph databases - an overview. Int. J. Comput. Sci. Inf. Technol. **5**(1), 657–660 (2014)
2. Angles, R., Gutiérrez, C.: Survey of graph database models. ACM Comput. Surv. **40**(1), 1–39 (2008)
3. De Virgilio, R., Maccioni, A., Torlone, R.: Model-driven design of graph databases. In: Yu, E., Dobbie, G., Jarke, M., Purao, S. (eds.) ER 2014. LNCS, vol. 8824, pp. 172–185. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12206-9_14
4. Shimpi, D., Chaudhari, S.: An overview of graph databases. In: International Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS 2012), Mumbai, India (2012)

5. The Apache Software Foundation: Apache TinkerPop. http://tinkerpop.apache.org/docs/current/tutorials/getting-started/. Last accessed 05 May 2018
6. DataStax: Titan Documentation. http://s3.thinkaurelius.com/docs/titan/1.0.0/getting-started.html. Last accessed 05 May 2018
7. Neo Technology: What is a Graph Database and the Property Graph? https://neo4j.com/developer/graph-database/#property-graph. Last accessed 05 May 2018
8. Iordanov, B.: HyperGraphDB: a generalized graph database. In: Shen, H.T., et al. (eds.) WAIM 2010. LNCS, vol. 6185, pp. 25–36. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16720-1_3
9. Shao, B., Wang, H., Li, Y.: Trinity: a distributed graph engine on a memory cloud. In: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, pp. 505–516 (2013)
10. Bernik, J.P.: A translation of the one-to-one relationship for introductory relational database courses. Newsletter ACM SIGCSE Bull. **35**(4), 66–67 (2003)
11. Batra, S., Tyagi, C.: Comparative analysis of relational and graph databases. Int. J. Soft Comput. Eng. **2**(2), 509–512 (2012)
12. Pinto, Y.: A framework for systematic database denormalization. Global J. Comput. Sci. Technol. **9**(4), 44–52 (2009)
13. Barker, R.: CASE Method: Entity Relationship Modelling (Computer Aided Systems Engineering). Addison-Wesley Professional, Massachusetts (1990)
14. Batini, C., Ceri, S., Navathe, S.B.: Conceptual Database Design: An Entity-Relationship Approach. Addison-Wesley Professional, Massachusetts (1991)
15. Vera, H., Boayentura, W., Holanda, M., Guimaraes, V., Hondo, F.: Data modeling for NoSQL document-oriented databases. In: Proceedings of the 2nd Annual International Symposium on Information Management and Big Data - SIMBig, vol. 1478, pp. 129–135 (2015)

Networking, Communications Engineering and Vehicular Technology

The Effects of Ionospheric Irregularities on the Navigational Receivers and Its Mitigation

Arslan Ahmed[1,2](✉), Rajesh Tiwari[2], Sunny Imam Ali[3], and Ghulam Jaffer[4]

[1] Department of Electrical Engineering, Sukkur IBA University, Sukkur, Pakistan
arslan.ahmed90@gmail.com, arslan-ahmed@iba-suk.edu.pk
[2] Department of Electrical Engineering, Newcastle University, Newcastle upon Tyne, UK
[3] Department of Informatics, King's College London, London, UK
[4] Department of Space Sciences, Punjab University, Lahore, Pakistan
http://www.iba-suk.edu.pk/ibasuk/aboutiba/wpMain.aspx

Abstract. The performance of navigational receivers using satellite-based navigation technology can be severely affected by the presence of time-varying electron-density fluctuations in the ionosphere which can cause amplitude and phase perturbations at the receiver resulting in loss of phase lock at the carrier tracking loop due to cycle slip and hence unavailability of the navigation services. This paper studies the effects of amplitude and phase fluctuations at high latitudes due to the irregular ionosphere and, their effects on the receiver performance by using real time raw data from the Global Positioning Systems (GPS) satellites. The paper also suggest the use of adaptive software-based receiver model or modified hardware receivers to mitigate the effects of amplitude and phase fluctuations due to irregular ionosphere.

Keywords: GPS · Software receiver · Ionospheric scintillation
Tracking jitter · Phase locked loop

1 Introduction

The Navigational receivers are widely used by both the civilians and military for location-based services. Almost, all of these receivers uses satellite-based navigation technologies such as GPS, GLONASS, Galileo etc. The satellites used by these navigation systems are placed into the outer space having an altitude of more than 20,000 kms. The signals from these satellites have to pass through the ionosphere (a heavily ionized medium) which is a layer of the Earths atmosphere which may contain time-varying electron density irregularities as a result of a geomagnetic storm or increased solar activity and therefore can cause amplitude and phase fluctuations in the trans-ionospheric signals such as those received by the navigational receivers [1–3]. These amplitude and phase fluctuations due to the irregular behaviour of the electrons movement in the ionosphere

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 87–97, 2018.
https://doi.org/10.1007/978-3-319-95450-9_7

are known as the ionospheric scintillation [4,5]. The amplitude fluctuations are termed as amplitude scintillation whereas phase fluctuations are termed as phase scintillation.

Phase scintillation is usually observed at high latitudes (above 60° geomagnetic latitude) due to the auroral phenomena, can occur any time of the day lasting from few minutes to several hours and may result in loss of phase lock at the carrier tracking loop resulting in degrading the receiver performance [5–8]. The phase scintillation does not affect the signal-to-noise ratio of the signal [9]. The amplitude scintillation, on the other hand, is more dominant near the equatorial latitudes (±20° geomagnetic latitude) which occurs due to plasma instabilities in the F-layer of the ionosphere. At low latitudes, the GPS signal passing through the ionosphere faces scattering and may add destructively to produce deep power fades which may result in dropping the signal-to-noise (S/N_o) ratio of the signal below the receivers lock threshold [10,11] and the satellites may be considered absent even if there are a number of satellites present [9]. The amplitude scintillation can introduce fading of upto 20 dB at L-band frequencies [6].

In this paper, the effects of amplitude and phase scintillations on the navigational receiver performance has been studied during geomagnetic storm conditions. For this study, raw GPS data have been used by installing GPS receivers at various parts of Europe based on their latitudinal positions. This paper also suggest a tracking phase jitter based carrier tracking loop technique which can be used in software receivers or hardware ones to improve the performance of all types of navigational receivers relying on satellite technology during disturbed ionospheric conditions which leads to strong amplitude and phase fluctuations.

2 Measuring the Ionospheric Scintillation

Ionospheric scintillation refers to rapid random fluctuations in the amplitude and/or phase of the received trans-ionospheric signals [12]. These fluctuations occur due to the disturbance in the Earths magnetic field whenever there is a geomagnetic storm which occurs when the heated plasma (electrically charged atoms and molecules) from the sun also known as the solar wind strikes the Earths magnetic field creating instabilities in the ionosphere plasma [6,13]. In order to measure the ionospheric scintillation due to disturbed ionosphere, NovAtel 4004B GPStation6 dual frequency receivers have been installed at different latitudinal positions around Europe as shown in Fig. 1.

The GPS transmit signals using several frequency bands such as $L1$ (1575.42 MHz), $L2$ (1227.6 MHz), $L3$ (1381.05 MHz), $L4$ (1379.913 MHz) and $L5$ (1176.45 MHz). At a particular time, there are usually 24 active satellites in the constellation of GPS. The $L1$ and the $L2$ frequency bands are widely used by both the civilians and military for navigation purposes. The $L1$ signal uses the Coarse acquisition (C/A) and the $P(Y)$ codes whereas the $L2$ signal uses only the $P(Y)$ code for GPS signal transmission [6].

The C/A code is a 1 ms long pseudorandom code (PRN) sequence having a chipping rate of 1.023 MHz and each satellite in the GPS has a unique C/A code.

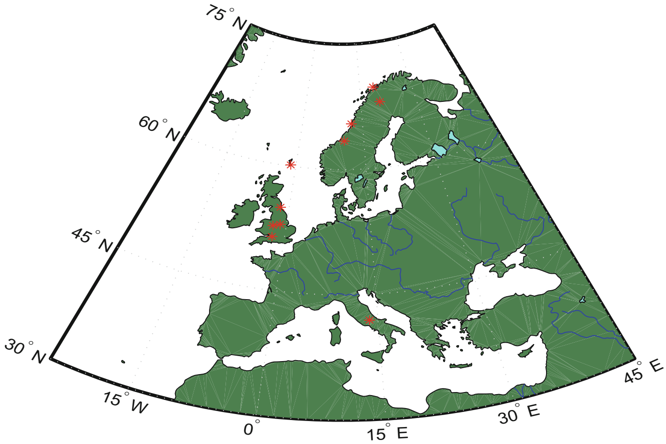


Fig. 1. GPS receiver stations installed at different latitudes around Europe for recording the scintillation activity [6].

The C/A code is available free of cost to all users. The $P(Y)$ code, on the other hand, is a 266 days long PRN code with a chipping rate of 10.23 MHz. This code is available to military only. The GPS signal on the $L1$ and $L2$ frequencies also contain the navigation data at 50 Hz which contains the information about the satellites orbit, time, position and the path that satellites follow when orbiting the Earth. This information is used at the receiver for position estimation.

The mathematical model of the GPS $L1$ and $L2$ signals [6] can be given as

$$S_{L1}(t) = A_c c(t) d(t) \cos(\omega t + \phi) + A_p P(t) d(t) \cos(\omega t + \phi) \tag{1}$$

$$S_{L2}(t) = A_{L2} P(t) d(t) \cos(\omega t + \phi) \tag{2}$$

where S_{L1} and S_{L2} are the $L1$ and $L2$ signals, A_c and A_p are the C/A and $P(Y)$ signal amplitudes at $L1$ and $L2$ frequencies respectively. A_{L2} is the $L2$ signal amplitude, $d(t)$ is the navigation data at 50 Hz, ω is the carrier frequency, $c(t)$ is the C/A code and $P(t)$ is the $P(Y)$ code. In case, the received signal is affected by the amplitude and phase fluctuations after passing through the ionospheric irregularities [6], (1) and (2) can be re-written as

$$\begin{aligned} S_{L1}(t) = A_c\ \delta A_c\ c(t) d(t) \cos(\omega t + \phi + \delta\phi) + \\ A_p\ \delta A_p P(t) d(t) \cos(\omega t + \phi + \delta\phi) \end{aligned} \tag{3}$$

$$S_{L2}(t) = A_{L2}\ \delta A_{L2}\ P(t) d(t) \cos(\omega t + \phi + \delta\phi) \tag{4}$$

where $\delta A_c, \delta A_p$ and δA_{L2} represents the fading in the amplitudes of the signals at the $L1$ and $L2$ frequencies respectively and $\delta\phi$ are the phase fluctuations. The phase scintillation is normally denoted by σ_ϕ index and is the square root of the standard deviation of the $\delta\phi$ over a certain time period usually taken as 60 s [6]. The amplitude scintillation, on the other hand, is the normalized standard deviation of the signal intensity ($\delta I = \delta A^2$) over a 60 s interval [6] given as

$$S_{4T} = \frac{\sqrt{E[\delta I^2] - (E[\delta I])^2}}{E(\delta I)} \tag{5}$$

where $E[]$ is the mean value. In the presence of ambient noise, the final equation for the amplitude scintillation index denoted by S_4 can be given as

$$S_4 = \sqrt{\frac{E[\delta I^2] - (E[\delta I])^2}{[E(\delta I)]^2} - \frac{100}{S/N_o}\left[1 + \frac{500}{19S/N_o}\right]} \tag{6}$$

where S/N_o is the signal-to-noise ratio. The signal intensity δI can be found as

$$\delta I = \frac{(NBP - WBP)}{(NBP - WBP)_{LPF}} \tag{7}$$

where NBP and WBP are the low pass filtered (LPF) narrowband and wideband powers respectively and can be given as

$$NBP = \left(\sum_{k=1}^{N} i_k\right)^2 + \left(\sum_{k=1}^{N} q_k\right)^2 \tag{8}$$

$$WBP = \sum_{k=1}^{N} \left(i_k^2 \; + \; q_k^2\right) \tag{9}$$

where i and q are the in-phase and quadrature components of the received signal generally summed over a 20 ms interval, i.e., $N = 20$, to find the NBP and WBP. Using (9), the carrier-to-noise ratio, C/N_o, of the received signal can be found as

$$C/N_o = 10 \log \left[\left(\frac{WBP}{N} - 1\right) \times 50\right] \tag{10}$$

3 Scintillation Effects on the Receiver Performance

In order to record the scintillation activity, the experimental setup used is shown below in Fig. 2. This is one of the receiver station out of several receiver stations around Europe which is setup in the Department of Electrical Engineering, Newcastle University, Newcastle Upon Tyne, UK. The similar setup has been used at other stations as well.

To determine the ionosphere effects on the receiver performance due to fluctuations in the amplitude and/or phase of the received signal, we selected one of our high latitude receiver stations installed at Trondheim, Norway (63.42° N, 10.4° E). The scintillation activity at Trondheim is shown in Fig. 3 on 24 April, 2012. Due to being located at high latitudes, there is a always high possibility of scintillation occurrence whenever there is a solar storm or a geomagnetic

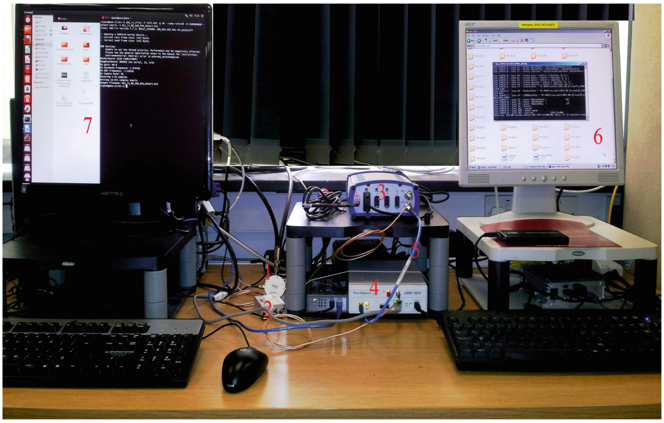


Fig. 2. Experimental setup for recording the scintillation activity using the NovAtel GSV4004B GPS Receiver. (1) Amplifier connected to the roof mounted GPS antenna; (2) splitter to split the signal between Novatel receiver and USRP2 N210 for raw data recording; (3) Novatel GPS receiver; (4) Universal Software Radio Peripheral 2 (USRP2) N210 front end device; (5) oscillator output from (3) to (4); (6) Scintillation data recording using the Novatel receiver; (7) GPS Raw data recording using USRP2 N210 device for signal acquisition and tracking manually by using a software receiver [12].

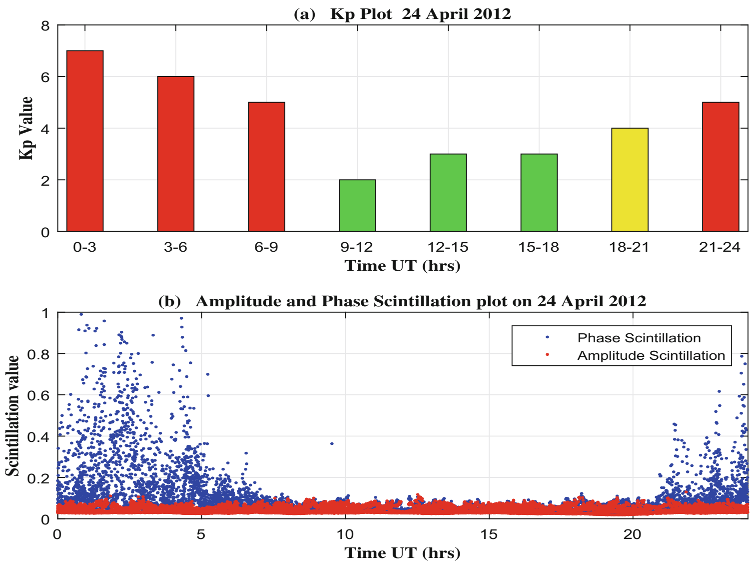


Fig. 3. Geomagnetic field activity and Scintillation observed at Trondheim, Norway on 24 April, 2012 (Color figure online)

storm. Figure 3(a) shows the geomagnetic field activity on 24 April, 2012 by using a planetary index Kp. This index is used to represent the disturbance in the Earth's magnetic field on a scale of 0 to 9 where Kp values of less than 4 means that there is no storm and no significant scintillation activity will occur, Kp value of 4 means that there might be a chance of amplitude and/or phase scintillation occurrence and Kp values of 5 or greater than 5 means that there is a high possibility of scintillation occurrence. The Kp index updates every 3 h with an estimate of the past 3 h values.

The red bars in Fig. 3(a) shows that there is a strong geomagnetic storm between 00:00 to 09:00 universal time (UT) and from 21:00 to 24:00 UT as the Kp index is either 5 or greater than 5. Between 09:00 to 21:00 UT, there was no storm while from 18:00 to 21:00 UT there was only a geomagnetic disturbance. Figure 3(b) shows the amplitude and phase scintillation activity for all the satellites on 24 April, 2012 that were locked by the GPS receiver. Strong phase scintillation was observed on all the satellites between 00:00 to 06:00 and from 21:00 to 24:00 UT as can be seen by the blue dots in Fig. 3(b). However, no significant amplitude scintillation was observed between these hours due to the station being located at high latitudes as explained earlier.

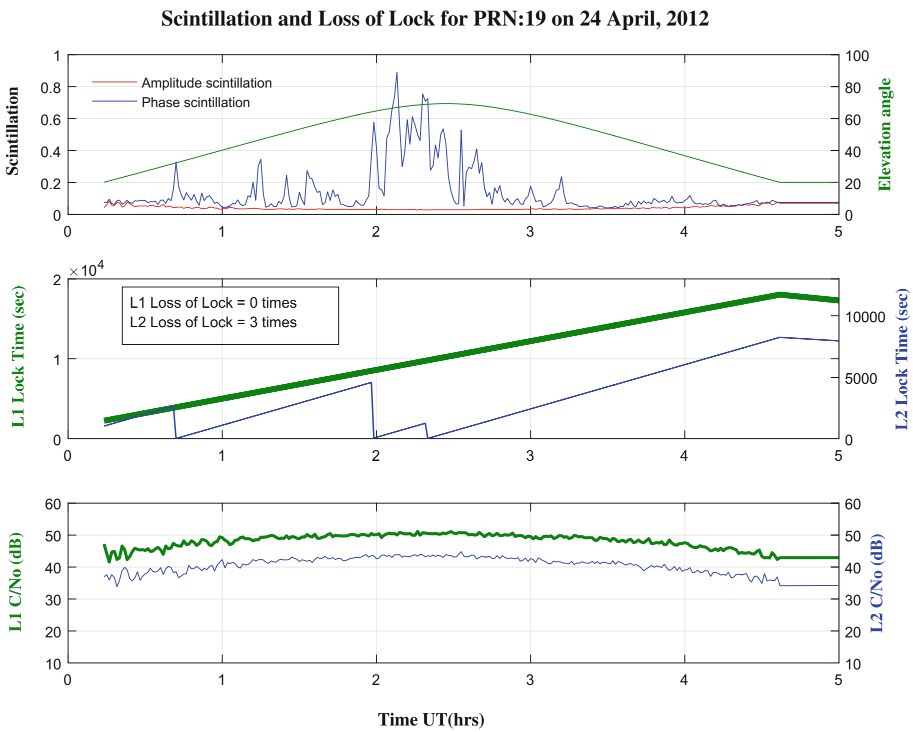


Fig. 4. Scintillation activity and the loss of lock occurrence for PRN 19 on 24 April, 2012.

Figure 4 shows the results of one of the satellites, i.e., PRN 19 that was locked by the Trondheim GPS receiver between 00:00 to 05:00 UT on 24 April, 2012 during the geomagnetic storm time. The top graph in Fig. 4 represents the amplitude and phase scintillation on PRN 19 along with the elevation angle. It should be noted that the satellite is considered to be locked when the elevation angle is greater than 20° in order to avoid spurious values which occured due to tall buildings or obstacles and does not contribute to amplitude and phase fluctuations introduced by the ionosphere. The middle graph in Fig. 4 shows the loss of tracking loop lock at the $L1$ and $L2$ frequencies whereas the bottom graph in Fig. 4 shows the C/N_o for the $L1$ and $L2$ signals. There are 32 satellites in the GPS system out of which only 24 are used for navigation. These 32 satellites are usually represented by $PRN1, PRN2, PRN3$ upto $PRN32$ where PRN stands for pseudo random number which is unique for each satellite in the GPS.

It can be seen in Fig. 4 that due to the strong phase scintillation, the $L2$ signal frequently lost lock whereas the $L1$ signal stayed in contact and provided the navigation services as usual. The reason for the $L2$ signal frequent loss of lock compared to the $L1$ signal is due to the critical frequency of the ionosphere. Frequencies close to the critical frequency are more affected by the ionosphere disturbance compared to the higher frequencies. The phase scintillation does not affects the signal-to-noise ratio as can be seen in the bottom graph in Fig. 4 which

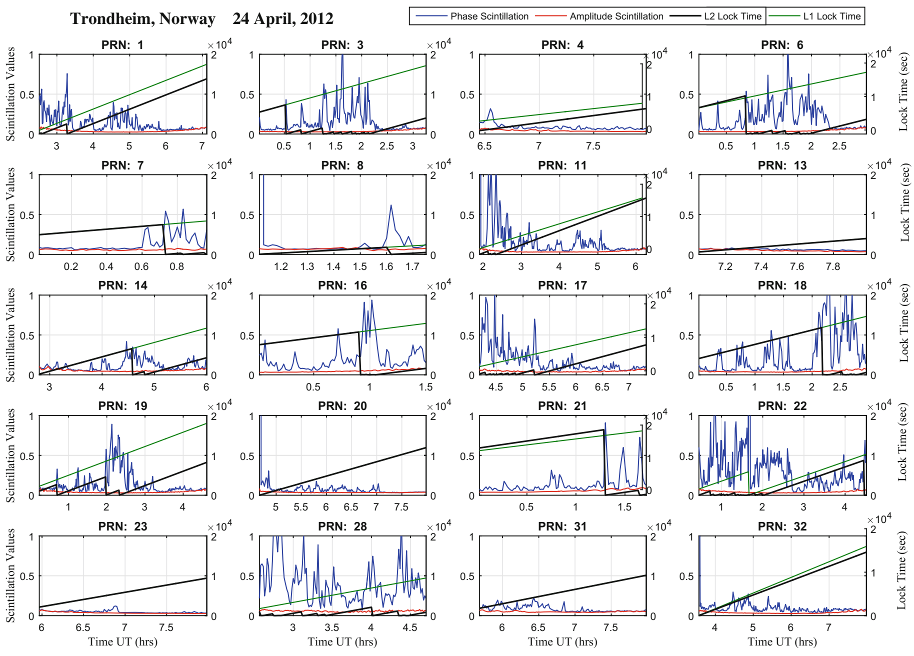


Fig. 5. Loss of signal lock by the tracking loop of the GPS receiver at the $L1$ and $L2$ frequencies for the satellites that were present between 00:00 to 09:00 UT.

is the C/N_o graph of the signal. The frequent loss of lock at the $L2$ frequency means unavailability of the navigation service during the time when the tracking loop is in reacquisition state after loss of lock. It is to mention that the $L2$ signal is used by the military for surgical and war related activities and for carrying out space based operations. Unavailability of the navigation services at the $L2$ frequency can affect the strategic activities of a nation and can lead to serious problems for the military and for carrying out space operations.

Figure 5 shows the loss of lock at the $L1$ and the $L2$ signals on all the satellites that were present between 00:00 to 09:00 UT. Some of the satellites were present only for a short period of time while others for a longer period of time. Strong phase scintillation was observed on some of the satellites such as PRN 3, 11, 16, 17, 18, 19, 22 and 28 whereas the other PRN's faced weak to moderate scintillation. It should be noted that the scintillation may not occur at all the satellites with equal intensity because only those satellites are disturbed which passes through the ionospheric irregularities which is a random occuring phenomenon due to the random movement of free electrons in a grouped form. Table 1

Table 1. Loss of Lock at the $L1$ and $L2$ frequency signals on the satellites that were present between 00:00 to 09:00 UT during the geomagnetic storm.

PRN	# of Loss of Lock at L1 frequency	# of Loss of Lock at L2 frequency
1	0	1
3	0	10
4	0	0
6	0	9
7	0	2
8	0	1
11	0	4
13	0	0
14	0	2
16	0	2
17	0	9
18	0	10
19	0	3
20	0	0
21	0	2
22	1	14
23	0	0
28	0	10
31	0	0
32	0	0

highlights the total number of loss of lock for each of the satellite in Fig. 5. It can be observed in Table 1 that the GPS receiver continuously lost lock to almost all the satellites at the $L2$ frequency but only PRN22 lost lock at the $L1$ frequency. The reason for the $L2$ signal being more susceptible to the ionospheric scintillation is due to its critical frequency close to the critical frequency of the ionospheric layer. The amplitude scintillation along with phase scintillation can cause more harm and has the ability to affect all kind of satellite communications as it not only affects the tracking loop but also introduces signal fading [14].

4 Scintillation Mitigation

The amplitude scintillation is more dominant at low latitudes and since most of our receivers are installed at high latitudes so, this paper will focus on mitigating the effects of scintillation for high latitude regions only. The scintillation effects on the navigational receivers can be mitigated by using the tracking phase jitter approach as suggested by [6, 12]. The phase jitter is the standard deviations of the phase fluctuations in the incoming signal. This approach updates the tracking loop parameters during the runtime when the signal tracking is in process by increasing or decreasing the noise bandwidth of the phase locked loop (PLL). [6, 12] methods estimates the tracking phase jitter using the formula given in (11). Once the tracking phase jitter is estimated, it can then be used to update the tracking loop parameters of a receiver either by using the software receiver or by using hardware modifications in a receiver [15] which could be able to update the tracking parameters during run time.

$$\sigma^2_{\phi e} = \frac{\pi T}{k f_n^{p-1} \sin\left(\frac{(2k+1-p)\pi}{2k}\right)} + \frac{\bar{B_n}}{C/N_o} \times \left[\frac{1}{1-S_4^2} + \frac{1}{2T_I C/N_o(1-3S_4^2+2S_4^4)}\right] + \sigma^2_{\phi osc} \quad (11)$$

where p is the phase power spectral density (PSD), T is the spectral strength of the phase PSD at 1 Hz, B_n is the noise bandwidth, f_n is the tracking loop natural frequency, k is the loop filter order and $\sigma_{\phi osc}$ is the phase variance due to oscillator noise. The advantage of using [6, 12] method is that the spectral parameters can be estimated using the amplitude and phase scintillation which was not possible by using these two parameters.

5 Conclusion

This paper has discussed the effects of the ionospheric irregularities also known as ionospheric scintillation on the navigational receiver performance using the GPS data from the high latitude regions. It is observed that the navigational receivers at high latitudes during the disturbed ionosphere are mostly affected by the phase scintillation which introduces rapid fluctuations in the phase of the received signal resulting in loss of lock at the tracking loop of the receiver due to cycle slip. During a geomagnetic storm, the irregular ionosphere was acting as the main cause of non-functionality of the receiver. It has been suggested in this paper that the effects of scintillation particularly phase scintillation can be mitigated by estimating the tracking phase jitter of the received signal which is not only simple in terms of implementation but can be used both in hardware or software receivers or a combination of both.

Acknowledgement. The authors wish to acknowledge; the EPSRC (Engineering and Physical Science Research Council) for financial support for some of this work through Grant EP/H004637/1. The first author also thanks Sukkur IBA University and HEC, Pakistan for providing the funding for this research work. The authors are also grateful to the Geospatial Institute at Nottingham University, UK for use of archived GPS data collected under linked EPSRC grants for scintillation modeling and mitigation between Newcastle, Nottingham and Bath Universities.

References

1. Jayachandran, P., Hamza, A., Hosokawa, K., Mezaoui, H., Shiokawa, K.: GPS amplitude and phase scintillation associated with polar cap auroral forms. J. Atmos. Solar Terr. Phys. **164**, 185–191 (2017). https://doi.org/10.1016/j.jastp.2017.08.030. http://www.sciencedirect.com/science/article/pii/S1364682617302341
2. Marques, H.A., Marques, H.A.S., Aquino, M., Veettil, S.V., Monico, J.F.G.: Accuracy assessment of precise point positioning with multi-constellation GNSS data under ionospheric scintillation effects. J. Space Weather Space Clim. **8**, A15 (2018). https://doi.org/10.1051/swsc/2017043
3. Paul, A., Paul, K.S., Das, A.: Impact of multiconstellation satellite signal reception on performance of satellitebased navigation under adverse ionospheric conditions. Radio Sci. **52**(3), 416–427 (2017). https://doi.org/10.1002/2016RS006076
4. Ahmed, A., Tiwari, R., Strangeways, H.J., Dlay, S., Johnsen, M.G.: Waveletbased analogous phase scintillation index for high latitudes. Space Weather **13**(8), 503–520 (2015). https://doi.org/10.1002/2015SW001183
5. Tiwari, R., Strangeways, H., Tiwari, S., Ahmed, A.: Investigation of ionospheric irregularities and scintillation using TEC at high latitude. Adv. Space Res. **52**(6), 1111–1124 (2013). https://doi.org/10.1016/j.asr.2013.06.010. http://www.sciencedirect.com/science/article/pii/S0273117713003517
6. Ahmed, A.: Scintillation on global navigation satellite signals and its mitigation. Ph.D. thesis, Newcastle University, UK (2015)
7. Strangeways, H.J.: Determining scintillation effects on GPS receivers. Radio Sci. **44**(1) (2009). DOIurl10.1029/2008RS004076
8. Datta-Barua, S., Doherty, P., Delay, S., Dehel, T., Klobuchar, J.: Ionospheric scintillation effects on single and dual frequency GPS positioning. In: Proceedings of the 16th International Technical Meeting of the Satellite Division of The Institute of Navigation, Portland, OR (2003)
9. Kintner, P.M., Ledvina, B.M., de Paula, E.R.: GPS and ionospheric scintillations. Space Weather **5**(9) (2007). DOIurl10.1029/2006SW000260

10. Pi, X., Iijima, B.A., Lu, W.: Effects of ionospheric scintillation on GNSS based positioning. Navigation **64**(1), 3–22 (2017). https://doi.org/10.1002/navi.182
11. Goswami, S., Paul, K.S., Paul, A.: Assessment of GPS multifrequency signal characteristics during periods of ionospheric scintillations from an anomaly crest location. Radio Sci. **52**(9), 1214–1222 (2017). https://doi.org/10.1002/2017RS006295
12. Ahmed, A., Tiwari, R., Shah, M.A., Yin, J.: GPS receiver rhase jitter during ionospheric scintillation. In: 2016 7th International Conference on Mechanical and Aerospace Engineering (ICMAE), pp. 605–608 (2016). https://doi.org/10.1109/ICMAE.2016.7549611
13. Coker, C., Bust, G.S., Doe, R.A., Gaussiran, T.L.: Highlatitude plasma structure and scintillation. Radio Sci. **39**(1) (2004). https://doi.org/10.1029/2002RS002833
14. Humphreys, T.E., Psiaki, M.L., Kintner, P.M., Ledvina, B.M.: GPS carrier tracking loop performance in the presence of ionospheric scintillations. In: Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation, Long Beach Convention Center, Long Beach, CA, pp. 156–167 (2005)
15. Ganguly, S., Jovancevic, A., Brown, A., Kirchner, M., Zigic, S., Beach, T., Groves, K.M.: Ionospheric scintillation monitoring and mitigation using a software GPS receiver. Radio Sci. **39**(1) (2004). https://doi.org/10.1029/2002RS002812, rS1S21

Automatic and Secure Wi-Fi Connection Mechanisms for IoT End-Devices and Gateways

Fu-Chiung Cheng(✉)

Tatung University, Taipei, Taiwan 104, Republic of China
fccheng@ttu.edu.tw

Abstract. Internet of Things (IoT) are developed rapidly in recent years and more than 50 billion of IoT devices are expected to be deployed worldwide in 2020. How to automatically and securely connect the tremendous number of IoT end devices to Internet is one of critical problems to be addressed. This paper proposes secure and automatic Wi-Fi (Wireless Fidelity) connection mechanisms for connecting IoT end devices and IoT gateways. Our design has the following advantages. First, IoT end devices, once powered on, can automatically connect to an IoT gateway without human intervention. Secondly, the SSID and password for high security strength WPA2 connection are randomly generated to enhance IoT security. Finally, the randomly generated password and SSID are automatically changed every day or when network attacking is detected.

Keywords: Automatic connection · IoT security
Wi-Fi wireless communication · IoT applications

1 Introduction

According to Gartner's forecast report [1] on the Internet of Things, the number of connected IoT devices in global use in 2017 will reach 8.4 billion, which is a 31% increase from 2016, and it will increase to 20.4 billion by 2020. The amount of expenditures related to service and endpoints will also reach $2 trillion in 2017. Internet of Things has become the most important research area in the industry as well as in academia. In addition, a newer Gartner's report [2] estimates worldwide spending on IoT security will reach $1.5 billion in 2018, a 28 percent increase from 2017. A more widespread and optimistic report from Cisco [3] estimates that the number of connected devices on the Internet will exceed 50 billion by 2020. It is no double that IoT has huge market values in industries and becomes very important research area in academia.

The current Wi-Fi technology provides high secure wireless communication mechanism such as WPA3 and WPA2 [4], but it may be inconvenience for people since setting passwords and SSID [12] is needed to connect to Access Points (AP). For machine to machine IoT applications, it is formidable and undesirable to set passwords by hand due to the large number of IoT clients. This paper proposes original algorithms

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 98–106, 2018.
https://doi.org/10.1007/978-3-319-95450-9_8

for automatically and securely connecting IoT clients to IoT gateways (APs) based on widely-used Wi-Fi technology. The contributions are as follows. First, Wi-Fi IoT clients can automatically connect to an IoT gateway without human intervention. Secondly, for better IoT security, the SSID and password for Wi-Fi connection are randomly generated. Finally, the randomly generated passwords and SSIDs are automatically changed every day or when cracking or attacking is detected.

2 Background Knowledge

This section introduces existing Wi-Fi connection technologies and Wi-Fi security and presents the problems and difficulties in the development of the IoT system platform as a reference for the theoretical basis and system design.

2.1 Wi-Fi Network

Wi-Fi network [5, 12] refers to the two basic service combinations defined in the IEEE 802.11 standard:

- Basic Service Set (BSS): BBS is mainly responsible for all message transmissions in the AP (Access Point) and wireless clients such as laptops, smartphones and IoT end devices in a local area. In IEEE 802.11 standards, IEEE 802.11b, IEEE 802.11a and IEEE 802.11 g are in the 2.4 GHz band while IEEE 802.11 h is in the 5 GHz band.
- Independent BSS (IBSS): IBSS supports peer to pear ad hoc network. Wireless clients directly communicate with each other without assistance of APs.

It is obvious that BSS-enabled Wi-Fi network is suitable for IoT applications. However, to connect to an AP, a Wi-Fi client needs to know the SSID and the password of the AP. It is formidable to connect billions of IoT devices to APs (gateways) by hand.

2.2 Wi-Fi Protected Access

Wi-Fi Protected Access [4], consisting of WPA, WPA2 and WAP3, is the security protocol developed by the Wi-Fi Alliance to secure wireless computer network. WPA and WPA2 are not secure and need to be improved [8]. A new protocol, WPA3, is released in 2018 to address the weak password problem with a 4-way handshake [6], as shown in Fig. 1 and 192-bit encryption.

2.3 Wi-Fi Security Attacks

Once a Wi-Fi network connection is established, it is vulnerable to attack. There are two common kinds of attacks:

- DoS (Denial of service) attacks: In DoS attacks, hackers (the attackers) try to make the system unavailable to normal users by flooding the targeted system with false requests. For example, there are three users and one attacker labeled as hacker in

Fig. 2. The hacker first monitors authentication (Auth) request packets. Once found, he can sends a great deal number of false Auth packets to disrupt network services, leading to no services available among the connected users and no new connection for the unconnected ones.

- Password cracking attacks: In password cracking attacks, hackers try to recover passwords of APs by repeatedly guess the passwords to gain unauthorized access. For example, the hacker in Fig. 2 may obtain the EAP and EAPOL packets in WPA2 and then uses a dictionary attack method to crack the passwords. It works for easy guessed meaningful password or simple password [7]. Currently the security of WEP, WPA and WPA2 are all problematic in Wi-Fi communication.

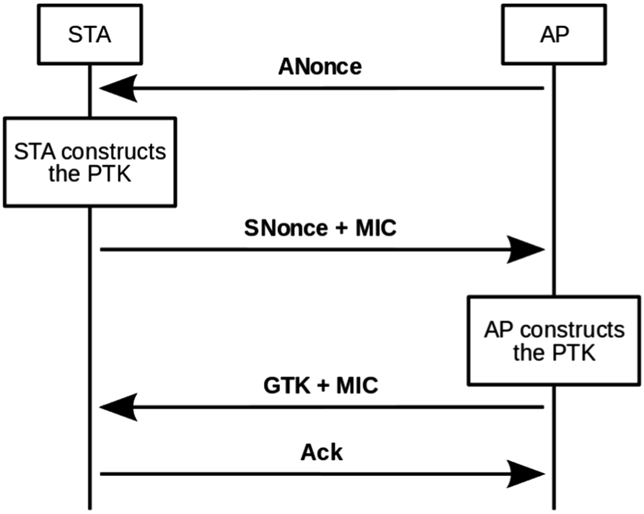


Fig. 1. 4-way handshake

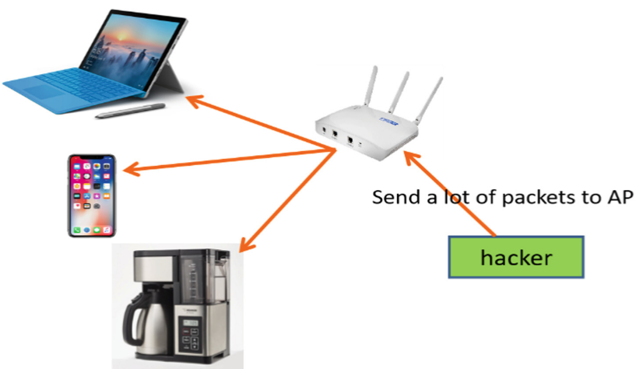


Fig. 2. Denial of service attacks

In [11], Zhang et al. proposed an intrusion prevention method for Wi-Fi clients in DoS attack. A Wi-Fi client is able to differentiate between legitimate and forged frames by using Medium Access Control filtering.

2.4 Wi-Fi Protected Setup

Wi-Fi Protected Setup (WPS) [9] is designed to ease the setup of secure Wi-Fi networks in home and small office environments by Wi-Fi Alliance. There are three supported methods:

- PIN method (PIN-WPS): Users read the PIN of the AP and enter the PIN number to the client device to connect to the AP.
- Push button method (PBC-WPS): Users push the setup buttons of the client device and the AP simultaneously to establish a Wi-Fi network.
- Near-field communication method (NFC-WPS): When a NFC-enabled client is closed to a NFC-enabled AP, PIN code is received and then a secure Wi-Fi network is established.

Neither PIN-WPS nor PBC-WPS are secure [9, 10]. PIN-WPS can be easily cracked in [9] and PBS-WPS can be cracked when the PBC buttons are pressed [10]. A related work in [14] also uses NFC to facilitate Wi-Fi setup. The AP is set in open mode and clients receive the password and AES key of the AP with NFC.

3 Automatic and Secure Connection

There are two cases to be considered in automatic and secure connection in a Wi-Fi network: (A) Pre-shared key auto-connection and (B) Keyless auto-connection. Both methods can be applied in WPA2 or WPA3 for secure connection. We define the following terms to be used in our algorithms:

- P1 is a randomly generated password of the IoT gateway.
- KEY1 is the pre-shared key stored in the IoT gateway.
- EP1 is the encrypted key by some encryption algorithm such as AES [13] and EP1 = AES(P1, KEY1)
- KEY2 is the pre-shared key stored in the IoT client devices and KEY1 = KEY2.
- $\text{Prefix}_{\text{ssid}}$ is the predefined prefix label as part of SSID.

3.1 Pre-shared Key Auto-Connection

In the pre-shared key auto-connection, the IoT gateway (GW) and the IoT End-devices (ED) have the preloaded shared keys, KEY1 and KEY2, where KEY1 = KEY2, respectively. The algorithm is shown in Fig. 3.

It works as follows: (1) GW loads the pre-shared key, KEY1, (S301) and then generates the random password, P1 (S302). (2) In step S303, GW generates the encrypted password EP1 by applying some encryption algorithm (e.g. AES) on P1. (3) GW generates a SSID by concatenating $\text{Prefix}_{\text{ssid}}$, EP1 and date in Step S304. (4) During steps

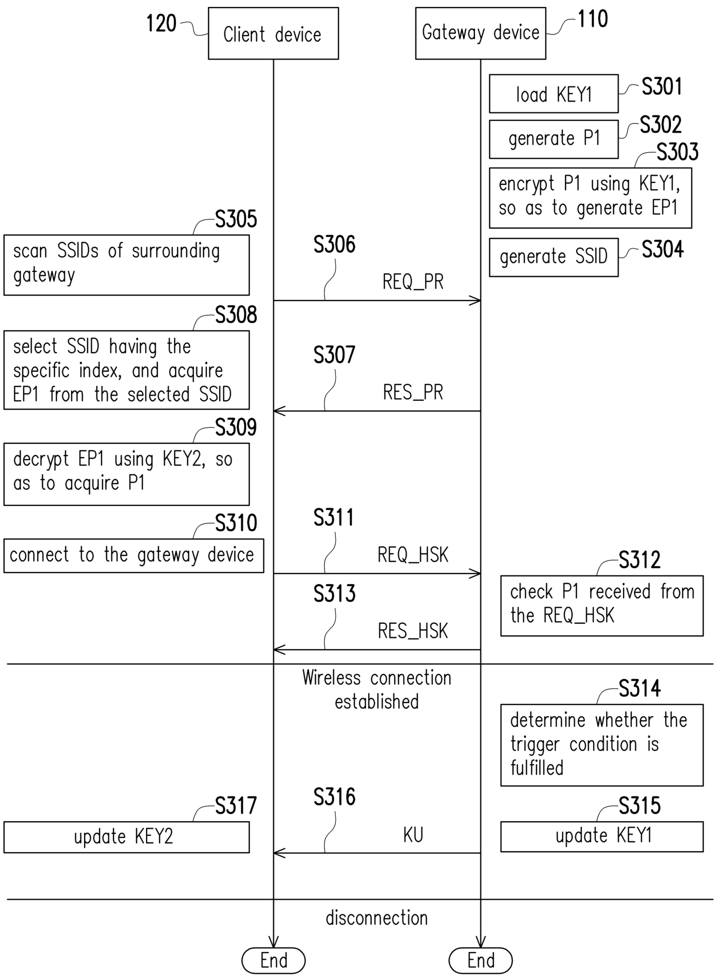


Fig. 3. Pre-shared key secure and automatic connection

305–307, ED sends a probe request (REQ_PR) packet to the surrounding GWs and receives probe response (RES_PR) packets from the responding GWs. (5) In step 308, ED selects the correct GW with identifying label, $\text{Prefix}_{\text{ssid}}$ and retrieves EP1 from SSID. (6) In step S309, ED decrypts EP1 with KEY2 and recovers P1. (7) During steps S310–S313, GW and ED perform 4-way handshake as shown in Fig. 1 to establish Wi-Fi connection. (8) Once connected, GW prepares a new shared key (or backup key), $\text{KEY1}_{\text{backup}}$ and sends it to connected EDs as a backup key, $\text{KEY2}_{\text{backup}}$. (9) Entire pre-shared key auto-connection process is repeated with the new keys (i.e. $\text{KEY1}_{\text{backup}}$ and $\text{KEY2}_{\text{backup}}$) at midnight or when some password cracking attack is detected.

3.2 Keyless Auto-connection

In the keyless auto-connection, GW and EDs do not have the preloaded shared key. Instead, some symmetric encryption/decryption algorithms (called keyless encryption/decryption algorithm) are preloaded in both GWs and EDs. The algorithm is shown in Fig. 4. It works as follows: (1) GW randomly generates the password, P1 and then encrypt P1 into the encrypted password, EP3, by applying keyless encryption algorithm, as shown in step S501 and S502. (2) In step S503, GW generates a SSID by concatenating $Prefix_{ssid}$, EP3 and date. (3) ED sends a probe request (REQ_PR) packet to the surrounding GWs and receives probe response (RES_PR) packets from the responding GWs during steps S504–S506. (4) In step S507, ED selects the correct GW

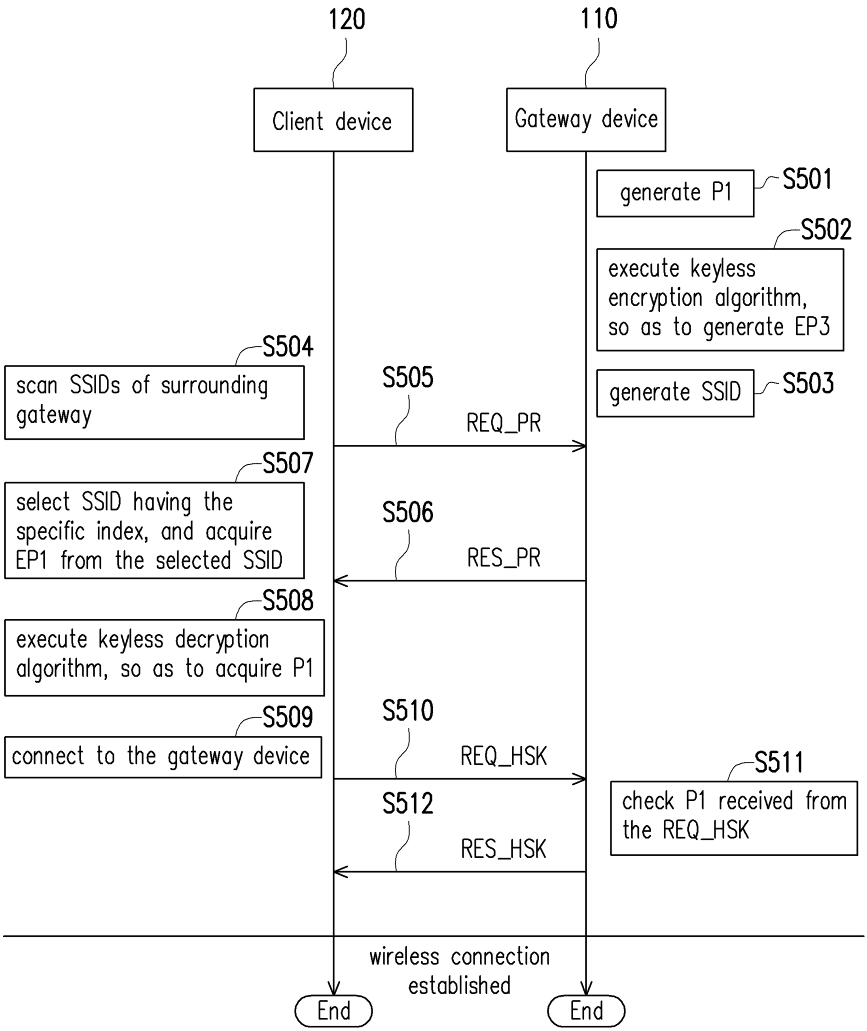


Fig. 4. Keyless secure and automatic connection

with identifying label, $Prefix_{ssid}$ and retrieves EP3 from SSID. (5) ED decrypts EP3 with KEY2 and recovers P1 in step S508. (6) GW and ED perform 4-way handshake to establish Wi-Fi connection. These steps are shown in steps S509– S512. (7) Once connected, GW prepares new shared key and sends it to connected EDs. (8) Entire pre-shared key auto-connection process with the new key is repeated at midnight or when the password cracking attack is detected.

3.3 Wi-Fi Validation Algorithm

The auto-connection algorithms presented in previous subsections can be further secure by applying the validation algorithm, as shown in Fig. 5. The idea is a valid ED can send a challenge request to validate the connected GW and vice versa. Assume all the valid EDs have a backup KEY2, $KEY2_{backup}$, and are connected to the valid GW having a backup KEY1, $KEY1_{backup}$, where $KEY1_{backup}$ = $KEY2_{backup}$, and an invalid

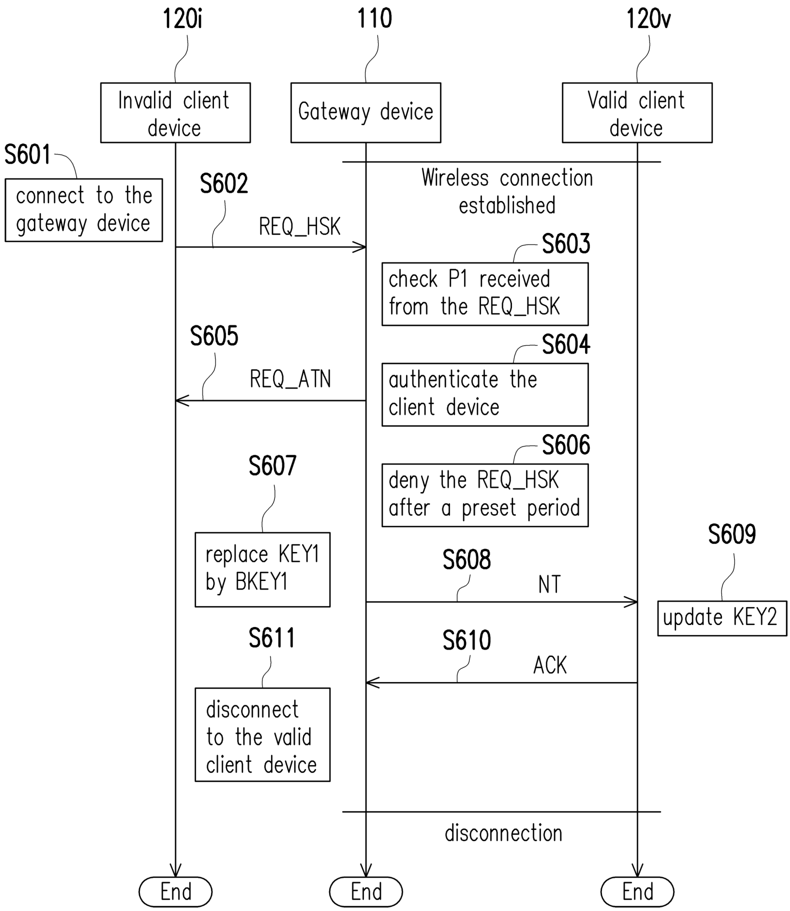


Fig. 5. Validation algorithm

ED cracks the password of the valid GW. The validation algorithm can prevent the invalid client from accessing network. It works as follows: (1) Since the invalid ED cracks the password P1, a 4-way handshake can be applied to connect to the valid GW as denoted in steps S601–S603 and an IP address may be granted from GW. (2) GW may further apply the validation algorithm to challenge the invalid ED in steps S604–S605. If the invalid ED fails to reply with a valid response, GW will disconnect the connection and record the MAC address of the invalid ED and thus no further connection is possible for the invalid ED in step S606. (3) In step S608, GW notifies all valid EDs that the network will be rebooted with the backup KEY for the incoming connection. (4) Once all the valid EDs are notified, GW reboots itself and starts a new auto-connection flow as shown in Fig. 3.

4 Discussions

The experimental results are conducted with a Raspberry Pi 2 as the AP (i.e. GW) and NodeMCUs [15] as the Wi-Fi clients. Four related works (i.e. PIN-WPS, NFC-WPS, PBC-WPS and [14]) are compared with our works with respect to automatic connection. The factors to be evaluated of the comparison table are "conn time" (the connection time required for an ED connecting to a GW), "conn mode" (the connection mode is classified into Easy (i.e. with human intervention) and Auto (without human intervention) connection) and "extra cost". The result is shown in Table 1. Only our methods provide secure and automatic Wi-Fi network connection with no extra hardware cost. Note that NFC-WPS and [14] may provide automatic Wi-Fi connection, however, both approaches require expensive NFC on both GWs and EDs. In addition, it may be not suitable for using NFC if the devices are too heavy to be moved or hard to reach (e.g. hang in high place).

Table 1. Wi-Fi auto or easy connection comparison table

	PIN-WPS	NFC-WPS	PBC-WPS	[14]	Proposed
conn time	60 s	30 s	30 dec	30 s	60 s
conn mode	Easy	Auto	Easy	Auto	Auto
extra cost	No	Yes	Yes	Yes	No

Four related works (i.e. PIN-WPS, NFC-WPS, PBC-WPS and [11]) are compared with our works with respect to Wi-Fi security. The factors to be evaluated of the comparison table are "DoS_{GW} secu" (DoS security on GW), "DoS_{ED} secu" (DoS security on ED) and "WPA2 secu" (better WPA2 security). The result is shown in Table 2. Only our methods provide DoS security for gateways and enhance WPA2 security.

Table 2. Wi-Fi security comparison table

	PIN-WPS	NFC-WPS	PBC-WPS	[11]	Proposed
DoS_{GW} secu	No	No	No	No	Yes
DoS_{ED} secu	No	No	No	Yes	No
WPA2 secu	No	No	No	No	Yes

5 Conclusion

We propose automatic and secure Wi-Fi connection mechanisms for IoT applications in which IoT gateways and end devices can automatically establish secure Wi-Fi connection without human intervention and with almost zero configuration cost. In addition, our works enhance WPA2 security by introducing randomly generated SSIDs and passwords. We also propose a validation algorithm to prevent password cracking and DoS attacks. To the best of our knowledge, our design is the only solution achieving secure and automatic Wi-Fi connection with no extra hardware cost.

References

1. Gartner IoT market report, February 2017. https://www.gartner.com/newsroom/id/3598917. Accessed 3 Apr 2018
2. Gartner IoT security report, March 2018. https://www.gartner.com/newsroom/id/3869181. Accessed 3 Apr 2018
3. Evans, D.: The Internet of Things: how the next evolution of the internet is changing everything? (2011). https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
4. Wi-Fi Protected Access (WPA). https://en.wikipedia.org/wiki/Wi-Fi_Protected_Access. Accessed 3 Apr 2018
5. Wi-Fi. https://en.wikipedia.org/wiki/Wi-Fi. Accessed 3 Apr 2018
6. 4-way handshake. https://en.wikipedia.org/wiki/IEEE_802.11i-2004. Accessed 3 Apr 2018
7. KRACK. https://en.wikipedia.org/wiki/KRACK. Accessed 3 Apr 2018
8. Liu, X.: The security analysis on Wi-Fi communication, Master thesis, Shanghai Jiaotong University (2009)
9. Wi-Fi Protected Setup. https://en.wikipedia.org/wiki/Wi-Fi_Protected_Setup. Accessed 3 Apr 2018
10. Chang, J.: Secure and automatic connection for IoT end-devices and gateways based on WiFi Technology, Master thesis, Tatung University (2015)
11. Zhang, Y., Sampalli, S.: Client-based intrusion prevention system for 802.11 Wireless LANs. In: IEEE Wireless and Mobile Computing, Networking and Communications, pp. 100–105 (2010)
12. Service set (802.11 network). https://en.wikipedia.org/wiki/Service_set_(802.11_network). Accessed 3 Apr 2018
13. Advanced encryption standard. https://en.wikipedia.org/wiki/Advanced_Encryption_Standard. Accessed 3 Apr 2018
14. Jie, M.A., Jin-long, E.: WiFi transmission connection scheme based on near field communication. Comput. Eng. **39**(6), 1–6 (2013)
15. NodeMCU. https://en.wikipedia.org/wiki/NodeMCU. Accessed 3 Apr 2018

Analysis and Evaluation of Wireless Networks by Implementation of Test Security Keys

Arifur Rahman[1] and Maaruf Ali[2](✉)

[1] PHASTAR, Unit 2, 2A Bollo Lane, London W4 5LE, UK
arif.rahman@phastar.com, arif.ovee2010@gmail.com
[2] International Association for Educators and Researchers (IAER), Kemp House, 160 City Road, London EC1V 2NX, UK
maaruf@theiaer.org, maaruf@ieee.org

Abstract. There are so many weaknesses found in the Wired Equivalent Privacy (WEP) key usage protocol and even in the improved Wireless Protected Access (WPA) security key generation algorithm that often mixed mode WPA-WPA2 or WPA2 are utilized - as they are considered a more secure way to obtain wireless security generated keys to date. This paper reports on a practical investigation to test the weaknesses of wireless network security keys, recommend more secure keys and provide a solution to increase the security level of the wireless network. Penetration tests are initiated using the Kali Linux operating system with the help of penetration testing tools to hack WPA-WPA2 mixed mode of access and then provide a solution to increase the security of wireless networks. This will greatly reduce the likelihood of the most common network attacks. The findings of the project will benefit users to both understand and to learn about possible loopholes within their wireless networks. Furthermore, the finding will also act as a guideline for the domestic Wi-Fi user about different security settings having implications on their Wi-Fi security.

Keywords: Wired Equivalent Privacy · WEP · Wireless Protected Access WPA · WPA2 · Penetration testing · Keys · Hacking · Kali Linux

1 Requirements for Cracking WPA-WPA2 Mixed Mode Security Key Protected Wi-Fi Password

WPA-WPA2 mixed mode security is the second most secured Wi-Fi [1, 6, 53, 55, 58, 59] security key and thousands of people make use of this security key in domestic environments – this being the rationale for undertaking the investigation. The following requirements in terms of the software tools used and the operating system, are shown in Table 1, below. These were used to initiate the successful attack, for this research.

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 107–126, 2018.
https://doi.org/10.1007/978-3-319-95450-9_9

Table 1. Software tools and requirements used for cracking the WPA-WPA2 Wi-Fi password.

Name of Software	Description
Kali Linux [19]	Linux operating system
Airmon-ng	Place cards in monitor mode
Airodump-ng	It captures raw frames
Aireplay-ng [3, 4, 14, 20]	It generates traffic used in aircrack-ng
Aircrack-ng [14]	It is a complete suite of tools used for monitoring, testing, cracking and attacking
Dictionary attack [57]	A type of Wi-Fi password attack
PWGen	PWGen is a software that generate passwords
Password list [13, 24]	A list of passwords with different combination to crack the Wi-Fi password

1.1 Details of the Requirement

Kali Linux [19]: is a Debian-based Linux [30] operating system (OS) distribution. It is a very advanced suite of tools [20] for penetration testing [43], also used for auditing. The Kali Linux OS comes with literally over six hundred different tools including those for: penetration testing, reverse engineering and forensics. Kali Linux is open source [30], more stable and gives the user greater freedom to carry out various tests. It was launched on 13th March, 2013. The six hundred penetration testing tools alone are fully customizable, supports multi-languages and can be developed and deployed in a secured environment (sandbox).

Airmon-ng: is used in the Linux penetration test to enable the monitor on the wireless interface. It is also used to turn on the monitor mode. The command has to be correct, otherwise the user will receive an invalid command result.

Airodump-ng: this captures the packets transmitted during an 802.11 session containing the standard frames. It is used with aircrack-ng [14] for cracking the Wi-Fi password. It generates and writes out numerous files that contain the details of all the clients and access points [5] over the intercepted air interface.

Aireplay-ng: this generates traffic and is used in aircrack-ng [14] for injecting frames used for cracking the WEP, WPA and WPA2 [2] security key. According to aircrack-ng.org [3, 4], there are several types of attacks that cause authentications to capture the handshake of the WPA data interactive packet reply including: fake authentications and during the ARP request reinjection.

Aircrack-ng [14]: is the network software suite, which consists of the: detector, packet sniffer to crack WEP, WPA and WPA2 and an analysis tool. It works with almost all wireless network interface controllers/cards [10, 33] that are compatible with raw monitoring mode and are able to sniff 802.11a, 802.11b, 802.11 g traffic. According to the official aircrack-ng website (2016) [4], it captures packets and exports data, reply attacks, checks for wireless cards and is able to crack Wi-Fi passwords. The software has not been tested on the latest 802.11ax standard.

PWGen: is a software which is able to generate millions of passwords very quickly. It gives the user the freedom to select different types of characters, phrases, formats and the amount of password that need to be generated. For example, it takes ten seconds to generate one million passwords, being dependent on the type of the central processing unit (CPU).

Password list [13, 24]: is a list of passwords with different combinations, which is used for cracking the Wi-Fi password. This type of file contains millions of passwords and if any password match with the target then it displays that specific password in the result menu.

Dictionary attack: is a method to break the password protected security system. In this method, a password-list is used which contain all types of passwords and tries to match the key with the target key. Dictionary attack is most often successful because many companies and organisations use very ordinary passwords or a default password. According to Rouse [45], Dictionary attack is also use to find out the key necessary to decrypt and encrypted a message or document.

1.2 The Plan of the Attack to Crack the Wi-Fi Password

Figure 1, below, is the set-up before initiating the Wi-Fi password crack attack.

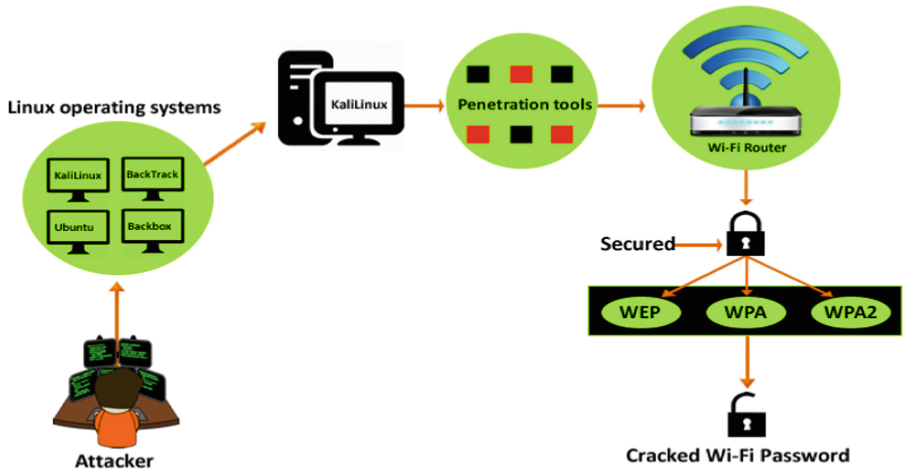


Fig. 1. Design of the Wi-Fi attack.

Linux has different operating systems. The most known Linux operating systems are: Kali Linux, BackTrack [43], Ubuntu and Backbox. The Kali Linux OS was chosen for this attack. The next step consisted of selecting the penetration tools. Linux offer lots of free Wi-Fi password cracking software. After researching the hacker scene: [2–5, 7–9, 12, 14–17, 19–22, 25, 27–29, 31, 34–45, 48–50, 54, 56, 57] these tools were chosen. The tools are: airmon-ng, airodump-ng, aireplay-ng which used in aircrack-ng. The Dictionary attack was executed with these penetration tools to crack the Wi-Fi router password. Acknowledging the fact that the Wi-Fi router can be secured with: WEP [27], WPA [27], WPA-WPA2 mixed mode or WPA2 security key. In the end, the

WPA-WPA2 mixed mode security encryption was cracked with this attack. The rationale being as this is the second most widely chosen way to protect (encrypt) the Wi-Fi password and Wi-Fi traffic – for its predominant use in the domestic environment.

1.3 WPA-WPA2 Mixed Mode Attack

The first step commences with the opening of a terminal on the Kali Linux OS and then executing the airmon-ng command to kill any processes that may interfere with the aircrack-ng suite. Then the network interface was turned down. If any interface is turned down then they have to be turned on again to continue the attack. Thus airmon-ng command is run to put the wireless card into monitor mode. Then the airodump-ng command is executed. This shows all the wireless networks including their: channel number, encryption key, BSSID (Basic Service Set Identifier) in the area under surveillance. Then a folder was created to save any intercepted handshakes. After that, the target was selected and the airodump-ng command was executed by specifying the channel number and BSSID of the selected target. Airodump-ng then collects more information about the target and shows the devices connected to the network and its station ID. In the next step, aireplay-ng was initiated where the client access point was specified, along with the BSSID, station ID and deauth (de-authentication) number. Upon hitting the enter button, aireplay-ng starts sending packets to the target device and tries to make a handshake. It takes time to capture the handshake. When the handshake has been acquired with the target clients, aircrack-ng is opened with the specified folder location where the handshake data is saved, along with the location where the password list has been saved (created earlier with different types of password combinations). The password list contains two million passwords. These different types of password lists were downloaded from online [56] and by using a software called PWGen. After generating passwords with PWGen, these were mixed together with the downloaded passwords to make one list that has all types of common and critical password combination. After running the aircrack-ng command, it started checking the target password with the already created password list. If aircrack-ng matches with the target password it shows that a key has been found. In this simulation, after just 56 min, a password from the list was indeed found to have matched with the target password.

1.4 Prevention of the Attack

It is very hard to stop cyber-attackers. If any actions are taken to stop them they will surely find new ways of attacking and circumventing the protection. To stop this type of attack explained above and strengthen the security of the Wi-Fi network, some important necessary steps can be taken. Figure 2, shows how to increase the security of a Wi-Fi network and stop penetration attacks. This is explained in Section Two.

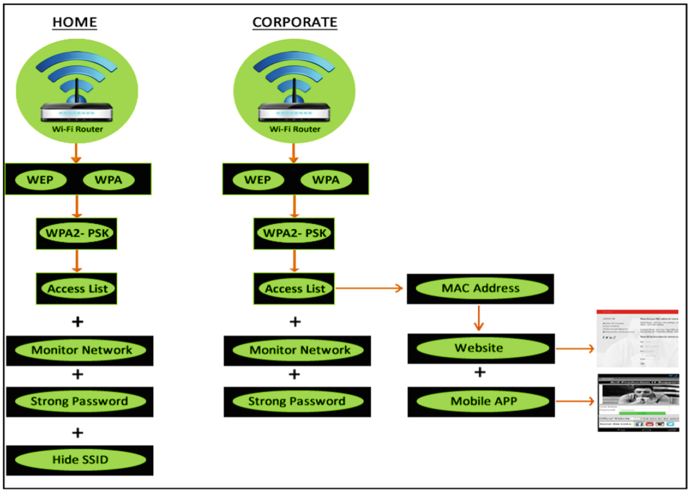


Fig. 2. Hardening of Wi-Fi networks to mitigate from the deleterious effects of attacks.

2 Implementation of the Processes of Hacking a Wi-Fi Router

This section explains hacking into a secured and password protected Wi-Fi router. The requirements, tools and techniques used during the attack have been mentioned in Section One. A step-by-step illustration of WPA-WPA2 mixed mode password hacking is explained below.

First, login to the router admin panel by typing the router IP address, which is usually 192.168.0.1 into the URL of a web browser. Then insert the username and password to gain access to the router admin panel. Figure 3, shows a typical router login admin panel.

Fig. 3. A typical router admin panel login page.

From the wireless network security options settings, the security option is changed from WPA2-PSK (AES) to WPA/WPA2-PSK (Mixed Mode), as shown in Fig. 4, by selecting the bottom radio button.

Fig. 4. Selection of WPA/WPA-2PSK (Mixed Mode) security.

The PWGen software is opened and different characters, phrases and formats are selected along with the number of passwords to generate. PWGen will then use these parameters to generate a list of passwords. This will take a few seconds. A mixture of different combination of passwords will then be used later on to crack the Wi-Fi password. Figure 5, shows the list of passwords that were generated by the PWGen software. The password list that may be used for the attack can be download from: https://www.dropbox.com/sh/pa53d91mvoo1q9o/AABenJGljphwXezSTNBCdJBPa?dl=0.

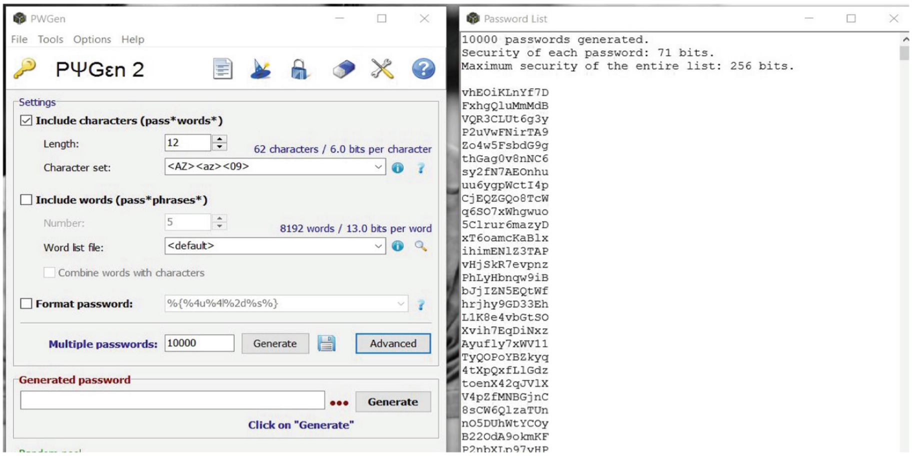


Fig. 5. The list of passwords that were generated by the PWGen software.

A new terminal is then opened in the Kali Linux OS, the command 'airmon-ng check kill' is issued in order to kill the interfaces that may interfere during the progress of the hack attack. This is shown in Fig. 6, below.

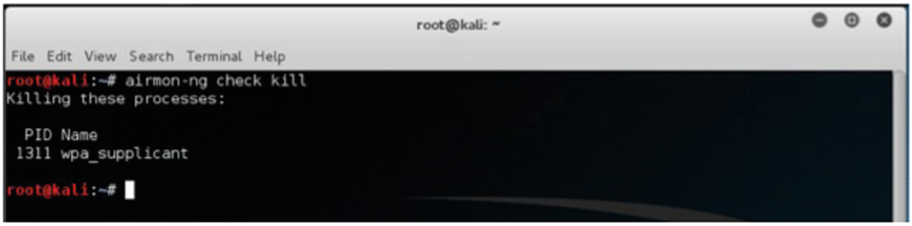


Fig. 6. The airmon-ng command running on the Kali Linux terminal.

On the next step, run the ‘ifconfig wlan0 down’ command, to turn down the network interface, as show in Fig. 7.

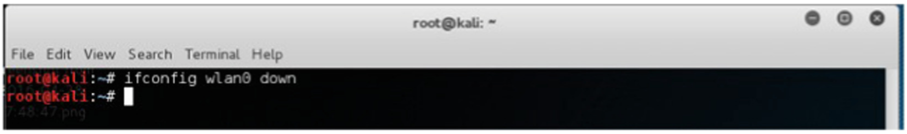


Fig. 7. The ‘ifconfig’ command to turn down the network interface.

After the network interface has been turned down, it is important to turn it back on again. This is achieved by typing ‘airmon-ng start wlan0’ to enable the monitor mode, as shown in Fig. 8.

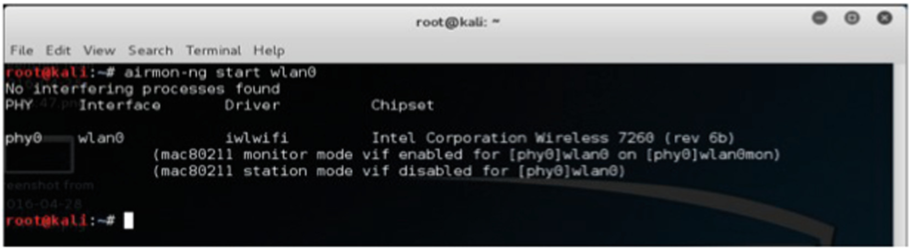


Fig. 8. The wireless interface wlan0 on monitor mode mon0.

After that, the ‘airodump-ng wlan0mon’ command is run to check the wireless network. Upon running the command, it starts searching the wireless network and shows the BSSID, channel number, encryption and all other necessary information needed to crack a network. The listing of the Wi-Fi networks found, that is output onto the screen, in shown in Fig. 9. All that is next required is to select the target from this list.

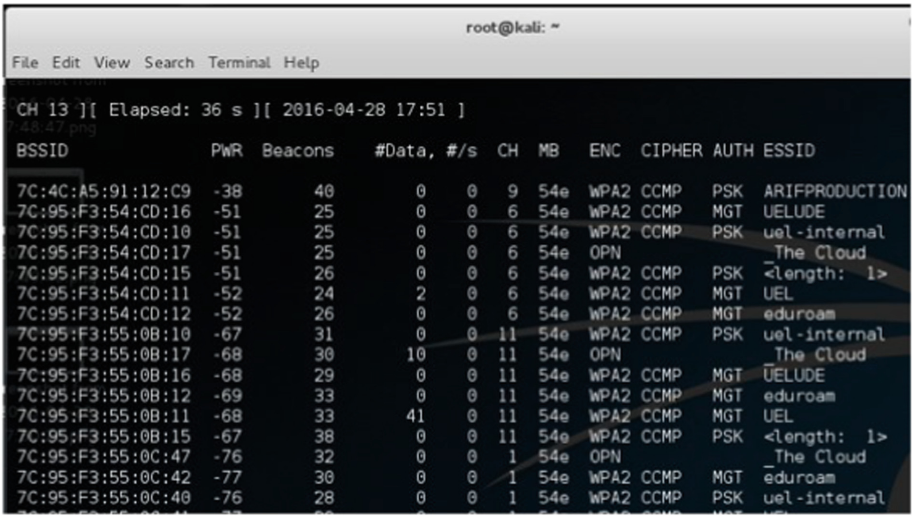


Fig. 9. The Wi-Fi networks found by executing the airodump-ng searching command.

After selecting the target, in this case from the first line of Fig. 9, type ‘airodump-ng wlan0mon -c -bssid 7C:4C:A5:91:12:C9 –w /root/Crack/wpa2psk’. This will display the devices connected to the wireless Internet and its station ID, as shown in Fig. 10.

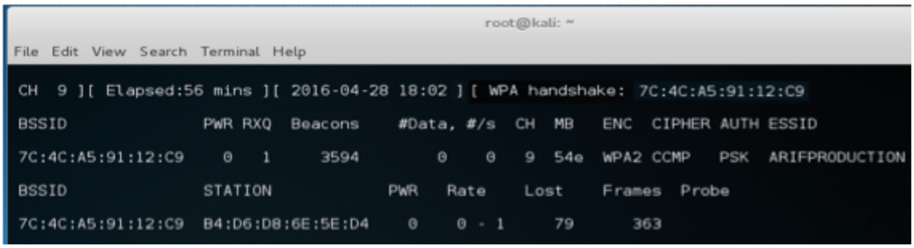


Fig. 10. Display of the information about the connected device to the target network.

Run ‘aireplay-ng wlan0mon -0 0 -a 7C:4C:A5:91:12:C9 -c B4:D6:D8:6E:5E:D4 wlan0mon’. This will make aireplay-ng start sending packets to the target device and try to make a handshake. This is shown in Fig. 11.

The last step is to type ‘aircrack-ng /root/Crack/wpa2psk-01.cap –w /root/darc0de.lst’. Here the location of the handshake data and the list of passwords are specified. Now aircrack-ng starts the process of trying to match the password on the target device but in this scenario, it has actually failed to crack the password - because of the target device connection being dropped. This situation is shown in Fig. 12.

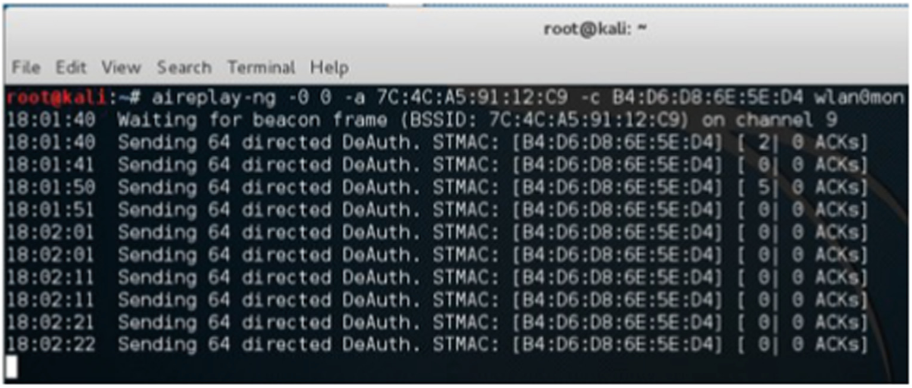


Fig. 11. Screen snapshot of the aireplay-ng command sending packets to the target device.

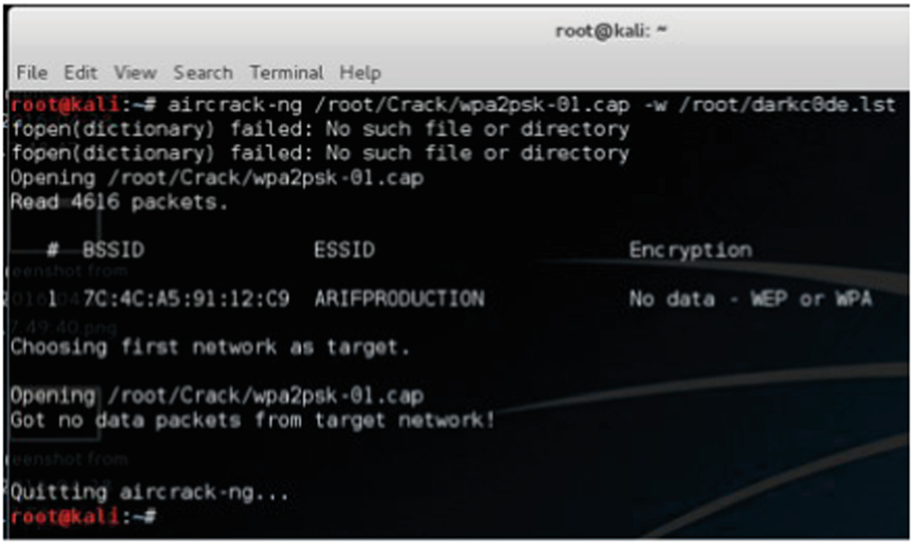


Fig. 12. The screen output showing how aircrack-ng tried to match the password but now a handshake has been received because of a connection drop.

When the target device connection is up and running, the same command is used again and this time a password from the password list has matched with the target device password. It took 41 min 06 s to crack the password. The success of the attack is shown in Fig. 13.

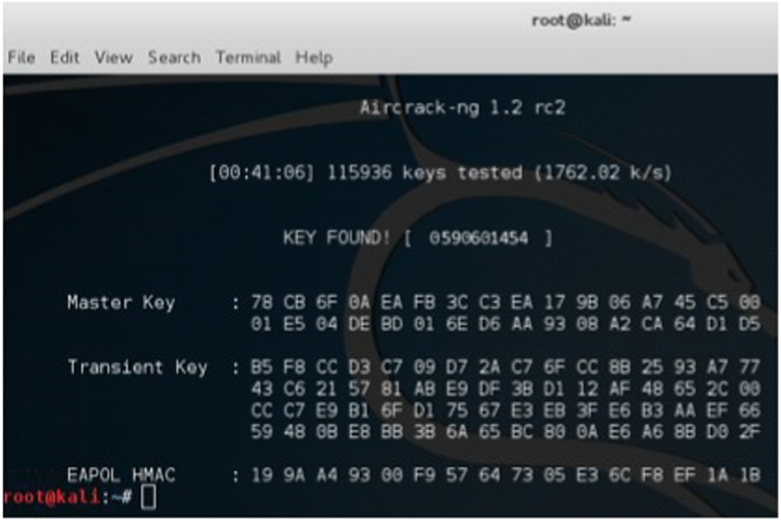


Fig. 13. The password has been cracked on the target device.

3 Solution of the Attack

Attackers always look for networks that are insecure and vulnerable and specifically attack those network to gain illegal access [46, 47]. Wi-Fi Internet is very popular but most of the people do not know how to configure their router and secure it to prevent attacks or even simple unauthorized access. People have to follow some necessary steps to secure their network, which is explained below.

3.1 Solution for the Home User

Home Wi-Fi networks are the cyber attacker's favourite type of network to target and gain access into. Most domestic consumers purchase their router and leave it with its default factory settings. The following nine steps, should be implemented in order to make a more secure home wireless network:

1. Change the router default admin panel password with a strong password.
2. Change the password of the wireless router from its default factory set password.
3. Use a password generator to create a strong password and change the password every three to four weeks.
4. Turn off the WPS (Wi-Fi Protected Setup) mode.
5. Turn on the router firmware and install the latest version of the firmware.
6. Monitor your network and logs periodically for anomalous entries or behaviour.
7. Use an access list from the router admin page - this will only give access to the wireless network of MAC addresses saved in the router access list.
8. Turn off the BSSID of your network, that is, make it invisible.
9. Change the security mode to WPA2-PSK - currently the most secure Wi-Fi key.

3.2 Solution for the Corporate Users

Corporate and enterprise networks have to be more secure because they have so many staff and corporate data which is important to protect and cannot be compromised. Everyday lots of visitors visit an organization for different purposes. Not all their intentions and motives can be ascertained beforehand. Attackers try to hack an organization's network for many reasons and often for big financial gain. The breakdown of the motivations are shown in Fig. 14.

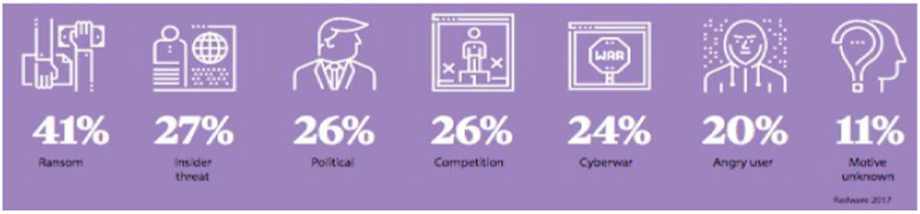


Fig. 14. Motives behind cyberattacks from a global study of large victim organisations [60].

Corporate users also have to follow the same steps that were explained for home users. Furthermore, they can adapt the following methods to make it easier for their staff and client. There will be two methods, traditional way and modern way, explained next.

Traditional Way of Request. In the traditional way of approach, there will be a form with all the required fields in the information zone of a company - visitors and staff who wants to use the wireless internet will complete the form and it will be sent to the IT department every hour. The form may typically look like that as shown in Table 2.

Table 2. Shows the traditional request form of wireless Internet usage request.

Name	Visitor/Staff	MAC Address of the device	Purpose of the request
…	…	…	…

Mobile App for Staff. Staffs are very important and they need Internet access for various company work. The network will be secured with access lists and staff can send their devices MAC addresses through the staff app to the IT department to authorize their device MAC address in order to connect with the internet. A staff app was created for the purpose of this research in order for the office staff to send their request for Internet access. To create this mobile platform app, an initial wireframe was created, as shown in Fig. 15.

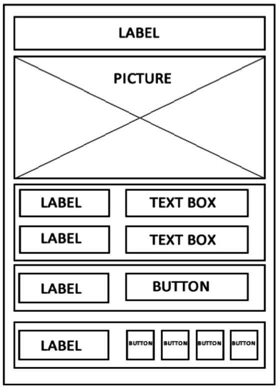


Fig. 15. The wireframe of the mobile app.

This wireframe in Fig. 15, shows how the mobile app will appear once built. There will be the name of the app at the top, followed by the picture of the company in the middle, staff user name and password for login, official website link and at the end, the social website links of the company. After building the app, it was run on an Android device. Staff can type in their username and password and login to the app to gain access to the corporate wireless network. The app is shown in Fig. 16.

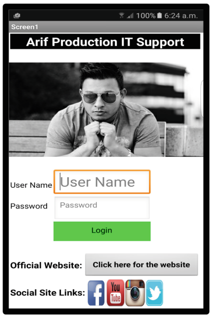


Fig. 16. The app design on an Android device.

After the login, the user is directed to the next page where the selection of user device type is requested, as shown in Fig. 17.

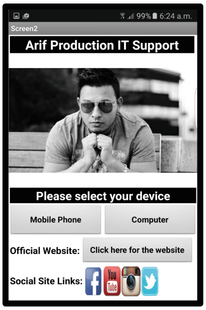


Fig. 17. Screen to prompt users to select their device.

After selecting the device, staff are prompted to follow the procedure to check their MAC address. After checking the MAC address, it is written in the "Enter your MAC address box". The "send request button" is then pressed to send a request message to the IT departmental phone. The screens for either devices are shown in Fig. 18.

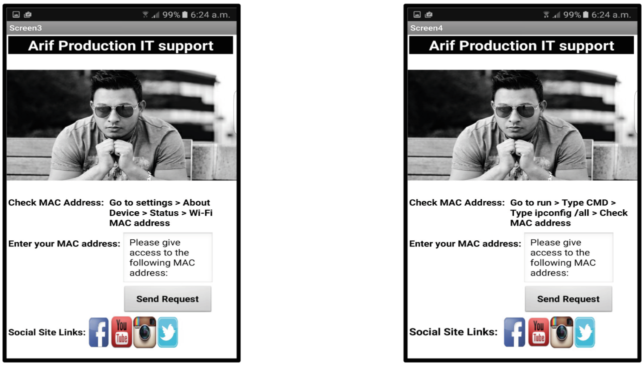


Fig. 18. Instructions to find the mobile device MAC address and send it via the 'Send Request'.

When a staff sends a request, this app immediately sends a text message to the IT department to give access to the staff device, as shown in Fig. 19. The IT department then logs into the wireless network admin panel and gives access only to that specified MAC address of the requested device and also only for that authorized staff member of the organization. The mobile app can be downloaded from the proposed website: http://www.arifproduction.co.uk/arif_production.apk.

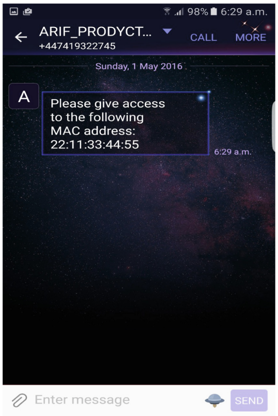


Fig. 19. The IT department has received a device add request.

3.3 Website for Clients

Using a website dedicated for clients is the easiest way for sending a request for wireless Internet access. An example of such a developed web page is shown in Fig. 20. When any clients/staffs connect to the organization network, it will take them directly to the wireless Internet access page. After filling the form, it will send a mail to the IT department to give access to the wireless Internet of the organization, shown in Fig. 21. Figure 22, shows the details of the client email sent to the IT department. To direct a user to the organization website, a method call “captive portal” have to be used. The use of this method, however, cannot be done using a normal home router. It is only possible using a highly advanced router. A website has been created to show a demo on how to send a request from such a website. All the client needs to do is to fill the details of the form and click the send button.

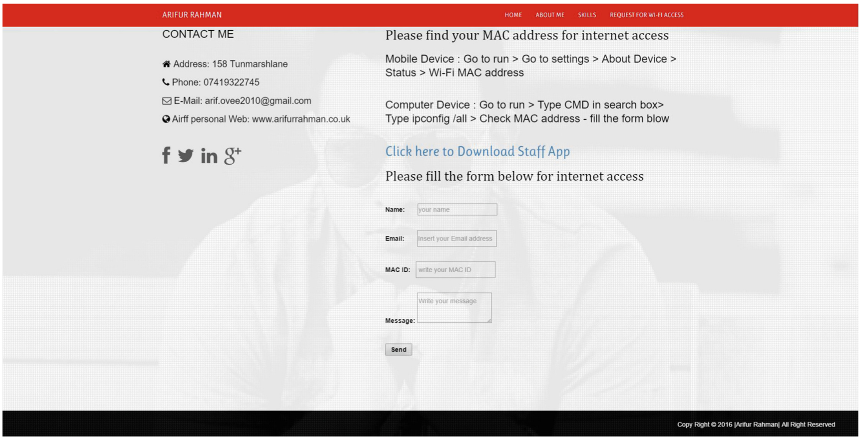


Fig. 20. The wireless internet request page from the organization website.

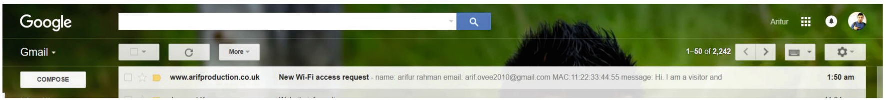


Fig. 21. A "new Wi-Fi access request" received in the mailbox as the subject header.

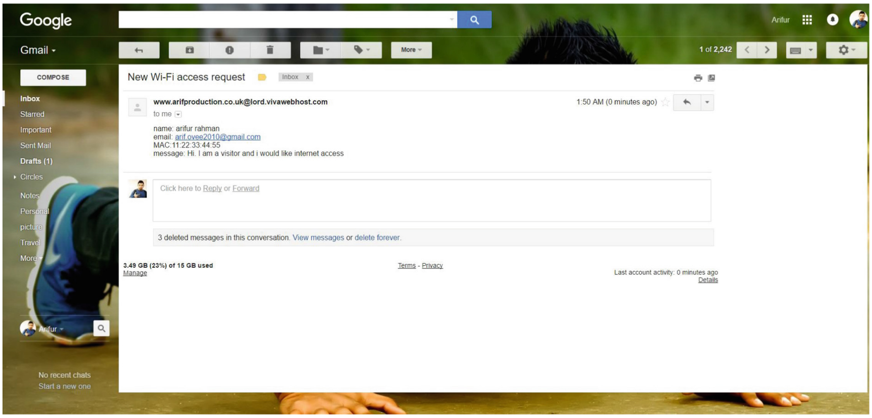


Fig. 22. The details of the request received in the email sent by the client.

The link for the website based on the proposed solution can be viewed on the following link: http://www.arifproduction.co.uk/.

The codes of the website have been uploaded on Dropbox and can be download from the following link: https://www.dropbox.com/sh/6wupjxr2glr3eby/AACzqeG2V1OY2k2CpZeUpv0ya?dl=0/.

4 Evaluation

The study was designed to find the different types of wireless networks [11], vulnerabilities of their security keys, comparing them, to attack a WPA-WPA2 mixed mode secured network and provide a solution. Previous researchers have tried to hack either WEP [17] or WPA secured networks only. They have not done anything on a WPA-WPA2 mixed mode secured Wi-Fi access network. Background information has been collected from previous research papers, journals, articles and websites. The penetration test hack was run by using the Linux OS. The aircrack-ng tool package and PWGen software were used in this test. All the software were downloaded legally and the knowledge of these tools and software has been gained from perusing their official websites and by watching YouTube tutorials.

The findings of this project gives a clear knowledge about the wireless network, their types, standard and security levels and the major attacks against them. Furthermore, the implementation gives an insight of the vulnerabilities of the wireless network and an overview of how a hacker can crack the wireless network password has also been presented.

The solution that has been produced in this research project offers lots of advantages [18, 23, 26, 32, 52] such as if any home user and corporate user use this solution, they will have lots of benefits and be able to secure their wireless network and prevent all major types of attacks.

The advantages of using this solution are as followed:

- By changing the default password of the router admin panel, only the network administrator will know the password and if the hacker now tries to login with the default password – she/he will fail her/his mission.
- Turning off WPS makes wireless routers more secure because having WPS setup enabled on the router makes it easy to hack with its easy and default configuration.
- Additionally a longer than nine digit pin for login should also be enabled and utilised.
- Installing the latest firmware will make the router more secure and harden it against future attacks. The router manufacturer's website should always be regularly checked for the latest firmware version and installed on the router.
- By turning off BSSID, attackers will not be able to see the network name.
- By using an access list, the user can ensure that only trusted devices are connected and given access to the wireless network. The wireless network will not now give access to any unauthorized device.
- By using WPA2-PSK security mode as the most secure and highly advanced encrypted security mode, will make it difficult to crack.
- A staff app and website is introduced to send wireless access requests. To authorize the staff device, companies can use the staff app and staffs can easily send their MAC address to the IT department in order to add it on the access list to give them Internet access. By using this website, clients can send wireless access requests directly to the IT department. All other unrecognized requests can be ignored by the IT Department. If any client connects to the company wireless device, it will direct them to the company's official website and ask them to fill a form. This form explains in detail how to find the device's MAC address and how to send the request.
- By adapting the traditional wireless request form method, companies can now save further time and money. This is so because users beforehand would have had to fill a paper form with all the necessary information in the information zone and then later on this would have to have been all collected and sent to the relevant IT department. This would have contained several points of weaknesses where information could have been lost, compromised or even spoofed.

During the evaluation, the solution designed in the form of the website application was not able to link with home wireless networks. To connect the website to the Wi-Fi router a method needed to be followed which is known as "captive portal". This is only available in modern highly advanced routers such as on the Cisco C3750 (layer three router switch). The penetration test has been done using only standard home routers so website were unable to be connected.

The evaluation is based on the methodology and implementation to achieve the required objectives. The attackers always search for vulnerable networks and gain access to it for something or for their pleasure. Wireless networks play a vital rôle in our daily lives and it needs to be secure in order to protect personal information and other sensitive data. Users can often become blasé about security when using a PAN (Personal Area Network) [51].

5 Conclusions

Wireless networks are one of the most popular technologies that has spread all over the world. However, a good number of users do not know about the safety status of their wireless network, that is, how vulnerable is it to being hacked by outside cyber intruders. The domestic user often just purchases their Wi-Fi router and leaves it with its default configuration. This can be potentially be very dangerous. They do not take any additional steps to secure it further. A router with a default setting is an easy target for attackers, who can crack the network very easily. For this reason it is very important to have a general knowledge about wireless networks before setting it up. Firstly, this project investigated the types of Wi-Fi wireless networks, the possible attacks they encounter and the weaknesses of WEP and WPA security keys.

This paper focused on security issues of wireless networks and as other researchers have already shown that WEP and WPA secured network can be cracked. This paper has taken the next stage and demonstrated that WPA-WPA2 mixed mode security key can be cracked. The step-by-step process of hacking this is explained in detail. This implementation will help the user to understand how the attackers try to crack their network and what tools they use. The solution has been proposed in this project to secure the wireless network that will not only help the home user but also help corporate users to secure their company network. The proposed solution is able to stop the attack that has been done in this project and furthermore it can prevent all other major attacks against wireless networks.

By reading this paper, readers can choose a suitable wireless network standard and security for their access points (routers) - ultimately securing their wireless routers.

5.1 Further Study

In the proposed design, a home network that was secured with WPA-WPA2 mixed mode was cracked. For further consideration, a network which is secured with WPA2-PSK security mode can also be cracked by exploiting the “HOLE 196” vulnerability which has been found recently. Furthermore, for the solution, a “captive portal” can be used to connect to the official website of a company to their router. An advanced app can be designed which will be available for visitors as well and they will be able to download it from the company’s website. WPA3 will also be released.

References

1. Gon. An introduction to 802.11(WI-FI) technologies (2015). http://www.4gon.co.uk/solutions/introduction_to_802_11_wifi.php. Accessed 18 Nov 2015
2. Agarwal, M., Biswas, S., Nandi, S.: Advanced stealth man-in-the-middle attack in WPA2 encrypted Wi-Fi networks. IEEE Commun. Lett. **19**(4), 581–584 (2015)
3. aircrack-ng: Cafe Latte attack (2010). http://www.aircrack-ng.org/doku.php?id = cafe-latte. Accessed 15 April 2016
4. aircrack-ng: Introduction (2016). http://www.aircrack-ng.org/doku.php?id = Main&DokuWiki = g339gdvjfup8lor0t66mv4ie63. Accessed 29 Apr 2016
5. Vijay, B.P., Pranit, S.T., Swapnil, D.D.: Protecting Wi-Fi networks from rogue access points. In: 4th International Conference on Advances in Recent Technologies in, Bangalore, India, 2012, pp. 119–122 (2012)
6. Beal, V.: WI-FI (2015). http://www.webopedia.com/TERM/W/Wi_Fi.html. Accessed 26 Nov 2015
7. Beaver, K.: Straightening Out the Hacker's Terminology (2016). http://www.dummies.com/how-to/content/straightening-out-the-hackers-terminology.html. Accessed 22 Jan 2016
8. Burbank, J.L., Andrusenko, J., Everett, J.S.: Wireless Networking Understand- ing Internetworking Challenges. Wiley, New Jersey (2013)
9. Choi, M.K., Robles, R.J., Hong., C.-H., Kim, T.-H.: Wireless network security: vulnerabilities, threats and countermeasures. Int. J. Multimedia Ubiquit. Eng. **3**(3) (2008). https://www.researchgate.net/publication/228864040_Wireless_Network_Security_Vulnerabilities_Threats_and_Countermeasures. Accessed 25 Apr 2016
10. Cioara., J.D., Cavanagh, M.J., Krake, K.A.: CCNA Voice Official Exam Certification Guide. Cisco Press, USA (2008)
11. Cities-Lyon. Wireless metropolitan area networks (Wi-MAN). http://www.cities.lyon.fr/en/wi-man.html. Accessed 19 Jan 2016
12. cmu95752: Wardriving: Legal or Illegal? (2012). https://cmu95752.wordpress.com/2011/12/12/wardriving-legal-or-illegal/. Accessed 28 Apr 2016
13. Dazzlepod: wordlist (2016). http://dazzlepod.com/site_media/txt/passwords.txt. Accessed 1 May 2016
14. Encarnacion, L.: How To Hack WPA/WPA2 Wi-Fi With Kali Linux & Aircrack-ng (2016). http://lewiscomputerhowto.blogspot.co.uk/2014/06/how-to-hack-wpawpa2-wi-fi-with-kali.html. Accessed 30 Apr 2016
15. Fowler, S., Zeadally, S.: Defending against Distributed Denial of Service (DDoS) attacks with queue traffic differentiation over Micro-MPLS-based wireless networks. In: International Conference on Systems and Networks Communications, ICSNC 2006, p. 8, October 2006
16. Guyot, V.: WEP-based security management in IEEE 802.11 wireless sensor networks. In: 3rd IEEE/IFIP International Conference in Central Asia on Internet, ICI 2007, 26–28 September 2007, pp. 1–4 (2007)
17. Hassan, H.R., Challal, Y.: "Enhanced WEP: an efficient solution to WEP threats. In: 2nd IFIP International Conference on Wireless and Optical Communications Networks, WOCN 2005, 6–8 March 2005, pp. 594–599 (2005)
18. Ipoint: Wireless Networking (Wi-Fi) – Advantages and Disadvantages to wireless networking (2016). http://ipoint-tech.com/wireless-networking-wi-fi-advantages-and-disadvantages-to-wireless-networking/. Accessed 11 Apr 2016
19. Kali Linux: What is Kali Linux? (2016). http://docs.kali.org/introduction/what-is-kali-linux. Accessed 28 Apr 2016

20. Kali Tools: Aircrack-ng Package Description (2014). http://tools.kali.org/wireless-attacks/aircrack-ng. Accessed 30 Apr 2016
21. Karnik, A., Passerini, K.: Wireless network security - a discussion from a business perspective. In: Wireless Telecommunications Symposium, 28–30 April 2005, pp. 261–267 (2005). https://doi.org/10.1109/wts.2005.1524796
22. Kaspersky: What is a Trojan Virus? – Definition (2016). https://usa.kaspersky.com/internet-security-center/threats/trojans#.Vyu8W_krKhd. Accessed 19 Apr 2016
23. Kazmeyer, M.: The Advantages & Disadvantages of Wi-Fi (2016). http://yourbusiness.azcentral.com/advantages-disadvantages-wifi-23878.html. Accessed 30 Mar 2016
24. Kearns, D.: Wordlist Package (2013). http://git.kali.org/gitweb/?p=packages/wordlists.git;a=summary. Accessed 3 May 2016
25. Kumar, K., Joshi, R.C., Singh, K.: A distributed approach using entropy to detect DDoS attacks in ISP domain. In: International Conference on Signal Processing, Communications and Networking, ICSCN 2007, 22–24 February 2007, pp. 331–337 (2007)
26. Lander, S.: Disadvantages or problems for implementing Wi-Fi technology (2016). http://smallbusiness.chron.com/disadvantages-problems-implementing-wifi-technology-61914.html. Accessed 20 Apr 2016
27. Lashkari, A.H., Mansoor, M., Danesh, A.S.: Wired Equivalent Privacy (WEP) versus Wi-Fi Protected Access (WPA). In: 2009 International Conference on Signal Processing Systems, 15–17 May 2009, pp. 445–449 (2009)
28. Borsc, M., Shinde, H.: Wireless security & privacy. In: 2005 IEEE International Conference on Personal Wireless Communications, ICPWC 2005, pp. 424–428 (2005)
29. Marco-Gisbert, H., Ripoll, I.: Preventing brute force attacks against stack canary protection on networking servers. In: 2013 12th IEEE International Symposium on Network Computing and Applications (NCA), 22–24 August 2013, pp. 243–250 (2013)
30. Marsh, J.: Linux: Advantages and Disadvantages of Open-Source Technology (2016). Accessed 29 Apr 2016
31. Mashhour, S.A., Saleh, Z.: Wireless networks security in Jordan: a field study. IEEE J. Sel. Areas Commun. **5**(4), 43–52 (2013)
32. Mckinney, E.: Disadvantages of Wireless Networks (2016). http://www.ehow.com/facts_4809373_disadvantages-wireless-networks.html. Accessed 28 Apr 2016
33. Microsoft: Wireless networking overview (2005). https://technet.microsoft.com/en-gb/library/cc784756(v=ws.10).aspx. Accessed 20 Jan 2016
34. Milton, K.: Can Viruses Spread Over Wi-Fi? (2016). http://smallbusiness.chron.com/can-viruses-spread-over-wifi-75136.html. Accessed 3 Apr 2016
35. Mitcheel, B.: wardriving - war drivin (2016). http://compnetworking.about.com/cs/wireless/g/bldef_wardrive.htm. Accessed 26 Jan 2016
36. Peer, D.: The Risks of Viruses, Worms and Trojan Horses on Wireless (2016). http://www.ehow.com/info_7869134_risks-worms-trojan-horses-wireless.html. Accessed 17 Apr 2016
37. Pinola, M.: What is Wi-Fi (2015). http://mobileoffice.about.com/od/glossary/g/wi-fi.htm. Accessed 26 Nov 2015
38. Poddar, V., Choudhary, H.: A comparative analysis of Wireless Security Protocols (WEP and WPA2). Int. J. AdHoc Netw. Syst. (IJANS). **4**(3) (2014). http://airccse.org/journal/ijans/papers/4314ijans01.pdf. Accessed 5 May 2018
39. Pospisil, J., Novotny, M.: Lightweight cipher resistivity against brute-force attack: Analysis of PRESENT. In: 2012 IEEE 15th International Symposium on Design and Diagnostics of Electronic Circuits & Systems (DDECS), 18–20 April 2012, pp. 197–198 (2012)
40. Practically Networked: How to Track Down Rogue Wireless Access Points (2016). http://www.practicallynetworked.com/support/030306wirelesssecurity.htm. Accessed 26 April 2016

41. No authors: Security (2018). https://www.wi-fi.org/discover-wi-fi/security. Accessed 5 May 2018
42. Rohrer, F.: Is Stealing Wireless Wrong? (2007). http://news.bbc.co.uk/1/hi/magazine/6960304.stm. Accessed 27 April 2016
43. Rosa, L.R., et al.: Analysis of security and penetration tests for wireless networks with backtrack Linux. In: IEEE International Conference on Communications, ICC 2013, pp. 1–6, June 2013
44. Rouse, M.: Ethical Hacker (2014). http://searchsecurity.techtarget.com/definition/ethical-hacker. Accessed 24 Jan 2016
45. Rouse, M.: Dictionary Attack (2005). http://searchsecurity.techtarget.com/definition/dictionary-attack. Accessed: 23 Apr 2016
46. Sahu, B., Sahu, N., Sahu, S.K., Sahu, P.: Identify uncertainty of cyber crime and cyber laws. In: 2013 International Conference on Communication Systems and Network Technologies (CSNT), 6–8 April 2013, pp. 450–452 (2013)
47. SANS Institute: How to Avoid Ethical and Legal Issues In Wireless Network Discovery. https://www.sans.org/reading-room/whitepapers/wireless/avoid-ethical-legal-issues-wireless-network-discovery-176. Accessed 28 Apr 2016
48. Sarmiento, O.P., Guerrero, F.B., Argote, D.R.: Basic security measures for IEEE wireless networks. Revista Ingeniería E Investigación, **28**(2), 89–96 (2008). http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092008000200012#fig3. Accessed 20 Apr 2016
49. Selim, G.; El Badawy, H.M.; Salam, M.A.: New protocol design for wireless networks security. In: The 8th International Conference Advanced Communication Technology, ICACT 2006, 20–22 February 2006, vol. 1, p. 776 (2006)
50. Shukla, R., Kolahi, S.S., Freeth, R., Kumar, A.: Educational institutes: wireless network standards, security and future. In: 2010 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), 30 November–2 December 2010, pp. 76–82 (2010)
51. Siep, T.M., Gifford, I.C., Braley, R.C., Heile, R.F.: Paving the way for personal area network standards: an overview of the IEEE P802.15 working group for Wireless Personal Area networks. IEEE Pers. Commun. **7**(1), 37–43 (2000)
52. Stone, D.: The Advantages & Disadvantages of Wi-Fi (2016). http://classroom.synonym.com/advantages-disadvantages-wifi-17344.html. Accessed 29 Mar 2016
53. TechTerms: Wi-Fi Definition (2014). http://techterms.com/definition/wi-fi. Accessed 27 Nov 2015
54. The Computer Gal: How to Get a Trojan Horse and How to Fix It (2010). http://www.thecomputergal.com/Programming/VirusAlert/VerizonWireless.html. Accessed 15 Apr 2016
55. The Economist: A brief history of Wi-Fi (2015). http://www.economist.com/node/2724397. Accessed 18 Nov 2015
56. Tradi: Darkc0de LST Password Dictionary Shared Files (2016). http://tradownload.com/results/darkc0de-lst-password-dictionary.html. Accessed 1 May 2016
57. Webopdia: Dictionary Attack (2016). http://www.webopedia.com/TERM/D/dictionary_attack.html. Accessed 22 Apr 2016
58. Wood, B.: The Evolution of WiFi (2014). http://www.purplewifi.net/history-wifi/. Accessed 18 Nov 2015
59. Hu, X.: Study on wireless local area network technology. In: 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), 21–23 April 2012, pp. 609–612 (2012)
60. http://res.cloudinary.com/yumyoshojin/image/upload/v1/pdf/cyber-risk-resilience-2017.pdf. Accessed 2 May 2018

Embedded System for Speed Estimation by Means of Sound Analysis in Three-Phase Induction Motors

Thyago Leite de Vasconcelos Lima[1(✉)], Julio César da Silva[2], José Anselmo Lucena Jr.[2], Filipe Vidal Souto[2], Thaís Christine Borges da Silva[2], Abel Cavalcante Lima Filho[2], Francisco Antônio Belo[2], and Marcéu Oliveira Adissi[1]

[1] Federal Institute of Education, Science and Technology of Paraíba, João Pessoa, PB 58015-020, Brazil
thyago.lima@ifpb.edu.br
[2] Federal University of Paraiba, João Pessoa, PB 58051-900, Brazil

Abstract. Electric motors consume a large portion of the electric power generated. Three-phase induction motors are the most used in industries, for their robustness, reliability and easy operation. They are inserted in the most diverse processes as the main electromotive force. Measuring speed directly on the motor shaft is no trivial task, because it requires time and additional cost due to adaptations of speed transducers to the axis, which causes costly stops to the process in which this motor is inserted. For this reason, manufacturers and research centers around the world have been developing speed estimation methods based on sensorless techniques. The speed measurement in motors can be used for various applications from vector control of the machine to failure analysis. In this work, a new method was developed and installed in an embedded system to estimate the speed in threephase induction motors through the FFT motor sound analysis. This technique proved to be reliable, showing good accuracy in comparison to the measured speed on the shaft, demonstrating the effectiveness of the method and applicability in other areas of technical and scientific relevance such as analysis and prevention of bearing failures or any mechanism involving shaft rotation.

Keywords: Three-phase inductors motors · Sound analysis · Speed estimation Embedded system · Sensorless measurements

1 Introduction

Three-phase induction motors (IM) are applied in the most diverse industrial sectors, from the petrochemical, sugaralcohol, mining, automotive, textile, among many others, being the most used driving force due to its low cost, robustness and efficiency.

Electric motors have a considerable share of the world's energy demand, with those operating in the industry responsible for the consumption of 60–70% of the world's electricity generated [1]. These operate below 60% of their rated load due to oversized installations, which ultimately increases the waste of electrical energy [2].

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 127–136, 2018.
https://doi.org/10.1007/978-3-319-95450-9_10

The commercialization of the IMs intensified exponentially with the advent of the Variable-Frequency Driver (VFD), since it was used in applications that were previously made only by motors of direct current, due to the advantage of being relatively simple its speed control. The monitoring of IMs in the industry is still almost always done in large machines, however it is estimated that 98% of the motors in service have a power of less than 200 HP [3].

The search for speed estimation methods in electric motors has been studied by several researchers, in virtue of their great importance in monitoring, fault prevention, efficiency estimation and process control, in which speed estimation with good precision and accuracy is necessary.

For direct speed measurement in IMs, there is a good range of speed transducers on the market such as; tachometers, encoders, tachogenerators, resolvers, among other types, all with good precision and accuracy. However, thanks to the difficulty of access, it is not always possible to couple such instruments to the motor shaft, especially in motors allocated in industrial operations. Another relevant aspect is the cost of these instruments, which makes their applications unviable if the number of machines is high.

The sensorless measurement method has been shown to be an attractive and efficient alternative when compared to other speed estimation methods. Such a technique consists in estimating speed without the need to couple measuring instruments to the motor shaft. This extends its application to difficult to reach motors, where the conventional direct measurement method would be impractical.

The method of slot harmonics is a sensorless technique that uses the signature of the IM phase current spectrum and the search for harmonics generated by the motor protrusions - which arise from the rotor slots and their eccentricity - to estimate the speed of the motor [4]. This is one of the most widely used methods to estimate the speed in induction motors non-invasively [5–7]. The Fast Fourier Transform (FFT) can be used to obtain the frequency spectrum, however other transforms can also be used (e.g., Wavelet transform, Hilbert transform). The choice of which transform to use will depend on the type of application and the signal to be acquired.

The main disadvantage of the slot harmonics method is the need for a high number of samples to obtain a satisfactory spectral resolution, which requires high processing costs. Furthermore, the susceptibility to noise, whether caused by mechanical or electrical means, makes the method not suitable for estimating the speed in real time.

Another innovative technique for speed estimation in IMs is to use the torque in the air gap (Air-Gap Torque - AGT). The method was initially proposed by [4] and consists of obtaining a direct relation between the desired angular velocity and the already known AGT. A linearization relationship is found by tracing a line between the point which comprises the synchronous speed and zero torque and the point with the nominal speed and torque. It is possible to estimate the speed with a good accuracy from this relation and the equation of the straight line; however it is necessary to measure the input power in order to estimate the torque in the motor air gap. This may lead to delays in the estimations, although much lower than method of the slot harmonics.

In spite of the fact that the methods mentioned above are a considerable advance in speed estimation in induction motors, they still require the acquisition of the power of the motor and a methodology of data processing to reach an estimated value, which demands operational cost and specific sensors.

This paper proposes a new approach to speed estimation in three-phase induction motors, based on the principle of acoustic signal acquisition and analysis through an embedded system.

2 System Description

2.1 Test Bench

It was designed and built an experimental workbench consisting of two motors: a three-phase induction motor and a DC motor, the latter functioning as a load simulator. The configuration of this test bench allows to apply a known torque, constant or not, to the axis of the three-phase induction motor, which in turn can be actuated in two different ways: conventional (direct, through the electric grid) or via VFD.

The experimental workbench for conducting measurement tests and estimation of speed, torque and efficiency is shown in Fig. 1. It consists essentially of a three-phase induction motor (manufacturer WEG, model W22 Plus, nominal power 4.9 HP, nominal speed of 1725 rpm) (1) that has as function to provide an action torque through its axis to the sets attached to its front; one bearing support with two bearings (2); a torque transducer (3); a DC motor (manufacturer Varimot, model 132S, nominal power 7.4 HP, nominal speed of 1800 rpm) (4), which generates a braking torque on its axis, by means of the direct voltage obtained by rectifying the alternating voltage of 220 V of the electric grid.

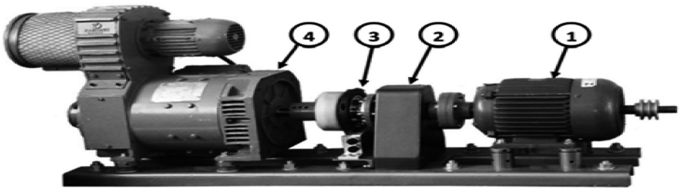


Fig. 1. Workbench for torque tests on three-phase induction motors [4].

2.2 Algorithm for FFT Calculation

The Fourier Transform (FT) is a tool for representing an aperiodic and time-continuous signal in terms of its frequency components, thus providing a spectral representation of the signal [8]. In a similar way, the Discrete Time Fourier Transform (DTFT) is used to represent a discrete-time aperiodic signal by means of its frequency components [9] which also leads to a spectral representation of the signal. According to [10] the main difference between FT and DTFT is that the FT represents signals with continuous sine-wave or exponential functions, while the DTFT uses sine-wave or exponential-time functions in discrete time.

The DTFT of a sequence x(n) is defined as [11]:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x(n)e^{-j\omega n} \quad (1)$$

As $X(e^{j\omega})$ is constructed only of complex exponential functions of periodicity 2π, it also presents this periodicity, that is, $X\left(e^{j\omega}\right) = X(e^{j(\omega+2\pi)})$.

However, while DTFT is an important tool for the analysis of discrete signals, it has certain practical limitations when used as a computational tool. A disadvantage is that a direct computation of $X(e^{j\omega})$ using the definition requires an infinite number of floating-point operations, this being aggravated by a second disadvantage in the computational aspect, the fact that the transformation itself must be calculated in an infinite number of frequencies [12].

In order to make the calculation of the DTFT possible in a computer, it is necessary to choose a finite number of frequency points, which is equivalent to sampling the FT at a certain number of points. Assuming some considerations about the interval in which the spectrum is effectively considered and that the acquisition process of the samples digitizes the relevant portion of the continuous signal by T_0 seconds, the next step is to assume that a periodic signal x(n) is cascaded from the N data samples acquired with the duration T_0 repeatedly [13]. In this way, the coefficients of the FT are determined using N data samples of a period, relating them by means of a multiplication, as explicit in (2).

$$X_k = \sum\nolimits_{n=0}^{N-1} x(n)W_N^{kn} \quad (2)$$

The expression in (2) can be understood as the Discrete Fourier Transform (DFT) of a sequence $x(n)$, $0 \leq n \leq N-1$. The factor W_N, called twiddle factor is defined as:

$$W_N = e^{\frac{-j2\pi}{N}} = \cos\left(\frac{2\pi}{N}\right) - j\mathrm{sen}\left(\frac{2\pi}{N}\right) \quad (3)$$

The DFT calculation, as expressed in (2), contains redundant complex products, and such replicates of these products can be eliminated to produce a faster execution [14]. In 1965, mathematicians Cooley and Tukey presented a rapid technique for calculating DFT [15] which became known as Fast Fourier Transform (FFT).

The computational effort of a DFT can be calculated by:

$$CE = N^2 \quad (4)$$

Where N is the number of elements of the signal under analysis. Thus, an 8-position vector would require the computation of 64 complex multiplications. If the vector of

N positions is divided into two parts, one referring to the even indexes and another to the odd indexes, the DFT of a signal can be written as [16]:

$$X_k = \sum\nolimits_{n=0}^{N/2-1} x(2n)W_N^{kn} + \sum\nolimits_{n=0}^{N/2-1} x(2n+1)W_N^{kn} \tag{5}$$

The calculation for the computational effort referring to the operation described in (5) will now not refer to the original length of the vector (N), but to twice that length. Therefore, we have:

$$EC = 2\left(\frac{N}{2}\right)^2 = 2\frac{N^2}{4} = \frac{N^2}{2} \tag{6}$$

It can be observed that dividing the sample in terms of even and odd indexes results in a 50% reduction of the computational effort to calculate a DFT. The main idea is that the process continues until there are only two values for the DFT computation. This operation is called decimation in time [16] and is illustrated for a better understanding in Fig. 2.

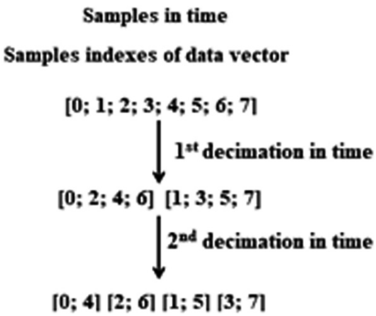


Fig. 2. Illustration of the decimation in time process.

The decimation in time performs the ordering of the samples at the input of the data in the time domain and it was implemented by the routine called bit reverse, which is illustrated in Fig. 3. Basically, the bit reverse operation consists of representing the indexes relative to the positions of the samples in a binary number vector and then doing the inversion in the order of the elements.

After the reordering step of the samples, the next step is to proceed with the operation known as Butterfly. The Butterfly uses the symmetry property of the samples present in the FFT logic [16], so that it allows a high computational gain. Figure 4 shows the butterfly process basic scheme.

For the evaluation of the algorithms developed for the analysis of the signals a comparison was made between the performance presented by them and the result of MATLAB® software, a widely used and accepted tool in the scientific environment. A sine wave with fundamental frequency of 1280 Hz was generated iteratively within

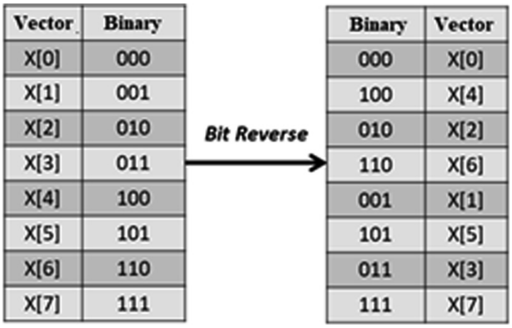


Fig. 3. Reorganization of a vector by bit reverse.

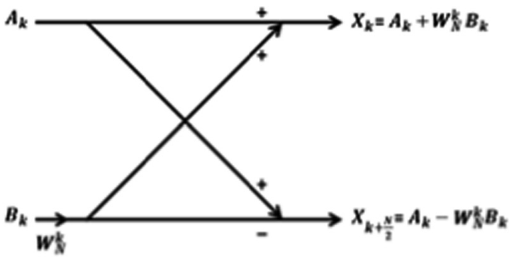


Fig. 4. Butterfly scheme illustration.

the FFT calculation code and then processing was performed normally. The sampling rate adopted was 40.960 Hz and the number of samples of the signal was 4096. In Fig. 5 the results of the algorithm running in MATLAB® and Arduino DUE, respectively, can be observed.

The error was calculated as the difference between the output value of MATLAB® and Arduino DUE. The mean value of the error was 1.0794×10^{-7}. The value of the standard deviation for the error signal was 36.568×10^{-7}. These measures were considered satisfactory for the research applications, since they reflect a good performance of the developed system. It is worth noting that the error occurred only with respect to the amplitude (dimensionless units), not identifying differences between the signals in relation to the information in the frequency.

2.3 Acquisition and Processing System

The embedded system used for acquiring and processing the sound signals used in the research is shown in Fig. 6. The system consists essentially of the development platform, Arduino DUE and the electret condenser microphone CMA4544PF-W.

Arduino DUE has high computational capabilities, as it has a 32 bits arm microcontroller, the Atmel SAM3X8E ARM Cortex-M3, which makes it a fast processing board. By prior configuration of its registers and internal timers, the acquisition frequency can be set to the value of 44.1 kHz.

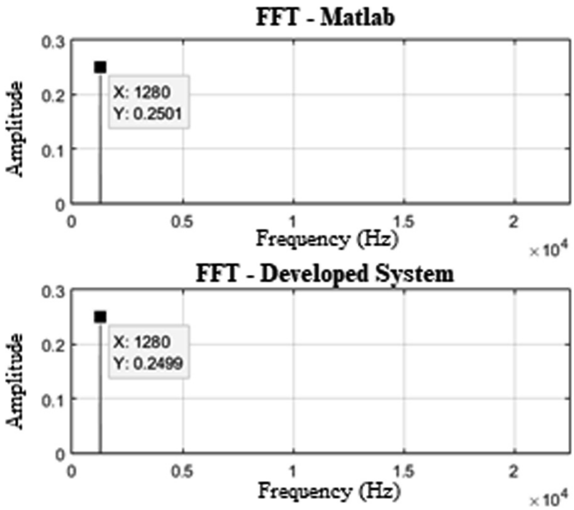


Fig. 5. FFT results – Matlab vs. proposed system.

Fig. 6. Embedded system for speed estimation by means of sound analysis.

The CMA-4544PF-W is an omnidirectional electret condenser microphone with sensitivity of −44 dB and with frequency of operation of 20 Hz to 20 kHz. Its frequency response curve (Fig. 7) shows to be quite stable, characterizing it as a good mechanism of transduction.

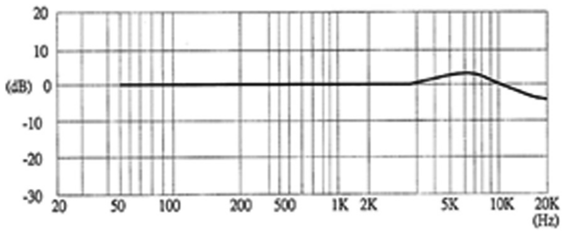


Fig. 7. Frequency response curve of the CMA-4555PF-W microphone.

Audio files are stored following the standard WAVE format, which ensures the storage of audio without any type of compression. Files are saved to a memory card so they can be viewed and analyzed. With the sound signal, the system proceeds with the speed identification, by detecting the peak of greater amplitude in the FFT (without considering its harmonics).

To measure the speed taken as reference for purposes of comparison with the proposed method, a tachometer from the manufacturer Minipa model MDT-2238B (Fig. 8) was used, which operates in two ways: by contact and by photo detection. Because it is practical, safe and has a smaller measurement failure, the speed was measured by the optical mode. In optical-read mode, the axis rotation speed can be measured in the range of 2.5 to 100,000 rpm with a detection distance between 50 to 500 mm and a resolution of 1 rpm (above 1,000 rpm in optical mode and an accuracy of ±0.05% reading +1 digit).

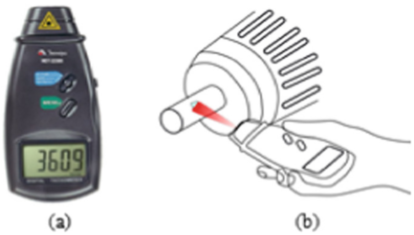


Fig. 8. Digital tachometer: (a) model MDT-2238B used at the research; (b) optical mode of operation.

3 Experimental Results

Using the controllable torque test bench, the necessary signals were acquired to verify the torque-to-speed ratio of the MIT and thus validate the speed estimation from the sound analysis through the embedded system. A torque scale ranging from 0 to 24 Nm was applied. During this interval, nine torque measurements and their respective speed were performed, as shown in Table 1.

Table 1. Torque, measured speed and estimated speed by the embedded system.

Torque (Nm)	Tachometer (rpm)	Embedded system (rpm)	Relative error %
0	1797	1795.716	0.071
3	1790	1789.59	0.022
6	1781	1781.634	0.035
9	1771	1773.06	0.116
12	1763	1762.038	0.054
15	1752	1751.634	0.020
18	1742	1740	0.114
21	1730	1728.978	0.059
24	1718	1716.732	0.073

The three-phase induction motor used has a power output of 4.9 HP and 4 poles, with a synchronous speed of 1800 rpm at 60 Hz with a slip of around 0.166%, considering the losses of mechanical or electrical origin negligible.

The torque vs. speed graph could be obtained (Fig. 9), allowing to verify that the torque and speed are inversely proportional and the relative error between the two curves presenting an average value of 0.0614 with a deviation of 0.038.

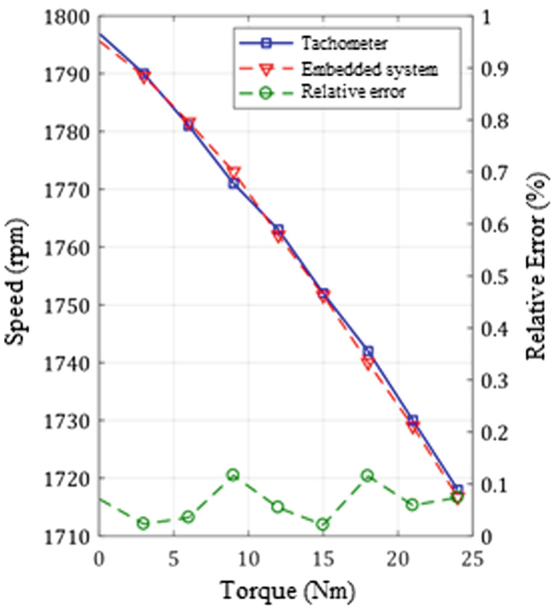


Fig. 9. Measured value vs. estimated value and relative error between the two measurements.

4 Conclusions

The embedded system presented minimal errors compared to the standard measurement system using the tachometer, which allows validating the embedded system for the application from which it was developed.

The method of speed estimation through the sound analysis was applicable, functional and without any type of motor invasion, which allows to be applied in the most varied circumstances.

The embedded system can be extended to other areas of interest and scientific relevance, including applications in analysis and prevention of failures in bearings and bearings supports in an industrial environment or in any mechanism that involves continuous rotary movement under normal working conditions.

Improvements to fix some of its limitations can be inserted, such as using an array of microphones and algorithms to identify different sources of sound emission, which would allow performing the measurement of speed simultaneously in several IM with only one device on industrial environments.

References

1. Kotak, V., Jaiwal, N.K., Patel, S.N.: Improvising strategies for efficiency of IE 4 SCIM 2.2 kW trough simulation. In: Proceedings of the International Conference on Electrical, Electronics and Optimization Techniques (ICEEOT), India, p. 3932 (2016)
2. Lu, B., Habetler, T.G., Harley, R.G.: A nonintrusive and in-service motor-efficiency estimation method using air-gap torque with considerations of condition monitoring. IEEE Trans. Ind. Appl. **44**(6), 1666–1674 (2008)
3. Lu, B.: Energy usage evaluation and condition monitoring for electric machines using wireless sensor networks. Doctoral Dissertation of the Georgia Institute of Technology (2006)
4. Adissi, M.O.: Study and development of energy efficiency monitoring system for electric motors in industry. Doctoral thesis of the PPGEM/UFPB (2015)
5. Blasco-Gimenez, R., Asher, G.M., Sumner, M., Bradley, K.J.: Performance of FFT-rotor slot harmonic speed detector for sensorless induction motor drives. IEE Proc. Electr. Power Appl. **143**(3), 258 (1996)
6. Hurst, K.D., Habetler, T.G.: A comparison of spectrum estimation techniques for sensorless speed and detection in induction machines. IEEE Trans. Ind. Appl. **33**(4), 898–905 (1997)
7. Roque, L.A.A.M.: Estimation of the induction motor speed through the stator current sinusoidal approach algorithm. Masters dissertation of the PPGEE/UNIFEI (2015)
8. Smith, S.W.: The Scientist and Engineer's Guide to Digital Signal Processing, 1st edn. California Technical Publication, USA (1997)
9. Haykin, S., Veen, V.: Signals and Systems, 1st edn. Wiley, New York (1999)
10. Lathi, B.P., Green, R.A.: Essentials of Digital Signal Processing, 1st edn. Cambridge University Press, New York (2014)
11. Mertins, A.: Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications, 1st edn. Wiley, Chichester (1999)
12. Schilling, R.J., Harris, S.L.: Fundamentals of Digital Signal Processing Using MATLAB, 2nd edn. Cengage Learning, Stamford (2011)
13. Tan, L., Jiang, J.: Digital Signal Processing – Fundamentals and Applications: MATLAB Examples, 2nd edn. Elsevier, New York (2013)
14. Stearns, S.D., Hush, D.R.: Digital Signal Processing with Examples in MATLAB, 2nd edn. CRC Press, Boca Raton (2011)
15. Cooley, J.W., Tukey, J.W.: An algorithm for the machine calculation of complex Fourier Series. Math. Comput. **19**(90), 297–301 (1965)
16. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C: The Art of Scientific Computing, 1st edn. Cambridge University Press, Cambridge (1988)

Analysis of Illumination Lamp's Performance by Retrofit at University Building

Shoaib Shaikh[1(✉)], Nareena Soomro[2], Fahad Razaque[2], Safeeullah Soomro[3], Najeebullah Shaikh[1], and Ghulam Abid[1]

[1] Hamdard University, Karachi, Sindh, Pakistan
skshaikh@outlook.com, Najeebullahshaikh09@gmail.com, engrabid1246@gmail.com
[2] Department of Computing, Indus University, Karachi, Sindh, Pakistan
nainee_soom@yahoo.com, fahad.indus1337@gmail.com
[3] College of Computer Studies, AMA International University, Salmabad, Kingdom of Bahrain
s.soomro@amaiu.edu.bh

Abstract. The research is comprised of possible energy saving, life cycle cost analysis and settlement period of the lighting system in the university building using of the fluorescent lamp which is an analysis of the dataset lighting system. Cost profit analysis of retrofit with extra efficient illumination scheme in terms of probable energy saves, the study of life cycle expenditure and settlement phase was steered. Comparison of existing such as the luminous lamp (IL), fluorescent lamp (FL) and compact fluorescent lamp (CFL), and retrofitting of illumination system based on power utilization is offered. LED (Light emitting diodes) based lamp technology is comparatively with conservative luminous and expulsion lamps. The investigation concluded that considerable quantities of energy will be saved by using an energy efficient illumination system as well as cost and to some extent reduces emission. The evaluation result illustrates that with present technology, FL's and LED lamp are used for utility as well as for university building.

Keywords: Illumination lamps · Life Cycle Assessment · Power consumption Cost benefit analysis

1 Introduction

Lighting system has been retained vital role to confirm efficiency, ease, and security of residents in university buildings. Consequently, correct designing of lighting system has become necessary to get desired illumination with least quantity of electricity use. Version of lighting used for able to one-third of intake university building electricity. International Energy Agency (IEA) consists of worldwide consumption of electricity for lighting was distributed about; residential 28%, service 48%, industrial 16%, street and extra lighting 8%. It was expected that energy lessening able to residential 27% and commercial 30% area might be attained by substituting to power effectual machineries which were development of power electrical based energy redeemable policies [9].

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 137–152, 2018.
https://doi.org/10.1007/978-3-319-95450-9_11

The terrific growth in unit electrical energy value was also one of causes for infiltration of such devices in existing power scheme. Though, extensive use of policy may possibly have a few injurious belonged on the quality of power. Capabilities require displayed that prevention was more economical next knowing point, instead of finding the solution to the next future problem [10]. Illumination is critical to students functioning in that it allows seeing things and performing activities. However, it is also important because it affects students and faculty beings psychologically and physiologically [20]. Almost overall domain as engineering, natural sciences, information technology, ICT, economics, trade, electronic commerce, atmosphere, health-care, and living science were roofed [21]. As period curriculums perception has become to a required assignment for learners in addition to stakeholder in software development and provides solutions to problems, that pictorial presentation in finest method is benefit to better understand it [22]. In the middle of dissimilar procedures for the decrease in power consumption, energy's standby wasteful lamps with energy well-organized lamp was one of the procedure. The investigation was revealed that illumination constitutes important portion in country's overall power consumption.

Replacement of Incandescent light with energy well-organized lamps decreases illumination power consumption besides helps in saving of energy throughout summer period by lowering icy necessities in buildings [11]. Illumination portion was about 23% of overall power consumption, according to the report of Swedish energy agency [12]. LED was ecological kindly, dissimilar compact fluorescent lamp (CFLs) because of inside mercury [2] which was poisonousness illustrious and produces the staid durability problem. The research was the selection of cost-effective and efficient lights in Indus university by comparing different factors of lights as efficacy, power

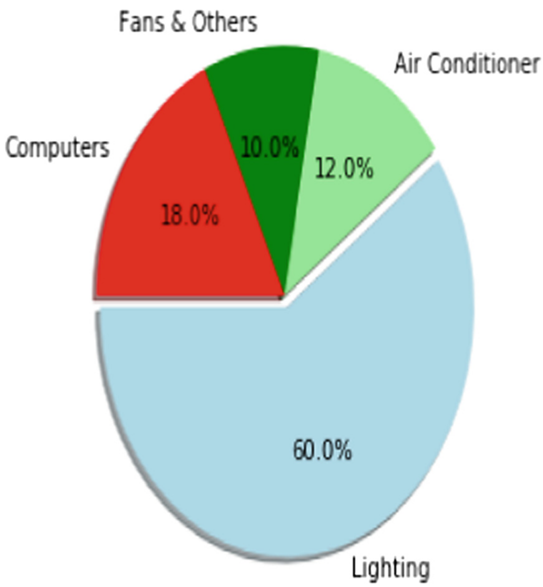


Fig. 1. Using different types of distributed loads in university building.

consumption, CRI, color temperature, life time, etc. Figure 1 shows that university building was utilized different loads such as 60% lighting and 18% computers.

Figure 2 shows that full diode bridge rectifier to transform AC into DC which was followed capacitor. It was delivered DC connection current, voltage source for the LED load. The constant source of current was used instead of voltage converter in LED ballast through minimalize voltage variant affected.

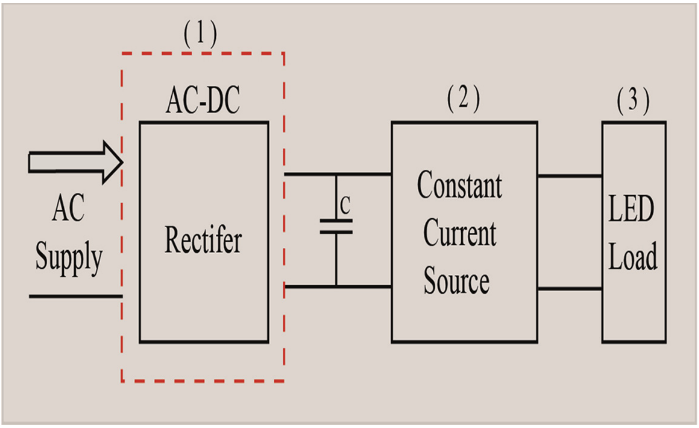


Fig. 2. Typical LED ballast

2 Methods and Material

2.1 Electrical Energy Consumption

Current fluorescent was expected for the entire department based on survey evaluated. EC (total energy consumption) of illumination scheme was resolute by N (number of lamps), W (power) recycled per fixture and operating period (OH) of lighting. Annual energy consumption for the current illumination scheme was designed based on the method as [8]:

$$\text{Electricity Consumption (EC)} = \frac{N \times W \times OH}{1000} \tag{1}$$

Electrical energy consumption is measured in Watt-time (watt-hours) which is the measure energy. ES (Energy saving) was the difference between current energy consumption (EC Current) and retrofit lighting (EC Retrofit) system.

$$\text{Energy Saving (ES)} = \text{EC Existing} - \text{EC Retrofitting} \tag{2}$$

BS (Bill saving) was calculated by multiplying ES with electricity tariff (ET).

$$\mathrm{BS} = \mathrm{ES} \times \mathrm{ET} \tag{3}$$

Operating cost (OC) was wanted on behalf of innovative retrofit scheme, which was N (total number of lamps), power consumption (in watts), OH (operating hours) and ET (electricity tariff) and considered as:

$$\mathrm{OC} = \mathrm{N} \times \mathrm{W} \times \mathrm{OH} \times \mathrm{ET} \tag{4}$$

PWF (Present worth factor) valued with upcoming money flow to be received in order to acquire existing value. Current value aspect was designed as

$$PWF = \sum_{1}^{N} \frac{1}{(1+r)^{t}} = \frac{1}{r}\left[1 - \frac{1}{(1+r)^{N}}\right] \tag{5}$$

PAY (payback period) calculated the quantity of time wanted to improve extra investment (increased cost ΔPC) on productivity development through lower operating costs which was solved by:

$$\Delta\mathrm{PC} + \sum\nolimits_{1}^{\mathrm{PAY}} \Delta\mathrm{OC}_{\mathrm{t}} = 0 \tag{6}$$

The payback period was established by interposing among two ages when above expression modifications symbol. If operating cost was constant as a solution [13]:

$$\mathrm{PAY} = -\frac{\Delta\mathrm{PC}}{\Delta\mathrm{OC}} \tag{7}$$

2.2 Comparison of Lamps Illuminations

The comparison was ended on the foundation of financial and technical grounded. Clients have been normally not knowledgeable of illumination design vocabularies and lamp physiognomies. Clients were choice lamp on basis of consumption's power (Watts), instead of considered how often LUX unit of illumination, it was received in a given region. Some terms have been discussed with the comparison as Lumens, Efficacy, CRI, Illumination and THD [6–8]. Lumen was a unit of luminous flux, which was the dimension of the total amount of brightness of light, radiated from the source of light that was influenced according to glow function human eyes sensitivity to different wavelengths. It was just based on the output of light rather than energy consumption, thus represent more accurate measurement unit. Lamp's conformist was the decrease in early lumens to B50 (50%) and lamp's life was usually measured in hours.

Illumination (LUX Level): It was measured brightness level in the particular area that was lumen's total number dipping on the specified zone and was measured for example lumens per meter square.

Total Harmonic Distortion (THD): It was a portion of RMS (root mean square) harmonic worth contented of blinking measure to RMS value of the central component. Total harmonic distortion (THD) was used to specified harmonic contented in distorted voltage and current, correspondingly as THDV (Total harmonic distortion Voltage) and THDI (Total harmonic distortion current).

Color Rendering Index (CRI): It was the facility of sunny basis to render color indeed with no misleading tone seen below occupied spectrum warmer and its variety from poor quality(0) toward near to natural daylight(100) [5]. The Fig. 3 depicts that CRI of Incandescent and halogen is near to natural daylight and white LED is poor quality.

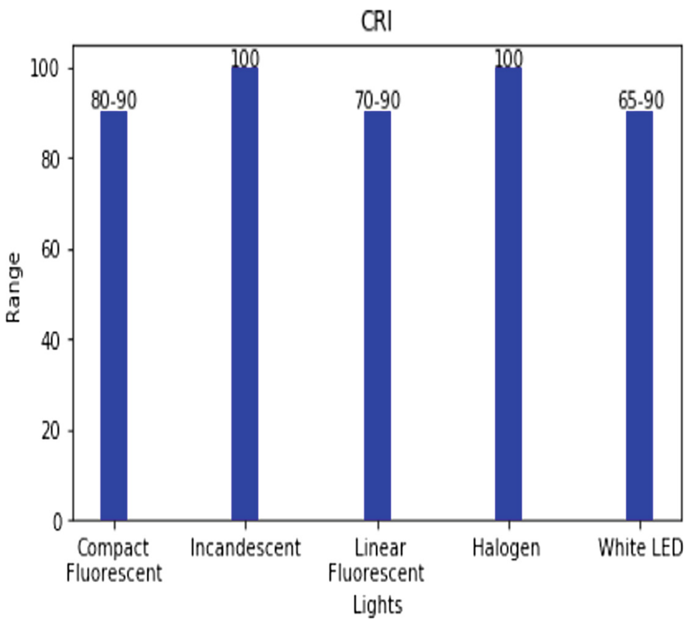


Fig. 3. Comparison of different lamps from color rendering index

Efficacy: It measured of how much lumen was emitted for given input power of electrical, which is precise in lumens per watt. Lamp's efficacy specified efficiency of lamp single. Energy consumption of regulator equipment such as transformers, ballasts, and further regulatory gear was not deliberate. The Fig. 4 shows that efficacy of linear fluorescent light is a high lumen per watt.

Color Temperature: It was the method to equate color of light from dissimilar kinds of lamps, which measured in Kelvin (K). It was frequently referenced as cool and warm light which was slightly blue and slightly orange, that evaluate by 5000 K as cool and 2700–3000 K as the warm light. Incandescent lamps and candles radiate warm, although sunshine and several fluorescent lamps emit cool. Figure 5 shows that color temperature of Compact Fluorescent, linear fluorescent light and White LED are warm light.

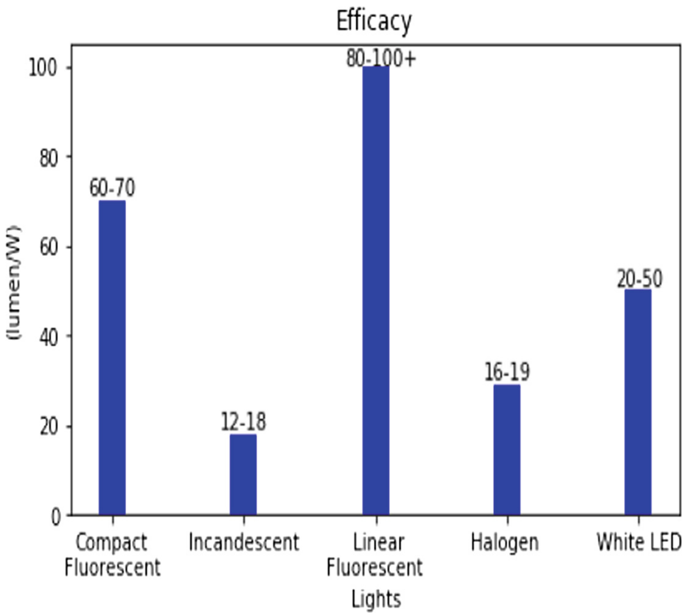


Fig. 4. Comparison of different lamps from Efficacy

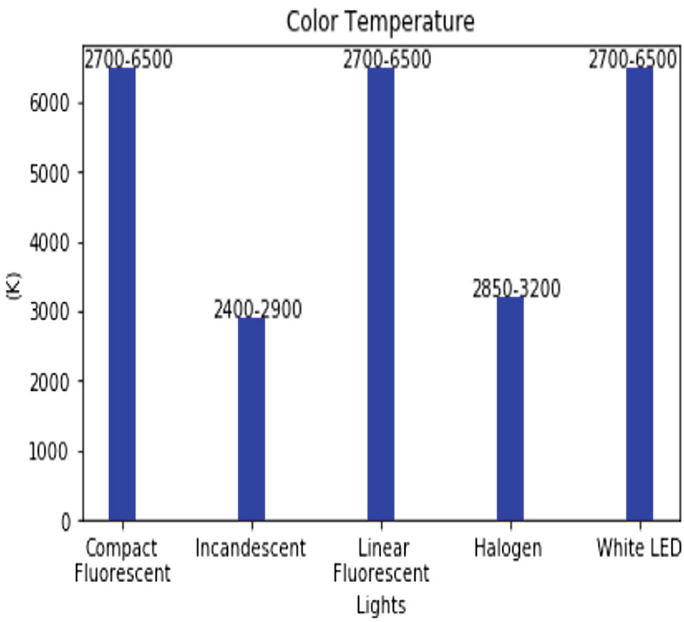


Fig. 5. Comparison of different lamps from color temperature (Color figure online)

Life Time: The middling life time of IL was typically about 1000 h which was much slighter than other illumination lamps. It was increased equal to 2000 h with addition, special unmotivated fumes within lamp tumbler. Conversely, release lamps have considerably upper life time than IL. LED package useful life was derived from lumens up keep until that was emitting of early output light L70 (70%) when it was quite high about 35×10^3–50×10^3 h. [14] Though, LED's life time luminaries was slighter than the life of LED because of the possible letdown in motorist, reflector. Now it was essential to differentiate between useful life and lifetime. The useful life of fixture referred to expected lumen upkeep of LED bases. Yet, fixture lifetime was associated with the consistency of gear's LED illumination fixture such that housing, wiring, electronics materials, connectors, seals. Whole LED luminaire lasts simply providing the critical element with shortest life exists [14]. Illumination investigators were stilled work to form some fresh procedure to estimation beneficial LED lifetime has taken all letdown instruments into deliberation [15].

Lamp Life: It was a pretentious using figure of on or off substituting maneuver. Frequent substituting maneuver lessens lamp's lifetime. Now halogen based ILs was unresponsive to substituting maneuver [15]. Though, in FL (fluorescent lamp), conductors misplace minute amount of covering each time lamp was exchanged on/off. Therefore ended stage of time, sufficient covering was scorched off and lamp miscarries to start. The motive, release lamps were not suggested where substituting rather repeated. CFL and LED lamps since equated to other lamps were affected via substituting maneuvers by reason of untimely breakdown in electric modules [16]. In EU standards for CFL and LEDs, the numeral of substituting maneuvers that illumination

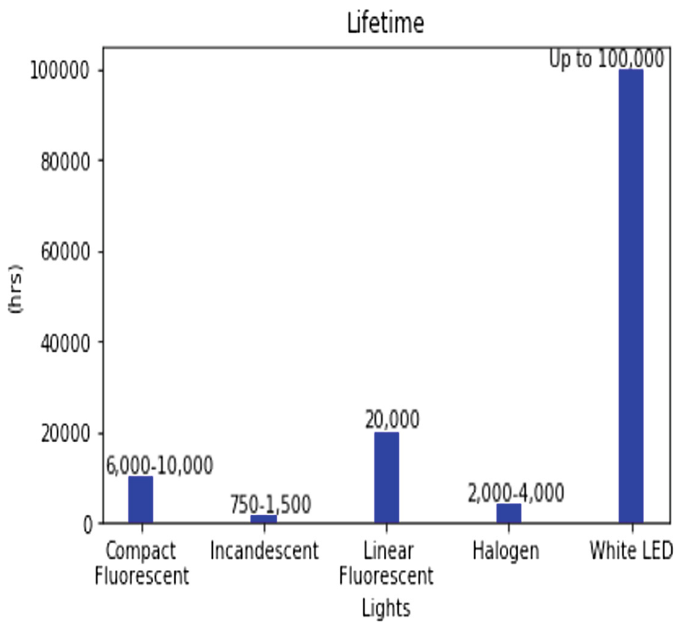


Fig. 6. Comparison of different lamps from lifetime

basis was weather beforehand untimely failure was larger than lamp life which in hours. Lamps claimed to endure recurrent substituting, numeral ought to be superiority than 60,000 buttons (on/off) cycles in standards of EU [17]. Figure 6 depicts that white LED is 100000 h.

Incandescent Light (IL): The incandescent light (IL) existence clean resistive load was not suffered by low power factor topics and so constantly current unity factor's power. It was not required a special electric driver for beginning purposes, hence no harmonic problem ascend. If the large number of low rating CFLs was used in giving position, the power factor necessity to be centrally rewarded [3]. LED lamps were modeled such as a voltage basis with low sequence resistance to boundary line present, which was a power factor usually much higher than discharge lamps. Energy's US Department vitality star program mandate least satisfactory power factors of 0.7 and 0.9 correspondingly for inland and profitable LED illumination [4]. Energy Saving Trust of UK was proven least a power factor of 0.7 with lengthy term impartial of 0.9 power factor aimed at integral LED bulbs [3].

3 Approach of Life Cycle Assessment

Life Cycle Assessment was a methodical approach that facilitates scholars to quantify ecological and sustainability influences across the variety of classifications for artifact ended the whole life cycle. It categorized and measured I/O (inputs, outputs) and ecological influences of a particular artifact on every phase of life cycle [1]. The overall technique for conducting an LCA was defined by ISO (International Organization for Standards) 14000 sequence. LCA's core stages according to ISO rules were goal and scope definition; IA (inventory analysis); impact assessment; and interpretation (Fig. 7).

3.1 Goal and Scope Definition

Goal's main aspects were an intentional application of study such as marketing, strategic planning, and product development; the purpose of study, for instance, to be published; intended viewers, comprised stockholders, managers, customers; and used as a relative analysis, whereby LCA outcomes were used to compare with other goods.

3.2 Inventory Analysis

Compiled and quantified of I/O for giving artifact scheme through the life cycle that contains procedural amounts for all pertinent unit procedures within system limitations. It was data excellence and processing steps that required activities to be finalized: data validation, linking data to unit procedures and to the functional unit.

3.3 Impact Assessment

It recognized and estimated extents and comparative significance of ecological influences arise from inventory investigation. I/O was allocated to influence kinds and

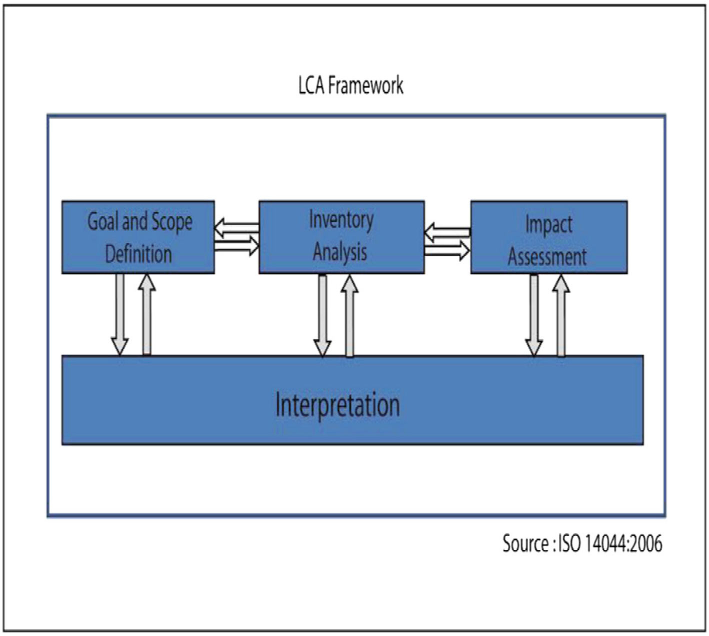


Fig. 7. Framework (ISO)

possible influences were quantified according to categorical factors which include: water, fossil fuels, chemicals, energy, and etc. by supply reduction, greenhouse gas emissions, land use, and water pollution. It was required elements have comprised as soon as conduct LCA for instance chosen of related influence kinds and classification.

3.4 Interpretation

Outcomes were assessed and checked to ratify that was reliable with goal of the study. As displayed in the figure, three other stages were linked to interpretation, which was a pivotal slice of the process and could lead to modifications in any fact of the process.

The valuation stage was attentive on augmenting reliability of study that comprises sensitivity checked on indecisions about data, calculations, allocation methods and assumptions. It comprises gap analysis, to ensure deficient part that has to be analyzed in classifying to meet the goal and scope of the study. Lastly, assessment contains a reliability test to make sure that methods and goal were gathered, for instance, data quality, time period, scheme boundaries.

4 Illumination's Life Cycle Assessment

US Department of Energy report, which synopses major findings of current appraisal on LCA further compared life cycle power utilization of LED lamps with IL and CFL in 2012 [18]. It represents 90% of entire lifecycle energy use on average from IL, CFL and LED lamps tracked by transport and manufacturing stages. Most of the doubts happen in the manufacturing stage of LED packaging supply to low as 0.1%, average as 6.6% and high as the 27% estimate specify by computing lifecycle energy consumption of LED lamps. The report covered energy in use was lifecycle ecological impact due to incandescent lamps as 60 W, CFL as 1.5 W and LED as 12.5 W. On behalf of comparison, the performance of LED lamp in 2012 was deliberated and proposed onward in 2017 with enhanced engineering approaches; performance and integrated circuit technology [19] which also included the impact of resource, soil, air and water.

Evaluated fifteen impact interest's measures transversely four lamp kinds deliberated, spider graphs were ready that was represented by spoke in the web, and comparative impact of all lamp variety were plotted on the diagram. Lamp category contains maximum influence of set analyzed describes scale symbolism by an external circle at the maximum distance from web center. Further products were next regularized to the distance from center signify severity of impact compared to the incandescent lamp. Folks sources with minimum impact were contained circle closer to the center and the maximum impact would be on web outer perimeter. The data plotted in the diagram was normalizing for the competence of illumination service measured in lumen hours.

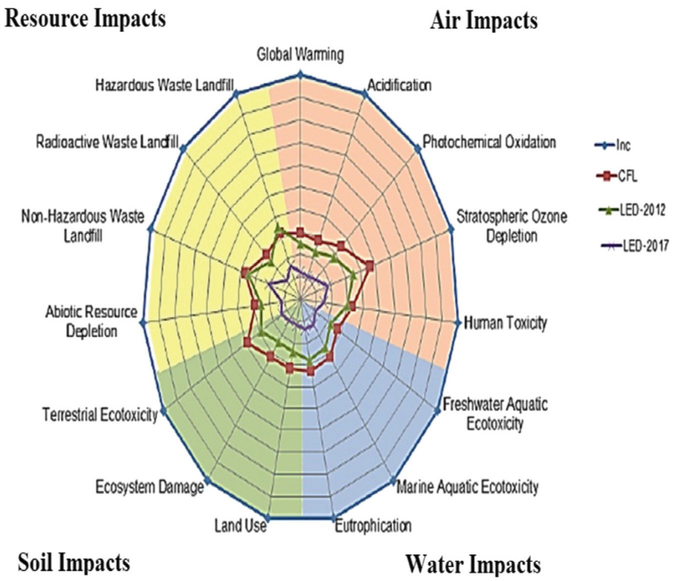


Fig. 8. LCA lamp's impacts analyzed comparative to incandescent

Figure 9 shows plots represent LED and CFL apparatus decrease fine within the outer circle, illustrated that IL was the uppermost impact per unit illumination service of everything lamps well thought-out. The discovery was not the purpose of the stuff content of single lamp; IL was the lowest mass and a minimum composite illumination system. It represented a low efficacy of a light source and resulted from big amounts of energy compulsory to produce light and various alternates were mandatory to length ranked life of LED lamp otherwise CFL. Producing an upper amount of electronic energy consumed per unit of light productivity reasons, considerable ecological impacts and outcomes in IL being more ecologically damaging across all fifteen impacts measured.

Although it was considerably lower impacts than incandescent, the dense fluorescent lamp was marginally further more damaging than 2012 inherently weight LED lamp against everything but harmful misuse landfill where mechanized of big aluminum warmth outcome used in LED lamp reasons collision to be a little larger for LED lamp than for CFL. The execution illumination source was estimated LED lamp in 2017, which obtain into version numerous potential enhancements in LED manufacturing, performance, and integrated circuit technology. Figure 9 indicates same results shown in Fig. 8, but diagram was adjusted to remove IL and deliver impacts comparative to CFL.

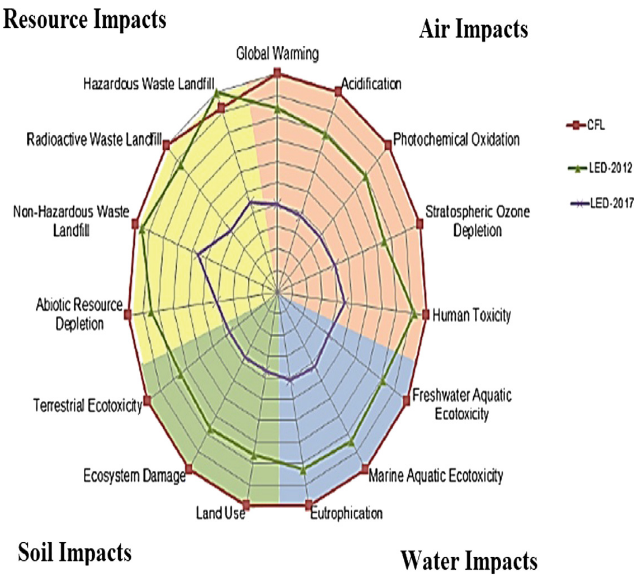


Fig. 9. LCA lamp's impacts analyzed comparative to CFL

5 Experimental Evaluations

In this research is a life cycle cost assessment to determine consumed lighting and numeral of possible retrofit of illumination scheme in the university building. Data is managed in all rooms of every block in the campus from the dataset which was contained numeral of illumination fixtures and operating hours. It was focused on fluorescent lamp in particular linear luminous, Compact fluorescent, Incandescent and Halogen then evaluate cost assessment which was a type of light cost effective and efficient. Data get from dataset was used to determine projected electrical energy investments and cost analysis of illumination retrofits. Figure 10 shows that the number of lamp, classroom, faculty room and corridor that have consumed compact fluorescent is 2454 and linear fluorescent is 2612.

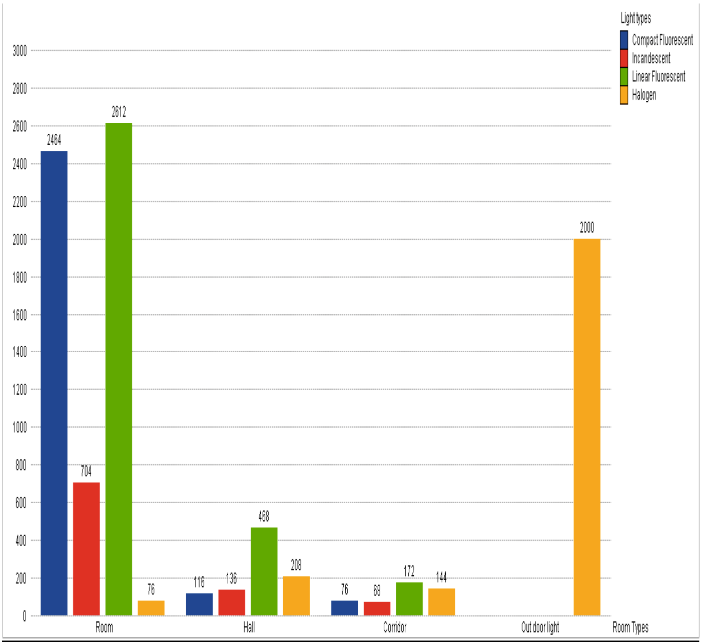


Fig. 10. A graph of different types of light used in university

Replacing each of four types of light lamp currently used at university building with an equivalent LED light lamp was effectively reduced power compulsory to illumination building, foremost to electric invoice investments over time. Table 1 considerable possible for investments using LED illumination lamp instead of Incandescent, Halogen, LFL, and CFL lamps which was maximum burning hours 100000 of LED lamps and measured formulas given as below which have been used in the table.

In research, 60 W/100 W (IL), 36 W (FL) and 26 W/18 W light was used for inside illumination. It velocity worth of overall lamps, establish from lamps' index was summarized in Table 1 illustrate LED lamps (12 and 30 watts) were also integrated into experiment result and recital data for IL, FL and CFL were experiencing.

$$n = N * \text{Cost/Lamp} \tag{8}$$

$$EC = OH * P \div 1000 \tag{9}$$

$$C = EC * R \tag{10}$$

$$LCC = C_{energy} * C_{lamp} \tag{11}$$

n = Number of lamp per cost
N = Number of lamp required for Operating Hours
EC = Energy Consumption (kWh)
OH = Operating Hours
P = Wattage Rating
C = Energy or Lamp cost
R = Tariff Rate
LCC = Life Cycle Cost

Table 1. Performance of IL, Halogen, LF and CF lamp and LED

Lamp	Burning hours	Watts	Bulb/lamp cost (Rs)	No. of bulbs	K-electric tariff rate (commercial)
Incandescent lamp	1500	100	40	908	Rs. 18
Halogen lamp	4000	60	80	2428	
LFL	20000	50	300	3252	
CFL	10000	25	200	2656	
LED lamp	100000	14	600	0	

Later mentioning burning hours of each lamp represented in Table 1 compared individually all lights with LED lamp. Table 2 shows the comparison of IL with LED lamp clearly shows that lifecycle cost of IL lamp is higher (216320) than LED lamp cost (33600) because running hours of 908 IL are equal to 14 lamps of LED. It is dangerous for the environment due to heating and green emission gases.

Table 2. Comparative of LED and IL lamp

Lamp	Burning hours	For 100000 h, no: of bulbs required	Cost of bulbs (Rs)	For 100000 h energy required (kWh) (100000 * W)/ 1000	Energy cost (Rs)	Life cycle cost (Rs)
IL	1500	908	36320	10000	180000	216320
LED	100000	14	8400	1400	25200	33600

Table 3 depicts the calculation and comparison of Halogen lamp with LED lamp has also limited users for usage of halogen lamp that higher (108000, 302240) in cost than LED lamp cost (25200, 84000) and burning hours (4000) of halogen are lesser

Table 3. Relative LED over Halogen lamp.

Lamp	Burning hours	For 100000 h, no: of bulbs required	Cost of bulbs (Rs)	For 100000 h energy required (kWh) (100000 * W)/ 1000	Energy cost (Rs)	Life cycle cost (Rs)
Halogen lamp	4000	2428	194240	6000	108000	302240
LED lamp	100000	98	58800	1400	25200	84000

than LED lamp (100000) burning hours. LED number of lamps required 98 lamps that are equal to 2428 Halogen Lamp by considering burning hours, in fact, have not economical for utilization purpose.

Table 4 depicts LF lamp which had led Incandescent lamp due to high (5000) energy consumption cost which is rarely used due to the higher cost of burning hours (20000) as compared to LED lamp 1400 energy consumption cost and burning hours (100000). LF lamp is 20000 Burning hours as the number of lamps required 3252 whereas LED lamp life time is 100000 h, as 651 numbers of lamp required.

Table 4. Comparative of LED and LF lamp

Lamp	Burning hours	For 100000 h, no: of bulbs required	Cost of bulbs (Rs)	For 100000 h energy required (kWh) (100000 * W)/ 1000	Energy cost (Rs)	Life cycle cost (Rs)
LF lamp	20000	3252	975600	5000	90000	1065600
LED lamp	100000	651	390600	1400	25200	415800

The evaluations of lamps calculated for CFL and LED lamp outcomes are depicted in Table 5 is a CFL lamp cost-effective due to burning hours and other factors that is the huge difference of cost such as 2656 CF lamps with the cost of 576200 can be equalized with 266 LED lamps having cost 184800 cannot feasible for normal use.

Table 5. Comparative LED over CF lamp

Lamp	Burning hours	For 100000 h, no: of bulbs required	Cost of bulbs (Rs)	For 100000 h energy required (kWh) (100000 * W)/ 1000	Energy cost (Rs)	Life cycle cost (Rs)
CF lamp	10000	2656	531200	2500	45000	576200
LED lamp	100000	266	159600	1400	25200	184800

Efficient light of source important factors mandated; minimum power consumption, high luminous efficacy and minimum life cycle cost that was resulted in concluded. Table 6 depicts that LED lamp life cycle cost (33600, 84000, 415800, and 184800) and power consumption is minimum (1400) that cost beneficial.

Table 6. Relative LED's advantages over IL, Halogen, LF and CF lamp

Lamp	Burning hours	For 100000 h, no: of bulbs required	Cost of bulbs (Rs)	For 100000 h energy required (kWh) (100000 * W)/ 1000	Energy cost (Rs)	Life cycle cost (Rs)
IL	1500	960	38400	10000	180000	216320
Halogen lamp	4000	2428	194240	6000	108000	302240
LF lamp	20000	3252	975600	5000	90000	1065600
CF lamp	10000	2656	531200	2500	45000	576200
LED lamp	100000	12	7200	1400	25200	33600
		98	58800			84000
		651	390600			415800
		266	159600			184800

6 Conclusion

This research was offered performance analysis of illumination lamps included incandescent lamp (IL), fluorescent lamp (FL), compact fluorescent lamp (CFL) and light emitting diodes (LED) lamps. Investigate the comparative impact of lamps towards power quality in terms of distortion was also accomplished. Present growth illustrated fabulous enhancement in performance of LED illumination and at the same instant decrease in cost growth. Comparison completed based on the retrofit of the power efficiency scheme extra outward. Life cycle cost of LED lamp is reduced that considered inexpensive compared to standard and further substitute technology. Because energy consumes LED reduced about half of CFL and LF. By retrofitting the existing system with LED did, power investments can be saved in the university building.

References

1. Linhart, F., Scartezzini, J.L.: Minimizing lighting power density in office rooms equipped with anidolic day lighting systems. Sol. Energy **84**(4), 587–595 (2010)
2. Pérez-Lombard, L.: A review on buildings energy consumption information. Energy Build. **40**, 394–398 (2008)
3. CIE, LED Bulb placement Meeting New Challenges. Components in Electronics (2012)

4. ECN Magazine: Power Factor Correction Techniques in LED Lighting (2011)
5. Philips, Technical Data (2012)
6. IEC: Electropedia: The World's Online Electro technical Vocabulary. International Electro technical Commission (2012)
7. Laughton, M.A., Warne, D.J.: Electrical Engineer's Reference Book. Newnes Publications, Boston (2002)
8. Philips, Philips Lamps & Lighting Electronics Catalogue. In: Philips (ed.) (2010)
9. Trifunovic, J., Mikulovic, J., Djurisic, Z., Djuric, M., Kostic, M.: Reductions in electricity consumption and power demand in case of the mass use of compact fluorescent lamps. Energy **34**, 1355–1363 (2009)
10. Watson, N.R., Scott, T.L., Hirsch, S.: Implications for distribution networks of high penetration of compact fluorescent lamps. IEEE Trans. Power Deliv. **24**, 1521–1528 (2009)
11. Saidur, R.: Energy consumption, energy savings, and emission analysis in Malaysian office buildings. Energy Policy **37**, 4104–4113 (2009)
12. Bladh, M., Krantz, H.: Towards a bright future? Household use of electric light: a microlevel study. Energy Policy **36**, 3521–3530 (2008)
13. Turiel, I., Chan, T., McMahon, J.E.: Theory and methodology of appliance standards. Energy Build. **26**(1), 35–44 (1997)
14. DoE, Lifetime of White LEDs. Solid-State Lighting Program: U.S. Department of Energy (2011)
15. DoE, LED Luminaire Lifetime: Recommendations for Testing and Reporting. Solid-State Lighting Program: U.S. Department of Energy (2011)
16. EUROPA, How to Read the Packaging. Europa—European Commission (2012)
17. Key World Energy Statistics. International Energy Agency, Paris, France (2007)
18. DoE, Life-Cycle Assessment of Energy and Environmental Impacts of LED Lighting Products (Part1: Review of the Life-Cycle Energy Consumption of Incandescent, Compact Fluorescent, and LED Lamps). Solid-State Lighting Program: U.S. Department of Energy (2012)
19. DoE, Life-Cycle Assessment of Energy and Environmental Impacts of LED Lighting Products (Part2: LED Manufacturing and Performance). Solid-State Lighting Program: U.S. Department of Energy. (2012)
20. Soomro, N., Soomro, S., Alansari, Z., Belguam, Z.M., Khakwani, A.B.K.: Development of UMLS based health care web services for android platform. Sindh Univ. Res. J. Sci. Ser. **48**(4), 5–8 (2016)
21. Soomro, S., Alansari, Z., Belgaum, M.R.: Path executions of java bytecode programs. In: Saeed, K., Chaki, N., Pati, B., Bakshi, S., Mohapatra, D.P. (eds.) Progress in Advanced Computing and Intelligent Engineering. AISC, vol. 564, pp. 261–271. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6875-1_26
22. Soomro, S., Alansari, Z., Belgaum, R.M.: Control and data flow execution of Java program. In: International Conference on Engineering Technologies and Social Sciences, ICETSS (2016)

Physarum Inspired Model for Mobile Sensor Nodes Deployment in the Presence of Obstacles

Abubakr Awad(✉), Wei Pang, and George Coghill

School of Natural and Computing Sciences,
University of Aberdeen, Aberdeen, UK
{abubakr.awad,pang.wei,g.coghill}@abdn.ac.uk

Abstract. Mobile wireless sensor networks (Mobile-WSN) are useful in harsh environments due to the presence of obstacles and/or dangerous for sensors to be deployed deterministically. In this paper, we proposed a Physarum inspired autonomous, model for dynamic deployment of sensor nodes where multiple Physarum (as representation of sensors) will compete over food resources (interest points) based on chemo-attraction, and repulsion forces exerted by competing Physarums and obstacles. Our simulation results have demonstrated the high coverage performance of the model with minimal move overhead in the presence of obstacles with the least number of sensors.

Keywords: Physarum polycephalum · Hexagonal cellular automaton
Mobile sensor network · Deployment · Coverage · Obstacles

1 Introduction

In mobile wireless sensor network (Mobile-WSN) the sensors are deployed randomly to gather the information from the environment. In different applications of Mobile WSN, it is not possible to deploy the sensors deterministically. After initial random deployment sensors are required to disperse autonomously without central control to maximize the coverage and re-establish the connectivity of the network [8]. Mobile WSNs are to collect ground data for various purposes such as such as battle field monitoring, bio-environmental surveillance, earthquake observation, and wildlife reservoir [4,7,9].

Environmental obstacles (building, lakes, mountains, ...) can form holes in the network, creating sets of isolated nodes and leaving uncovered areas. Sensor coverage and connectivity problems were investigated thoroughly, and several techniques were proposed with various capability and limitations [2,12].

Physarum may not have brain, but they are capable of solving many significant problems. Physarum senses gradients of chemo-attractants and repellents and forms a yellowish vascular network which expands up to tens of centimeters in search of nutrition. The Physarum foraging behavior consists of two

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 153–160, 2018.
https://doi.org/10.1007/978-3-319-95450-9_12

simultaneous self-organized processes of expansion (exploration) and shrinkage (exploitation) [11].

Many mathematical models have been proposed to simulate Physarum foraging behavior [1,5,6,10,14]. Using these models Physarum Polycephalum is capable of solving many NP-hard problems, such as finding the shortest path in directed or undirected network [16], simulating transport network [15]. Physarum can sense its environments as in maize labyrinth model [10], the applied approach allow mobile nodes to navigate in unknown environments avoiding obstacles.

Our aim is to use unconventional computational power of Physarum polycephalum to provide an energy aware distributed self-deployment algorithm for dynamic deployment of sensor nodes after initial random deployment to avoid obstacles, enhance coverage and re-establish the connectivity of the network. In this model multiple Physarum (as representation of sensors) will be competing for target points (chemo-attractants as food), and avoiding boundaries and obstacles (repellent as light). Each Physarum will consider all chemo-attractive forces (nutrient sources) and repulsive forces due to the presence of competitors (neighbor sensor), (obstacles, and field boundary) to determine its movements. Up to our knowledge this is the first paper to simulate multiple Physarums in hexagonal CA to solve the problem of node deployment in mobile WSN.

2 The Proposed Model

WSN is an example of graphically expressed problem. Given an initial random deployment of n mobile sensor nodes over a 2-D area, we formulated a hexagonal CA reaction diffusion model for dynamic relocation of sensors using multiple Physarums as a representation of mobile sensors and food sources as interest points. We have designed energy aware algorithm where the sensor energy decrease by 1% with every movement step this will give a priority to less used sensors to compete over interest points.

2.1 The Model State of Cellular Automaton (CA)

In order to model mobile-WSN, we considered a CA grid in the two-dimension space, which is divided into a matrix ($X \times Y$) of identical hexagon cells, in which every cell $c_{(i,j)}$ has six neighbours. In this grid a set of m sensors ($S = s_1, s_2, \ \dots \ , s_m$) are competing on a set of n interest point ($IP = ip_1, ip_2, \ \dots \ , ip_n$). The state of a cell $c^t_{(i,j)}$ at time t located at position (i, j) is described by its type as in Eq. 1, whether it is an interest point, a sensor, an empty cell, or an obstacle cell (Ex:- physical obstacle, boundary wall).

$$CT_{(i,j)} = \{\text{“}FREE\text{”}, \text{“}OBSTACLE\text{”}, \text{“}INTEREST\_POINT\text{”}, \text{“}SENSOR\text{”}\} \tag{1}$$

An interest point is defined by its mass, and a sensor is defined by its energy, similarly to the original Physarum competition model, where chemical is defined by its mass, and Physarum is defined by its mass respectively.

2.2 Area Hexagonal Tessellation

We considered hexagonal deployments that minimize redundancy, and avoiding that more than one node senses and processes the same event [13]. The area is dynamically tessellated by regular hexagons with its side equal communication radius (Rc). The vertices and the centers of the regular hexagons will be identified as the interest points to be filled up with food source to attract Physarums. It has been proved that such node placement technique maximizes the area coverage using a minimum number of nodes [3].

2.3 Cellular Automaton (CA) Model Rules

In our model, each sensor is a self organized computational unit. Each of them aims to achieve the maximum utility based on its local environment by choosing appropriate behaviors. The CA model rules are mainly based on the diffusion equations combined with Physarum heuristics in competition settings, where multiple sensors (Physarums) will compete for these interest points (Food resources). Each sensor will execute the diffusion process (as defined in Eqs. 2, 3) to explore its neighborhood within its communication radius (R_c). Each sensor at iteration (t) uses the values of its six neighbours cell to calculate the value of the energy at the next iteration ($t + 1$).

$$SE^{t+1}_{(i,j)} = SE^{t}_{(i,j)} + \sum_{(k,l)} \begin{cases} \left(SF * SD * SE^{t}_{(k,l)}\right) - SE^{t}_{(i,j)}, & \text{if } S\_AA_{(i,j),(k,l)} = 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\forall (k,l): \; i-1 \leqslant k \leqslant i+1, \quad j-1 \leqslant l \leqslant j+1, \quad k \neq l$$

$$SF = 1 + S\_AttForce^{t}_{(i,j),(k,l)} + S\_RepForce^{t}_{(i,j),(k,l)} \tag{2}$$

where,

$SE^{t+1}_{(i,j)}$ defines the diffusion of sensor energy for the next generation $(t + 1)$ at cell $c_{(i,j)}$.

$SE^{t}_{(i,j)}$ is the current energy of the sensor at iteration (t) for cell $c_{(i,j)}$.

SF is the forces affecting a sensor.

SD is the sensor diffusion coefficient.

$$S\_AA_{(i,j),(k,l)} = \begin{cases} 1, & \text{if } CT_{(k,l)} = \text{“}FREE\text{”}\ OR\ \text{“}INTEREST\_POINT\text{”} \\ 1, & \text{if } CT_{(k,l)} = (\text{“}SENSOR\text{”})\ AND\ (SID_{(i,j)} = SID_{(k,l)}) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where,

$S\_AA_{(i,j),(k,l)}$ defines whether a sensor at cell $c_{(i,j)}$ is available to diffuse towards a neighbouring cell $c_{(k,l)}$.

$SID_{(i,j)}$ is the ID of the sensor.

In our model, we have addressed a 1% decrease in sensor energy (Physarum mass) with each movement step. Simply, sensor superiority in competition is directly proportional to sensor energy, a key point for load balancing and will give a priority to less used sensors to process messages and replace failed nodes. The process of searching for interest points will be executed for several rounds until all interest points are filled or other stopping conditions are met. Sensors failed to occupy interest points will not move and stay in stand-by for fault repair. This will minimize node displacement, and help to enhance the network lifetime.

2.4 Modelling Multiple Physarum and Multiple Food Resources

We created a new formula to compute two forces (attraction/repulsion) acting on Physarum: The first is chemo-attraction force to food sources (interest points), and the second is the repulsion negative forces the competing Physarums exert on each other based on its mass (sensor energy), and repulsion forces exerted by obstacles and boundary wall.

The attraction/repulsion forces as described in Eqs. 4, 5 determine the movement of Physarum towards the food and away from other competitors and obstacles.

$$S\_AttForce_{(i,j),(k,l)} = \begin{cases} \dfrac{IPM_{(k,l)}}{Total\_IPM}, & \text{if } IPM_{(k,l)} = MAX(IPM_{(i,j)}) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where,
$S\_AttForce_{(i,j),(k,l)}$ defines the value of attraction force of $SE_{(i,j)}$ towards its neighbouring cell $c_{(k,l)}$.
$IPM_{(i,j)}$ is the current mass of the interest point for cell $c_{(i,j)}$.
$Total\_IPM$ is the total sum of all interest points mass on the grid.

$$S\_RepForce_{(i,j),(k,l)} = \begin{cases} \dfrac{SE^t_{opp(k,l)}}{Total\_SE}, & \text{if } SID_{(i,j)} \neq SID_{opp(k,l)}, \\ & \quad SE^t_{opp(k,l)} > Rep\_Limit \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where,

$S\_RepForce_{(i,j),(k,l)}$ defines the value of repulsion force of $SE_{(i,j)}$ towards its neighbouring cell $c_{(k,l)}$.

$SE^t_{opp(i,j)}$ is the neighbor sensor energy at the opposite direction.

Rep_Limit is a limit where sensor must reach to repel neighboring sensor.

Each Physarums will execute the algorithm, to explore its neighborhood within its Rc and will find shortest pass to nearest interest point.

Algorithm 1. Slime Mould Diffusion

Formal Name: SM_DIFF
Input : $cell^{t}_{(i,j)}.sm$ (A slime mould in cell(i,j) at time 't')
Ensure : $cell^{t}_{(i,j)}.sm.mass > Diffusion\ Limit$

for $dir \in HEX\_Directions$ **do**
  $SM\_Forces = 1 + SM\_AttForce(cell^{t}_{(i,j)}.sm, dir) + SM\_RepForce(cell^{t}_{(i,j)}.sm, dir)$;
  $diffused\_mass = SM\_AA(cell^{t}_{(i,j)}.sm, dir) * SM\_Forces * cell^{t}_{(i,j)}.sm.diffusion\_factor * cell^{t}_{(i,j)}.sm.mass$;
  $cell^{t+1}_{(dir)}.sm.mass = cell^{t+1}_{(dir)}.sm.mass + diffused\_mass$;
  $cell^{t+1}_{(i,j)}.sm.mass = cell^{t+1}_{(i,j)}.sm.mass - diffused\_mass$;
end

Physarums failed to occupy interest points will not move and stay in stand-by for fault repair. This will minimize maximum node displacement, and help to enhance the network lifetime since node movement exhausts energy.

3 Experimental Results

The core model was implemented in Java with Processing package https://processing.org/ being used for graphical simulation. All the experiments were repeated for 30 times using the same parameters of diffusion equation as in [14] (Table 1).

Table 1. Parameters values for the experiments

Parameter	Value
SD	0.1
SE	3000
IPM	3000
REP_LIMIT	5

In this experiment, Physarum as a representation of sensors were placed in a 2D (50×50) hexagonal grid with obstacles for sensor communication (lake, mountains, etc. ...). In harsh environment even sensors random deployment over all the interest area may not be feasible. In this research sensors were deployed in selected areas away from obstacles (Fig. 1-a). The sensors are homogeneous, they have same sensing and communication radii, where $R_s = 2$, and $R_c = 5$. All the sensors move with the same speed and the same energy.

The area is virtually tessellated by regular hexagons with its side equal communication radius. All the vertices and the center of each hexagon (interest points = 82) will be filled with food source to attract Physarums (Fig. 1-a). We conducted three experiments scenarios; one folds (82 sensors), 1.5 folds (123 sensors) and two fold (164 sensors) the number of interest points. Physarums will execute the the proposed algorithm in Sect. 2. Physarum will sense its surrounding environment and define obstacles within its R_s, and will communicate with other Physarums in its R_c range. The algorithm will be repeated until at least 90% of interest points are filled or after 30 rounds are executed (Fig. 1-b).

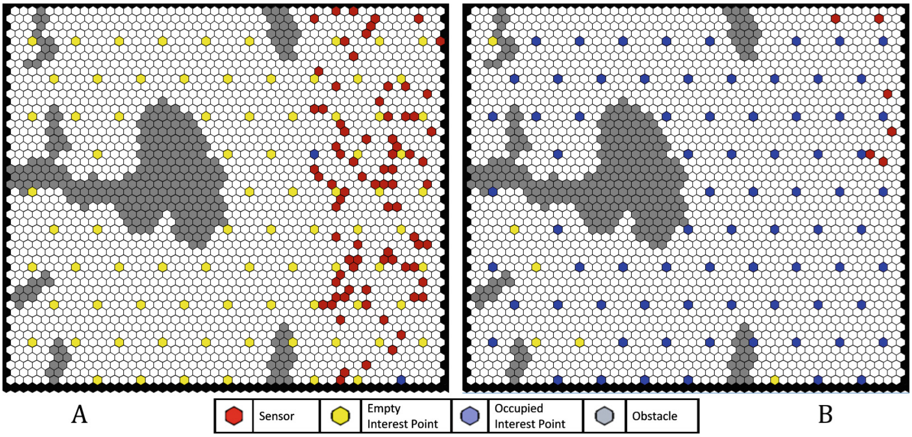


Fig. 1. (A) The initial sensor deployment and (B) the relocation of sensors after 30 round execution of the proposed model for the first scenario (82 sensors).

The outcome of the experiment showed that percentage of coverage, and the total number of moves to fill interest points were nearly similar in the three scenarios (Fig. 2). After the first round about 50% 0f coverage is achieved with an average one step/sensor. There after the sensors disseminate all over the area away from wall boundaries and obstacles with least number of sensors (one fold) and with minimal number of movement (Fig. 2).

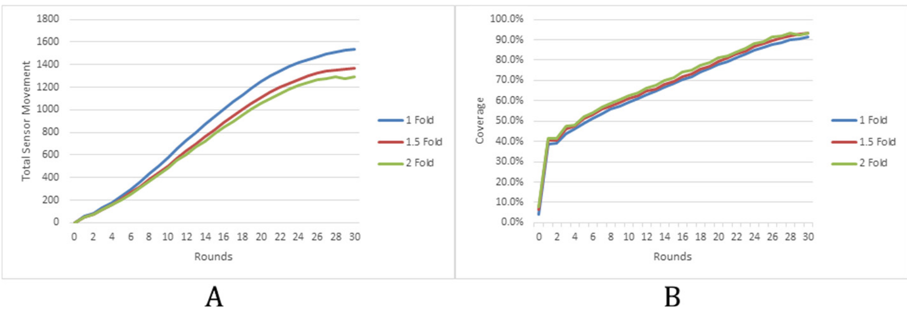


Fig. 2. (A) Movement steps and (B) Area coverage versus number of rounds for the three scenarios.

4 Conclusion

In this study, we presented a novel Physarum inspired energy aware model for dynamic deployment of mobile sensor nodes in the presence of obstacles. This model simulated Physarum complex foraging behavior based on hexagonal cellular automata and reaction diffusion system, where multiple Physarums will sense the surrounding environment, and will compete over the interest points based on chemo-attraction and repulsion forces exerted by obstacles, and competing Physarums. Simulation results have demonstrated the high coverage performance of the model with minimal move overhead even in the presence of obstacles with the least number of sensors.

Acknowledgement. Abubakr Awad research is supported by Elphinstone PhD Scholarship (University of Aberdeen). Wei Pang and George Coghill are supported by the Royal Society International Exchange program (Grant Ref IE160806).

References

1. Adamatzky, A.: From reaction-diffusion to physarum computing. Nat. Comput. **8**(3), 431–447 (2009). https://doi.org/10.1007/s11047-009-9120-5
2. Beghdad, R., Lamraoui, A.: Boundary and holes recognition in wireless sensor networks (2016). https://doi.org/10.1016/j.jides.2016.04.001. ID: 311969
3. Brass, P.: Bounds on coverage and target detection capabilities for models of networks of mobile sensors. ACM Trans. Sens. Netw. **3**(2), 9 (2007). https://doi.org/10.1145/1240226.1240229
4. Goubier, O.N.P., Huynh, H.X., Truong, T.P., Traore, M., Pottier, B., Rodin, V., Nsom, B., Esclade, L., Rakoroarijaona, R.N., Goubier, O., Stinckwich, S., Huynh, H.X., Lam, B.H., Vinh, Udrekh, Muslim, H., Surono: Wireless sensor network-based monitoring, cellular modelling and simulations for the environment. ASM Sci. J. **2017**(Special issue 1), 56–63 (2017)
5. Gunji, Y.P., Shirakawa, T., Niizato, T., Haruna, T.: Minimal model of a cell connecting amoebic motion and adaptive transport networks. J. Theor. Biol. **253**(4), 659–667 (2008). https://doi.org/10.1016/j.jtbi.2008.04.017
6. Jones, J.: Influences on the formation and evolution of physarum polycephalum inspired emergent transport networks. Nat. Comput. **10**(4), 1345–1369 (2011). https://doi.org/10.1007/s11047-010-9223-z
7. Lam, B.H., Huynh, H.X., Pottier, B.: Synchronous networks for bio-environmental surveillance based on cellular automata. EAI Endorsed Trans. Context-Aware Syst. Appl. **16**(8) (2016). https://doi.org/10.4108/eai.9-3-2016.151117
8. Li, X.: Improving area coverage by mobile sensor networks. Ph.D. thesis (2009). AAINR47481
9. Malaver, A., Motta, N., Corke, P., Gonzalez, F.: Development and integration of a solar powered unmanned aerial vehicle and a wireless sensor network to monitor greenhouse gases. Sensors (Switzerland) **15**(2), 4072–4096 (2015). https://doi.org/10.3390/s150204072
10. Nakagaki, T., Yamada, H., Tóth, Á.: Maze-solving by an amoeboid organism. Nature **407**(6803), 470 (2000). https://doi.org/10.1038/35035159
11. Reid, C.R., Latty, T.: Collective behaviour and swarm intelligence in slime moulds. FEMS Microbiol. Rev. **40**(6), 798–806 (2016). https://doi.org/10.1093/femsre/fuw033
12. Rout, M., Roy, R.: Dynamic deployment of randomly deployed mobile sensor nodes in the presence of obstacles. Ad Hoc Netw. **46**, 12–22 (2016). https://doi.org/10.1016/j.adhoc.2016.03.004. ID: 272922

13. Saha, D., Das, N.: Self-organized area coverage in wireless sensor networks by limited node mobility. Innov. Syst. Softw. Eng. **12**(3), 227–238 (2016). https://doi.org/10.1007/s11334-016-0277-7
14. Tsompanas, M.-A.I., Sirakoulis, G.C., Adamatzky, A.: Cellular automata models simulating slime mould computing. In: Adamatzky, A. (ed.) Advances in Physarum Machines. ECC, vol. 21, pp. 563–594. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26662-6_27
15. Tsompanas, M.A.I., Sirakoulis, G.C., Adamatzky, A.I.: Evolving transport networks with cellular automata models inspired by slime mould. IEEE Trans. Cybern. **45**(9), 1887–1899 (2015). https://doi.org/10.1109/TCYB.2014.2361731
16. Zhang, X., Gao, C., Deng, Y., Zhang, Z.: Slime mould inspired applications on graph-optimization problems. In: Adamatzky, A. (ed.) Advances in Physarum Machines. ECC, vol. 21, pp. 519–562. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26662-6_26

On the Equivalence Between Eigen and Channel Inversion Based Precoders

Khalid W. Hameed[1], Yasir Al-Yasir[1], Naser O. Parchin[1], Raed A. Abd-Alhameed[1,2], and Peter S. Excell[1,2(✉)]

[1] Faculty of Engineering and Informatics, University of Bradford, Bradford BD7 1DP, UK
r.a.a.abd@bradford.ac.uk
[2] Wrexham Glyndwr University, Wrexham LL11 2AW, UK
p.excell@glyndwr.ac.uk

Abstract. Multi-user MIMO precoding is crucial in modern and next generation wireless communication systems. In this paper the equivalence between two linear precoding methods using closed form solutions is investigated. The first one is the regularized zero forcing (RZF) algorithm and the second one is signal to leakage and noise ratio (SLNR). Three studies are presented: (1) comparison between the regularized and non-regularized versions; (2) finding a good regularization factor that can fit with all methods; (3) to present the equivalence of the methods in certain cases and the superiority of SLNR over RZF for user cases with more than a single antenna. A simple mathematical proof of the equivalence between RZF and SLNR beamformer implementations for the single antenna user case in a multi-user transmission scenario is presented: this matches simulation results.

Keywords: Beamforming · Channel inversion · Eigenvalue decomposition Multi-user-MIMO · Regularized zero forcing · Signal to leakage and noise ratio

1 Introduction

To enhance the capacity of a communication system, various approaches have been applied, such as the use of multiple antennas or the smart antenna, where spatial diversity is used to mitigate the channel condition without increasing the transmitted power or bandwidth [1]. Increasing the capacity and reliability of wireless communications systems through the use of multiple antennas has been an active area of research for over 20 years [2] and modern multi-antenna systems can take several configurations [3]: multi-input single output (MISO), single-input multi-output (SIMO) and multi-input multi-output (MIMO). The MIMO configuration can operate in two modes, spatial diversity and spatial multiplexing. The first mode enhances the performance of the bit error rate (BER) while the latter mode is used to increase the capacity. A more advanced configuration is multi-user MIMO [4–6] (MU-MIMO). This configuration works in a spatial multiplexing mode. It differs from single user MIMO (SU-MIMO) or what is also called point-to-point MIMO in that it does not allow user co-operation in decoding, whereas cooperation between antennas is essential in detection with

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 161–172, 2018.
https://doi.org/10.1007/978-3-319-95450-9_13

SU-MIMO. MU-MIMO has several advantages over SU-MIMO [7]: it allows for a direct gain in multiple access capacity, better immunity performance against system and environment impairments, and it can achieve high capacity with a single antenna at the users' terminals, meaning smaller, cheaper handsets. In the analysis of MU-MIMO there are two main scenarios studied in the literature [8]: the multiple-access channel (MAC) or the reverse link, where signals are transmitted from users' terminals simultaneously to the base station, and the broadcast channel (BC) where the base station transmits signals to the users using the same time-frequency resource.

The precoders that are used in MU-MIMO are divided into two categories, non-linear and linear. Although the non-linear category achieves higher data rates, it has a higher complexity in comparison with linear ones. This becomes a significant restriction in next generation networks. In these networks some large scale regimes (massive MIMO and dense small cells) are proposed to deliver the required capacity [9]. The lower complexity linear precoder categories include maximum ratio transmission (MRT) [10], channel inversion or zero forcing (ZF) [10, 11], regularized zero forcing (RZF) [11], which is also known as minimum mean square error (MMSE) [10], and signal to leakage ratio (SLR) [12] or its regularized form, the signal to leakage and noise ratio (SLNR) [13] which has also been adopted recently to support multiple streams per user [14].

The ZF/RZF category has simpler equations and is easier to implement, but it has a dimension restriction in that the number of antennas at the base station should be larger than the total number of active users' antennas. To mitigate this condition, optimization of criteria such as the signal to interference per user is desirable: however this is constrained by a problem with coupling of variables and gives no closed form. On the other hand, the SLR/SLNR category gives an optimized precoder with a closed-form solution.

The authors of [15] and [16] show, in two different approaches, the equivalence between the RZF (or MMSE) and the SLNR precoders. In the present work, a hybrid approach between the methods used in these two references was utilized to achieve the same result with simpler mathematics.

2 System Mathematical Model

Consider a communication system with K active users served by a base station with M antennas, as shown in Fig. 1.

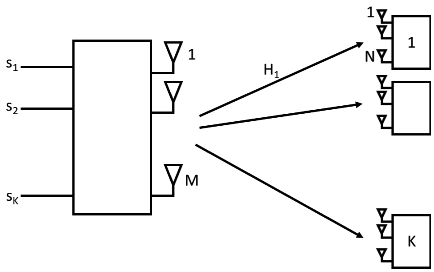


Fig. 1. The system model for MU-MIMO.

A time-frequency resource block is utilized to serve the active users simultaneously. The channel from the base station to user i is given by [4]:

$$H_i = \begin{bmatrix} h_{1,1,i} & \cdots & h_{1,M,i} \\ \vdots & \ddots & \vdots \\ h_{N,1,i} & \cdots & h_{N,M,i} \end{bmatrix} \tag{1}$$

Where $h_{n,m,i}$ is the channel from the mth transmitter antenna at the base station to the nth antenna at the ith user. The elements of H_i are assumed to be Rayleigh channels (i.e. unity variance with zero mean independently identically distributed (i.i.d.) complex Gaussian random variables), hence slow-flat fading channels. The aggregated channel for all users is given by [7]:

$$H = \left[H_1^T H_2^T \cdots H_K^T\right]^T \tag{2}$$

where H_i is the total channel matrix between the base station and the ith user. The leakage channel for user i (the channel from the base station to all other users except the intended user) is given by [13]:

$$\hat{H} = \left[H_1^T \cdots H_{i-1}^T H_{i+1}^T \cdots H_K^T\right]^T \tag{3}$$

The received signal by user i is

$$y_i = H_i X + n_i \tag{4}$$

Where n_i is the noise vector at user i with variance equal to σ^2, X is the transmitted vector from the base station and equals the sum of the transmitted vectors for all of the users:

$$X = \sum_{i=1}^{K} w_i s_i \tag{5}$$

Where $w_i \in \mathbb{C}^{M \times 1}$ is the precoder vector for user i and s_i is the data symbol for the same user.

3 Proposed Proof of Equivalence

For simplification of the mathematics, some definitions need to be established first. a is the user channel and b is the leakage channel. Now define A and B as follows:

$$A = a^H a \tag{6}$$

$$B = b^H b \tag{7}$$

$$Q = \left(B + \sigma^2 I\right) \tag{8}$$

From [16], the SLNR weights are given by

$$w_i^o \propto \max.EV\left(\left(B + N\sigma_i^2 I\right)^{-1} A\right) \tag{9}$$

which can be rewritten as:

$$w_{SLNR} \propto Q^{-1} a^H = \left(B + \sigma^2 I\right) a^H \tag{10}$$

From Lemma 1 in [15]:

$$C^{-1} d \propto \left(C + dd^H\right)^{-1} d \tag{11}$$

Where C is a matrix and d is a vector, then by letting $C = (\sigma^2 I + B)$ and $d = a^H$ we get:

$$\left(\sigma^2 I + B\right)^{-1} a^H \propto \left(\sigma^2 I + B + A\right)^{-1} a^H \tag{12}$$

which leads to:

$$\left(\sigma^2 I + B\right)^{-1} a^H \propto \left(\sigma^2 I + H^H H\right)^{-1} a^H \tag{13}$$

Where $H = A + B$. Now from [17] the RZF precoder is given by:

$$w_{RZF} \propto \left(H^H H + \alpha I\right)^{-1} H^H \tag{14}$$

4 Results

In this section results are presented to give a general perspective and to prove the equality between RZF and SLNR in a certain case. The non-regularized version of the methods is that where the effect of the channel only is considered in the optimization of the beamformer weights, while the regularized version takes the effect of the additive white Gaussian noise into account by adding a factor related to this noise. The first three figures, Figs. 2, 3 and 4, present a comparison between two approaches of zero forcing, the first one by using the pseudo-inverse (pinv) function in Matlab: this is equivalent to $H^H * (HH^H)^{-1}$. The second one is $(H^H H)^{-1} H^H$. Three observations can be made from these figures. Firstly, there is equivalence between ZF1 and RZF2. The second observation is that this equivalence still holds for multi-antenna users, especially in the low SNR region. The third observation is that the capacity tends to saturate when we use 8 antenna elements per user for the same scenario. The next two sets, including the figures from Figs. 5, 6, 7, 8, 9 and 10, present the performance of RZF

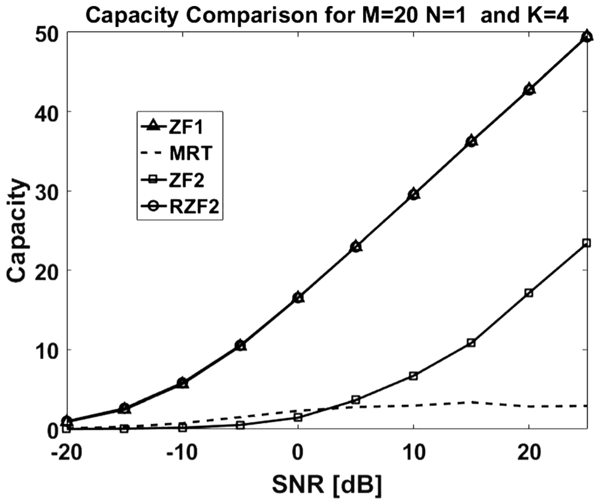


Fig. 2. Performance comparison between MRT, two versions of ZF, and RZF for a single antenna at user's location.

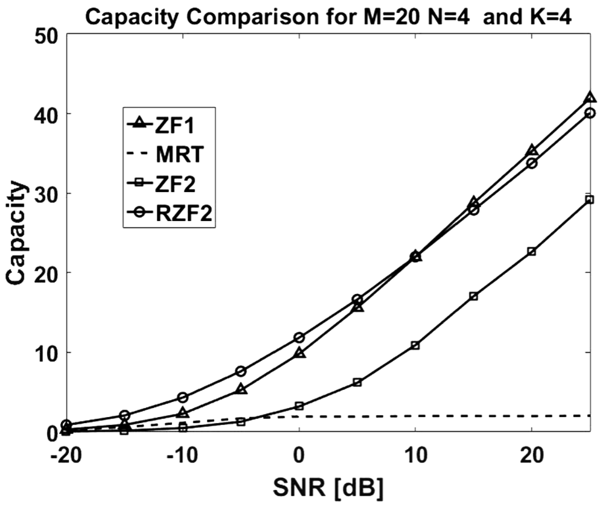


Fig. 3. Performance comparison between MRT, two versions of ZF and RZF for 4 antennas at users' location.

and SLNR with different regularization terms ($\sigma^2 I$, $\sigma^2 MI$, $\sigma^2 KI$, $\sigma^2 NI$ and I) in each figure for different numbers of antennas at the users' ends. The same behavior mentioned in the first set when the number of antennas per user was changed was also noticed with different regularization factors. The conclusion from these figures is that the regularization factor $\sigma^2 I$ is the best choice as it gives better performance compared with the others, for both beamformers.

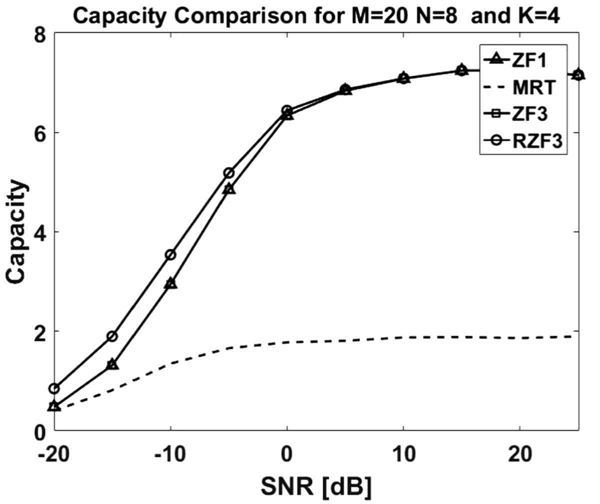


Fig. 4. Performance comparison between MRT, two versions of ZF and RZF for 8 antennas at users' location.

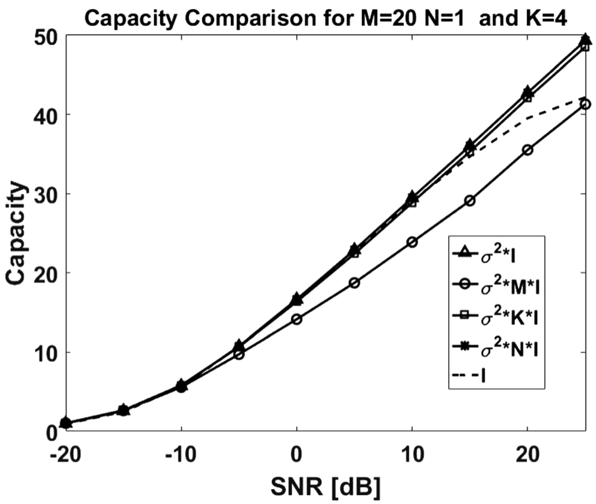


Fig. 5. Performance comparison for RZF with different regularization term for single antenna users.

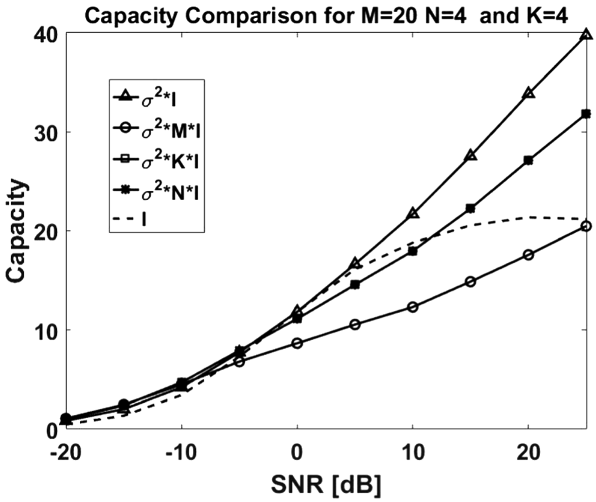


Fig. 6. Performance comparison for RZF with different regularization term for users with 4 antennas.

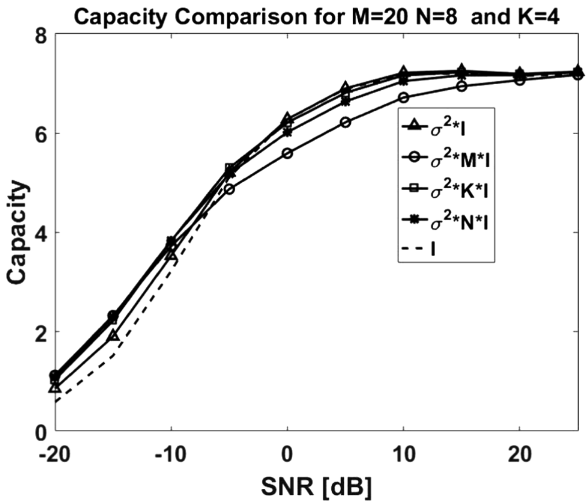


Fig. 7. Performance comparison for RZF with different regularization term for users with 8 antennas.

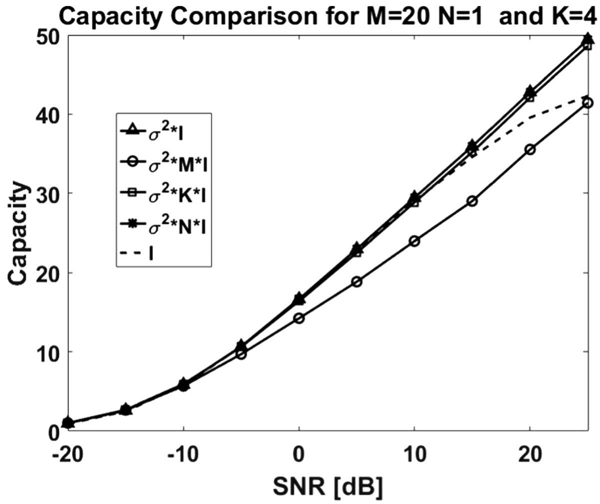


Fig. 8. Performance comparison for SLNR with different regularization term for single-antenna users.

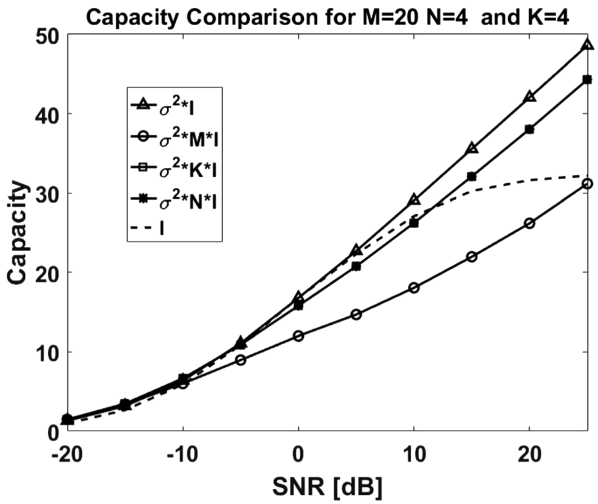


Fig. 9. Performance comparison for SLNR with different regularization term for users with 4 antennas.

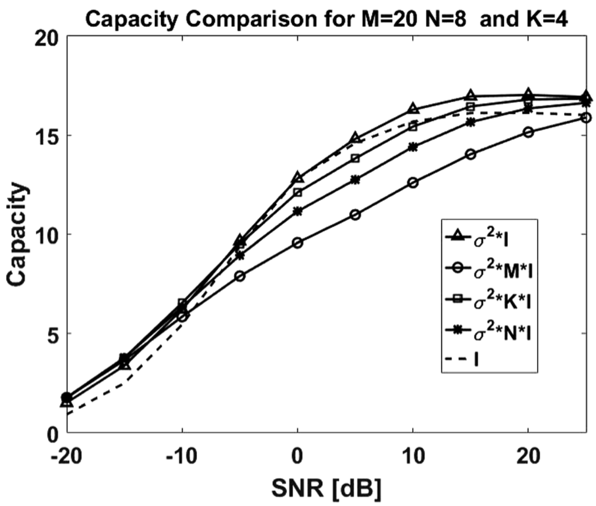


Fig. 10. Performance comparison for SLNR with different regularization term for users with 8 antennas.

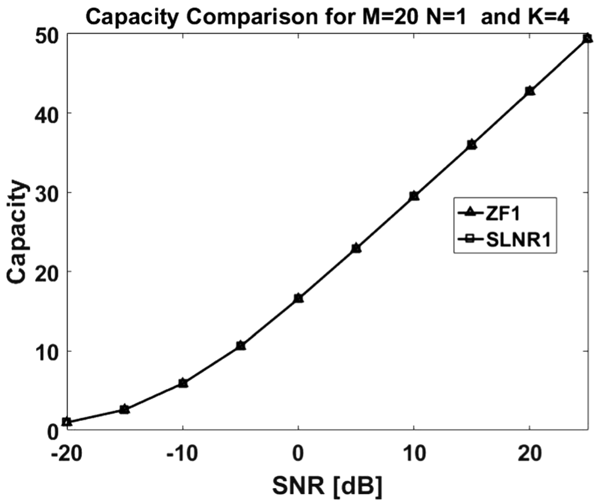


Fig. 11. Equivalence between RZF and SLNR with single-antenna users and regularization factor = $\sigma^2 I$.

Figures 11 and 12 show the relation between RZF and SLNR. Figure 11 reveals the equivalence between RZF and SLNR for the case of single-antenna users, using the regularization factor $\sigma^2 I$, however when the numbers of antennas at the user are increased the two methods start to diverge. It should noticed that the regularization factor used in Fig. 13 is $\sigma^2 NI$, which leads to lower performance than that of RZF with regularization factor $\sigma^2 I$ (Fig. 13 and Table 1).

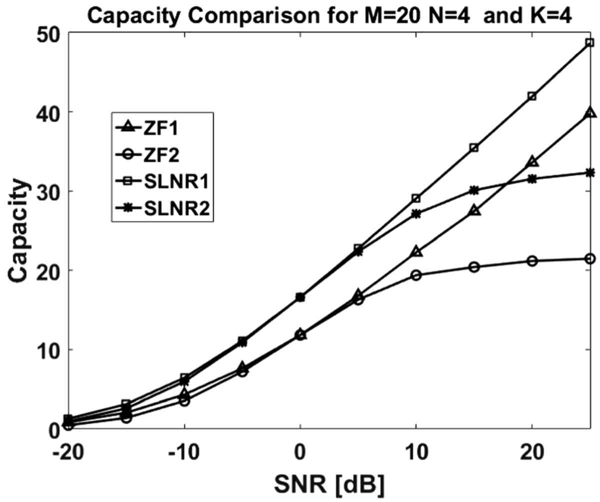


Fig. 12. Performance comparison between RZF and SLNR for 4 antennas at user side and two regularization factors.

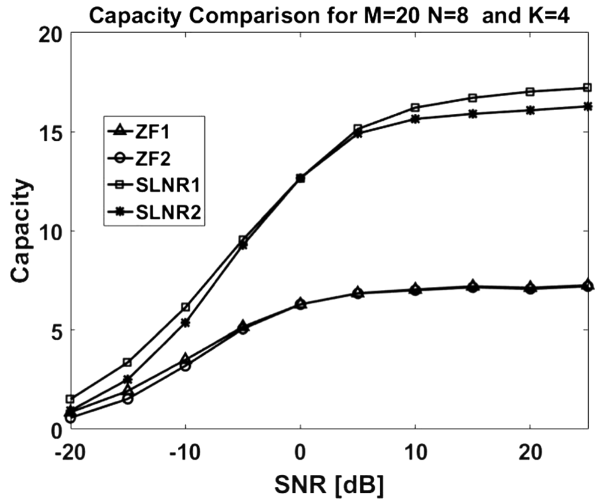


Fig. 13. Performance comparison between RZF and SLNR for 8 antennas at user side and two regularization factors.

Table 1. Capacity saturation for different cases

Method	Antennas per user		
	1	4	8
ZF	No	Factor dependent	Yes
SL	No	Factor dependent	Yes

5 Conclusion

In this paper the equivalence between RZF and SLNR precoders for the MU-MIMO transmission scheme has been presented. The equivalence between the two methods was proven by simulation. It was observed that the equivalence was intrinsic with single antenna users, meaning that channel inversion is another form of Eigenvector for simple cases where the user channel is a vector rather than a matrix. For more complex cases, where each user channel is a matrix, the SLNR performed better than RZF for the same regularization factor and the maximum performance was attained through the usage of ($\sigma^2 I$) as a regularization factor.

Acknowledgment. This work is partially supported by the EU Innovation Programme under grant agreement H2020-MSCA-ITN-2016 SECRET-722424 and financial support from the UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/E022936/1.

References

1. Wong, K.-K., Cheng, R.-K., Letaief, K.B., Murch, R.D.: Adaptive antennas at the mobile and base stations in an OFDM/TDMA system. IEEE Trans. Commun. **49**, 195–206 (2001)
2. Swindlehurst, A.L., Ayanoglu, E., Heydari, P., Capolino, F.: Millimeter-wave massive MIMO: the next wireless revolution? IEEE Commun. Mag. **52**, 56–62 (2014)
3. Paulraj, A., Nabar, R., Gore, D.: Introduction to Space-Time Wireless Communications. Cambridge University Press, Cambridge (2003)
4. Choi, L.-U., Murch, R.D.: A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach. IEEE Trans. Wirel. Commun. **3**, 20–24 (2004)
5. Serbetli, S., Yener, A.: Transceiver optimization for multiuser MIMO systems. IEEE Trans. Sig. Process. **52**, 214–226 (2004)
6. Spencer, Q.H., Peel, C.B., Swindlehurst, A.L., Haardt, M.: An introduction to the multi-user MIMO downlink. IEEE Commun. Mag. **42**, 60–67 (2004)
7. Gesbert, D., Kountouris, M., Heath Jr., R.W., Chae, C.-B., Salzer, T.: Shifting the MIMO paradigm. IEEE Sig. Process. Mag. **24**, 36–46 (2007)
8. Goldsmith, A., Jafar, S.A., Jindal, N., Vishwanath, S.: Capacity limits of MIMO channels. IEEE J. Sel. Areas Commun. **21**, 684–702 (2003)
9. Rajoria, S., Trivedi, A., Godfrey, W.W.: A comprehensive survey: small cell meets massive MIMO. Phys. Commun. **26**, 40–49 (2018)
10. Björnson, E., Bengtsson, M., Ottersten, B.: Optimal multiuser transmit beamforming: a difficult problem with a simple solution structure [lecture notes]. IEEE Sig. Process. Mag. **31**, 142–148 (2014)

11. Peel, C.B., Hochwald, B.M., Swindlehurst, A.L.: A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization. IEEE Trans. Commun. **53**, 195–202 (2005)
12. Tarighat, A., Sadek, M., Sayed, A.H.: A multi user beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005 (ICASSP 2005), vol. 3, pp. iii/1129–iii/1132 (2005)
13. Sadek, M., Tarighat, A., Sayed, A.H.: A leakage-based precoding scheme for downlink multi-user MIMO channels. IEEE Trans. Wirel. Commun. **6**, 1711–1721 (2007)
14. Lopes, P.A., Gerald, J.A.: Leakage-based precoding algorithms for multiple streams per terminal MU-MIMO systems. Digit. Sig. Proc. **75**, 38–44 (2018)
15. Yuan, F., Yang, C.: Equivalence of SLNR precoder and RZF precoder in downlink MU-MIMO systems. arXiv preprint arXiv:1202.1888 (2012)
16. Seo, B., Shin, J.: Equivalence between SLNR and MMSE precoding schemes in the K-user MISO interference channel. WSEAS Trans. Sig. Process. **11**, 23–28 (2015)
17. Sung, H., Lee, S.-R., Lee, I.: Generalized channel inversion methods for multiuser MIMO systems. IEEE Trans. Commun. **57**, 3489–3499 (2009)

Carbon Nanotube Technology as an Option for Future Computing Devices

Nataliia Luhyna[1(✉)], Peter Excell[1], Richard J. Day[1], Alison J. McMillan[1], Fawad Inam[2], and Ardeshir Osanlou[1]

[1] School of Applied Science, Computing and Engineering, Wrexham Glyndŵr University, Wrexham LL11 2AW, UK
{n.luhyna, p.excell, r.day, a.osanlou}@glyndwr.ac.uk, alison.mcmillan@physics.org

[2] Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
fawad.inam@northumbria.ac.uk

Abstract. Carbon nanotubes (CNTs) offer the potential for radical transformation of future electronics as they exhibit conductivity, semiconductor and mechanical properties that far exceed those of traditional materials. They thus offer wide opportunities for novel computing devices, especially those required to be small and light, such as “wearables”. They also offer environmental advantages by reducing energy demand in manufacture and use. A practical investigation is reported, focusing on microwave processing of a CNT composite sample. A significant reduction in energy consumption in manufacture while increasing mechanical and electrical performance is deduced, suitable for low power wearable, wireless computing devices. However, CNTs inherently suffer from flaws such as the tendency for nanoparticle agglomerations to form due to strong covalent bonds, and other uncertainties in nanoparticle behaviour. This experimental study demonstrates that microwave processing of CNT composites can reduce the power consumption and the temperature generated during the manufacturing process.

Keywords: Future computing devices · Wearable devices · Carbon nanotubes Epoxy nanocomposites · Microwave curing

1 Introduction

The discovery of carbon nanotubes (CNTs) can be traced back to 1951 [1], although this was not widely publicised and their interesting properties were not discovered until much later. In 1992, a paper on their potentially very useful electrical properties was published [2], and subsequently the work on them has expanded in the 21st century, in parallel with the related work on graphene. The nanotubes, which are formed of tubes of graphene, have the potential to be very highly conducting, rivalling superconductivity or, in some circumstances, to have semiconducting properties. In addition, they have mechanical properties that are greatly in advance of traditional materials, as is also

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 173–182, 2018.
https://doi.org/10.1007/978-3-319-95450-9_14

the case with graphene and carbon fibres more generally. They thus have great potential for transformation of many aspects of the electronics and computing industries.

Already, a basic CNT computer has been constructed [3] and, even though it will take some time for this to overtake the advantages of silicon, the much-publicised limitations that silicon is approaching mean that searches for alternative materials are actively in progress: CNTs are strong candidates in this competition.

CNTs also offer great potential in the construction of hardware [13–16]. Their great strength means that thinner and lighter structures can be developed, and this is particularly relevant to the development of wearable technologies because less material is involved than is the case with metal or traditional plastics, there is less environmental demand on the sources of materials and, further, the environmental stresses in the manufacturing process are greatly lessened compared with metal (at least) due to the much lower processing temperature, as well as the likely reduced volume.

Furthermore, significant achievements include 1 GHz operation of CNT interconnects with silicon transistors [17], organic interconnects and spintronic switches [18], Cu-CNT composite interconnects as potential replacements of global Cu interconnects [19, 20], demonstration of Cu-like resistivity [21, 22] and integration scheme for CNT based Through-Silicon-Vias (TSVs) [23] as key enablers of 3D integration, etched in the silicon substrate.

For all these reasons it was considered timely to investigate carbon nanotubes in a laboratory specialising in carbon composite technology and hence some sample test pieces were created. The manufacture and processing of these were investigated and refined and some of the properties investigated. So far this has focused on mechanical properties, but electrical properties have also been studied via the context of the bulk microwave conductivity, this being immediately relevant to the processing of the nanotubes within an epoxy bonding matrix. This gives some indication of the future potential of the material in electronic applications, for which the project reported here is an initial stepping stone.

2 Experimental Work

2.1 Material

The epoxy resin system used in this study was Araldite LY 5052/Aradur 5052. Araldite LY 5052 is a low viscosity multifunctional epoxy system supplied by Huntsman, USA. Epoxy resin was produced from a reaction of bisphenol A resin and epichlorohydrin [4]. The hardener for this system was Aradur 5052 which is mixture of polyamines. Commercially available highly purified multi-walled CNTs (MWNTs) supplied by Electrovac, Austria (95% as per thermogravimetric analysis, having traces of metal and metal oxide) were added, having a density of 0.98 g/cm^3 (as per He pycnometery), specific surface area of 26 m^2/g, average length up to 2500 nm and average diameter of 50 nm. The synthesis method used was Chemical Vapour Deposition (CVD). The mix ratio of epoxy resin to hardener was 100:33 parts by weight which corresponds to 24.8% of hardener and 75.2% of resin, according to amine/epoxy ratio, due to high chemical activity of amine groups [5].

2.2 Specimen Preparation

The components of resin and hardener were weighed accurately according to the processing data and hand mixed. Pre-calculated amounts of CNTs and epoxy resin were carefully weighed and manually mixed together. MWNTs in the amount of 0.01 wt.%, 0.1 wt.% and 0.2 wt.% were infused in the matrix and dispersed via bath sonication (Ultra 7000, ultrasonic frequency: 42 kHz, power consumption: 50 W) for 1 h. Afterwards each epoxy system was divided into 6 parts and poured into glass tubes. The tubes were put into a vacuum oven for 1 h to remove the presence of air before microwave curing.

2.3 Curing Procedure

The microwave setup used in this study was MARS 6 supplied by CEM Corporation, USA (magnetron frequency 2.45 GHz, power output 1800 W). It was used with vessels having self-regulating control of the temperature and pressure. MARS 6 automatically recognises the type and number of vessels that have been loaded, and adjusts the output power and other parameters. Neat epoxy resin (Epoxy 1) and an epoxy system infused with 0.01 wt.% CNTs (Epoxy 2), with 0.1 wt.% CNTs (Epoxy 3), and 0.2 wt.% CNTs (Epoxy 4) were cured under the same microwave conditions. The initial parameters used were: ramp time: 10 min, hold time: 1 min, temperature: 40 °C and maximum power: 500 W.

Fig. 1. MARS 6 microwave apparatus

2.4 Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) is a useful tool to determine post-curing processes as well as temperature transitions (melting, crystallisation and glass transition temperatures) in materials. DSC of all specimens was carried out using Perkin Elmer

Pyris 1 apparatus. The samples were cut into small pieces weighing from 5 mg to 10 mg using an engineering blade machine Labotom-3 supplied by Struers, Australia. Specimens were placed in aluminium plates containing a crimped lid with a small hole. The hole is necessary to maintain the constant pressure in the system and prevent deformation or rupture of aluminium pans. The DSC measurements were carried out from 30 °C to 250 °C at a high heating rate of 10 °C/min for three cycles under a nitrogen atmosphere. DSC was performed for the neat epoxy and all nanocomposite systems under the same conditions.

2.5 Scanning Electron Microscopy

Scanning Electron Microscopy (SEM) in one of the most widely used methods for inspecting topographies of CNT nanocomposites and CNT distributions. SEM of all specimens was carried out using a field emission MIRA 3 TESCAN apparatus. The samples were cut into small pieces to fit into the machine using the engineering blade cutting machine Labotom-3. A conductive platinum sputter coating was used to coat samples prior to the testing. All SEM images were taken at 5 kV voltage.

3 Results and Discussion

Figure 2 shows the power profile as a function of time applied by the microwave curing process in order to maintain the set minimum temperature of 40 °C. The power profile has the same characteristics for the epoxy and nanocomposite samples during the curing process. Three stages in Fig. 2 can be noted. Stage 1 (up to 12 s) of the graph shows a sharp decrease in the power as the sample heats up. As shown in Fig. 2, during this stage the power decreased rapidly as the default initial power (500 W) was relatively high. The fall of the curve extends until the consumed power reaches approximately 110–120 W. Afterwards, the epoxy and nanocomposite systems start to react and generate heat within the sample. As can be seen from Fig. 1, the chemical reaction occurred over a small period of time, from 12 s to 154 s, which may be referred as stage 2. The epoxy-amine reaction is the dominant reaction during the microwave curing process [6, 7], whereas the epoxy-hydroxyl groups are more dominant during conventional heating. The reaction can only take place from a certain minimum energy called the energy threshold of the reaction [7]. In this study, it can be considered that the energy threshold occurs at 12 s, after which an active chemical reaction of the curing process started. During stage 2, the microwave energy is absorbed by the epoxy system and maximum power consumption was found to be around 220 W. At the third stage of curing (after 154 s), in all cases, no energy consumption was observed (Fig. 2). The epoxy resin system ceased to consume microwave power and no further external energy was required for curing.

The area under the power-time curve was calculated by integration (Eq. 1), where a and b limits are 0 and 160 s. Theoretically, it defines the amount of potential energy absorbed during the microwave processing. The obtained values are presented in Table 1: this indicates the amount of energy consumed for curing neat epoxy and nanocomposites. It was found that with the smaller addition of CNTs, more energy was required for

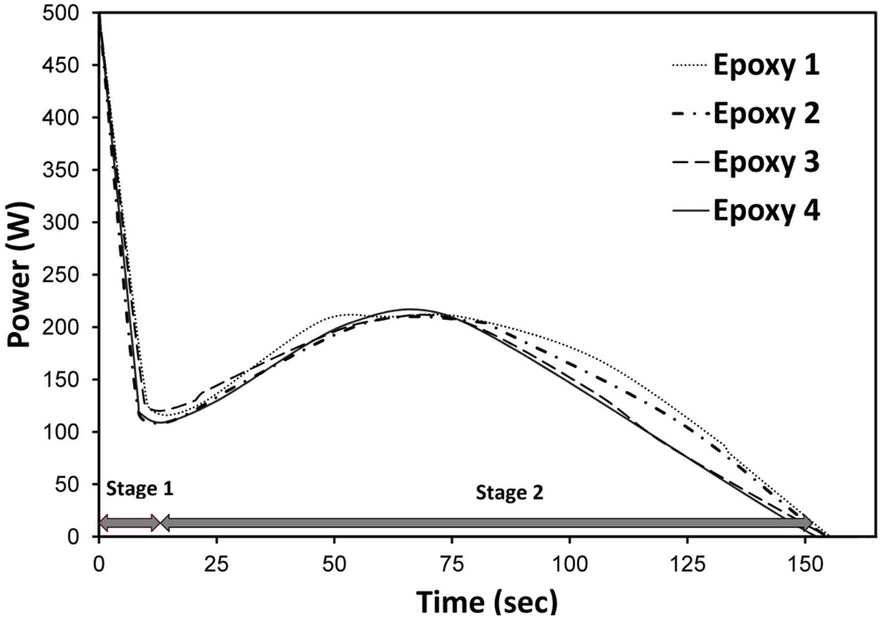


Fig. 2. Power profile for the samples obtained using microwave curing

microwave curing. The areas under the curves were evaluated and compared. The experiments allowed a 12.5% energy reduction by employment of 0.2 wt.% of CNTs in the epoxy matrix (Table 1). Therefore, it can be concluded that CNTs can significantly lower the microwave energy consumption for curing epoxy nanocomposites.

$$Area = \int_a^b f(x)dx \tag{1}$$

A series of DSC tests were conducted in order to observe the degree of cure for epoxy and nanocomposite samples. DSC analysis confirmed that all microwaved epoxy and nanocomposites were completely cured since there was no evidence of chemical reaction during testing (post-curing process). If the system is fully cured, the absorption of heat will not occur and the epoxy system can be re-heated and re-cooled reversibly below its glass transition temperature Tg [8]. A cyclic execution of this test is an accurate way for analysing the degree of cure compared to a single heating cycle. In this work, three heating/cooling cycles were conducted prior to the reporting of results (Table 1). All microwaved samples were fully cured significantly before the curing time specified by the supplier of the epoxy system [9]. The technical data sheet recommends curing (at room temperature) and post curing (at 100 °C) times of at least 4 h 40 min.

The temperature of the microwave process could not be controlled or changed manually during the curing because the microwave oven is programmed to be self-controllable. The temperature-time profile of the microwave heating of epoxy and nanocomposite systems is presented in Fig. 3. In contrast to the power-time curves (Fig. 2), the temperature profile was divided into two stages of the process. Stage A (up to 325 s) shows a gradual increase of temperature of the microwave curing process. By analysing this stage, it was possible to observe gradual heat release from the chemical reaction. This was attributed to the dipole moments and their polarisation in the electromagnetic field [10–12]. Stage B shows gradual reduction of curing temperature due to the process of crosslinking and the progressive curing of the composites. By the end of this stage, the material was fully cured and hardened. As can be implied from the power-time graphs (Fig. 2), consumption of power took place during stage 2 of the curing (until 154 s), and Fig. 3 shows temperature rising at the same time. Based on these results (Figs. 2 and 3), it can be concluded that after 154 s of the heating process, heat was generated from inside the system during exothermic reaction and no external heating was needed for curing.

At higher temperatures, above 50 °C, it was found that the monolithic Epoxy 1 (Fig. 3) required a higher amount of heat as compared to any other system. Monolithic Epoxy 1 had the highest curing temperature of 73 °C as compared to 58 °C for the Epoxy 4 (0.2 wt.% CNTs) system. Here, a reduction of 20.5% in temperature can be seen because of the addition of 0.2 wt.% CNTs. Moreover, the higher content of CNTs lowered the temperature required for full curing of epoxy nanocomposites (Fig. 3). This may be attributed to the high electrical and thermal properties and selective heating of CNTs which allow the production of materials at lower temperatures. CNTs create conducting paths and absorb microwaves efficiently and this may lead to rapid curing: it also indicates the potential for CNTs in advanced electronics. The microwave energy was applied directly to the material and as a result, no further side reactions or

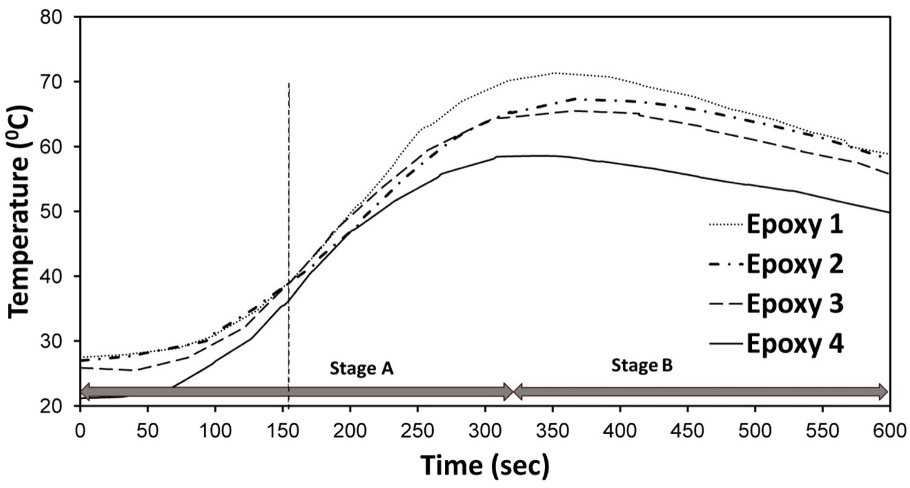


Fig. 3. Temperature profile for epoxy and nanocomposite samples. The dotted line shows the point where the energy consumption was 0 W

Table 1. Summary table for microwave cured samples

Epoxy	wt.% of CNTs	Energy, J	Maximum temperature reached during curing, °C
Epoxy 1		416	71.3
Epoxy 2	0.01	390	67.4
Epoxy 3	0.1	380	65.5
Epoxy 4	0.2	365	58.5

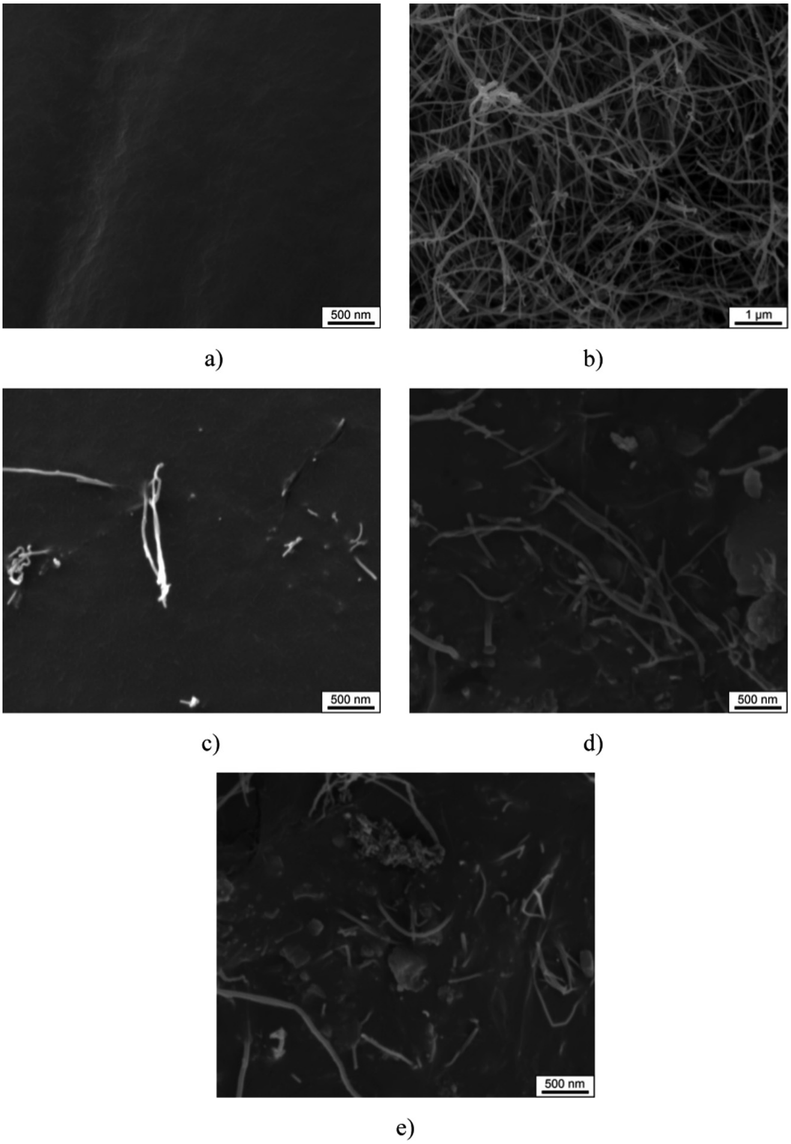


Fig. 4. SEM images of studied materials: (a) Epoxy 1; (b) CNTs; (c) Epoxy 2, (d) Epoxy 3 and (e) Epoxy 4

heat losses were observed. These observations are very important from a manufacturing perspective because they indicate significant energy and time savings.

Figure 4 depicts a series of SEM images obtained for CNTs, neat epoxy resin (Epoxy 1) and CNT composites (Epoxy 2–4). CNT agglomerates can be clearly seen as well as individually dispersed CNTs (Fig. 4, c–e). This may be attributed to an inefficient sonication process via bath sonication, as described earlier. The ultrasonic technique was not strong enough to break covalent bonds and fully eliminate CNT clusters: this is the main obstacle in creating advanced CNT composites: a tendency of nanoparticle agglomerations to form due to strong covalent bonds. These agglomerations indicate an uneven volumetric CNT dispersion and could lead to unbalanced distribution of properties in greater material volumes. However, it can be observed that the individual CNTs are not broken after 1 h sonication and therefore should retain their original structures and properties.

Despite the agglomeration effects, the SEM images demonstrate that CNT networks have a great potential to create a 3D conductive platform via an efficient sonication process (Fig. 4). Further investigations into an improvement of the sonicating process are in progress.

4 Conclusions

Experiments to refine manufacturing procedures for carbon nanotube composite materials have been reported. Specifically, the effect of CNT addition on the microwave curing of epoxy resin composites was investigated. CNT filled epoxy nanocomposites can be manufactured in a few minutes with significantly reduced energy consumption compared to a monolithic epoxy resin without CNTs. The experiments conducted showed reductions of 12.5% in energy and 20.5% in temperature by adding 0.2 wt.% of CNTs to an epoxy matrix. Differential Scanning Calorimetry analysis confirmed that microwaved CNT-filled epoxy nanocomposites were completely cured. This may be attributed to the good electrical and thermal conductivity and microwave absorbing properties of CNTs. However, SEM images showed varying degrees of unevenness in the distribution of CNTs in the samples, indicating a need for further investigations into the sonication processing: these are being conducted. However, the enhanced conductivity shown by the microwave properties gives confidence that CNT-based materials are close to being ready to be implemented in electronic and computing devices and are already an advantageous choice for the mechanical parts. Being lightweight, the mechanical parts would be especially suitable for wearable wireless devices, which have a great potential for many applications. Further, CNTs in the electronic parts offer the possibility of greatly reduced power consumption, hence reducing battery drain, which is a major problem for wearable devices.

References

1. Radushkyevich, L.V.: On the structure of carbon formed by the thermal decomposition of carbon oxide at an iron contact. J. Phys. Chem. **26**, 88–95 (1952). (in Russian)
2. Mintmire, J.W., Dunlap, B.I., White, C.T.: Are fullerene tubules metallic? Phys. Rev. Lett. **68**(5), 631–634 (1992)
3. Shulaker, M., Hills, G., Patil, N., Wei, H., Chen, H.-Y., Wong, H.-S.P., Mitra, S.: Carbon nanotube computer. Nature **501**, 526–530 (2013)
4. Chernin, I.Z., Smekhov, F.M., Zherdyev, Yu.V.: Epoxy Polymers and Compositions. Khimiya publishing house, Moscow (1982). (in Russian)
5. Pinprayoon, O.: Microwave thermal analysis of epoxy resins. M.Sc. Thesis, University of Manchester, Manchester (2007)
6. Wallace, M., Attwood, D., Day, R.J., Heatley, F.: Investigation of the microwave curing of the PR500 epoxy resin system. J. Mater. Sci. **41**(18), 5862–5869 (2006)
7. Mezzenga, R., Page, S., Manson, J.A.: Enthalpic, entropic, and square gradient contributions to the surface energetics of amine-cured epoxy systems. J. Colloid Interface Sci. **250**(1), 121–127 (2002)
8. PerkinElmer: Thermal Analysis in F1 Composites and Adhesives [Presentation]. Advance composite training and development centre, Hawarden (2011)
9. Huntsman: Araldite LY 5052/Aradur 5052. Cold curing epoxy systems (2010). http://www.chemcenters.com/images/suppliers/169257/Araldite%20LY5052,%20Aradur%205052.pdf
10. Yusoff, R., Aroua, M.K., Nesbitt, A., Day, R.J.: Curing of polymeric composites using microwave resin transfer moulding (RTM). J. Eng. Sci. Technol. **2**(2), 151–163 (2007)
11. Boey, F., Lee, W.: Microwave radiation curing of a thermosetting composite. J. Mater. Sci. Lett. **9**(10), 1172–1173 (1990)
12. Zhou, J., Shi, C., Mei, B., Yuan, R., Fu, Z.: Research on the technology and mechanical properties of the microwave processing of polymer. J. Mater. Process. Technol. **137**, 156–158 (2003)
13. Kim, W., Oh, H., Kwak, Y., Park, K., Ju, B.-K., Kim, K.: Development of a carbon nanotube-based touchscreen capable of multi-touch and multi-force sensing. Sensors **15**, 28732–28741 (2015)
14. Shulaker, M.M., Rethy, J.V., Hills, G., Wei, H., Chen, H.-Y., Gielen, G., Wong, P.H.-S., Mitra, S.: Sensor-to-Digital interface built entirely with carbon nanotube FETs. IEEE J. Solid-state Circ. **49**(1), 190–201 (2014)
15. Zhou, Y., Azumi, R.: Carbon nanotube based transparent conductive films: progress, challenges, and perspectives. Sci. Technol. Adv. Mater. **17**(1), 493–516 (2016)
16. Todri-Sanial, A., Ramos, R., Okuno, H., Dijon, J., Dhavamani, A., Wislicenus, M., Lilienthal, K., Uhlig, B., Sadi, T., Georgiev, V., Asenov, A., Amoroso, S., Pender, A., Brown, A., Millar, C., Motzfeld, F., Gotsmann, B., Liang, J., Gonçalves, G., Rupesinghe, N., Teo, K.: A survey of carbon nanotube interconnects for energy efficient integrated circuits. IEEE Circ. Syst. Mag. **17**(2), 47–62 (2017)
17. Chen, X., Akinwande, D., Lee, K.-J., Close, G.F., Yasuda, S., Paul, B.C.: Fully integrated graphene and carbon nanotube interconnects for gigahertz high-speed CMOS electronics. IEEE Trans. Electron Devices **57**(11), 3137–3143 (2010)
18. Sharad, M., Roy, K.: Spintronic switches for ultra low energy on-chip and inter-chip current-mode interconnects. IEEE Electron Device Lett. **34**(8), 1068–1070 (2013)
19. Yoo, J.J., Song, J.Y., Yu, J., Lyeo, H.K., Lee, S., Hahn, J.H.: Multiwalled carbon nanotube/nanocrystalline copper nanocomposite film as an interconnect material. In: Proceedings of 58th Electronic Components and Technology Conference (2008)

20. Aryasomayajula, L., Rieske, R., Wolter, J.: Application of copper–carbon nanotubes composite in packaging interconnects. In: Proceedings of 34th International Spring Seminar on Electronics Technology (2011)
21. Chai, Y., Zhang, K., Zhang, M., Chan, P.C.H., Yuen, M.M.F.: Carbon nanotube/copper composites for via filling and thermal management. In: Proceedings of 57th Electronic Components and Technology Conference (2007)
22. Chai, Y., Chan, P.C.H.: High electromigration-resistant copper/carbon nanotube composite for interconnect application. In: Proceedings of IEEE International Electron Devices Meeting (2008)
23. Wang, T., Chen, S., Jiang, D., Fu, Y., Jeppson, K., Ye, L., Liu, T.: Through-silicon vias filled with densified and transferred carbon nanotube forests. IEEE Electron Device Lett. **33**, 420–422 (2012)

Adaptive Threshold Technique for Spectrum Sensing Cognitive Radios Under Gaussian Channel Estimation Errors

Syed Affif Raza Naqvi[1], Aamir Zeb Shaikh[2(✉)], Krishan Lal Khatri[2], Altaf Ahmed Mugheri[2], and Shabbir Ahmed[2]

[1] MS Computer Networking, Ryerson University, Toronto, Canada
affifraza93@gmail.com
[2] Department of Electronic Engineering, NED University of Engineering & Technology, Karachi, Pakistan
{aamirzeb,krishan.lal}@neduet.edu.pk, mugherialtaf@live.com, shabbir_umrani@yahoo.com

Abstract. Spectrum sensing helps cognitive wireless users to gather RF information regarding presence or absence of spectral holes. These spectral holes are not permanent in nature. These are exploited by cognitive users in secondary fashion in such a way that they do not create harmful interference for primary users (PU). Thus, on sudden arrival of a PU, secondary user must vacate those bands for PU because they are high priority users in comparison to cognitive users. The receiver circuit of cognitive radio estimates the received signal and noise parameters and computes a test statistic. This statistic is compared with a pre-set threshold. However, under realistic scenarios, wireless communication channels behave as time-varying entities. Hence, received signal as well as noise varies significantly. The variation in estimated receiver parameters results in deteriorated detection performance for fixed-threshold sensors. In this paper, it is assumed that there are Gaussian estimation errors in received signal. Under this case, an adaptive threshold based testing rule is applied to explore the performance of spectrum sensing radios under adaptive threshold rule. The results clearly recommend the use of proposed algorithm for received signal with Gaussian channel estimation errors. The results show that the proposed method significantly improves the detection performance of the considered cognitive radio i.e. for a false alarm rate of 0.1, the detection probability of the proposed system improves more than 3 times in comparison to the classic cognitive radio under Gaussian Channel estimation errors. The proposed technique can be utilized for future intelligent radios for 5G wireless networks.

Keywords: Adaptive spectrum sensing · Collaborative spectrum sensing Gaussian channel estimation errors

1 Introduction

Evolution of new wireless standards and services require additional bandwidth. However, RF spectrum is almost packed as it is reserved to different users on permanent basis. Hence, the new standards can look towards exploiting unlicensed RF

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 183–189, 2018.
https://doi.org/10.1007/978-3-319-95450-9_15

bands. These bands include both ISM and UN-II. However, due to license-exempt nature of these bands, they are also overtly congested [1]. This apparently poses a great threat to execution of 5G services as well, with one of the prime motive to increase traffic quantity and quality [2]. The spectrum occupancy evaluations by many researchers in various parts of the world [3–7] report that more than 75% of the spectrum remains unutilized. This spectrum can be utilized in secondary fashion to produce virtual unlicensed bands [8]. This suggests one possible way to cater the bandwidth requirements of 5G users. The notion of using RF spectrum in secondary fashion is also supported by Federal Communications Commission (FCC) and also issued a notice to implement the idea through opportunistic secondary usage [9, 10].

The successful implementation of the proposed technology depends on gathering RF information about unused spectral bands and then using those identified bands in opportunistic fashion. The RF information can be shared by a secondary Base Station (BS) or Spectrum Sensing Detection. In the first method, secondary BS collects the information about available white spaces and transmits to secondary users. And in the latter case, all the cognitive sensors sense the spectrum in a distributed fashion. RF sensing can be performed by various algorithms, including coherent detector such as matched filter, cyclostationary feature, energy, autocorrelation, multi-taper spectral estimation method, radio-identification method, waveform transform based estimation, time-frequency analysis, Hough transform and covariance [11, 12]. Of these algorithms, energy detection based spectrum sensing is one of the widely used methods for the case when you do not have the exact knowledge of PU [13, 14]. This is a blind algorithm that requires only the noise variance to compute the probabilities of detection and false alarm. It works faster than coherent sensors (i.e. matched filter) that require complete knowledge of PU to compute detection results.

Conventional detectors used to compare test statistic with fixed value of threshold. However, realistic wireless channels result in significant variation of noise power. Hence, the performance of conventional detectors is highly compromised under low SNR [15]. Because, the threshold depends on various factors that include sensing time [16–18], transmission power of PU [18, 19], and target error probabilities i.e. false alarm and missed-detection [18].

To improve the detection performance under realistic variable-noise environments, many authors have recommended proposals to shift the attention to dynamic threshold setting in comparison to fixed threshold sensing. In [20], authors present an adaptive double threshold based spectrum sensing algorithm to improve detection performance. The decision region is divided into two regions. One that produces confirmed results of 1 and 0 and the other is marked as confused region. Through simulation results, it is shown that double threshold based detection outperforms the conventional energy spectrum detector as well as cooperative spectrum sensor. In [21], authors proposed a multi-threshold spectrum sensing scheme based on phototropism. It is shown that the proposed multi-threshold driven energy detection algorithm performs better than fixed-threshold energy detectors. In [22], authors proposed a three-threshold based energy detection algorithm. The performance of the proposed detector for cognitive radios is compared with conventional CFAR detector as well as adaptive double threshold detector that do not take the account of confused region of operation. In [23, 24], authors optimized the detection threshold of the spectrum sensing algorithms. In [25], authors propose adaptive threshold algorithm for multichannel cognitive radio and in [26] for wideband cognitive radio applications. Additionally, an excellent recursive estimator based spectrum sensor is utilized under cluttered environments to improve performance of the sensing equipment [27].

In this paper, we assume the cognitive sensor to result in sensing values of SNR with Gaussian channel estimation errors. The proposed model considered in this paper is introduced by [27]. We incorporate adaptive threshold based detection rule to improve the performance under estimation errors. The simulation results reveal that the proposed setup gives improved values of detection probability in comparison to the conventional methods.

The Rest of the paper is organized as follows: System model of the proposed setup is presented in Sect. 2 of the paper. Section 3 presents the performance evaluation of the proposed algorithm under IEEE standard for TV White Spaces. Section 4 presents the conclusion of the paper.

2 System Model

Consider, the spectrum sensing framework introduced by [27]. This setup includes the operation of a cognitive sensor under unlicensed RF band. The proposed setup includes one PU and k secondary users. These secondary users are located at the same distance from PU. It is assumed that the wireless channel between PU and sensing node is both fast and slow fading i.e. Rayleigh and Lognormal Shadowing [27]. Authors in [27] use fixed threshold based detection algorithm for detection of spectral holes. We incorporate adaptive detection threshold into the proposed framework and analyze its impact for the performance improvement of the sensing device.

Assuming, the received signal follows simple binary hypothesis i.e. H_0 representing occupied channel and H_1 representing spectral hole. Thus, CR user is only allowed to operate if the sensor results in H_1.

$$y(n) = \begin{cases} \eta(n); & H_1 \\ x(n) + \eta(n); & H_0 \end{cases} \tag{1}$$

In the above equation, $\eta(n)$ represents additive white Gaussian noise while x(n) represents the power transmission by PU. Under null hypothesis, the CR senses the presence of a PU, hence it avoids transmitting on the band, while under alternative hypothesis; it detects the presence white space, hence exploits the RF band in secondary fashion.

Using the derived results of [27], the probabilities of detection and false alarm can be given by

$$\mathrm{P_d} = \mathrm{Q}\left(\Delta\mu \cdot \frac{\sqrt{\mathrm{n}}}{\sqrt{\sigma^2 + \frac{100}{\mathrm{N}\,\ln^2 10}}} + \mathrm{Q}^{-1}(\mathrm{p_{fa}})\right) \tag{2}$$

$$P_{fa} = Q\left(\frac{\lambda - \sigma_n^2}{\sigma_n^2/\sqrt{N/2}}\right) \tag{3}$$

In the above equations, σ_n^2 represents the noise variance. Q(.) is the Gaussian Q function defined by (4)

$$Q(y) = \frac{1}{\sqrt{2\pi}}\int_y^{\infty} e^{-\frac{t^2}{2}}dt \tag{4}$$

$\Delta\mu$ in Eq. (2) represents the difference between the mean values of received signal samples under H_1 and H_0 hypotheses; n shows the number of samples. Under both hypotheses, the variance of signal is represented by $(\sigma^2\sum + \frac{100}{N\ln^2 10}I)$ [27]. Where ln(.) represents the natural logarithm operation, $\sum$ represents the covariance matrix of **y** and **1** is the n × n identity matrix and the $\frac{100}{N\ln^2 10}$ is the Cramer-Rao Lower Bound (CRLB) [27, 28]. Equation (4) shows the Gaussian Q Function. Probability of miss detection is computed as $P_{md} = 1 - P_d$. The higher value of missed detection results in lower chances of exploiting the RF spectrum in secondary manner and higher values of false alarm results in producing harmful interference for PU activity. In the proposed setup it assumed that cognitive sensor periodically senses the noise variance of the channel and incorporates those values in computing threshold of detection. This feature enables the cognitive user to perform better in noise-varying environments. Additionally, based on the estimated data the sensor modifies the threshold (λ) for detection accordingly on periodic basis. Thus, under the proposed setup the detection performance of the sensor results in improved performance as compared to fixed threshold energy detection rule.

3 Performance Evaluation

The performance of the proposed spectrum sensing algorithm is presented in this section. We evaluate the performance of proposed adaptive threshold based collaborative spectrum sensing with Gaussian estimation errors. The proposed adaptive algorithm for spectrum sensing cognitive radios is simulated and analyzed under IEEE standard for TV white spaces i.e. IEEE 802.22. This standard provides useful data that can be incorporated by unlicensed users in broadcast bands for interference-free secondary transmissions. This standard opens up great opportunities for unlicensed wireless users to operate in license-exempt regime. Probability of false alarm is assumed to be 0.01, σ = 2.5, N = 20 as presented by [27].

Figure 1 compares the performance of proposed adaptive threshold sensor with classic fixed detector under Gaussian channel estimation errors. Under fixed threshold energy detection rule, with false alarm rate of 0.1, the sensor results in 5% detection probability. However, adaptive detection threshold produces 25% detection probability. Similarly, for P_{fa} = 0.4, fixed threshold scheme produces 15% detection probability and adaptive threshold rule produces 30%. This shows a consistently improved performance of proposed sensing scheme, producing double values of detection probability in

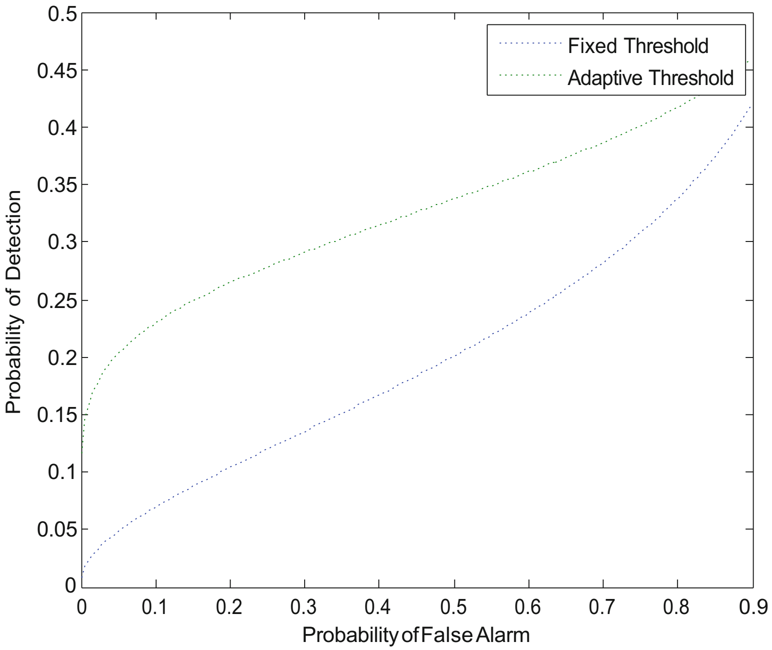


Fig. 1. ROC comparison between Adaptive and Fixed Threshold

comparison to fixed detection rule. This massive increase in detection probability will increase the chances of opportunistically exploiting the RF spectrum in secondary fashion.

4 Conclusion

Adaptive Threshold based detection rule is applied to a collaborative spectrum sensing algorithm with Gaussian channel estimation errors. This problem is incurred in a real wireless environment as compared to the perfect sensing results that is considered to be a theoretical or ideal sensing environment. The performance of the proposed algorithm shows a marked improvement over classical fixed threshold detection rule for sensing with Gaussian estimation errors. For a false alarm rate of 10%, the proposed technique performs more than 3 times better than the classic fixed threshold technique. The results are highly useful as the considered environment is IEEE standard for White Spaces. This environment allows the secondary unlicensed users to access the wireless spectrum in secondary fashion until they don't produce interference to the primary activity of the channel.

Acknowledgment. The authors would like to thank the Administration of NED University of Engineering & Technology, Karachi, Pakistan for providing resources to complete this research. Part of this paper was submitted towards final year project of principal author at NED University of Engg. & Technology, Karachi, Pakistan. Additionally, Usama Bin Ali at Affiniti, TRG Pakistan, Fizza Irfan and Namra Saleem for their valuable discussions with the first author of the paper.

References

1. Zakaria, O.: Blind signal detection and identification over the 2.4 GHz ISM band for cognitive radio (2009)
2. Hong, X., et al.: Cognitive radio in 5G: a perspective on energy-spectral efficiency trade-off. IEEE Commun. Mag. **52**, 46–53 (2014)
3. McHenry, M.A.: NSF spectrum occupancy measurements project summary, Shared spectrum company report (2005)
4. Wellens, M., et al.: Evaluation of spectrum occupancy in indoor and outdoor scenario in the context of cognitive radio. In: 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CrownCom 2007, pp. 420–427 (2007)
5. McHenry, M.A., et al.: Chicago spectrum occupancy measurements & analysis and a long-term studies proposal. In: Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum, p. 1 (2006)
6. Islam, M.H., et al.: Spectrum survey in Singapore: occupancy measurements and analyses. In: 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CrownCom 2008, pp. 1–7 (2008)
7. Hoven, N., et al.: Some Fundamental Limits on Cognitive Radio. Wireless Foundations EECS. University of California, Berkeley (2005)
8. Čabrić, D., et al.: A cognitive radio approach for usage of virtual unlicensed spectrum. In: Proceedings of 14th IST Mobile Wireless Communications Summit, pp. 1–4 (2005)
9. Spectrum Policy Task Force: Spectrum policy task force report et docket no. 02-135, US Federal Communications Commission (2002)
10. FCC, ET Docket No 03-222 Notice of proposed rule making and order, December 2003
11. Yücek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutorials **11**, 116–130 (2009)
12. Axell, E., et al.: Overview of spectrum sensing for cognitive radio. In: 2010 2nd International Workshop on Cognitive Information Processing (CIP), pp. 322–327 (2010)
13. Cabric, D., et al.: Spectrum sensing measurements of pilot, energy, and collaborative detection. In: Military Communications Conference, MILCOM 2006, pp. 1–7. IEEE (2006)
14. Zeng, Y., et al.: A review on spectrum sensing for cognitive radio: challenges and solutions. EURASIP J. Adv. Signal Process. **2010**, 2 (2010)
15. Tandra, R., Sahai, A.: SNR walls for signal detection. IEEE J. Sel. Top. Sign. Process. **2**, 4–17 (2008)
16. Liang, Y.-C., et al.: Sensing-throughput tradeoff for cognitive radio networks. IEEE Trans. Wireless Commun. **7**, 1326–1337 (2008)
17. Nair, P.R., et al.: An adaptive threshold based energy detector for spectrum sensing in cognitive radios at low SNR. In: 2010 IEEE International Conference on Communication Systems (ICCS), pp. 574–578 (2010)
18. Ling, X., et al.: Adaptive threshold control for energy detection based spectrum sensing in cognitive radios. IEEE Wireless Commun. Lett. **1**, 448–451 (2012)
19. Choi, H.-H., et al.: Adaptive sensing threshold control based on transmission power in cognitive radio systems. In: 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CrownCom 2008, pp. 1–6 (2008)

20. Bagwari, A., Tomar, G.S.: Cooperative spectrum sensing with adaptive double-threshold based energy detector in cognitive radio networks. Wireless Pers. Commun. **73**, 1005–1019 (2013)
21. Zhao, X., et al.: A multi-threshold spectrum sensing algorithm based on the phototropism theory for cognitive radio. In: 2014 4th IEEE International Conference on Information Science and Technology (ICIST), pp. 136–139 (2014)
22. Tri, S.C., Jang, J.: A New Spectrum Sensing Scheme using Three Adaptive Thresholds Based on Energy Detection
23. Liu, Y., et al.: Energy detection threshold optimization for cooperative spectrum sensing. In: 2010 2nd International Conference on Advanced Computer Control (ICACC), pp. 566–570 (2010)
24. Chatziantoniou, E., et al: Threshold Optimization for Energy Detection-based Spectrum Sensing over hyper-Rayleigh Fading Channels
25. Gorcin, A., et al.: An adaptive threshold method for spectrum sensing in multi-channel cognitive radio networks, In: 2010 IEEE 17th International Conference on Telecommunications (ICT), pp. 425–429 (2010)
26. Wang, N., et al.: Spectrum sensing using adaptive threshold based energy detection for OFDM signals. In: 2014 IEEE International Conference on Communication Systems (ICCS), pp. 359–363 (2014)
27. Chen, Y., Beaulieu, N.C.: Collaborative spectrum sensing with imperfect gaussian channel estimation. In: Wireless Communications and Networking Conference, WCNC 2009, pp. 1–5. IEEE (2009)
28. Steven, M.K.: Fundamentals of Statistical Signal Processing. PTR Prentice-Hall, Englewood Cliffs (1993)

Survey on Fuzzy Logic Enabled Cognitive Radios

Shabbir Ahmed, Aamir Zeb Shaikh(✉), and Altaf Ahmed

Department of Electronic Engineering, NED University of Engineering and Technology, Karachi, Pakistan
shabbir_umrani@yahoo.com, aamirzeb@neduet.edu.pk, mugherialtaf@live.com

Abstract. Fuzzy Logic is an excellent method to incorporate for making decisions at various levels of cognitive radio under uncertain, incomplete and nonlinear environments. This is one of the most important methods to employ. It is due to the inherent characteristics of wireless channels that produce mostly inaccurate and incomplete information. Thus, the coexisting radios in a particular RF band have to make many decisions using incomplete information especially under cognitive radio access regime. This paper elaborates the concept of fuzzy logic and also investigates a review of the fuzzy logic in the domain of spectrum sensing, power control, resource management for cognitive radio applications. The survey presents the key applications as well as the benefits offered by this technology in comparison to the hard decision making logic i.e.1 and 0 for future wireless communications.

Keywords: Fuzzy logic · Spectrum sensing · Radio resource management Power control

1 Introduction

Fuzzy Logic is employed in various applications of wireless communications to improve decision making especially under uncertain and ambiguous situations and environments [1]. Fuzzy Logic based systems also find useful applications in smart control, decision analysis, signal classification, pattern recognition, transmission systems, and knowledge driven systems for optimization of power and computer vision. This technology converts the subjective knowledge into fuzzy logic that is used to drive systems.

Fuzzy logic introduces rule based system for incorporating human knowledge. Additionally, it can also be used by driving the fuzzy system through an intelligent algorithm such as neural network and evolutionary techniques [2].

Cognitive radio is an evolutionary technology that promises to fulfil imminent need of RF spectrum as part of the next generation 5G wireless systems [3]. This radio uses cognitive cycle to transmit over a white space (spectral hole) [4]. The cognitive cycle defines a set of principle steps that should be followed for successful exploitation of RF spectrum in secondary fashion. These steps include RF spectrum sensing, spectrum decision making and spectrum analysis. Additionally, various other sub steps are also

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 190–196, 2018.
https://doi.org/10.1007/978-3-319-95450-9_16

required for successful utilization of RF spectrum in secondary fashion. The secondary access refers to the fact that the primary access of spectrum is related with the licensed access of the spectrum and the secondary access is related with unlicensed access. Spectrum sensing is the method of detection of RF spectrum bands for secondary access. This step can be employed through various detection algorithms including coherent and non-coherent techniques. The common algorithms which are extensively used include matched filter detection, cyclostationary feature detection and energy detection [5]. Various parameters are considered for selection of a detection algorithm. These include complete information about the primary user, complete information about the noise, partial or incomplete knowledge about primary transmitter or noise and no knowledge of primary user or its noise. Thus, selecting an algorithm is based on various factors. Hence, deciding to use an algorithm is also a problem based on incomplete information.

Classically, one algorithm is used extensively in the literature i.e. energy detector [6]. This method assumes the noise power of the primary environment is known in the absence of the knowledge about the primary user. However, this algorithm in practice cannot be applied to different signal detection scenarios. Hence, fuzzy logic driven decision making processor can be attached to the cognitive radio to help it in making better decision among usage of detector type based on received signal power, noise information and other information i.e. using under licensed band or unlicensed bands. Thus, performance of such a radio can be improved tremendously just by incorporating soft learning into the intelligent secondary radios. In addition to this example there are various other applications where fuzzy logic can be incorporated successfully to improve the performance of next generation radio systems.

The rest of the paper is organized as follows. Section 2 presents the application of fuzzy logic in transmits power control, spectrum sensing techniques, and spectrum analysis and spectrum decision process. Section 3 concludes the paper.

2 Fuzzy Logic Enabled Cognitive Radios

Decision making process finds vital role in the future 5G and 5G+ wireless communication systems. It is due to the changing role of mobile stations from least intelligent in classic 2G communications to the smart and intelligent radios in 4G and beyond. Thus, decision making in 5G radios can be incorporated through fuzzy logic based decision process. There are many key areas where fuzzy logic based radios can be exploited.

2.1 Spectrum Handoff Management

Spectrum Handoff is a channel switching process in which mobile station switches the cellular channel due to coverage or other reasons. For instance, at the boundary of a cell station, handoff process enables a user to continue a call in seamless fashion. It is also known as horizontal handoff. Another type, vertical handoff enables a mobile user to switch between two separate networks to enable the user to continue the active call. Typically, the handoff process is initiated once the received signal becomes weaker than a threshold. Two mechanisms can be enabled by a handoff user i.e. hard handoff and soft handoff. In hard handoff the mobile user disconnects its connection from a base station and connects to another. This feature is employed by 2G communication users. However, soft handoff communications exploited by CDMA users are connected with more than one base station.

In [7], authors use fuzzy logic based algorithm for handoff mechanism in indoor communications. Three parameters of interest are selected to decide the channel. These include received power, cellular user population and the bandwidth of each base station. The result of the proposed scheme is also considered with the classic power enabled handoff strategy. The results advocate the proposed scheme results in uniform population among various base stations. This equal load distribution among base stations results in better utilization of radio resources.

In [8], authors use various QoS parameters to decide regarding selection of a Base station using fuzzy logic based mechanism. The results are also compared with classical power based decision mechanism. The proposed scheme gives better selection of a network for the end user. Additionally, this method also results show an improved performance in the Quality of Service domain, in addition to avoiding repetitive switching among the networks. A QoS enabled fuzzy logic controller system is also proposed by authors in [9] for micro cellular mobile stations. In this system, authors introduce three input parameters for reliable fuzzy output. These include distance between mobile station and base station, received signal power and population on a specific cell.

In [10], authors propose a fuzzy logic based algorithm for handoff scheme that targets to get reduced packet-loss and probability of forced termination in addition to not increasing call-blocking significantly. This system also increases the system accuracy. Additionally, the proposed system provides reduction in location update cycle time.

In [11], authors propose a fuzzy logic based rule to involve a handoff mechanism in future heterogeneous networks. These networks are composed of other networks that include WAN, MAN, LAN and WPAN. Thus, seamless connectivity of a user passing through various wireless networks is an important task. In this paper, researchers also compute the accurate time of initiating a handoff through fuzzy logic. Genetic Algorithm (GA) is also involved in prediction of the rules for the proposed fuzzy logic based system.

In [12], authors propose a fuzzy logic based spectrum handoff algorithm that relies on several features that include received signal strength, traffic load of Base Station, path loss and signal to noise ratio. The proposed algorithm balances the traffic among Base Stations of the cellular network by selecting the best segment of the network. The simulation results show a remarkable improvement in the fair distribution of the load.

2.2 Spectrum Sensing

Spectrum sensing is a key step towards successful implementation of cognitive radio technology. This step enables a cognitive user to gather information about RF spectrum. Sensing can be performed locally as well as in cooperative fashion. Local sensing refers to the individual spectrum sensing performed by a sensor. The techniques for detection in local fashion include energy detection, cyclostationary detection and matched filter detection. In cooperative method, several users coordinate in sensing combining sensing results. This method is highly useful especially to combat hidden node issue in wireless communications [13].

This cooperation can be of two categories i.e. soft combining and hard combining. In soft combining method, complete sensing results are communicated to a fusion center (FC) that combines the results and computes final decision regarding presence or absence of a primary user. In hard combining approach, only one bit results i.e. 1 or 0 are transmitted to a FC that combines these bits to produce final decisions regarding presence or absence of a primary user in a given network. This benefit of improved results is achieved at the cost of cooperative communication overload among different sensing radios. Thus, to improve local spectrum sensing results, without bearing additional overload of cooperation, intelligent algorithms can be involved in spectrum sensing step to provide improved results. In such cases, fuzzy logic based controller can play a vital role.

In [14], authors propose a two-step spectrum sensing technique for improved detection of spectral holes in a cognitive radio network. In first step cognitive sensors detect the presence of primary activity through spectrum sensing algorithms i.e. energy detection, matched filter and feature detector. In the second step all the collected information is combined using fuzzy logic to produce improved decisions regarding presence of a primary user. The simulation results show that the proposed technique produces better detection probability and lower false alarm rate in comparison to the existing algorithms.

In [15], authors propose a fuzzy logic based cooperative spectrum sensing technique for improved detection results. This technique gets benefit of exploiting both cooperative features of the cognitive radios as well as fuzzy logic. Local spectrum sensing is performed through energy detectors. The sensing results are then combined using fuzzy logic system to produce final results regarding presence or absence of a primary user.

In [16], authors exploit the fuzzy logic approach in cooperative spectrum sensing radios to improve spectrum sensing efficiency in cognitive radio applications. This paper assumes the cooperative nature of local sensors to combine the sensing results using selection combining and maximum ratio combining by incorporating fuzzy logic based system. Fuzzy system is introduced to evaluate the energy of sensing radios. This method helps in avoiding the use of channel state information that is essentially required to compute local spectrum sensing algorithms. Additionally, the results of the proposed method suggest an improvement in terms of sensing accuracy.

In [17], authors use cooperative spectrum sensing technique for detection of free spectral slot. An additional benefit of the proposed system is employing trust of local sensors that take part in local spectrum sensing process. Thus, trust established by the sensors is incorporated in computing decisions regarding presence or absence of a primary user through fuzzy logic based system. The computed results show improved performance in comparison to the classical spectrum sensing techniques for cognitive radio applications.

In [18], authors integrate the HMM algorithm with fuzzy-C means clustering to devise an efficient model for the prediction of RF spectrum in Cognitive Radio networks. The simulation results suggest that the proposed technique works better even under higher percentage of failure of secondary users in the cognitive radio network environment.

2.3 Power Control

Power control in wireless communications is a challenging task. By transmission of data over higher power can cause harmful interference to the other users of the network in addition to creating many other issues. However, transmission of data over low power will not close the wireless link. Additionally, power control mechanisms have more important role in cognitive radio applications as well. It is because the cognitive radios are secondary radios and they have to transmit over controlled levels of transmission power. Because, in case these radios transmit over higher transmit powers, the primary radios will be highly affected due to the interference from these radios. In such cases, transmit power control methods will be highly important for future cognitive radio applications.

In [19], authors consider the power control scenario by establishing a cognitive radio secondary network through low power radios that coexists with primary network. The proposed network can work simultaneously without creating any harmful interference to the primary activity as per the guidelines of Federal Communications Commission (FCC). The proposed setup helps in maximizing energy efficiency of secondary network with an added benefit of guaranteed-QoS for both primary and secondary users. Authors also derive the solutions for centralized and distributed setup. Additionally, a joint power and admission control techniques are also provided so that the priority of primary activity shall always be ensured. The simulation results of the proposed technique show the effective utilization of power in secondary users can improve efficiency of radios significantly.

In [20], authors present a theoretical framework for power allocation techniques in cognitive radio applications. The key goals in designing any power control algorithm include QoS protection to licensed activity, opportunities for secondary activity, admissibility of secondary users into primary network and autonomous operation of individual users. Additionally, two more goals which are required rather than mandatory are also considered. These include licensing and adaptability. Furthermore, a duo priority (DPCPC) driven policy is also presented that satisfies the required goals. The performance of the proposed setup results in improved performance in interference-aware paradigm over classical methods of operation in power control mechanisms.

In [21], authors present a power control mechanism for cognitive radios under uncertain channel conditions. Typically, the channel between primary user and the cognitive sensor is unknown. In this paper, authors consider both small scale fading and lognormal fading scenarios. Additionally, authors also introduce uncertainty into the system by considering the fact that primary user is not active all the time. It may switch on and off during the complete cycle of activity. Furthermore, centralized network utility maximization (NUM) problems are presented and solved through sequential programming. Authors also consider a specific case where the channel between two secondary users is considered as uncertain. For such network, outage probability is also computed.

3 Conclusion

A survey of fuzzy logic enabled techniques for future cognitive radios is presented. The applications selected for the purpose are power control, spectrum sensing and handoff management. This investigation summarizes the achievements of several authors in this area. As the incomplete information gathered by the wireless sensors create a hurdle in decision making process, the fuzzy logic enabled techniques can perform faster as well as more reliable. The presented techniques may take a fair portion of the Cognitive Radio equipment in the future generation of wireless radios.

Acknowledgment. The authors thank the university administration of NED University, Karachi that provided required research facilities, without those the research could not be completed.

References

1. Milanés, V., Pérez, J., Onieva, E., González, C.: Controller for urban intersections based on wireless communications and fuzzy logic. IEEE Trans. Intell. Transp. Syst. **11**, 243–248 (2010)
2. Erman, M., Mohammed, A., Rakus-Andersson, E.: Fuzzy logic applications in wireless communications. In: IFSA/EUSFLAT Conference, 2009, pp. 763–767 (2009)
3. Badoi, C.-I., Prasad, N., Croitoru, V., Prasad, R.: 5G based on cognitive radio. Wirel. Pers. Commun. **57**, 441–464 (2011)
4. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**, 201–220 (2005)
5. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutorials **11**, 116–130 (2009)
6. Urkowitz, H.: Energy detection of unknown deterministic signals. In: Proceedings of the IEEE, vol. 55, pp. 523–531 (1967)
7. Dang, M.S., Prakash, A., Anvekar, D.K., Kapoor, D., Shorey, R.: Fuzzy logic based handoff in wireless networks. In: Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo, 2000 IEEE 51st, pp. 2375–2379 (2000)
8. Jie, S., Liangrui, T.: A triangle module operator and fuzzy logic based handoff algorithm for heterogeneous wireless networks. In: 12th IEEE International Conference on Communication Technology (ICCT), 2010, pp. 488–491 (2010)
9. Sati, D.C., Kumar, P., Misra, Y.: Fuzzy logic based handoff controller for microcellular mobile networks. IJCEM Int. J. Comput. Eng. Manag. **13**, 52–62 (2011)
10. Kim, J., Cho, J.-D., Jeong, J., Choi, J.-Y., Song, B.-h., Lee, H.: Fuzzy logic based handoff scheme for heterogeneous vehicular mobile networks. In: International Conference on High Performance Computing & Simulation (HPCS), 2014, pp. 863–870 (2014)
11. Fayyazi, H., Sabokrou, M.: An evolvable fuzzy logic system for handoff management in heterogeneous wireless networks. In: 2nd International eConference on Computer and Knowledge Engineering (ICCKE), 2012, pp. 94–97 (2012)
12. Girma, S.T.: Real time traffic balancing in cellular network by multi-criteria handoff algorithm using fuzzy logic. In: JKUAT-PAUSTI (2018)
13. Arslan, H., Yücek, T.: Spectrum sensing for cognitive radio applications. In: Arslan, H. (ed.) Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems, pp. 263–289. Springer, Dordrecht (2007). https://doi.org/10.1007/978-1-4020-5542-3_9

14. Ejaz, W., ul Hasan, N., Aslam, S., Kim, H.S.: Fuzzy logic based spectrum sensing for cognitive radio networks. In: 5th International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST), 2011, pp. 185–189 (2011)
15. Kieu-Xuan, T., Koo, I.: A cooperative spectrum sensing scheme using fuzzy logic for cognitive radio networks. KSII Trans. Internet Inf. Syst. (TIIS) **4**, 289–304 (2010)
16. Bharatula, S., Murugappan, M.: An intelligent fuzzy based energy detection approach for cooperative spectrum sensing. Circuits Syst. **7**, 1042 (2016)
17. Taghavi, E.M., Abolhassani, B.: A two step secure spectrum sensing algorithm using fuzzy logic for cognitive radio networks. Int. J. Commun. Netw. Syst. Sci. **4**, 507 (2011)
18. Das, S., Acharya, T.: Faulty node detection in HMM-based cooperative spectrum sensing for cognitive radio networks. Comput. J. (2018)
19. Qian, L., Li, X., Attia, J., Gajic, Z.: Power control for cognitive radio ad hoc networks. In: 15th IEEE Workshop on Local & Metropolitan Area Networks, 2007. LANMAN, 2007, pp. 7–12 (2007)
20. Sorooshyari, S., Tan, C.W., Chiang, M.: Power control for cognitive radio networks: axioms, algorithms, and analysis. IEEE/ACM Trans. Netw. (TON) **20**, 878–891 (2012)
21. Dall'Anese, E., Kim, S.-J., Giannakis, G.B., Pupolin, S.: Power control for cognitive radio networks under channel uncertainty. IEEE Trans. Wirel. Commun. **10**, 3541–3551 (2011)

AI, Expert Systems and Big Data Analytics

A Novel Supervised Learning Model for Figures Recognition by Using Artificial Neural Network

Zeyad M. Alfawaer[1(✉)] and Saleem Alzoubi[2]

[1] CCD, Department of Computer Science, Imam Abdulrahman Bin Faisal University, Dammam 31441, Kingdom of Saudi Arabia
zmalfawaer@iau.edu.sa
[2] CSIT, Department of Computer Science, Irbid Private University, Irbid, Jordan

Abstract. Supervised learning has been considered as an important topic as it is used in different fields to exploit the advantages of artificial intelligence. This research introduces a new approach using Artificial neural networks (ANN) to supervise machine learning that enables the machine to recognize a figure via calculating values of angles of the figure, as well as area and length of the line. The research also introduces a processor that would be suitable for the algorithm that uses rotation techniques to specify the best situation in which the figure will be identified easily. This algorithm can be used in many fields such as military and medicine fields.

Keywords: Supervised learning · Figures recognition · Neural network

1 Introduction

Many researchers conducted their work in image recognition of flat regular figure, which was widely used in deferent areas such as robotics, space, communication, telecommunication, medicine, transportation and others [1–8]. Currently, there are plenty of developed methods of image treatment and recognition based on deferent approaches, which have advantages and imperfections [9–16]. The earliest applications of ANNs were published by Pugh [17, 18]. Smith trained an ANN to identify mean and variance shifts [19]. Much of the early research focused on detecting mean and variance shifts using similar approaches to Pugh [17, 18] and Smith [19], including Guo and Dooley [20] and Cheng [21]. Ho and Chang [22] developed an integrated neural network approach for monitoring process mean and variance shifts. Velasco and Rowe [23] demonstrated the potential of ANN application in the analysis of quality control charts. Perry et al. [24] developed two back propagation ANNs for the detection of trends, mixtures, cycles and systematic variation.

Supervised learning has been a great success in real-world applications. This type of learning is analogous to human learning from past experiences to gain new knowledge in order to improve our ability to perform real-world tasks. However, since

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 199–208, 2018.
https://doi.org/10.1007/978-3-319-95450-9_17

computers do not have "experiences", machine learning learns from data, which are collected in the past and represent past experiences in some real-world applications.

We identified various mathematical tools to study the figure recognition by using neural network collected from various sources. The mathematical tools identified was implemented using VC++ programming language.

2 Supervised Learning Model for Figures Recognition

2.1 Processing and Recognition of Images

We consider the principle of parallel functioning of devices in processing and recognition the figures in the environments of Cellular neural network (CNN). Process of recognition of flat figure image is carried out by its transformation at the entrance of the device to a form most convenient distinguishing necessary attributes, and formation vector V of attributes from them. The given process is presented by the following model.

$$\boldsymbol{I} \overset{T_0}{\Rightarrow} \boldsymbol{I}_{cp} \overset{T_1}{\Rightarrow} \boldsymbol{I}_M \overset{T_M}{\Rightarrow} \boldsymbol{V} \tag{1}$$

Where $\boldsymbol{I}$ - the initial image at the entrance of the device; $\boldsymbol{I}_{\mathrm{cp}}$ - the initial image, which is written down in electronic multiprocessing matrix environment; $\boldsymbol{I}_{\mathrm{M}}$ - set of images $\{\boldsymbol{I}^1, \ldots .\boldsymbol{I}^{\mathrm{k}}\}$, any of which is intended for obtain of information on the chosen attribute; V – a vector which contains necessary attributes for obtaining the most complete information on the image which is recognized.

$\mathbf{T}_0$ - Operation of transformation of initial image **I** into the image of CNN environment $\boldsymbol{I}_{\mathrm{cp}}$, T_1 – operation of transformation of $\boldsymbol{I}_{\mathrm{cp}}$ (Iav) into set of image $\boldsymbol{I}_{\mathrm{M}} = \{\boldsymbol{I}^1, \ldots .\boldsymbol{I}^{\mathrm{k}}\}$, which form provides the most or that class and within a class is determined. T_{M} – set of operations over the set of images $\boldsymbol{I}_{\mathrm{M}}$, which are oriented on determination of corresponding attribute. Obtained vector V is compared with reference and its identity with this or that class and within a class is determined. Accuracy of recognition depends on a choice of the set of necessary attribute, and also on the accuracy of determination of their quantitative characteristics. The vector of the attributes necessary for recognition of images looks like.

$$V = \left\langle S, N, \alpha_i, l_{1,2}, l_{2,3}, \ldots, l_{N-1,N}, l_{N,1} \right\rangle \tag{2}$$

Where S – the area of the figure at the input of the system; which is measured in corresponding dimensions of individual discrete environments; N – the number of peaks of the figure at the input of the system. It is measured in corresponding dimensions of individual discrete environment; $\mathbf{l}_{\mathrm{i,i+1}}$

The relation between i-th and (i + 1)-th peaks which is expressed by the length, taking into account the geometrical sizes of individual discrete environment; α_{i} - An angle between two neighboring sides in i-th peak i = 1, N. For exact allocation of element of the contour of the figure is used.

$$\boldsymbol{I}_K = \{a^1_{i,j} \,/\, a^1_{i,j}\ \boldsymbol{R}_K\, a^K_{i,j}\} \tag{3}$$

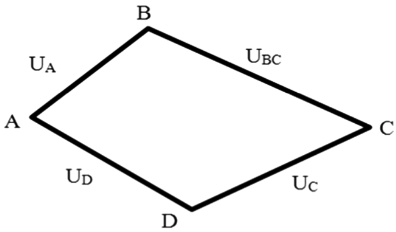
Where $a^k_{i,j}$ – element of the environment, which belongs to limits of the element $a_{I,j}$; R_k - relations which is set between discrete count $a^1_{i,j}$ and element $a^k_{i,j}$. The side is presented as the set of points, which belong to the contour, between two next peaks.

$$\boldsymbol{I}_p(m) = \left\{a^1_{i,j}(m)\right\} \tag{4}$$

Where $a^1_{i,j}(m)$ – points, which belong to the contour of the image between next m-th and (m + 1)th peaks. The peak is presented as

$$\boldsymbol{I}_b(m) = \{\boldsymbol{I}_p(m-1) \cap \boldsymbol{I}_p(m)\} \tag{5}$$

The figure example which is shown in Fig. 1, can be expressed as the set of relations R*i* that represent the contour of the figure in matrix form.



R_l	A	B	C	D
A		l_{AB}		l_{AD}
B	l_{BA}		l_{BC}	
C		l_{CB}		l_{CD}
D	l_{DA}		l_{DC}	

R_α	U_{AB}	U_{BC}	U_{CD}	U_{DA}
U_{AB}		α_B		α_A
U_{BC}	α_B		α_C	
U_{CD}		α_C		α_D
U_{DA}	α_A		α_D	

Fig. 1. Representation of the quadrangle by sides and peak.

The algorithm of determination of image peaks of the figure in CNN, are considered and also the data for determination of peaks at different approaches are studied.

2.2 Supervised System Structure

The structure of the supervised system as shown in Fig. 2 mimic the principle of learning in humans where the nerve cells conduct primary treatment only, and then send it to the processor to supervise and store in main memory if it was the first time you enter into the system, this method is the best and the easiest and least expensive in addition to being supported the principle of parallelism.

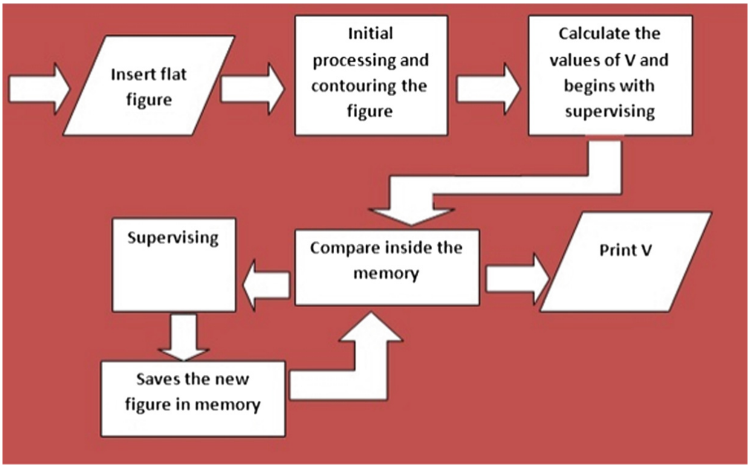


Fig. 2. Supervised system structure

Definition 1. Peak of the image of the flat figure designed in CNN environment is PE, that submits the point of the contour provided that the sum of values of the neighboring PE, which submit internal area of the figure is not equal to the sum of values of the next PE, that do not belong to the image.

$$b_{i,j} = \begin{cases} b^b_{i,j} = 1, & \textit{if } \sum\limits_{n=1}^{8} \left(a^{1,k}_{i,j}(n) - a^{0,k}_{i,j}(n) - a^{1,k'}_{i,j}(n) \right) \neq 0 \\ b_{i,j} = 0, & \text{in the other case} \end{cases} \tag{6}$$

Where $a^{1,k}_{i,j}(n)$, $a^{0,k}_{i,j}(n)$, - elements of environment which belong to limits of PE and take value accordingly, of logic "1" and "0"; $a^{1,k'}_{i,j}(n)$ - elements of limits $PE_{i,j}$, that belong to the contour; $b^{b}_{i,j}$ - value $PE_{i,j}$ in the peak. But the given definition is valid in case of presence of two neighboring cells of the contour. For filled figures the following definition is also valid.

Definition 2. Peak of the image of flat figure projected in CNN environment is PE, which submits the point of contour, provided that the sum of values of neighboring PEs, which present internal area of the figure equals 1.

$$b_{i,j} = \begin{cases} b^b_{i,j} = 1, & \textit{if } \sum\limits_{n=1}^{8} \left(a^{k}_{i,j}(n) - a^{1,k'}_{i,j}(n) \right) \neq 0 \\ b_{i,j} = 0, & \text{in the other case} \end{cases} \tag{7}$$

For definition of convex and concave peaks the following definition are introduced.

Definition 3. The cell in convex peak, if the sum of values of all eight neighboring cells of environment, which belong to Moore's limit is less than 5.

$$b_{i,j} = \begin{cases} b^b_{i,j} = 1, & if \sum_{n=1}^{8} a^k_{i,j}(n) < 5 \\ b_{i,j} = 0, & \text{if in other case} \end{cases} \quad (8)$$

Definition 4. The cell is a concave peak in case when the sum of values of all eight neighboring cells of environment, which belong to Moore's limit is greater than five.

$$b_{i,j} = \begin{cases} b^b_{i,j} = 1, & if \sum_{n=1}^{8} a^k_{i,j}(n) > 5 \\ b_{i,j} = 0, & \text{if in other case} \end{cases} \quad (9)$$

At the account of aliasing for exact allocation of peak the additional layer Fig. 3 is introduced and the following models are used. Cell is a peak which is determined in the second layer in case if the following condition in valid

$$C = \begin{cases} C^b = 1, & if \quad \mathrm{AB}(\mathrm{X}_1(\mathrm{X}_6 + \mathrm{X}_8) + \mathrm{X}_2(\mathrm{X}_6 + \mathrm{X}_7) + \mathrm{X}_{12}(\mathrm{X}_7 + \mathrm{X}_8)) = 1 \\ C^b = 0, & if\ the\ other\ case \end{cases} \quad (10)$$

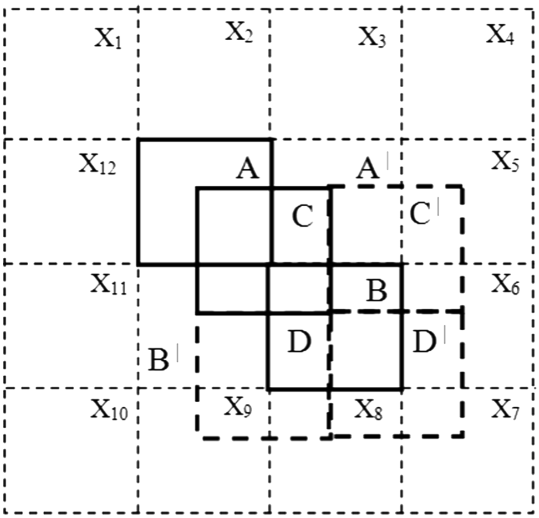


Fig. 3. The example of 2-layered environment with marked all.

The cell B (A) is peak which is determined in the first layer in case if the following condition is valid.

$$C = \begin{cases} \mathrm{B}^b = 1, & if\ C \cdot C' + D \cdot D' + C \cdot D + C' \cdot D' = 1 \\ \mathrm{B}^b = 0, & in\ the\ other\ case \end{cases} \quad (11)$$

The example of allocated peak is presented in Fig. 4 by the result of operations of preliminary processing of images occurrence of jamming cells which are divided into individual. Any of them is presented by logic depending on which is applied. Taking this into algorithm of removal of jamming cells in CNN are presented. Individual jamming cells are eliminated by realization of the following logic expression.

$$
\begin{aligned}
b_{ij}(t+1) &= b_{i+1,j}(t) \vee b_{i+1,j-1}(t) \vee b_{i+1,j+1}(t) \vee b_{i-1,j}(t) \vee b_{i-1,j-1}(t) \vee b_{i-1,j+1}(t) \\
&\quad \vee b_{i,j-1}(t) \vee b_{i,j+1}(t) \\
&= 0
\end{aligned} \tag{12}
$$

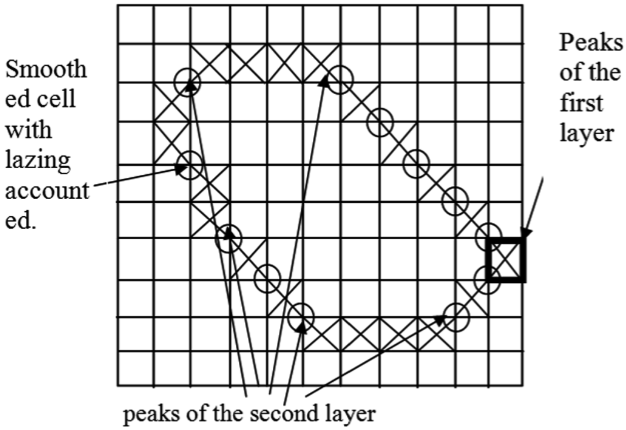


Fig. 4. The example of elimination of peaks in 2-layered structure.

2.3 Determination of Jamming Cells

Removal of the connected jamming cell is determined by the sum of values of the neighboring cells which is described in the following way.

$$
b_{i,j}(t+1) = \begin{cases} 0, & if \ \sum_{n=1}^{8} a^{k}_{i,j}(n) = 1 \\ b_{i,j}(t), & in\ other\ case \end{cases} \tag{13}
$$

Definition 5. Cell P is the contour jamming cell in case when: it's neighboring to two cells of the contour (one of these cells is activated) vertically and horizontally, and they (two cells of the contour) are neighboring by diagonal between each other, and have other neighbors; It has three neighboring cells that belong to the contour and they are orthogonal; If has three neighbors, one of them is diagonally activated, and two non-diagonal neighbors (they are neighbors by diagonal) are not activated. It has only one neighbor that belongs to the contour. To eliminate such cells the limits of the second order are used, the cells of this limit are the closest neighbors of the cells of the limits of the first order, and the cells that are in the shortest neighborhood are determined.

Definition 6. Three cells which belong to the contour of the image of the figure are in the shortest neighborhood if one of them is neighboring for others which are not neighbors between themselves and have no neighbors that are not neighbors for the common cells processing.

Definition 7. The cell is a jamming cell in case when it's in the neighborhood with cells which are in the shortest neighborhood between themselves. Definitions which characterize relations between the sides consist in development of algorithm s of definition of angles in peaks for this purpose the figure is divided into the triangles, two sides of any of them belong to the contour, and the following formula is applied.

$$\cos\,\alpha = \frac{(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)}{l_1 \cdot l_2} \tag{14}$$

For definition of angles the methods of definition of the neighboring peaks which form the angles is used and the method implies the shift of figure image to the extreme left column, where neighboring peaks are determined. The algorithm is the most effective in case when peaks are known, but their location is unknown.

3 Simulation Result and Discussion

The insertion of the flat figure to the nerve cells in deferent ways to recognize it quickly and accurately by using the algorithm as shown in Fig. 5 then contouring the figure to recognize it easily as shown in Fig. 6. Then rotate the figure to insure that the rotation process did not change the characteristics of the initial figure as shown in Fig. 7 so the changed cells after the rotation will be processed by the 2-layered structure as we mentioned before. In Fig. 8 the rotations was tested after each 5° and then the result table has been printed to choose all possible vector values in the neural network.

Fig. 5. The inserted figure to the network

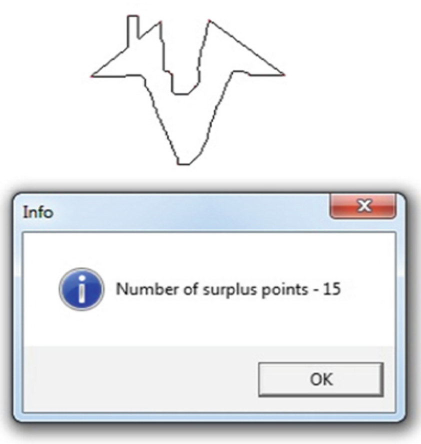


Fig. 6. The figure after contouring process.

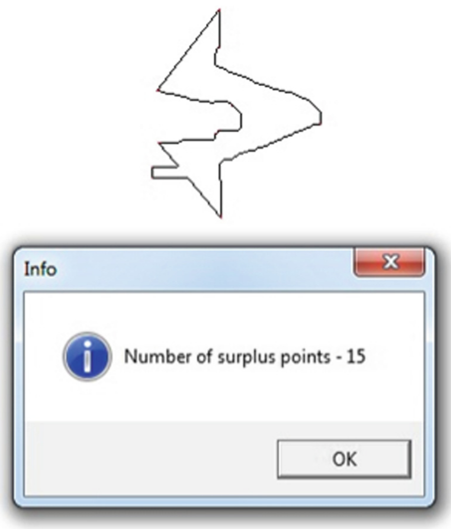


Fig. 7. The figure after rotation process.

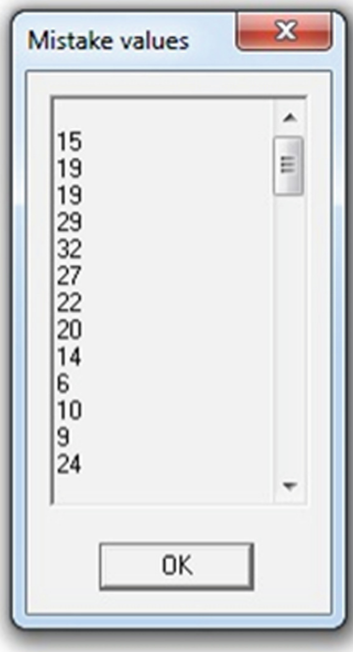


Fig. 8. Rotate the figure 360° after the contouring.

4 Conclusion and Future Enhancements

This research introduces a new approach using neural networks to supervise machine learning that enables the machine to recognize a figure via calculating values of angles of the figure, as well as area and length of the line. The proposed algorithm uses rotation techniques to specify the best situation in which the figure will be identified easily. The proposed algorithm can be used in many fields such as military and medicine fields. The research also introduces a proposed processor that would be suitable for the proposed algorithm. The study is limited by one type of figures and should make enhancement to convert any kind of image to be examined.

References

1. Sylwester, R.: Human and machine intelligence: an intriguing perspective, February 2005
2. Jain, S., Osherson, D., Royer, J.S., Sharma, A.: Systems That Learn: An Introduction to Learning Theory, 2nd edn. The MIT Press, Cambridge (2005)
3. Nilsson, N.J.: Artificial Intelligence: A New Synthesis. Morgan Kaufmann Publishers (2008). ISBN 1558604677
4. Zytkow, J. (ed.): Machine Discovery. Reprinted from FOUNDATIONS OF SCIENCE, 1:2 (1997). 150 p., Hardcover. ISBN 978-0-7923-4406-3

5. Fukushima, K.: Neural network for visual pattern recognition. Computer **21**(3), 65–115 (1983)
6. Hecht-Nielsen, R.: Replicator neural networks for universal optimal source coding. Science **269**, 1860–1863 (1995)
7. Galushkin, A.I., Luskinovich, P.N., Nesmeyanov, S.S., Nikishin, V.I., Frolov, V.D.: The quantum neurocomputer. J. Br. Interplanet. Soc. **47**, 331–333 (1994)
8. Cosman, P.C., Gray, R.M., Olshe, R.A.: Evaluating quality of compressed medical images. In: Proceedings of the IEEE "SNR, Subjective Rating, and Diagnostic Accuracy", vol. 82, no. 6, pp. 919–932 (1994)
9. Dash, L., Chatterji, B.N.: Adaptive contrast enhancement and de-enhancement. Pattern Recogn. **24**(4), 289–302 (1992)
10. Impedovo, S., Dimauro, G., Pirlo, G.: Off-line signature verification by fundamental components analysis. In: Proceedings of the 7-th ICIAP, Bari, Italy (1993)
11. Bartenek, N.: The role of handwriting recognition in future reading systems. In: Proceedings of the Fifth International Workshop on Frontiers in Handwriting Recognition, Univ. of Essex, England, pp. 147–165 (1996)
12. Shumann, J., et al.: Document analysis – from pixels to contents. Proc. IEEE **80**, 1101–1119 (1992)
13. Bilan, S.M., Koval, D.M.: Filling regions on the base of cellular aperiodical neuroautomatons. In: Bulletin of MITI, no. 10, pp. 166–173 (1999)
14. Bilan, S.M., Motornyuk, R.L.: Homogeneous cellular structures for images' segmentation and marking centers of segments out and modelling them into the medium Active –HDL. In: Bulletin of VPI, no. 5, pp. 55–58 (2002)
15. Montessori, M., Gutek, G.L.: The Montessori method: the origins of an educational innovation (2004)
16. Fu, K.: Structural methods in recognition of images. Mir, M., 320 p. (1977)
17. Pugh, G.A.: Synthetic neural networks for process control. Comput. Ind. Eng. **17**(1–4), 24–26 (1989)
18. Pugh, G.A.: A comparison of neural networks to SPC charts. Comput. Ind. Eng. **21**(1–4), 253–255 (1991)
19. Smith, A.E.: X-bar and R control chart interpretation using neural computing. Int. J. Prod. Res. **32**(2), 309–320 (1994)
20. Guo, Y., Dooley, K.J.: Identification of change structure in statistical process control. Int. J. Prod. Res. **30**(7), 1655–1669 (1992)
21. Cheng, C.S.: A multi-layered neural network model for detecting changes in the process mean. Comput. Ind. Eng. **28**(1), 51–61 (1995)
22. Ho, E.S., Chang, S.I.: An integrated neural network approach for simultaneous monitoring of process mean and variance shifts – a comparative study. Int. J. Prod. Res. **37**(8), 1881–1901 (1999)
23. Velasco, T., Rowe, R.: Back propagation artificial neural networks for the analysis of quality control charts. Comput. Ind. Eng. **25**(1–4), 397–400 (1993)
24. Perry, M.B., Spoerre, J.K., Velasco, T.: Control chart pattern recognition using artificial neural networks. Int. J. Prod. Res. **39**(15), 3399–3418 (2001)

Cluster and Logistic Regression Distribution of Students' Performance by Classification

Nareena Soomro[1(✉)], Fahad Razaque[1], Safeeullah Soomro[2], Shoaib Shaikh[3], Natesh Kumar[4], Ghulam e Mustafa Abro[3], and Ghulam Abid[3]

[1] Department of Computing, Indus University, Karachi, Sindh, Pakistan
nainee_soom@yahoo.com, fahad.indus1337@gmail.com
[2] College of Computer Studies, AMA International University, Salmabad, Kingdom of Bahrain
s.soomro@amaiu.edu.bh
[3] Hamdard University, Karachi, Sindh, Pakistan
skshaikh@outlook.com, mustafa.abro@hamdard.edu.pk, engrabid1246@gmail.com
[4] Usman Institute of Technology, Karachi, Sindh, Pakistan
nateshsolanki1992@yahoo.com

Abstract. In the research cluster based logistic regression model on student result, performance at computing department, and other demographics to predict whether or not student will annually enroll if admitted that help the campus administrators to manage registrations. In this study, deals with performance and analysis of examination results' performance of students from computing department by also establishing general assessment. However, it cannot be stand-alone and only serves to compliment campus administrator of decision making procedure to manage registrations effectually. Predict students of educational performance are critical for scholastic departments because planned program can be scheduled in maintaining performance of students during their period of studies in departments. The demographic profile of students and fourth year of academic are used as predictor variable for performance of students educational in academic program.

Keywords: Logistic regression · Performance · Cluster · Probability

1 Introduction

Research purpose is data mining method's efficiency apply in education and benefit educational department better use this method to notify student graduation policies. Advance university standing, improved student retention foremost to graduation recovers admission executive and reduce on recruiting costs. From side of student, retention chief to qualification was social, own and financial insinuations. This study's purpose was a variable's prediction that needs influence on educational student's performance was significant as auxiliary programs could be applied to avoid failures. It perceived relationship between the possibility of educational failure and level of

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 209–219, 2018.
https://doi.org/10.1007/978-3-319-95450-9_18

knowledge in software engineering and computer science [9]. In this research, logistic regression model to predict computing students' performance in their four years using educational, mental and professional learning and motivational policies as variables. It was examined interactions between student and personal contextual characteristics, educational preparation and performance traits, quantitative education by logistic regression was conducted to scrutinize association between variables and ability to predict student persistence in academic. It should do a better job of summit students' wants and sighted them down to degree achievement. It is significant for the future of students, higher education, and society as a whole.

Educational organizations are progressively interested in intensive care act of their students, which contributes increase to necessity to investigation, collate, scrutinize and interpret data, in order to have proof to notify an educational strategy that was formulated to progress student's performance, excellence training and support resources; producing involvement policies to mitigate factors that will definitely influence student performance.

The core research of department of Computing and Technology is to deliver quality learning to students and to advance quality of decision-making. The approach to accomplish quality level in academic is by learning information from instructive data to study key attributes that might affect performance of students. It could be used to suggestion supportive and positive references to educational organizers in department to improve decision making process, educational performance of students, teaching and reduce failure ratio, to well recognize behavior of students, to assist lecturers, and various other profits [13].

Study on predicts feature causative to students of educational success would be helpful to educational area, communal and others who is concerned with improving performance of students throughout universities times.

Academic Data Mining was one of emerging field which comprise procedure of examined students' details by different elements such as earlier semester marks, attendance, assignment, discussion, lab work were of used to improved bachelor academic performance of students, and overcome difficulties of low ranks of students [14]. It was extracted useful knowledge from academic students data collected from department of Computing. Subsequently preprocessing data, which was applied data mining techniques to discover classification and clustering and outlier detection. In this study, classification method was described which based on K mean algorithm and Cluster based logistic regression model for predicting the students' efficiency of academic.

2 Methodology

Figure 1 demonstrates of data preparation and data pre-processing contain data set that taken student's data from department of computing. The data preparation determination was examined and transformed raw data in order to create them mean more and improved data quality. Without data preparation, hidden information was not easily accessed using data mining models [15]. Data Pre-processing step was executed to develop excellence of data set through removed incomplete values. Data set considered, 61 records were removed from 660 entire data and simply 591 records were prepared for data mining method later. When pre-processing method applied on data set, 91 records were eliminated from data set of 591 records which left only 500 clean records. The total, data of student comprised 160 missing values in numerous parameters from 660 records was ignored from data set. The total numbers of records was reduced to 500 [14].

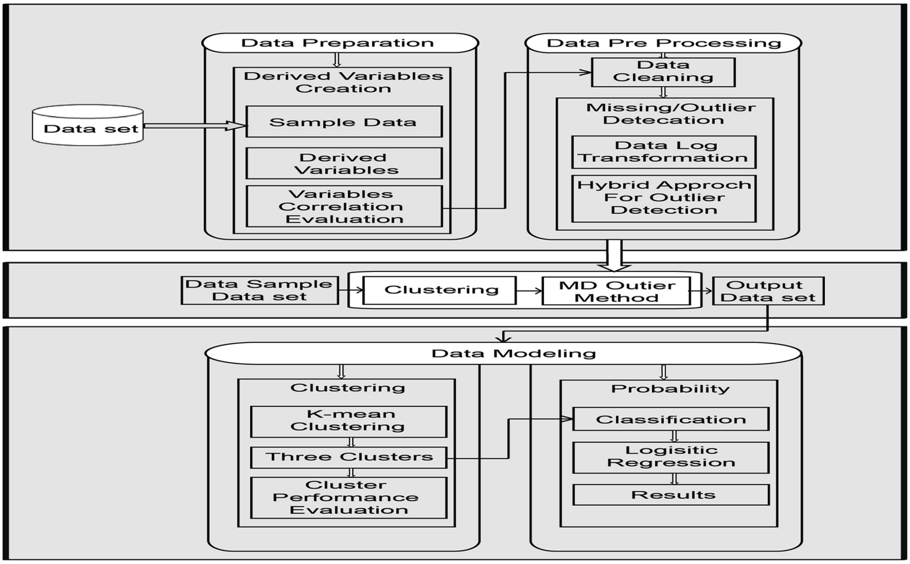


Fig. 1. Research schema of data analysis

It shows that data preparation to expose further concealed info in data by making variables, and to clean raw data set by using the method of hybrid outlier detection. Data set is organized for modeling; adapt clustering and predictive models to execute the analysis. K-means clustering method is modified to divider sessions into three player groups, at same the time, the target variable used in predictive models is also generated. Model of logistic regression is practical to recognize which behavioral pointers are highly related with gambling addiction because of its highest total accuracy.

Clustering was recognized descriptive model, which dividers data into clusters set, such as remarks with parallel appearances were assembled. Hence, the cluster was collected of data items which were parallel to each one but disparate to those facts in further clusters [2]. A good cluster model was hypothetical to ensure that intro cluster resemblance was high, however intercluster resemble ought to be low [3]. Numerous diverse cluster procedures were developed, but furthermost extensively used was k-means algorithm that was also used in the study.

K-means attempted to divider n remarks in a data set into k number of clusters in which every remark belongs to cluster with nearest centroid [2]. Additionally, clustering was frequently significant beginning point to other forms of data modeling [6].

In this study, outlier detection and deletion is most significant task in phase of data preparation as researchers identify numerous irrational items in the data set. The presence of those irrational data points will introduce complexity into data models, and finally reach specious deductions. Because of conduct review and compare different methods in order to discover a most appropriate technique. Outlier referred to those data points that were substantial unrelated in data set. While many outlier detection approaches was proposed, most of them can be classified into four kinds, which are distributed, density, distance and clustering based. Distribution method, for example Standard Deviation is mainly applied to deal with univariate data set [5], but the data set is multivariate with several variables.

MD analysis follows Chi-Square distributions, which have critical values table is used as a means to determine threshold that is decided by significance level (p) and degrees of freedom (df). The Significance level is usually set at 0.05 ($p = 0.05$), which is most commonly used number and has already been accepted as a standard by researchers [12]. Different Chi-Square test, MD is evaluated with degrees of freedom equal to the number of independent variables involved in the calculation ($df = n$) [8]. Though the MD approach has been commonly used, some researchers pointed out that it is not appropriate to deal with outliers in a large data set, since the distance between observation and center of the whole data set needs to be calculated which increase the computation time but decrease the accuracy [10].

Probabilistic classifier that is able to predict, input observation, a probability distribution over classes set, instead of output most likely class that observation has to belong. It delivered classification that could be beneficial in own right [4].

Regression models were typically adapted to discover which predictors were highly linked to variable, and how modifications of predicators affect aim variable. It was most operative when it was used to predict data set taking big quantity of observations but trivial variables number. Moreover, regression models effort fine to predict data set when predicators and variable have underlying association and modifications between them was estimated to be predictable [1].

A regression model was linear regression and logistic regression model usually used. In investigation, logistic regression instead of linear regression as latter was primarily applied to predict association between single input variable and aim variable with category [6].

3 The Graduate Students Data Set and Preprocessing

The data set comprises graduate students' information composed of Department of computing and Technology. The student's data set as sample data contains about 500 records and 13 attribute. Table 1 shows the attributes, description and the possible values that exist in the data set.

The department of Computing and Technology awarded their graduate bachelor degree included two areas for bachelor degree in Computer Science and Software

Table 1. The graduate students data set description

Variable	Description	Possible values
Faculty	The name of the faculty	Computing
Computing program	The name of the program	BS (Computer Science), BS (Software Engineering)
Bachelor academic year	The year of academic	1st Year, 2nd Year, 3rd Year, 4th Year
H.S.C or equivalent study medium	The type of medium	Urdu, English
1st year semester 1	Semester 1 (GPA)	GPA (1.00 to 4.00)
1st year semester 2	Semester 2 (GPA)	GPA (1.00 to 4.00)
2nd year semester 1	Semester 1 (GPA)	GPA (1.00 to 4.00)
2nd year semester 2	Semester 2 (GPA)	GPA (1.00 to 4.00)
3rd year semester 1	Semester 1 (GPA)	GPA (1.00 to 4.00)
3rd year semester 2	Semester 2 (GPA)	GPA (1.00 to 4.00)
4th year semester 1	Semester 1 (GPA)	GPA (1.00 to 4.00)
4th year semester 2	Semester 2 (GPA)	GPA (1.00 to 4.00)
Classes are mostly	The procedure of class	Lecture & discussion, lecture & lab, lecture, discussion & practical lab, lecture based

Engineering. The data preparation and preprocessing of data set and to get better input data for data mining techniques, It was some preprocessing for composing data earlier loaded data set of software of data mining, immaterial attributes was removed. The elements as selected as shows in Table 1 were treated by the rapid miner software to apply the data mining approaches on them.

Figure 2 shows that two axis X and Y which was X include academic year and y contains that distinguished actions of the centers for all three clusters (C1, C2, C3) using K-mean algorithm in same dataset during the academic years. It is showing that movements of the three clusters' centers (shown as symbols circle, cross and triangle) are volatile and they heavily depend on the random choice of academic year (shown as colored circles, blue, red, green and orange). C1 Student ratio highest in first year, C2 Student ratio highest in second year and fourth year, and C3 Student ratio highest in fourth year from academic year.

Predictor of educational failure or success such as statistically important ($p \leq 0$) Fig. 3 indicate probability of achievement according to different values of continuous variables where probability of achievement is directly proportional to score obtained in

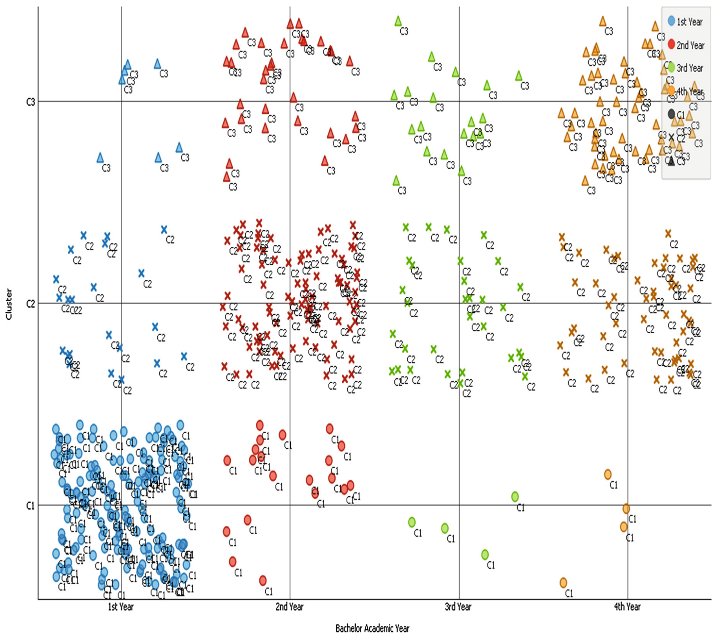


Fig. 2. Strength of students clusters wise using k mean. (Color figure online)

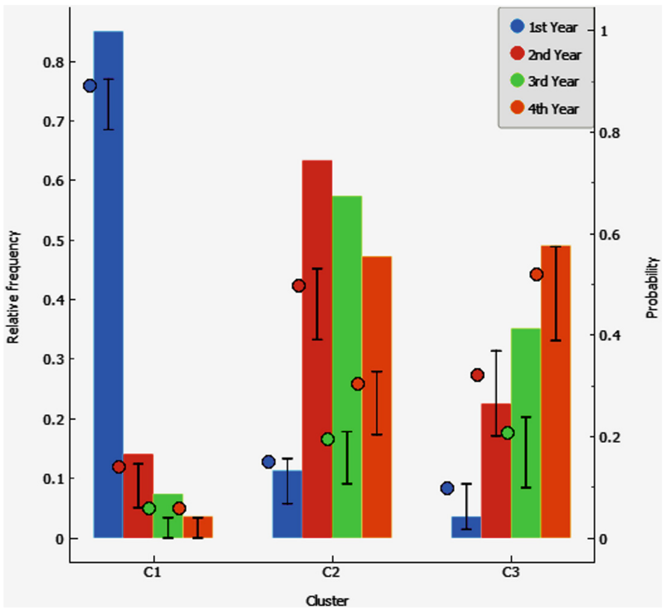


Fig. 3. Distribution of 'Cluster' grouped by 'Bachelor Academic Year' (relative probability)

academic. Examining three clusters is different importance in cluster analysis that distinguishable performing successfully [7]. Essential assumption in relation to investigation is that data must be approximately distributed in year wise using K mean cluster as C1 contain first year probability was 0.8, C2 have second year 0.7 probability and third year 0.64 probability, and C3 in fourth year 0.59 probability.

Figure 4 illustrate that number of students was 500 in clusters as logistic regression classifiers predicted in percentage of student as C1 96.9%, C2 98.5% and C3 97.3%. C2 have highest percentage that was best cluster. It helps to find out which behavior are highly related to dependence by identifying which predictor variables contribute more to target variable. Figure 5 shows that two axes X and Y which was Y include academic year and X contains cluster (C1, C2, C3) that was proved as C2 best cluster because student ratio is high.

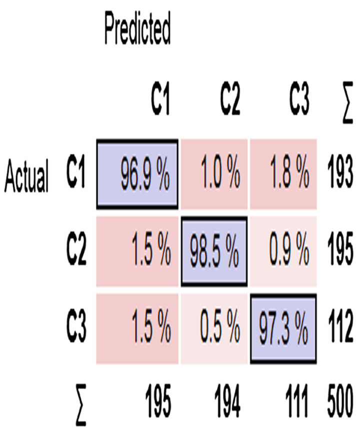


Fig. 4. Logistic regression (showing proportion of predicted)

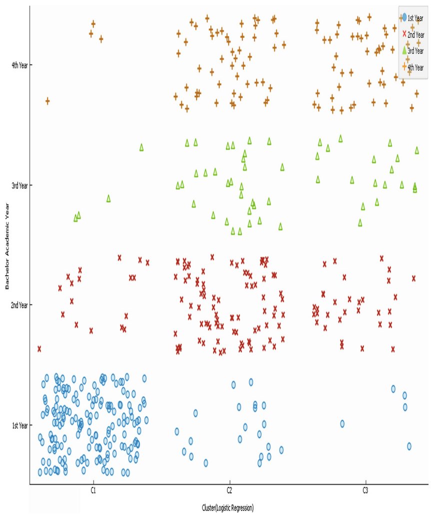


Fig. 5. Classify clusters (c1 c2 c3) using logistic regression

However, it needs to check whether the logistic regression is statistically significant by using the relative frequency ratio. If the probability in the graph is less than 0.05, the Cluster is statistically significant and the predicator variables have an impact on the target variable. Based on the result, the regression is finally determined to be applied for cluster analysis that was C2 cluster probability ratio high (Fig. 6).

The comparison of clusters by performance indicators assists to find out unique behavioral characteristics of each cluster and therefore to differentiate them. Typically, researchers compare the cluster center of each variable within each cluster. But we consider that comparison of cluster center is not sufficient since cluster center only indicates the behavior of academic years in program and is highly affected by extreme

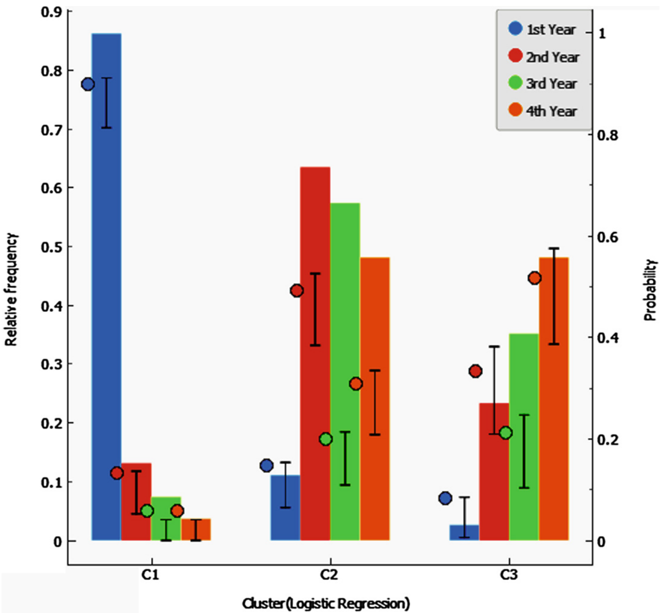


Fig. 6. Clarify the clusters (c1 c2 c3) of probability through logistic regression

values within each cluster. Figure 7 shows that Number of clusters as logistic regression classifiers reaches almost the same outcomes in terms of C1, C2 and C3. By further looking at the logistical regression C1 is slightly better than C3. However, it is hard to determine which cluster is better on basis of slight difference of cluster between each logistic regression classifier.

		Predicted			
		C1	C2	C3	Σ
Actual	C1	97.2 %	0.9 %	2.3 %	212
	C2	0.9 %	96.5 %	0.6 %	112
	C3	1.9 %	2.7 %	97.1 %	176
	Σ	213	113	174	500

Fig. 7. Probability measured cluster wise (C1, C2, C3).

In Fig. 8 logistic regression C1 (blue circle) at height zero and C2 (red cross) at height one. Every cluster pushes on distribution, though not with equal force. The C2 push dividing line towards the C1 and C1 push it back towards the C2, that logistic

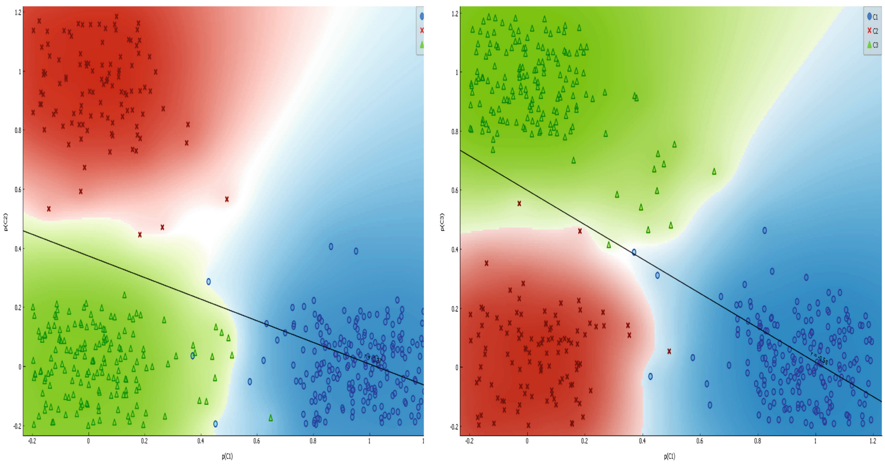


Fig. 8. Logistic regression clusters difference of P(C1) and P(C2), P(C1) and P(C3). (Color figure online)

regression algorithm selects could be thought of as equilibrium clusters of all these forces. The distribution on lift includes blue circle in green part of distribution and vice versa. The distribution on the right does a much better job of matching blue to blue and green to green, so this would be closer to probability of cluster 1 (P(C1)) and probability of cluster 3 (P(C3)) chosen by logistic regression.

Distribution is close to plane on one side of line, but below plane on other side, cross section of distribution is curve as logistic. Logistic regression is blue dots line at height one, green and red dots lines points at height zero, distribution that minimizes distances from distribution function based on logistic curve to dots lines and in plane.

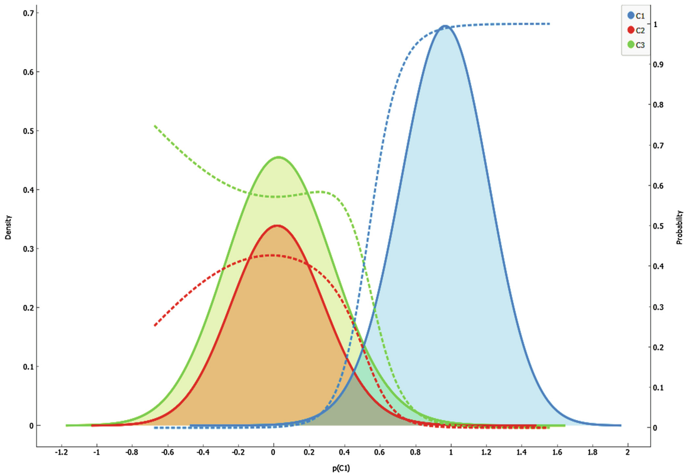


Fig. 9. Logistic regression clusters difference of P(C1), P(C2) and P(C3). (Color figure online)

It uses logistic function to fit output values between zero and one, just like a probability [11]. Figure 9 shows that were three cluster difference in three clusters include P(C1) logistic regression was R = 0.95 approximately 1.0 best score by regression line.

4 Conclusion

In this paper, students' performance is known problem which has its importance in found education and research strategies in order to raise quality level. The paper of such phenomena is benefit to society; sought to analyze association between student variables related with student determination and their capability to predict constant enrollment. It requires fundamental measures to be taken to increase not only financial situations but standards for research environment in university. Incessant development of commercial situations of students and creation of non-stressful situations are important aspects in growing performance of student. Consuming historic enrollment information predictive model was made to estimate enrollment probability of future student. Logistic regression classifier, related four year bachelor academic and demographic data on students to relative probability, was estimated. Subsequently enrollment pattern may modification such as in Campus policies, model needs to be always modified and validated year after year to improve its predictive power. This study cannot be used as stand-alone but helps to admissions administrators in decision making process to competently succeed enrollments.

References

1. Armstrong, J.S.: Illusions in regression analysis. Int. J. Forecast. **28**, 689–694 (2012)
2. Correa, A., González, A., Nieto, C., Amezquita, D.: Constructing a credit risk scorecard using predictive clusters. In: SAS Global Forum, p. 128. SAS Institute Inc. (2012)
3. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2006)
4. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning, p. 348. [I]n: Data Mining Applications the Interest is Often More in the Class Probabilities Themselves, Rather Than in Performing a Class Assignment (2009)
5. Jayakumar, D.S., Thomas, B.J.: A new procedure of clustering based on multivariate outlier detection. J. Data Sci. **11**, 69–84 (2013)
6. Linoff, G.S., Berry, M.A.: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley Publishing Inc., Indianapolis (2011)
7. Mooi, E., Sarstedt, M.: A Concise Guide to Market Research. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-12541-6
8. Northern Arizona University: Multiple Regression (2002). Northern Arizona University. http://oak.ucc.nau.edu/rh232/courses/EPS625/Handouts/Regression/Multiple%20Regression%20-%20Handout.pdf. Accessed 8 Dec 2013
9. Oliver, T.A.L., Smith, C., Winston, S.J., Geranmayeh, F., Behjati, S., Kingston, O., Pollara, G.: Impact of UK academic foundation programmes on aspirations to pursue a career in academia. Med. Educ. **44**, 996–1005 (2010)
10. Pachgade, S.D., Dhande, S.S.: Outlier detection over data set using cluster-based and distance-based approach. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **2**(6), 12–16 (2012)

11. Allison, P.D.: Logistic Regression Using the SAS System: Theory and Application. SAS Institute and Wiley, Cary (2003)
12. Taylor, C.: What Is the Difference Between Alpha and P-Values? (2013). About.com. http://statistics.about.com/od/Inferential-Statistics/a/What-Is-The-Difference-Between-Alpha-And-P-Values.htm. Accessed 20 Nov 2013
13. Kumar, V., Chadha, A.: An empirical study of the applications of data mining techniques in higher education. Int. J. Adv. Comput. Sci. Appl. **2**(3), 80–84 (2011)
14. Razaque, F., Soomro, N., Shaikh, S.A., Soomro, S., Samo, J.A., Kumar, N., Dharejo, H.: Using Naïve Bayes algorithm to students' bachelor academic performances analysis. In: 4th IEEE International Conference of Applied Science and Technology, ICETAS (2017)
15. Pyle, D.: Data Preparation for Data Mining. Academic Press, Norwell (1999)

Face Recognition Analysis Using 3D Model

Muhammad Sajid Khan[1,2](✉), Muhammad Jehanzeb[2], Muhammad Imran Babar[2], Shah Faisal[3], Zabeeh Ullah[4], and Siti Zulaikha Binti Mohamad Amin[5]

[1] College of Computer Science, Sichuan University, Chengdu 610065, People's Republic of China
sajidpk48@yahoo.com
[2] Department of Computer Science, Army Public College of Management & Sciences (APCOMS), Rawalpindi, Punjab, Pakistan
[3] Department of Computer Science, University of Haripur, Hattar Road Near Swat Chowk, Haripur 22620, Khyber Pakhtunkhwa, Pakistan
[4] Federation University Australia, Mt Helen, Ballarat, VIC 3350, Australia
[5] Western China Earthquake and Hazards Mitigation Research Centre, College of Architecture and Environment, Sichuan University, Chengdu 610065, People's Republic of China

Abstract. Facial Recognition is a commonly used technology in security-related applications. It has been thoroughly studied and scrutinized for its number of practical real-world applications. On the road ahead of understanding this technology, there remain several obstacles. In this paper, methods of 3D face recognition are examined by measuring quantifiable applications and results. In facial recognition, three Dimensional Morphable Model (3DMM) techniques have attracted more and more attention as effectiveness in use increases over time. 3DMM provides automation and more accurate image rendering when compared to other traditional techniques. The accuracy in image rendering comes at a cost; as 3DMM requires more focus on texture estimation, shape-controlling limits, and extrinsic variations, accurately matching fitting models, feature tracking and precision identification. We have underlined different issues in comparison based on these methods.

Keywords: Reconstruction · Recognition · Morphable model
3D model

1 Introduction

Human face modelling in three dimensions is a challenging topic for researchers in the field of graphic design and pattern recognition. In the last two decades, several techniques have been reported successful for recognitions such as identifications, using geometrical models for verification. Although some algorithms have performed well both in accuracy and speed, improvements are still needed. The development of 3DMM has its challenges but reconstructing according to this software will without a doubt show more accurate results. This issue is the fundamental problem in computer vision. It uses growth in software sectors like plastic surgery, face tracking, face

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 220–236, 2018.
https://doi.org/10.1007/978-3-319-95450-9_19

morphing, animations and 3D games, fields that are not developed to the standard of 3DMM. Before this technology came along, facial reconstruction was limited to producing real faces, and it also faced the need to focus on algorithms to make them simpler, faster and more accurate [26]. In the 20th century, facial features, bone reconstructions and technical development began. The first attempt was made in the US and was considered remarkably successful [41]. In the same year, neck and nose were reformed where facial bones were finished professionally. Faces were reconstructed in two steps: the basic reconstruction and final modelling. The beginning of the 21st century marks the introduction of digital face; various software systems were developed to reconstruct digital faces. The first computerized technique was developed at University College London in Great Britain for forensic purposes; the system was developed for 3D surface data acquisition of the human face [34]. The 3D faces reconstruction examples and application given in Table 1. The algorithms used for 3D face reconstruction are divided into two main categories briefly described in Fig. 1.

Table 1. Example of 3D face reconstruction applications

Domain name	Application	Input data	Purpose
Recognition	Associate-predict model [32]	Identity data set	Intra personal variation
Synthesis	Partial Least Squares (PLS) method [63]	Pie data set	Multi-modal face recognition
Features detection	Conditional regression forests system [12]	Raw images	Processing images in real time
Resolution	3D variant of patch match approach [19]	Quantized depth map	Super resolution for colour
Image matching	Energy based multi-model [50]	Features pairs	Improves the accuracy of models

Table 2. Three-dimensional face reconstruction models

Approach	Representation	Function	Typical criterion
Deformable [31]	Corners, geometrical, photometrical images	Fitting, changes in shape	Classification error
Statistical [4]	Voxels, structure, type and position of data set	Detection and diagnosing	Classification error
Mesh model [68]	Texture, shape density of shape	Accurate and fast	Acceptance error
Morphable [18]	Regions or cells	Appearance	Classification error

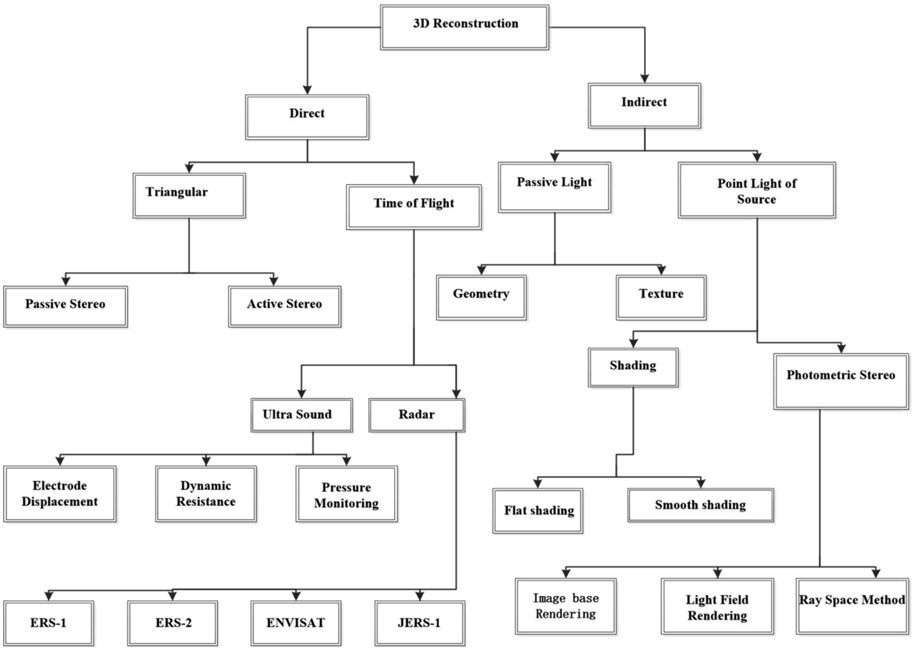


Fig. 1. Classification model of 3D face reconstruction, [36, 41, 42, 54, 57, 58]

Face image processing is commonly used in various practical life applications such as cosmetic planning surgery, security applications, and for machine interaction by human and robotic animation. But what is 3D facial reconstruction? 3D reconstruction is the phenomenon of capturing the shape and appearance of real objects. The procedure may be completed by active or passive methods. 3D Face terminologies are divided into two categories based on reconstruction and features. Active methods deal with reconstruction of radio metric and mechanical objects like laser and visible light, ultrasound, whereas the passive method concerns the measurement of radiation and emission of light in 3D structure such as image sensors [52]. Reconstruction of a 3D object requires, first, the gathering of 3D information about that object. This process is called data acquisition. It is the fundamental part of the reconstruction process and has a very important role in computer vision applications. After an accurate data acquisition, a registration is needed to fit and manipulate a generic face model with those results to complete reconstruction. Reconstruction of human faces is necessary to generate human face models looking as realistic as possible. This process requires a conversion from two-dimensional (2D) spaces to three-dimensional (3D) spaces. 3D face modelling is currently receiving a lot of attention in the computer industry developing computer graphic communities and is a thriving research field that can yield various applications such as virtual reality, animation, face recognition, facial synthesis, video meeting and games. 3D models have advanced along with the headway of PC applications, such as 3D films, on line amusements, security applications, restorative surgery, PC vision and mathematic rules. It’s difficult to overcome the critical issues and enhance the development systems of displaying three-dimensional images. Several improvements of 3D-displaying frameworks have been achieved using 2D response efforts to create an advanced output of 3D images. Examples of this include the advancement of savvy cam and camcorders. Moreover, the easy adaptability and affordability for the use of these new outputs grant everyone access to advanced applications. Face reconstruction is briefly presented in this article, and we focus on the different strategies of 3D displaying, especially providing various approaches about fitting of three-dimensional face mode [14, 20].

1.1 Deformable Model Approach

The recovery and segmentation of shapes, 3D cases, corners, geometrical and photometric images, parameterizations of models and representation of shapes are based on simple deformable model techniques of 3D face reconstruction. Deformable models can be cited as changes made in shape of any object as per set of instruction or parameters [40]. All deformable models are independent in a wide range of applications and the conversion of these models in any shape of objects may increase the intention of researches in that area. These models are utilized in many applications such as morphing and texturing, pose performance and illumination in face recognition (Table 2).

Additionally, they can be applied in the deformation of boarder-converging models, twisting and taper bending, functions of local deformation and shearing. These applications are available in (human soft surface model, tracking model, animations and surgery simulation) a variety of shapes and structures. These variables, represent compactness, linearity and other convergence theorems have been used to acquire the different requirements of 3D modelling from 3D deformable lines. Powerful and high-resolution specification techniques are usually required for better performance of 3D deformable models, and it is computationally expensive when combining many operations for one model.

1.2 Morph Able Model Approach

In the 3D Morphable approach, the spatial reconstruction of face is briefly mentioned in the geometrical part of vector space representation. Shape, texture and density of natural faces in space were addressable issues of morphing models. We may be able to introduce new arbitrary faces by controlling the parameters of texture and shape. In many cases linear transformation is adopted to simplify the mapping between morphing model and 3D images. Fitting methods reduce the computational time and improve the fitting performance. The diversity of the human face (shape, texture, appearance) makes the analysis of facial imaging more critical and complex. Appearance and variation can be categorized into four basic sources: (1) pose changes (2) lighting sources (3) facial expression (4) aging. A large amount of approaches, algorithms, analyses and techniques have been cited for fitting models, computational time and mapping purposes. Blaz and Vetter reported the issues of 3D morphable model including texture and estimation of 3D, data sets from human faces and facial comparison for recognition purposes. This system can be further implicated for facial automation, feature detection and faster-fitting production. Examples are mentioned in Table 3.

Table 3. Applications of three morphable models

Models	Property	Purpose
Patel [47]	Shape and texture	Face shape recovery
Heo [25]	Features derivation from input images	Improving the accuracy and efficiency of fitting models
Moghaddam [48]	Silhouettes computing from input images	Edges and specular highlights
Knothe Model [22]	Local dominance and model feature analysis	Fitting improvement
Volker and Vetter [65]	Texture extraction	Matching of various morphable models
Vetter [35]	Synthesis approach	Extraction human face from single image
Cootes [9]	Aligning model to image	Optimize parameters, illumination and rotation of rigid body

1.3 Statistical Model Approach

In the biomedical era, statistical shape models act as a catalyst for interpreting and segmented images. Variation in voxels and shape structure information are obtained from probabilistic atlases and landmarks in most of the algorithms required for same place. The main concept of statistical model shape has derived from training sets and how variation, structure, type and position of data sets can be changed with statistical analysis. Currently, analysts take more interest in statistical shape models because these models are used in detecting and diagnosing diseases. Shape and landmark based correspondence are both challenging problems (non-linear description and non-rigid shape) in constructing models and are considered time consuming as well as leaving room for error [29, 39]. Prior research demonstrates that a small error in shape correspondences affects the accuracy and structure of final statistical shape model.

Minimum description length approach (MDL) defines how bit length is used to measure the correspondence error and how it arranges the training set to build this model. The Bayesian approach makes an improvement in face tracking such as through translation, rotational, positional movement of face. It is impossible to ensure a one hundred per cent Error-Free model, however. The verification process is important to specify all the aspects of the model's validity. Both factors are very essential for proposing the new statistical shape model.

1.4 Mesh Model Approach

Mesh models can be applied in many complex cases of representation. Mesh is a geometrical representation assigning the smaller region or cell over which the flow is solved. It is more precise in tracking and compressing images. Meshing models lay on optimization process to reduce the error occurrences between actual images and mesh model images [56]. Neighbourhoods are an important concept in mesh applications as it shows faster and more accurately the non-uniformed samples of images consisting of irregular patches. Mesh models are divided into three categories, depending on the skewness and smoothness: (1) structure grids (2) unstructured grids (3) hybrid grids. The quality depends on the convergence rate, accuracy, required time and shapes. The main problem in mesh modelling is the determination of an accurate shape for the desired images. Image derivative of the first and second degree evaluate the performances of models. Content-based methods provide fast and efficient mesh modelling without high computational costs and other optimization processes. Multiple features of 3D models were (uniform scaling, robust, rotational, and translational, re-meshing simplification) scientists considered as a perfect mesh geometrical model approach among other models [56].

1.5 Texture Mapping and Classification Approach

Texture mapping has become a well-known method in computer vision and graphics. It is a straightforward method to bring realism into imaginary images (Fig. 2).

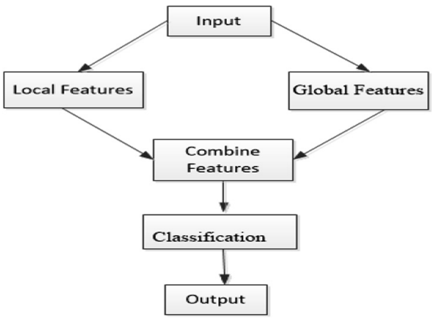


Fig. 2. The overview of texture process [21]

A various number of changes in transparency, colour, security, displacement and surface normality, motion blur and lens refocus known as texturing and are divided into sub categories, deterministic texture and s-stochastic texture. Texture mapping is defined as the low-cost method for faking the surface, or it is the transformation process (array or algebraic form) of a three-dimensional object and is considered as one the most important and valuable process of image processing and computer graphics. The actual applications of texture classification are applicable in pattern recognition, medical imaging and industrial investigation. In many cases the texture is relative to the arrangement of surrounding environmental images. Currently, texture-mapping methods are primarily used to make variations in images and movies. Texture mapping might be more commonly known as Image Perspective Transformation (IPT). The real applications of texture classification are applicable in pattern recognition, medical imaging and industrial investigation. Three main issues appear in texture classification, the use of wide ranging features to characterize the texture and the statement of variations as well as the measured distances of two of the same textures. In the early stages, features are limited to autoregressive models, Markov random fields and co-occurrences matrices. The extraction of local and global features can be assembled into many techniques to do texture classification and is considered one of the simplest and easiest methodologies. Local and global features can be jointly called as fusion features. Fusion features are classified by different classifiers; it has provides a high level of accuracy, improvement and efficiency in classification to process [23, 24].

1.6 Scope and Organization

In remaining part of the paper, we primarily review morphable models for 3D face reconstruction and classification, focusing on recent developments. In other sections we describe the different algorithms and techniques in table form for the ease of the reader. The point of which is to define those methods that have been briefly mentioned and permitting their application along with new amendments and ideas. Face reconstruction is a very dispersive field, and various topics are covered in many journals in (face detection, recognition, alignment and texture) several fields. In Sect. 2 we provide a detailed introduction of 3D reconstructed methodologies and briefly explain the classification diagram of techniques. In Sect. 3 we discuss 3DMM and survey the reconstruction issues based on face recognition with other methods. The articles available in area of face models are divided into sub topics: Correspondence Estimation, robustness, representation power, albeit with parameters, PCA for appearance Compositional Image Alignment (ICIA) algorithm and texture shape error. The initial focus on our review is the basic concept of 3D for face reconstruction. Section 4 describes the classification of face recognition methods and provides the evaluation performances of many algorithms.

2 Three Dimensional Morphable Model

Many applications involved in generating real human faces emphasize other characteristics such as changes in age, physical appearance, body weight and other facial features. The morphable model is a multitasking function that consists of large number of 3D scans using linear combinations. Simply models were developed in three stages process: pre-processing, model building and user interaction. To understand different algorithms that are used for fitting cover shape, it is important to get basic idea of morphable 3D faces. A morphable 3D face model is a vector space of 3D shapes and textures spanned by set examples (Sattar and Kang, 2006) [33]. The morphable model arises from cyberspace layer scan and captures any variations and common properties found in the set. Shape and texture can be defined by the following linear combinations of standard deviations [46].

$$S = \sum\nolimits_{i=1}^{m} a_i s_i \tag{1}$$

where $i = 1$ and $T = \sum b_i T_i$

When the laser scans are converted into S and T vectors of shape and texture, there should be point-to-point correspondence of all scans with reference to the face. The vector S is usually stored in terms of x, y and z coordinates of all the vertices of a 3D mesh. Therefore, we have the following vector:

$$S_i = (X_1; Y_2; z_1; X_2;\ X_2; ::::: Zn)^T \tag{2}$$

In the same way, we can texture vectors form red, blue and green of all vertices colours as shown below:

$$T = (R_i; G_i; B_i; :::: Rn; B_n)^T \tag{3}$$

Having covered the basic morphable 3D shapes, we can now look at the various types of methods. Based on morphable models and the understanding 3D, the available information and its comparisons from different statistical algorithms were applicable. Generally, the reconstruction process is accomplished by establishing a detailed analysis in shape alignment and finally is statistically approached. The well-known models of 3DMM are described in Table 4. In face recognition, faces from images are not solely based on similarity, but there are many factors involved in tracking faces from images i.e. poses, illumination, expression, various parameters, aging and different sources. The major aim of recognition algorithms is to categorize the properties of faces that are demonstrated by texture and intrinsic shape of facial surface. Face recognition is divided into models based on appearance; many statistical techniques in appearance are based on methods used in different applications. These image-based techniques are then sub-divided into 2D images and 3D images. 3DMM is a type of 3D image-based technique, details form different issues related to the recognition of 3D face models and mention survey reports on fitting of models limited around shape and texture. All these methods will be published in sequence and describes the shape of table of contents, which summarizes work of different researchers. Image identification given in Fig. 3.

Table 4. Summary of research on 3DMM based on recognition

Models	Property	Purpose
Ganar A N, Gode C S [16]	Recover texture and shape parameters in term of model represents the identity of face from single image	High score of fitting and identification performance useful for measuring
Wang L Y, Liu B, Su S et al. [61]	Introducing multidimensional models in 3DMM for face recognition	Fast and efficient for matching
Chu, Baptiste, Sami Romdhani [11]	Using 3DMM to recognize faces in videos	Improve the accuracy of fr sdks in video dealing the face images

(*continued*)

Table 4. (*continued*)

Models	Property	Purpose
Huber [30]	Newton optimization and newton multi stage method used to accurate the shape and texture	Recognize the facial expression of human face in real scenario
I. Choi, and D. Kim et al. [10]	Better performance with extrinsic variation	Fully automatic algorithm and better accuracy for matching
Papazov C, Marks T K [45]	Model used large database of Korean faces	Show real changes and controlling of parameters for comparison of similarities
C. Mayer, M. Wimmer et al. [34]	Fitting algorithm is automatic and properly for facial expression	Fast and most applicable for real time application in recognition of face
Yongli Hu, Baocai Yin et al. [49]	Mesh resampling method is used to avoid the errors in model reconstruction	Multi lighting model is performed on different images to match facial image
B. Amberg, R. Knothe et al. [3]	Fully automatic and accurate, bit slower	Consider for high level of recognition
Rekik, Ahmed et al. [53]	Morphable expression model that demonstrates emotions of face by different parameters	Detection and verification of human face
Nathan Faggian, Andrew P. Paplinski [15]	Labelling the key features in automatic models of morph able models	Suitable for real time applications

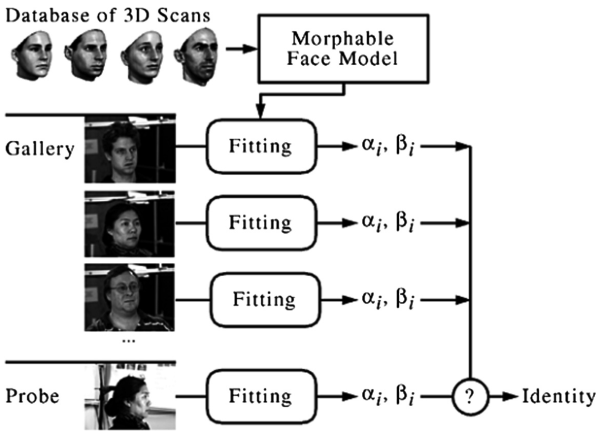


Fig. 3. Tracking model of 3DMM [46]

3 Classification and Performance Evolution of 3DMM

In this section we prefer many choices during the development of 3D morphable algorithms, like fitting of shape, representation, dissimilarity and correspondence between points. All these together affect the performance, speed, efficiency, applicability and accuracy of algorithm. In this section we discuss the basic issues in 3DMM techniques, which influence the performance. The results of many algorithms mentioned in form of tabular as given in Tables 5, 6 and 7. The region-based methods are applied above in this section the algorithms applied to use databases consist of data training sets, range images, galleries and points. The calculated performances of various algorithms depend on dissimilarities are measured by many measurement methods. The results are considered the best in [40, 42, 43]. The region-based methods are used to calculate the similarities for matching purposes. Some different cases are measured by using Hausdorff, which features vectors, Euclidean distance and Iterative closet (ICP). These algorithms worked to the managed missing data for the processing.

Table 5. Evolution of 3DMM database

Database	Type	Approach	References
FRGC	M-s	Gabor	Hsu [26]
USF3D	PCA-ST	PCA	Abiantun [1]
USF3D	PCA-S	LBP	Niinuma [44]
USF3D	M-s	LBP	Hassner [23]
BJUT3D	PCA-S	Gabor	Li [37]
USF3D	PCA-S	Pixel	Prabhu [51]

Table 6. Evolution of 3DMM with other models based on recognition

Approach	Models	Members	Data type	Error ratio	Algorithms	References
3D morphable models techniques						
Frame work	3D	05	Training data set	6.8849	Fitting algorithm learning-based	Zhu et al. [69]
Local features	3D	05	Real images	0.86	Cascaded regression	Huber et al. [30]
Bilinear programming	3D	03	Multi-dimensional model	0.8	BMMMs	Wang et al. [61]
Surface mesh	3D +2D	02	Training sets	0.5	Iterative Closet Point (ICP)	Tsalakanidou et al. [20]
View-based AAM	3D	04	Feature points	3.76	Particle swarm optimization	Lin et al. [38]
Image formation	3D	03	Spare set	0.524	Multi-linear equation	Aldrian et al. [2]

(*continued*)

Table 6. (*continued*)

Approach	Models	Members	Data type	Error ratio	Algorithms	References
Shape models	3D	06	Silhouette facial landmarks	0.82	Active shape model	Lale [59]
Albedo	2D +3D	05	Texture mapping images	0.5	AB3DMM	Hu [28]
Dense registration	3D	01	Internet photos	0.5	IBMM	Kemel [32]
Deformable models						
Template	2D +3D	04	Blend-shaped models	7.05	Template based	Rui [66]
3D shape template	3D	02	Variational frame work	3.1–5.8	Template based isometric	Adrien [5]
3D volumetric observations	3D +4D	04	Non-rigid moving object	2.12 to 12.35	On line algorithm	Xu [64]
3D scanning system	3D	05	Aligning input sequence	0.94 to 1.21	Bundle adjustment	Zhu [70]
3D variability	3D +4D	04	Frame work	43.2	Mean pose inference model	Zhu [70]

Table 7. Evolution of 3DMM with other models based on recognition

Approach	Models	Members	Data type	Error ratio	Algorithms	References
Statistical models						
Landmark localization	3D	04	Frontal images	0.0817 to 0.1025	RSF	Sagonas [55]
Laplace-Beltrami operator	3D	02	Deformable shapes	96.3 precision	Riemannian metric isometric	Gasparetto A [17]
Minimum Description Length (MdL)	3D	02	3d scan faces	0.43 to 1.26	Quasi-Newton method	Bolkart [8]
Wavelet transform	3D	03	3D Surfaces	0.189 to 4.896	Multilinear wavelet	Bolkart [7]
Mesh models						
FAUST	3D	04	Partial Scan	3 mm	Benchmarking	Bogo [6]
LBP	3D	02	Frame work	99.0% Accuracy	Descriptor	Werghi [62]

(*continued*)

Table 7. (*continued*)

Approach	Models	Members	Data type	Error ratio	Algorithms	References
Texture mapping						
Template matching	3D	04	Monocular images	0.1 to 10	Descriptor	Ngo [43]
Examples based	3D	04	Deficient images	3	Face texture	Dessein [13]
Super resolution	2D +3D	04	View points	2.1 to 2.55	Bayesian framework	Tsiminaki [60]
Co-segmentation	3D	03	Texture shapes	0.2	Bayesian spare annotation	Yumer [67]

4 Classification of 3DMM

In Fig. 4, the complete tree diagram of facial recognition, approaches image-based terminology. The theme driving this model combines the existing and new methods to show the sub-types and updates exposing strengths and weaknesses of our current understanding. The proposed figure is used to summarize the specific methods of face recognition. It will show reasonable results when applying on texture, shape, extrinsic

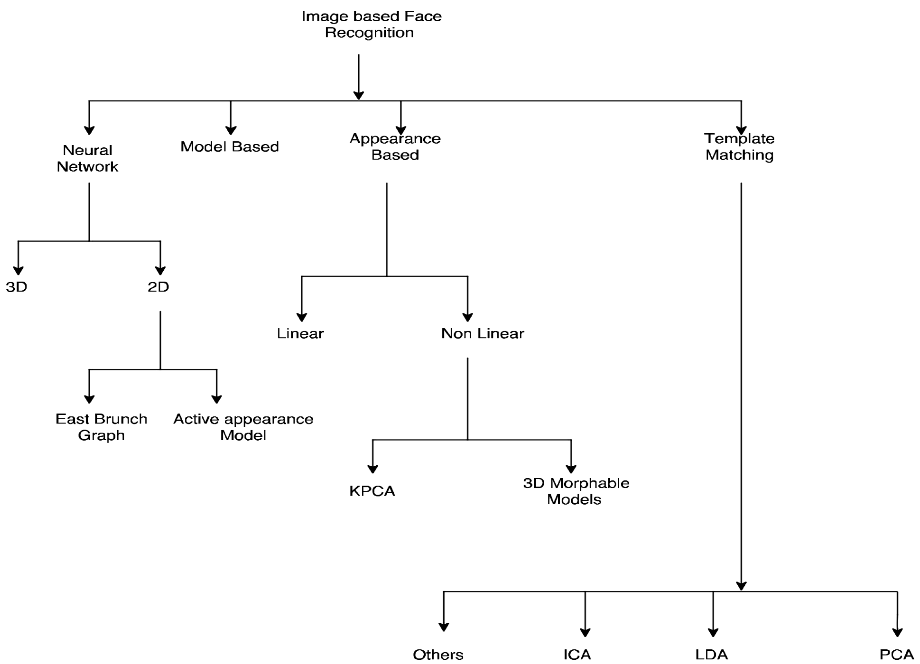


Fig. 4. Tree diagram of 3D face image-based face recognition [27].

and intrinsic variation in three-dimensional rendering. This model is used for 3D face analysis and reconstruction based on geometric recognition. The idea essentially classifies the methodologies in a visual layout to aid the understanding various techniques and easily decided which method is more suitable for face recognition. On the other hand, we provide the analysis of different techniques proposed by researchers for better performance and high improvements in different areas of face reconstruction for comparison and various purposes.

5 Conclusion

This article has provided an overview of three-dimensional face reconstruction based on recognition to construct and understanding of our current stage of development. The analysis of the methods of 3D face reconstruction combined with a list of 3D morphable face models helps highlight current success as well as problems in the industry. Many sectors ranging from graphic design, medical reconstruction, and facial identification software are well suited for the advancement in this technology. With the growing demand to take safety precautions whether they are cyber or physical, the emergence of facial recognition biometrics has, and will grow to supply safety measures to consumers. A key challenge in adequate methods has been shown through face-fitting issues in 3DMM models. There is a balance we must strike with the current state of our software development as well as the computational processing power of today's hardware. The types of facial recognition and specifically listed algorithmic complexities show efficiency discrepancies through multiple data types used in each approach.

Acknowledgement. The author is very grateful to everyone for their recommendations and guidance in research, and for their continuous support, motivation and immense knowledge. The author also thanks to colleagues and lab mates for stimulates discussions and encouragement.

References

1. Abiantun, R., Prabhu, U., Savvides, M.: Sparse feature extraction for pose-tolerant face recognition. IEEE Trans. Anal. Mach. Intell. **36**(10), 2061–2073 (2014)
2. Aldrian, O., Smith, W.A.: Inverse rendering of faces with a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. **35**(5), 1080–1093 (2013)
3. Amberg, B., Knothe, R., Vetter, T.: Expression invariant 3D face recognition with a morphable model. In: 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008, pp. 1–6. IEEE (2008)
4. Ashburner, J., Friston, K.J.: Voxel-based morphometryłthe methods. Neuroimage **11**(6), 805–821 (2000)
5. Bartoli, A., Collins, T.: Template-based isometric deformable 3D reconstruction with sampling-based focal length self-calibration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1514–1521 (2013)

6. Bogo, F., Romero, J., Loper, M., Black, M.: FAUST: dataset and evaluation for 3D mesh registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3794–3801 (2014)
7. Bolkart, T., Brunton, A., Salazar, A., Wuhrer, S.: Statistical 3D shape models of human faces (2013). http://statistical-face-models.mmci.uni-saarland.de
8. Bolkart, T., Wuhrer, S.: A groupwise multilinear correspondence optimization for 3D faces. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3604–3612 (2015)
9. Bustard, J.D., Nixon, M.S.: 3D morphable model construction for robust ear and face recognition. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2582–2589. IEEE (2010)
10. Choi, I., Kim, D.: 3D face fitting using multi-stage parameter updating in the 3D morphable face model. In: Tenth IEEE International Symposium on Multimedia, ISM 2008, pp. 274–279. IEEE (2008)
11. Chu, B., Romdhani, S., Chen, L.: 3D-aided face recognition from videos. In: 2014 5th European Workshop on Visual Information Processing (EUVIP), pp. 1–6. IEEE (2014)
12. Dantone, M., Gall, J., Fanelli, G., Van Gool, L.: Real-time facial feature detection using conditional regression forests. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2578–2585. IEEE (2012)
13. Dessein, A., Smith, W.A., Wilson, R.C., Hancock, E.R.: Example-based modeling of facial texture from deficient data. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3898–3906 (2015)
14. Dikmen, M.: 3D face reconstruction using stereo vision. Ph.D. thesis, Middle East Technical University (2006)
15. Fanelli, G., Dantone, M., Gall, J., Fossati, A., Van Gool, L.: Random forests for real time 3D face analysis. Int. J. Comput. Vis. **101**(3), 437–458 (2013)
16. Ganar, A.N., Gode, C., Jambhulkar, S.M.: Enhancement of image retrieval by using colour, texture and shape features. In: 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies (ICESC), pp. 251–255. IEEE (2014)
17. Gasparetto, A., Torsello, A.: A statistical model of Riemannian metric variation for deformable shape analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1219–1228 (2015)
18. Groeber, M., Ghosh, S., Uchic, M.D., Dimiduk, D.M.: A framework for automated analysis and simulation of 3D polycrystalline microstructures: Part 1: statistical characterization. Acta Mater. **56**(6), 1257–1273 (2008)
19. Gruen, A., Akca, D.: Least squares 3D surface and curve matching. ISPRS J. Photogramm. Remote Sens. **59**(3), 151–174 (2005)
20. Gu, Y., Cao, Z., Zhang, Y.: Three-dimensional reconstruction of multiplatform stereo data with variance component estimation. IEEE Trans. Geosci. Remote Sens. **52**(7), 4211–4226 (2014)
21. Guan, Y.: Automatic 3D face reconstruction based on single 2D image. In: International Conference on Multimedia and Ubiquitous Engineering, MUE 2007, pp. 1216–1219. IEEE (2007)
22. Hagihara, T., Hanawa, M.: Multivariate identification of low-loss sampled fiber bragg gratings by downhill simplex method. In: 2013 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), pp. 758–763. IEEE (2013)
23. Hassner, T., Harel, S., Paz, E., Enbar, R.: Effective face frontalization in unconstrained images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4295–4304 (2015)

24. Heimann, T., Meinzer, H.-P.: Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal. **13**(4), 543–563 (2009)
25. Heo, J., Savvides, M.: In between 3D active appearance models and 3D morphable models. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, pp. 20–26. IEEE (2009)
26. Khan, M.S., Ullah, Z., Jaffri, U.A.: A proposed (FRMS) 3D face reconstruction method from stereo images. In: Proceeding of the 9th International Conference on Computer and Automation Engineering, ICCAE 2017, 18–21 February 2017, Sydney, Australia, pp. 150–154 (2017)
27. Hu, G.: Face analysis using 3D morphable models. Ph.D. thesis, University of Surrey (2015)
28. Hu, G., Chan, C.H., Yan, F., Christmas, W., Kittler, J.: Robust face recognition by an albedo based 3D morphable model. In: 2014 IEEE International Joint Conference on Biometrics (IJCB), pp. 1–8. IEEE (2014)
29. Hu, Y., Zheng, Y., Wang, Z.: Reconstruction of 3D face from a single 2D image for face recognition. In: 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 217–222. IEEE (2005)
30. Huber, P., Feng, Z.-H., Christmas, W., Kittler, J., Ratsch, M.: Fitting 3D morphable face models using local features. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 1195–1199. IEEE (2015)
31. Hur, S.C., Henderson-MacLennan, N.K., McCabe, E.R., Di Carlo, D.: Deformability-based cell classification and enrichment using inertial microfluidics. Lab Chip **11**(5), 912–920 (2011)
32. Kemelmacher-Shlizerman, I.: Internet based morphable model. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3256–3263 (2013)
33. Kemelmacher-Shlizerman, I., Basri, R.: 3D face reconstruction from a single image using a single reference face shape. IEEE Trans. Pattern Anal. Mach. Intell. **33**(2), 394–405 (2011)
34. Kim, Y., Chung, S.-T., Kim, B., Cho, S.: 3D face modeling based on 3D dense morphable face shape model. Int. J. Comput. Sci. Eng. **2**(3), 107–113 (2008)
35. Knothe, R., Romdhani, S., Vetter, T.: Combining PCA and LFA for surface reconstruction from a sparse set of control points. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 637–644. IEEE (2006)
36. Lee, T.-Y., Sum, Y.-N., Lin, Y.-C., Lin, L., Lee, C.: Three-dimensional facial model reconstruction and plastic surgery simulation. IEEE Trans. Inf. Technol. Biomed. **3**(3), 214–220 (1999)
37. Li, S., Liu, X., Chai, X., Zhang, H., Lao, S., Shan, S.: Morphable displacement field based image matching for face recognition across pose. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7572, pp. 102–115. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33718-5_8
38. Lin, K., Wang, X., Li, X., Tan, Y.: Self-adaptive morphable model based multi-view non-cooperative 3D face reconstruction. In: 2014 IEEE Congress on Evolutionary Computation (CEC), pp. 320–325. IEEE (2014)
39. Liu, Z., Zhang, Z., Jacobs, C., Cohen, M.: Rapid modelling of animated faces from video. J. Vis. Comput. Anim. **12**(4), 227–240 (2001)
40. Ming, Y., Ruan, Q., Li, X.: 3D face reconstruction using a single 2F face image. In: 2010 International Conference on Educational and Information Technology (ICEIT), vol. 3, pp. V3–V32. IEEE (2010)
41. Minoi, J.-L., Jupit, A.J.R., Gillies, D.F., Arnab, S.: Facial expressions reconstruction of 3D faces based on real human data. In: 2012 IEEE International Conference on Computational Intelligence and Cybernetics (CyberneticsCom), pp. 185–189. IEEE (2012)

42. Moeslund, T.B., Granum, E.: A survey of computer vision-based human motion capture. Comput. Vis. Image Underst. **81**(3), 231–268 (2001)
43. Ngo, T.D., Park, S., Jorstad, A.A., Crivellaro, A., Yoo, C., Fua, P.: Dense image registration and deformable surface reconstruction in presence of occlusions and minimal texture. In: International Conference on Computer Vision, No. EPFL-CONF-211260 (2015)
44. Niinuma, K., Han, H., Jain, A.K.: Automatic multi-view face recognition via 3D model-based pose regularization. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–8. IEEE (2013)
45. Papazov, C., Marks, T.K., Jones, M.: Real-time 3D head pose and facial landmark estimation from depth images using triangular surface patch features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4722–4730 (2015)
46. Park, J.-H., Baasantseren, G., Kim, N., Park, G., Kang, J.-M., Lee, B.: View image generation in perspective and orthographic projection geometry based on integral imaging. Opt. Express **16**(12), 8800–8813 (2008)
47. Patel, A., Smith, W.A.: 3D morphable face models revisited. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1327–1334. IEEE (2009)
48. Patel, A., Smith, W.A.: Simplification of 3D morphable models. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 271–278. IEEE (2011)
49. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. In: Sixth IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2009, pp. 296–301. IEEE (2009)
50. Poultney, C., Chopra, S., Cun, Y.L., et al.: Efficient learning of sparse representations with an energy-based model. In: Advances in Neural Information Processing Systems, pp. 1137–1144 (2006)
51. Prabhu, U., Heo, J., Savvides, M.: Unconstrained pose-invariant face recognition using 3D generic elastic models. IEEE Trans. Pattern Anal. Mach. Intell. **33**(10), 1952–1961 (2011)
52. Ullah, Z., Mumtaz, I., Khan, M.S.: Analysis of 3D face modeling. Int. J. Sig. Process. Image Process. Pattern Recogn. **8**(11), 7–14 (2015)
53. Rekik, A., Ben-Hamadou, A., Mahdi. W.: 3D face pose tracking using low quality depth cameras. In: VISAPP (2), pp. 223–228 (2013)
54. Remondino, F., El-Hakim, S., Gruen, A., Zhang, L.: Turning images into 3D models. IEEE Sig. Process. Mag. **25**(4), 55–65 (2008)
55. Sagonas, C., Panagakis, Y., Zafeiriou, S., Pantic, M.: Robust statistical face frontalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3871–3879 (2015)
56. Santamaría, J., Cordon, O., Damas, S., Botella, M., et al.: 3D forensic model reconstruction by scatter search-based pair-wise image registration. In: 2006 IEEE International Conference on Fuzzy Systems, pp. 1086–1092. IEEE (2006)
57. Stylianou, G., Lanitis, A.: Image based 3D face reconstruction: a survey. Int. J. Image Graph. **9**(02), 217–250 (2009)
58. Suen, C.Y., Langaroudi, A.Z., Feng, C., Mao, Y.: A survey of techniques for face reconstruction. In: IEEE International Conference on Systems, Man and Cybernetics, ISIC, pp. 3554–3560. IEEE (2007)
59. Tsalakanidou, F., Malassiotis, S., Strintzis, M.G.: Integration of 2D and 3D images for enhanced face authentication. In: Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 266–271. IEEE (2004)
60. Tsiminaki, V., Franco, J.-S., Boyer, E.: High resolution 3D shape texture from multiple videos. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1502–1509. IEEE (2014)

61. Wang, L., Liu, B., Su, S., Cheng, Y., Li, S.: An improved 3D bilinear multidimensional morphable models used in 3D face recognition. In: 2014 International Conference on Information Science, Electronics and Electrical Engineering (ISEEE), vol. 3, pp. 2052–2056. IEEE (2014)
62. Werghi, N., Tortorici, C., Berretti, S., Del Bimbo, A.: Representing 3D texture on mesh manifolds for retrieval and recognition applications. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2521–2530. IEEE (2015)
63. Wold, H.: Partial least squares. In. In: Kotz, S., Johnson, N.L. (eds.) Encyclopedia of Statistical Sciences. Wiley, New York (1985)
64. Xu, W., Salzmann, M., Wang, Y., Liu, Y.: Deformable 3D fusion: from partial dynamic 3D observations to complete 4D models. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2183–2191 (2015)
65. Yoshiki, K., Saito, H., Mochimaru, M.: Reconstruction of 3D face model from single shading image based on anatomical database. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 4, pp. 350–353. IEEE (2006)
66. Yu, R., Russell, C., Campbell, N., Agapito, L.: Direct, dense, and deformable: template-based non-rigid 3D reconstruction from RGB video. In: IEEE International Conference on Computer Vision (ICCV 2016). University of Bath (2015)
67. Yumer, M.E., Chun, W., Makadia, A.: Co-segmentation of textured 3D shapes with sparse annotations. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 240–247. IEEE (2014)
68. Zhang, C., Chen, T.: Efficient feature extraction for 2D/3D objects in mesh representation. In: Proceedings of 2001 International Conference on Image Processing, vol. 3, pp. 935–938. IEEE (2001)
69. Zhu, X., Yan, J., Yi, D., Lei, Z., Li, S.Z.: Discriminative 3D morphable model fitting. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 1, pp. 1–8. IEEE (2015)
70. Zhu, Y., Zhang, Y., Bonev, B., Yuille, A.L.: Modelling deformable gradient compositions for single-image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5417–5425 (2015)

Named Entity Recognition System for Sindhi Language

Awais Khan Jumani[1(✉)], Mashooque Ahmed Memon[2], Fida Hussain Khoso[3], Anwar Ali Sanjrani[4], and Safeeullah Soomro[5]

[1] Department of Computer Science, Shah Abdul Latif University, Khairpur Mirs, Pakistan
awaisjumani@yahoo.com
[2] Department of Computer Science, Benazir Bhutto Shaheed University, Layari, Karachi, Pakistan
pashamorai786@gmail.com
[3] Department of Basic Sciences, Dawood University of Engineering and Technology, Karachi, Pakistan
fidahussain.khoso@duet.edu.pk
[4] Department of Computer Science, University of Baluchistan, Quetta, Pakistan
anwar.csd@gmail.com
[5] Department of Computer Science, AMA International University, Salmabad, Bahrain
s.soomro@amaiu.edu.pk

Abstract. Named Entity Recognition (NER) System aims to extract the existing information into the following categories such as: Person's Name, Organization, Location, Date and Time, Term, Designation and Short forms. Now, it is considered to be important aspect for many natural languages processing (NLP) tasks such as: information retrieval system, machine translation system, information extraction system and question answering. Even at a surface level, the understanding of the named entities involved in a document gives richer analytical framework and cross referencing. It has been used for different Arabic Script-Based languages like, Arabic, Persian and Urdu but, Sindhi could not come into being yet. This paper explains the problem of NER in the framework of Sindhi Language and provides relevant solution. The system is developed to tag ten different Named Entities. We have used Ruled based approach for NER system of Sindhi Language. For the training and testing, 936 words were used and calculated performance accuracy of 98.71%.

Keywords: NER · Sindhi NER · Gazetteer based approach · Rule based model

1 Introduction

During our past school days it is being taught us that a proper noun is a "specific person, place, or thing," so this definition has been taken from a concrete noun. Unfortunately, it is observed that simple mnemonic and computational linguistic tasks are extremely complex, the retrieval system of named entities like that Person's Name, Organization, Location, Date and Time, Term, Designation and Short Term. Actually,

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 237–246, 2018.
https://doi.org/10.1007/978-3-319-95450-9_20

the classification of named entity system can be termed as the identification of Named Entities in computer understandable text through information retrieval system can be categories with annotation [1, 2].

It is not only observed that information retrieval system is a subtask of NER but it can play a vital role for reference resolution, categories of disambiguation, and meaning representation with many of other NLP applications. Parts of speech tagging, semantic parsers, and thematically meaning representation can be enlarged with tagging system to achieve a better results. On the other hand, specific application of NER system exist in large amount of question answers system, automatic forwarding content, textual requirement and news searching. Even the understanding of NER system provides a better platform for analytical frameworks and cross-referencing. Named entity contains top three level categorizations according to Defense Advanced Research Projects Agency's message to understand the approach of named entities, temporary expressions and number of expressions but the categories of named entities can be described as a unique identifiers of persons' names, locations, events and organizations, it can be considered as entities and a lot of others.

1.1 Gazetteer Based Approach

NER System provides many annotations to candidates, and assure to a certain amount, can list the probability of a candidate which is joining to a group or sub-group of a NER. But this is not required as a complete solution of machine learning methods, some knowledge is required for untrained candidacy tokens. Furthermore classification of candidates is required as other issues or problems can be resolved with gazetteer based approach. Gazetteer based approach should be developed to supply external knowledge for learners, or changed to supply unannotated data with their training material.

Therefore, the researchers have come towards the development of gazetteers based encyclopedia of named entities, otherwise, some special applications can be developed on the basis of gazetteer. The Systems can be defined in [3–5] usages of a combinational rule-based devices, parts of speech (POS) tags, and some word frequency can be analyzed to propose these candidates who have no any approach to learn machine learning methods.

2 Literature Review

Many researchers have been working with information extraction during my research for the NER system, I have selected this kind of literature. Some of the researchers just give the results of their respective research regarding NER.

Shrimad Hinal, [6] has focused on the NLP, which is being used in different Indian languages and also compared that language with each other through conditional random fields. So, they have proven that which is better for Indian languages to extract the named entity information. Moreover Tarek, [7] has introduced the new method of extraction information for Arabic languages from the news articles. They have been using the two methods for extracting information (RenA and ALDA) which is better than previous tools, for such kind of these methodologies they have been taking accurate results to extract Name, Organization, and Location from online resources. Also Nita, [8] has surveyed that NER system used in different Indian languages and non-Indian languages. They observed that different kinds of NER especially in Indian languages, which techniques and approaches are best for Indian languages. Similarly Ridong, [9] has recognized that many researchers have developed different kind of NER system, it is quite difficult to which is best NER system for new user. So, they have constructed the hybrid NER system for our interest. Likewise, Maithilee, [10] has researched that different type of named entity has been introduced with different languages, this study shows (NERC) NER and classification. Many of the researchers just used rule based approaches which is to perform and this study related to learning based approaches. As well as, Seth, [11] has explored the ways and limitations of data extraction in NER and termed recognition for getting meaningful concept. Also they perform digital humanities research in searching and browsing operations. It is understanding the value of NER system. Correspondingly, Khaled, [12] has described the recent activities and growth regarding Arabic NER study and the importance of Arabic NER characteristics of languages are highlighted. Mainly common tools features can be used in Arabic NER and illustrated the evaluation of their classification. Respectively, Maksim, [13] has explored many combinations of NER features and compared the performance with each other. So, they have built conditional based approach and collected the results, statistical importance of their boost performance with their previous top performance system. Similarly, Sherief, [14] has described the evolution of (NERA) NER system in Arabic language, also focusing the integrating machine learning with rule based approach for NERA. They have implemented the methods with another taking the results regarding NERA and collect best approach with rule based approach and machine learning system. Likewise, Ronan, [15] has discovered the flexibility of NLP in their specific task of engineering and considering a lot of their prior information. Each task can be measured and optimized the features of NLP, so, this system can acquire the internal representation on the basis of huge amount of unlabeled data sets. Also, Darvinder, [16] has surveyed that NER system used in various languages like Chinese and Spanish and so on. In English language a lot of work has been done specially capitalization is more important part of NER system. Secondly, in India Punjabi is official language of Punjab and many of the tagging and information extraction work has been done in it. Respectively, Wenhui, [17] has implemented the semi- supervised algorithm learning methodology with conditional random fields for NER, also this algorithm used better efficiency and redundant the data. It has improved algorithm for the next iteration. Correspondingly, Alireza, [18] has recognized the named entity for extracting the information like person, organization and so on. They have compared the portable message understanding conference which is highly used everywhere, it can be used robust and novel learning based with fuzzy technology. Furthermore, Trian, [19] has presented the experimental results which has been taken with help of support vector machine and applied on Vietnamese language. Through the comparison of conditional random fields, the support vector machine gives better results as compared to CRF. The identification and classification of proper nouns in plain text is of key importance in numerous NLP applications. It is the first step of a desktop application as proper names generally carry important information about the text itself, and thus are targets for extraction. Moreover Sindhi NER (SNER) can be a stand-alone application. It includes proper nouns, dates, identification numbers, phone numbers, e-mail addresses and so on.

3 Problem in NER System

Sindhi language is part of Asian Language and many other Asian Languages being a part of this language and they do not need any concept of capitalization. It is noticed European Language that in English this kind of feature is commonly used to identify Named Entity in text therefore all the names of European languages are always start with capital letter. Deficiency of capitalization tool makes the NER task more for Sindhi Language. Sindhi Names consist of lot of confusion so it can be used as a proper noun or common noun. The actual goal of any NER system is to separate or remove proper noun in place of common noun. For instance: شفقت (Shafqat), رحمت (Rehmat), حڪمت (Hikmat) or سداقت (Sadaqat) can be counted as a person name or it can be counted as a place. Many other problems are being occurred in Sindhi Language as a standardization of Sindhi spelling as well as word. Multiple spelling formats are available for one Sindhi word. For example tablet can be written in Sindhi with different ways like (گوري، ٽڪي) (Gori/ Tiki). In another example word like House can be written in with different ways like (گھر، جڳه) (Ghar/ Jagah) etc. It is too difficult task for NER System. The Approach of any language is more important for any resources either it is Statistical or Rule based. As consider these kind of problems in Sindhi Language there is no any mechanism or any Gazetteer and annotated data available in Sindhi Language.

Limitation of Sindhi NER system, we analyze some of the fundamental design challenges and misconceptions that underlie the development of an efficient and robust Sindhi NER system. Rule based systems are usually best performing system but suffers some limitation such as language dependent, difficult to adapt changes.

4 Rule Based Approach

Sindhi text can be identified by means of these different Rules.

1. Some of the rules should be applied for the better recognition for date and time tags. Such kinds of tag can easily be recognized by regular expressions as it may be generated for particular forms like 05.06.2016 or 05/06/2016 and it is better known as 10:40 or 02:30. The entire system has capability to find the date like 07 جوالء 2016 or 07 جوالء and سال 2016.
2. Many of the locations' names and terms may be used with different identifications for suffix matching. In Sindhi Language and some of the other Asian languages that contain many locations which end on "PUR" just like (Khairpur, Ranipur), "STAN" (Baluchistan, Afghanistan) "GARR" (Muzafargarr, Khangarr), "NAGAR" (Shantinagar, Naseemnagar) and some cities that end on "ABAD" (Islamabad, Nooriabad). Persons' names, Terms and Organizations can be used by Suffix Matching, Just Like:

Persons' name that end on "DAD" (Allahdad, Saindad), "ALLAH" (Hidayatullah, Naimatullah). Some Terms that end on "YAT" (Hayatyat, Falkyat), Some Person's last name ends with "Hassan" or "Hussain" after that, we can identify them as Organization or it may be Persons' names like Zahid Hussain and Ali Hassan.

3. This system uses most common Gazetteer based approach for the common Persons' names for their identification. This system has capability to tag the words of three length as counted one Named Entity just like: محمد علي جماڻي (Muhammad Ali Jumani). Here Form implementation of the Named Entity System we have composed 10000 Sindhi names and 7000 Urdu persons' names.
4. Some of Sindhi and Punjabi Surnames have been stored in this system for the better recognition of the Named Entity, the system can search out His/Her first name, Just like surname (Jumani) جماڻي the system has ability find out the word before the Surname it may be the first name of person like (Awais Jumani). اويس جماڻي.
5. Title of the person can help to find out the designation of people just like وزير اعظم (Wazir-e-Azam) and مسس (Mrs) that contains proper name next to it as to help of searching the title and surname of the person. The system has ability to detect Persons' Named Entity easily and also recognize those names which are not included as a part of gazetteer. We have collected 60 Title Persons and 200 Designations.
6. During the implementation of NLP, our system can easily resolve many problems which occurred for identity of true person. Many of the rules are applied, for instance: the system can find out the opacity in persons' names. So system finds out the word اويس (Awais), then it can easily solve the problem of sentence structure or detect either it is noun or preposition where it may search Title, Person or Designation. If there is no clue to recognize as Named Entity then at last it checks out the post or position of ambiguity word like اويس جي گهر ۾ ڪتاب پيو آهي (Awais je ghar me kitab payo ahe.) the word (je) جي indicates that it is persons' name, Hence the system tag it as Person Named Entity (PNE).
7. One of the rules is to detect the numbers that are not numerals just like سيون، ايٽ، نائن (Seven, Eight, and Nine). Our system can calculate three words as One Number Entity just like ڇهسو پنج (Six Hundred Five).
8. Persons' Name can be written/searched by the abbreviations just like جي اي مهر (J A Mahar). The system have should search out surname only then it should continuously try to find out the short form of persons' name.
9. This system also have the ability to find out the short form tags like (DDR) ڊي ڊي آر (KTN) ڪي ٽي اين etc.
10. Organizations can be tagged where gazetteer based approach is used. During the testing of our system, we have stored a lot of related organizations' data and some heuristics to search ORG: tags. Suppose, if text includes Org: (Uho Sindh university me parhe tho.). اوهو سند يونيورسٽي ۾ پڙهي ٿو and it does not found out this Organization in gazetteer. The system applies rules to search and tag org: as "Sindh University".

5 Flowchart of NER System

The Approach of any language is more important for any resources either it is Statistical or Rule based. Our total works depend upon the given below diagram Fig. 1 shows the flow of working Sindhi NER System.

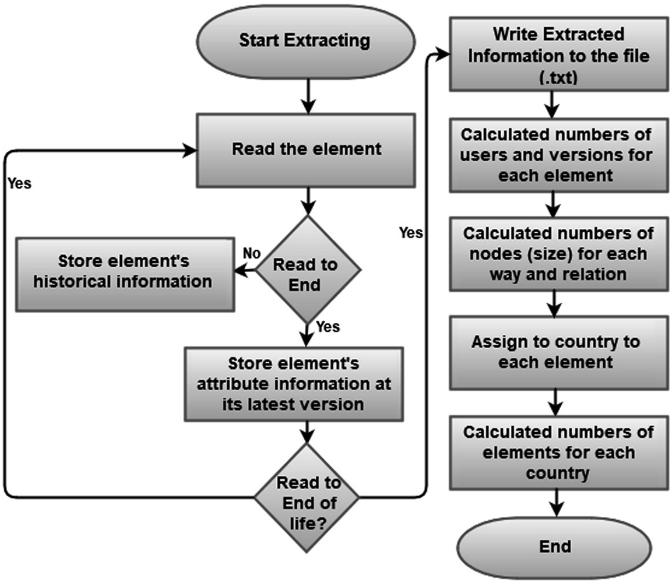


Fig. 1. Shows framework of Sindhi NER system

6 Algorithm of NER System

The entire system has been developed on visual studio.net platform through the implementation of latest features of named entity, it uses different classes and C sharp programming just like: linked list technique and tokenize the classes and functions.

1. Input text, or file uploading
2. Standard Control of Input Text
   1.1 Eradicate unwanted space
   1.2 Eradicating the special characters from the end of strings.
3. Gazetteer based search
   3.1 Gazetteer based search can be Places, Brands, Abbreviations, Terms and Organizations Tags.
4. Tokenized and Standardized
   4.1 Each word can be tokenized and searched regarding Gazetteer.
5. Searching tags of Date and Time
   5.1 Numeral numbers can be searched and also Date, Time, URL and Email tags.
6. Rules of Person Name tags

   6.1 It can detect the persons' names with Title, Surname and Designation without any Usage of Gazetteer Based Technique.
7. Removing the Suffix rules
   7.1 Removable of suffixes rules it can detect the places' names, organizations, Izaafats and another types of persons' names
8. Searching the names of persons and numbers
   8.1 Gazetteer Based can search many of the persons' names with various techniques
   8.2 It can be applied for Persons' names equal to three words writing.
   8.3 Short form of names can be detected.
   8.4 Ambiguity in names can be resolved.
   8.5 Numbers can be found in non-numeral form.
9. Some Abbreviations which were not found in the Gazetteer Based Lookup.
10. Organizations Searching Rules
    10.1 Some organizations which are not be possible to search in Gazetteer Based Technique, the entire rule can be found and tagged them.
11. Output of the tagged and untagged data

This algorithm is self-explanatory, where some of the steps which are concerned with Gazetteer lookup to extract the various Named Entities in each text. Collection of several Named Entities related to different fields like legislation, commerce and etc. In algorithm's step 3 gazetteer look up can be concerned with Locations, Brands, Terms, Abbreviations and Organizations Tags. These tags are not ambiguous that's why the system can easily work without any rule. Many persons' names have patterns just like suffixes or prefixes of the word, so, the step 6 has capability to detect the same kind of words. It has been taken ideas from another researchers [19].

7 Data Collection

In the computer age or digital world, it is almost common to collect data through two sources—primary sources and secondary sources. The data collect through the primary sources is called as primary data; and the data collected through the secondary sources is termed as secondary data according to experts of this field.

Primary Data is also said to be as 'raw data'. This data is actually collected by means of genuine source in a controlled or uncontrolled environment. It means that a controlled environment is based upon experimental research in which the researcher directly controls some variables. On the other hand, data collected by means of observations or questionnaire survey in a natural–cum–practical settings is good example of the data obtained in an uncontrolled environment [6].

Whereas, Secondary Data availed through the secondary sources such as: reports, journals magazines, books, documents, research papers, articles, dictionaries – soft and hard copies and websites etc. The simplest method to guess either a typical phrase is a named entity or not simply to look it up in a gazetteer. Look-up symptoms work prettily only with large entity lists. In case of ambiguous entities, the approach is usually competitive against machine learning algorithms. Generally, in machine learning approaches, Gazetteer features are also most common and performance of identification systems can be further more developed and progressed gradually. Presently, in the computer world and much striving digital world where all living and non-living things are shaping themselves in accordance with the customs of global village, there are a lot of websites resources that are, with less efforts, adaptable and accessible to NER, for instance: Britannica, Wikipedia, video libraries, programing, software and encyclopedia. Surely which are helpful and useful to materialize any dream dreamt by any outstanding and knowledge full person. Even, there will be so many other supplementary digital sources to assist global man in this digital world in coming years [7].

8 Layout of Sindhi NER System

Sindhi NER System shows the following extraction and it is connected to backend database. In this system user can search and mark the given below tags so our system will show their respective data Fig. 2 shows the dashboard of Sindhi NER System.

Fig. 2. Illustrates the Sindhi NER dashboard

Figure 3 shows the searching results using their tags like name, places, surname, designation and organization. This is the first Sindhi NER system which work like a google search engine and this is the desktop application. User can easily find out their desired data from Sindhi NER system.

Fig. 3. Shows the searching results

9 Conclusion

NER System aims to extract the existing information into the following categories such as: Person's Name, Organization, Location, Date and Time, Term, Designation and Short forms. Now, it is considered to be important aspect for many natural languages processing tasks such as: information retrieval system, machine translation system, information extraction system and question answering. Even at a surface level, the understanding of the named entities involved in a document gives richer analytical framework and cross referencing. It has been used for different Arabic Script-Based languages like, Arabic, Persian and Urdu but, Sindhi could not come into being yet. This paper explains the problem of NER in the framework of Sindhi Language and provides relevant solution. The system is developed to tag twelve different Named Entities. We have used Ruled based approach for NER system of Sindhi Language. For the training and testing, 936 words were used and calculated performance accuracy of 98.71%. It is a desktop application which recognize the words from database and in future we will work on web application using support vector machine (SVM) approach.

References

1. Shah, H., Bhandari, P., Mistry, K., Thakor, S., Patel, M., Ahir, K.: Study of NER for Indian languages. Int. J. Inf. Sci. Tech. (IJIST) **6**(1/2), 15–20 (2016)
2. Kanan, T., Ayoub, S., Saif, E., Kanaan, G., Chandrasekar, P., Fox, E.A.: Extracting named entities using named entity recognizer and generating topics using Latent Dirichlet allocation algorithm for Arabic news articles. In: Proceedings of the International Computer Sciences and Informatics Conference (ICSIC) (2016)

3. Patil, N., Patil, A.S., Pawar, B.V.: Survey of NER systems with respect to Indian and Foreign languages. Int. J. Comput. Appl. **134**(16), 21–26 (2016)
4. Jiang, R., Banchs, R.E., Li, H.: Evaluating and combining NER systems. In: Proceedings of the Sixth Named Entity Workshop, Joint with 54th ACL, Berlin, Germany, pp. 21–27 (2016)
5. Patawar, M.M.L., Potey, M.M.A.: Approaches to NER: a survey. Int. J. Innov. Res. Comput. Commun. Eng. **3**(12), 37–42 (2015)
6. van Hooland, S., Wilde, M.D., Verborgh, R., Steiner, T., Van de Walle, R.: Exploring entity recognition and disambiguation for cultural heritage collections. J. Digital Sch. Humanit. **30**, 262–279 (2014)
7. Shaalan, K.: A survey of Arabic NER and classification. Assoc. Comput. Linguist. **40**(2), 469–510 (2014)
8. Tkachenko, M., Simanovsky, A.: NER: exploring features. In: Proceedings of KONVENS 2012, Vienna (2012)
9. Abdallah, S., Shaalan, K., Shoaib, M.: Integrating Rule-Based System with Classification for Arabic NER, pp. 311–322. Springer-Verlag, Heidelberg (2012)
10. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (Almost) from scratch. J. Mach. Learn. Res. **12**, 2493–2537 (2011)
11. Kaur, D., Gupta, V.: A survey of NER in English and other Indian languages. Int. J. Comput. Sci. Issues **7**(6), 89–95 (2010)
12. Liao, W., Veeramachaneni, S.: A simple semi-supervised algorithm for NER. In: Proceedings of the NAACL HLT Workshop on Semi-supervised Learning for Natural Language Processing, pp. 58–65 (2009)
13. Mansouri, A., Affendey, L.S., Mamat, A.: NER approaches. Int. J. Comput. Sci. Netw. Secur. **8**(2), 67–71 (2008)
14. Tran, T., Pham, T., Hung, T.X., Dinh, D., Collier, N.: NER in Vietnamese documents. Natural Institute of Informatics (2007)
15. Singh, U.P., Goyal, V., Lehal, G.S.: NER system for Urdu. In: Proceedings of Cooling Mumbai, pp. 2507–2518 (2012)
16. Kazama, J., Torisawa, K.: Exploiting Wikipedia as external knowledge for NER. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 698–707 (2007)
17. Alexander, E., Richman, P., Schone, P.: Mining Wiki resources for multilingual NER. In: Proceedings of the 46th Annual Meeting of the Association of Computational Linguistics: Human Language Technologies, Stroudsburg, PA, pp. 1–9 (2008)
18. Nadeau, D., Sekine, S.: A survey of NER and classification (2008). http://nlp.cs.nyu.edu/sekine/papers/li07.pdf
19. Belgaum, M.R., Soomro, S., Alansari, Z., Alam, M.: Ideal node enquiry search algorithm (INESH) in MANETS. Ann. Emerg. Technol. Comput. (AETiC) **1**(1), 26–33 (2017)

Exploring the Potential Benefits of Big Data Analytics in Providing Smart Healthcare

Salma Al Mayahi, Ali Al-Badi, and Ali Tarhini(✉)

Information Systems Department, Sultan Qaboos University, Muscat, Oman
{salmam, aalbadi, alitarhini}@squ.edu.om

Abstract. Big Data is an emerging technology in different sectors. It refers to massive amount of heterogeneous data produced from many sources. Big data analytics is the process of analyzing a huge set of data to build and discover meaningful patterns, correlation that will add value to the corresponding business through predictive decisions and other useful information.

The health industry every years generate big data in different formats. The healthcare data need analysis to make decisions and forecasting but there is lack of understanding of the potential of big data in health industry.

This paper aims to explore the potential values of big data analytics in healthcare to enhance the efficiency and smartness of healthcare services. In addition, conducting an experiment on the dataset exported from an online healthcare research repository on a big data analytical topic.

The study has been carried out by conducting literature review big data analytic and referring datasets from online healthcare research repository.

The research concludes that providing evidence-based treatment, making predictive analysis and providing efficient healthcare service are the main potential benefits of applying analytics in healthcare. However, ensuring anonymity of patients' information and educating healthcare staff about the role of analytics in healthcare are essential steps before adopting such technologies.

This research is conceptual in nature based on existing literature reviews and secondary data. In future primary data would be used to understand the relevance of big data analytics in healthcare.

Keywords: Big data · Big data analytics · Healthcare · Predictive analysis
Biomedical informatics · Data mining

1 Introduction

Digitalization is becoming part of our lives penetrating every aspect of our normal life. Through the evolution of digitalization, a massive amount of data had been generated over the past years which remained beyond the capabilities of the available data storages and managements referred to as "Big Data" as described by IBM [1]. Big data includes the legacy enterprise data, machine generated data like sensors and weblogs and the data generated by social media like Twitter, Facebook and YouTube, etc. [2]. In 2001, big data defined by three main characteristics: (1) Volume which is the massive amount of data generated by industries, (2) Velocity which describes the fast movement of data among parties, (3) Varity which is the various type of data sources

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 247–258, 2018.
https://doi.org/10.1007/978-3-319-95450-9_21

and types that includes social, mobile data, machine geographic data and biometrics [2]. In 2014, new characteristics of big data were proposed [3]:

- **Value**: How many values extracted from the data,
- **Veracity**: How accurate the data and how reliable is the source of data
- **Variability:** Consistency of the data and its continuity and availability
- **Viscosity:** Latency of data to the corresponding topic
- **Virality:** Rate of data spreading and how often the data is recurrent by another partner

This research aims to explore the potential benefits of big data analytics in healthcare by reviewing the previous studies and applications of this concept in health industry. The research is based on the study conducted between 2010 and 2017.

The paper is structured as follows: Sect. 2 presents the literature review. Section 3 describes the research methodology used. Section 4 presents the results. Section 5 provides a detailed discussion on the findings.

2 Literature Review

Big data will not be useful unless it provides information and meaningful and readable insights. Big data analytics is the process of constructing valuable patterns, useful information, and descriptive trends from the pool of data. Health industry is one of the major area that is growing exponentially and producing big data in the form of patient information, clinical notes, X-ray imaging and pharmaceutical data. In 2012, Bonnie Feldman said that it is expected that data over the world in healthcare will be 50% more than the current data and it will reach 25000 petabytes whereas in 2012 it was only 500 petabytes [4]. In addition, the unstructured data like written clinical notes, video and audio streams will increase 15 times more than the structured data along with the new forms of data beside the existing types like human genetics data, radiology images and biometric, and genomics readings [4].

There are tremendous benefits and huge advantages of big data analytics in healthcare sector will reach the level of disease investigation to the level of treatment. In US, there is a high demand to healthcare big data analytics since the expenses has been increasing rabidly in last decades [5]. In 2011, Manyika from McKinsey Global Institute proposed that if USA healthcare applied effective and innovative big data analytics, USA could save $300 billion every year [6]. Furthermore, big data analytics creates transparency and easy accessibility to relevant data in order to create more values and facilitates analytical experimentation to investigate needs and supports leader's decisions as an evidence [6]. The new innovative pathway of applying big data analytics in healthcare include the following benefits [5]:

- **Right living:** The advancing the lifestyle by engaging the end users in the health care process will eventually minimize the needed care by industry.
- **Right Care:** Care is provided to the patients based on evidence and this will ensure more safety and enhance the expected results.

- **Right provider:** The provider selection will be more accurate and give insights toward quality.
- **Right Value:** The analytics will help reducing the care cost and yet maintain the quality.
- **Right innovation:** The analytics will help encouraging innovative health solution, discovering new correlations and making new trends.

There are many promising applications of big data analytics proving the potential values of this technology in healthcare [2].

2.1 Challenges of Big Data Analytics

There are many challenges in implementing big data analytics due to the high complexity and diversity of the healthcare data sets [7]. As stated by Ward, Marsolo and Froehle in their study in 2104, the challenges to the applications of big data analytics include (1) no standard protocols of data structures (2) data collection obstacles (3) lack of qualified big data analysts [8]. A number of researchers from University of Otago, New Zealand, stated that they faced big challenge in managing the big data within the scope of the project in collaborative way [7]. Another challenge raised by LaValle from MIT Sloan Management Review is that the adoption process of data analytics might face data quality issues, unproductive data governance and management barriers [9].

2.2 Advantages of Big Data Analytics to Healthcare

LaValle from MIT Sloan Management Review partnered with the IBM stated that better utilization of the available technologies and tools is essential to leverage the healthcare data effectively and to help organization realize what is currently happening around and predicts what is most likely to happen to take proactive actions and be well prepared [9]. MIT Sloan Management Review conducted a wide survey among 3000 executives and managers in 100 countries with different sectors and interests and the key finding they confirmed that top-performing business utilizes the analytics in their data five times more than their opponents who perform less [9].

Raghupathi and Viju in their review in 2014 are supporting LaValle from MIT Sloan Management Review research outcomes of the survey by stating that big data analytics potential benefits are promising the healthcare industry with valuable outcomes and valuable results [10]. In addition, the use of analytics help organization converting challenges into opportunities by constructing future strategies and making day-to-day processes guide [9]. Manyika and colleagues from McKinsey Global Institute (MGI) stated in their report in June 2011 that big data analytics in healthcare have potential values and extraordinary results in reducing the cost, increasing the revenues and the efficiency and improving effectiveness of patient treatment [6]. Manyika proposed the benefits of big data analytics in healthcare into five categories. The first category is improving the use of clinical operations in better treatment and effective diagnosis of the diseases. The second category is reducing the treatments' cost and overall healthcare expenditure. The third category is raising the value of research and development by building predictive modules and developing new algorithms that will improve the clinical design. The forth category is building new business models in healthcare industry. The fifth category is improving public health surveillance and responses [6].

In 2014, a group of researchers from Health Affairs in US conducted a research by applying big data analytics to identify the opportunity of reducing the healthcare costs. The research goals achieved by predicting the number of readmissions and high cost patients in six different use cases of inpatients records and they conclude that such analytics is considered as powerful tool to be adopted in healthcare industry [11]. Moreover, a recent literature review conducted in 2016 among 209 articles provided an obvious evidence of an exponential positive impact of big data analytics in increasing the accuracy and quality in healthcare services and reducing the costs of clinical analysis [12].

In 2014, a study conducted to experiment the potential benefits of big data analytics on electronic health records (HER) to build predictive models using three databases from different health systems and proved that the efficiency of building research models increased and the reuse of the health data helped in creating useful research objectives [13]. In 2016, Srinivasan Suresh stated that predictive analysis can be utilized to help healthcare providers to better prescribe medication to children and to give better awareness to the patients about their health which will lead to improvement in the collaboration between the physicians and patients and improve their life style habits [14].

2.3 Worldwide Applications of Big Data Analytics in Healthcare

Nowadays, health organizations around the world are starting to realize the value of big data analytics to enhance their productivity in healthcare. In 2013, United Kingdom initiated a project called care.data which aims to link all patients health records with the social healthcare in a centralized place owned by the Health and Social Care Information Centre (HSCIC) [15]. The main objectives of care.data project is to make predictive analysis for the various disease among UK citizens and to help the government to provide better healthcare services with evidence based treatments [15, 16]. In 2012, National institute of health in US lunched BigData to Knowledge (BD2K) and aimed to develop innovative tools and methods to utilize the biomedical big data into useful knowledge in healthcare. In 2015, a team from Oxford university conducted a project with the BD2K Center for Causal Discovery to test and train biomedical data consisting of cancer driver, lung disease information and human brain data by developing a new algorithm with a user-friendly system based on Bayesian algorithms “a datamining technique” [17]. In 2013, Ahumada (2013) proposed a medication alert fatigue application which aimed to discover and assess the medication alerts in Children’s Hospital of Philadelphia [18]. This application implemented an analytical dashboard with a user-friendly interface to investigate drug-allergy and maximum dose alerts that helps in clinical decision supports and real time management.

In 2015, IBM Watson Health project is launched and considered to be one of the innovative approaches toward raising up the level of efficiency in healthcare in more simple and creative way [19]. IBM Watson is a cloud based project which is capable of storing a large size of structured and unstructured data, evaluate them and provide evidence based results [19]. Furthermore, Watson health is a cognitive system that play a significant role to uncover the value hidden in the unstructured data in order to provide evidence based reasoning to support healthcare staff to make proper decisions [20]. As an evidence, more than 80% of Watson health's executives show positive influence in their running business [20]. In 2015, a practical study conducted by Househ, Hasman and Mantas from Germany Heilbronn university to predict the survival rate of the colon cancer from a predefined attribute based on medical experts' opinions. The objective of the project is to compare the data mining algorithms' accuracy with the physician's accuracy and their results shows that data mining algorithms an accuracy of about 67.7% comparing to 59% accuracy of the physician's [21]. There are many other projects and practical researches done in big data analytics nowadays around the world. Table 1 summarizes studies on valuable projects, areas, limitations and benefits in healthcare.

Table 1. Worldwide applications of big data analytics in healthcare

Project name	Project place	Technology used	Limitation	Benefits
Care.data	UK	NA	Mismanagement, miscommunications, inadequate protections for patient anonymity, conflicts with doctors	1. Predict diseases pattern 2. Future plans for health services for each area in UK 3. Propose better treatment 4. Make evidence based treatment
BD2K	US	Casual Bayesian algorithms	Might not fit in all biomedical domains	1. Predict the drivers of cancer disease 2. Discover factors leading to lung disease
Medication alert fatigue	Philadelphia	SQL queries of Children's Hospital of Philadelphia EHR database		1. Better surveillance on the medication alerts 2. Helps in Day-to-Day management
IBM Watson Health		Deep QA		1. Provide evidence based result 2. Support decision making for healthcare staff

(*continued*)

Table 1. (*continued*)

Project name	Project place	Technology used	Limitation	Benefits
Survival prediction	Germany	Data mining algorithms		1. Good prediction analysis 2. Support decision making 3. Can fit to other problems in health prediction 4. High accuracy

3 Research Methodology

The research started by conducting a thorough investigation using systematic literature review to explore the potential benefits achieved by applying big data analytics in healthcare domain. Next step after reviewing was building the conceptual understanding, discussion, and evaluation of the main benefits. This step is followed by preparing the dataset, the analytical tool and the techniques to be used in conducting the experiment. After that, a pilot study on a small amount of data to test the methodology and to make any adjustment in the experiment methodology before conducting the main study performed. This step helps to visualize constrains and obstacles that might show up during the main study. Finally, conducting the main study, analyze the results, discuss and map the output with the findings from the literature review.

This research focuses on the main potential benefits of big data analytics in healthcare collected from various research papers, reviews and projects. Figure 1 shows the considered benefits in the present study.

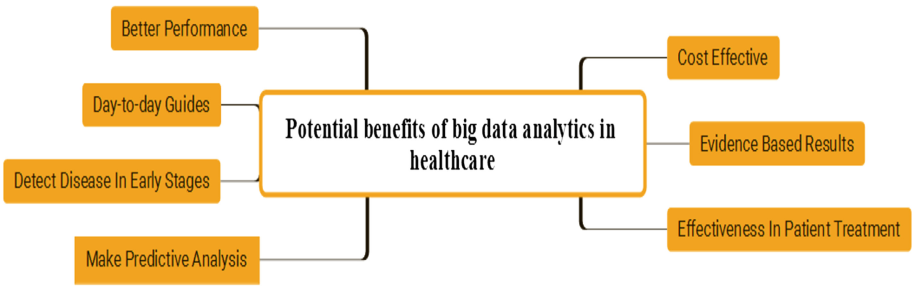


Fig. 1. Potential benefits of big data analytics in healthcare

4 Experimental Investigation on Big Data Analytics in Healthcare

This experiment aims to elaborate deeply on the idea of big data analytics in healthcare, to investigate the anticipated potential benefits in healthcare industry, and to support the findings from the literature review. The experiment is conducted using a real dataset about chronic kidney disease exported from UC Irvine Machine Learning Repository [22] and analyzed using an open source tool called Weka explorer version 3.8. This tool is used to apply two classification algorithms: naïve Bayesian and J48 trees. Classification is a data mining technique used to build a model that describes the data by analyzing and supervised learning of training data in which the class label is known and the resulted model in this case is then used for predicting whether the patient has a chronic kidney disease or not [23].

The exported dataset of chronic kidney disease is used to predict the disease and it consists of 25 attributes and 400 instances [22]. The dataset originally contains 400 as taken from the online repository, but when applied data mining techniques it reduced to 160 due to pre-possessing of the records. The list of attributes, description and the used abbreviations shown in the Table 2.

Table 2. Dataset of chronic kidney disease description

S. No	Abbreviation	Description
1	age	Age
2	bp	Blood Pressure
3	sg	Specific Gravity
4	al	Albumin
5	su	Sugar
6	rbc	Red Blood Cells
7	pc	Pus Cell
8	pcc	Pus Cell Clumps
9	ba	Bacteria
10	bgr	Blood Glucose Random
11	Bu	Blood Urea
12	Sc	Serum Creatinine
13	sod	Sodium
14	pot	Potassium
15	hemo	Hemoglobin
16	pcv	Packed Cell Volume
17	wc	White Blood Cell Count
18	rc	Red Blood Cell Count
19	htn	Hypertension
20	dm	Diabetes Mellitus
21	cad	Coronary Artery Disease

(*continued*)

Table 2. (*continued*)

S. No	Abbreviation	Description
22	appet	Appetite
23	pe	Pedal Edema
24	ane	Anemia
25	class	Class

The first classification algorithm to be applied is naïve Bayesian which is considered to be a very powerful algorithm which deals with each attribute independently and it performs well in the dataset that contains missing values [24]. After applying this algorithm, the resulted output is shown in the Fig. 2.

```
=== Summary ===

Correctly Classified Instances         152               95      %
Incorrectly Classified Instances         8                5      %
Kappa statistic                          0.8943
Mean absolute error                      0.0493
Root mean squared error                  0.2043
Relative absolute error                 10.5446 %
Root relative squared error             42.5665 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.922    0.000    1.000      0.922    0.960      0.899    1.000     1.000     ckd
                 1.000    0.078    0.877      1.000    0.934      0.899    1.000     1.000     notckd
Weighted Avg.    0.950    0.028    0.956      0.950    0.951      0.899    1.000     1.000

=== Confusion Matrix ===

  a  b   <-- classified as
 95  8 |  a = ckd
  0 57 |  b = notckd
```

Fig. 2. Naïve Bayesian algorithm result

From the results of this algorithm, the achieved percentage of the corrected classified instances is 95%, which indicates a very good results and it gives the anticipated results using percentage split 60%. From the resulted confusion matrix, only 8 records were incorrectly classified. After increasing the percentage splits to explore more accuracy, the following Table 3 shows the results achieved. With 60% percentage split, the accuracy is more and the recall is better.

Table 3. Percentage spilt and achieved results

Percentage split	Accuracy	Error rate	Sensitivity	Specificity	Precision	Recall
60%	95%	5%	0.950	$1 - 0.028 = 0.972$	0.956	0.950
80%	92.5%	7.5%	0.925	$1 - 0.040 = 0.960$	0.938	0.925
90%	90%	10%	0.900	$1 - 0.067 = 0.933$	0.920	0.900

Figure 2 represents the readings when the splits percentage set to 60% only but the other readings of Table 3 shows the results when we change the split percentage to different one but at the end we achieve a better result in 60% split percentage which is represented in Fig. 2. In Fig. 2 also we refer to the readings of weighted average. The second classification method is J48 which is a decision tree algorithm that is constructed from a learning process of the given data and used widely in medicine, financial and biology [23]. After applying the algorithm to the chorionic kidney disease dataset, the following Fig. 3 shows the great result obtained by this algorithm. Obviously, this algorithm performed perfectly in chronic kidney disease dataset and 100% accuracy achieved. As shown from the confusion matrix, all the instances were classified correctly. Figure 4 shows resulted decision tree.

```
=== Summary ===

Correctly Classified Instances         160              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0.0218
Root mean squared error                  0.0856
Relative absolute error                  4.6561 %
Root relative squared error             17.8392 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     ckd
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     notckd
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000

=== Confusion Matrix ===

   a   b   <-- classified as
 103   0 |   a = ckd
   0  57 |   b = notckd
```

Fig. 3. J48 algorithm result

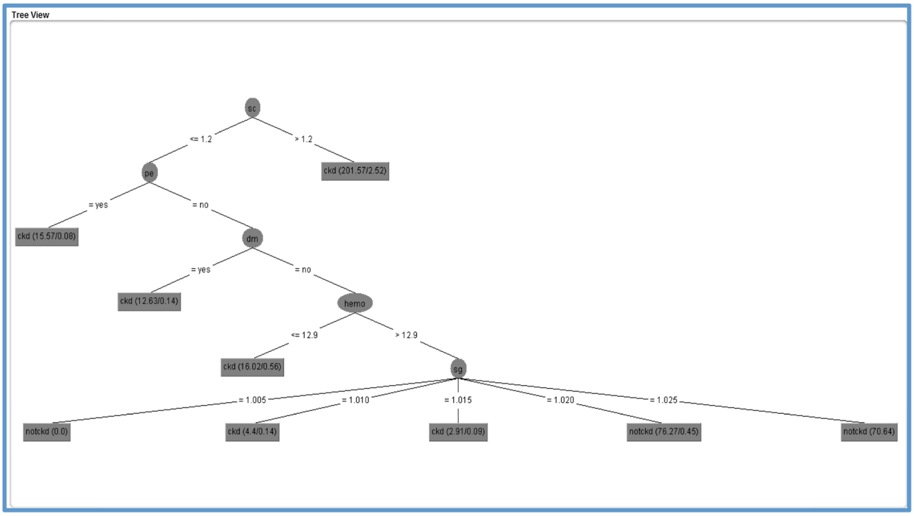


Fig. 4. Decision tree result

All the 25 attributes from Table 2 attributes are included in the data mining process but the tree have been created based on the significance of the attribute in the result whether to have kidney disease or not. From the tree, the main factors of the chronic kidney disease to be taken into consideration. The performance of the two algorithms is very strong and could be used to train different health problems datasets. As Patil and Sherekar proved in their research in 2013, both algorithms are efficient and give high accuracy [24].

5 Analysis and Discussion

From the exploration of big data analytics in healthcare in the literature review, the main benefits that most of the researchers have reached consensus are on providing evidence based treatment, improving patients' treatment, making predictive analysis and supporting decision making in healthcare. As a clear evidence, the experimental investigation conducted supports the findings of the literature review. The Table 4 explains the mapping between the benefits from the literature review and the experimental investigation of big data analytics using chronic kidney disease dataset.

Table 4. Mapping between the benefits from the literature review and the experimental investigation

SN	The benefit from literature review	References	Experimental investigation
1	Better performance	[6, 9, 19, 21]	The experiment shows high accuracy and the running time in that dataset was efficient. It took 0.01 s only but it may vary depending on the size of the dataset. Moreover, the graphical representation of the disease factors helps the doctor to grasp the knowledge in more efficient way
2	Day-to-Day guides	[6, 9, 18]	The medical staff could benefit from the classification results in their daily diagnosis process with different patients' cases
3	Detect disease in early stages	[14, 18, 21]	There is no obvious evidence from the experiment to achieve this benefit but it might be possible after using other type of analytics
4	Make predictive analysis	[6, 11, 12, 16, 21]	The dataset is trained by the classification algorithms to provide predictive analysis based on its attributes
5	Cost effective	[7, 11, 12]	The experiment shows very good performance comparing to reviewing the historical data using manual routines. This helps in saving medical staff time and hence, it is cost effective

(*continued*)

Table 4. (*continued*)

SN	The benefit from literature review	References	Experimental investigation
6	Evidence based results	[6, 15, 16, 19]	From the historical dataset used in the experiment, it might be used as an evidence to prove the achieved results and the decision taken by the medical staff
7	Effectiveness in patient treatment	[6, 10, 12, 21]	From the machine learning in this experiment, it shows high accuracy for prediction and this helps in reducing the errors while treating the patients

In addition, the literature review shows that big data analytics in healthcare is widely applied in different areas around the world. It shows that healthcare analytics provide positive impact and great benefits in healthcare sector. As highlighted in the review, there are some limitations that must be considered in adopting analytics in healthcare. The most important concern is maintaining the anonymity of the patients' information. Another concern is to avoid adoption's conflicts that can be resolved by developing well-managed project documentations and educating the healthcare staff on the potential benefits and application of big data analytics. In addition, analysts must have eligible skills and experience in analyzing healthcare datasets and medical expertise to understand the dataset's variables and correlations.

6 Conclusion

Big data analytics has a tremendous potential benefits in healthcare that will smarten the services and increase the services' efficiency. This research explored the anticipated benefits of applying the analytics in healthcare and looked into the various applications around the world. Moreover, an experiment conducted to investigate the potential benefits using a real dataset and it showed very promising results. Adopting the analytics in healthcare is essential but the highlighted limitations and challenges must be well addressed and resolved. Moreover, there are research topic, which has scope in healthcare such as apply different data mining algorithms and measure their accuracy; apply the experiment on different datasets in healthcare with supervision of a medical expert in the selected domain; analyze unstructured dataset like doctors' clinical notes in a local hospital; select different analytical tool and try to investigate other results.

References

1. Dewey, J.: Big Data. Salem Press Encyclopedia (2014)
2. Priyanka, K., Kulennavar, N.: A survey on big data analytics in health care. Int. J. Comput. Sci. Inf. Technol. **5**(4), 5865–5868 (2014)
3. Vorhies, W.: How many V's in big data? The characteristics that define big data. Data Science Central, October 2014. http://www.datasciencecentral.com/profiles/blogs/how-many-vs-in-big-data-the-characteristics-that-define-big-data

4. Feldman, B., Martin, E.M., Skotnes, T.: Big data in healthcare hype and hope. Dr. Bonnie **360** (2012)
5. Groves, P., et al.: The ‘big data’ revolution in healthcare. McKinsey Q. **2** (2013)
6. Manyika, J.: Big Data the Next Frontier for Innovation, Competition & Productivity. McKinsey & Company, S.L. (2011)
7. Frost, S.: Drowning in big data? Reducing information technology complexities and costs for healthcare organizations (2015)
8. Ward, M.J., Marsolo, K.A., Froehle, C.M.: Applications of business analytics in healthcare. Bus. Horiz. **57**(5), 571–582 (2014)
9. LaValle, S., et al.: Big data, analytics and the path from insights to value. MIT Sloan Manag. Rev. **52**(2), 21 (2011)
10. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. **2**(1), 1 (2014)
11. Bates, D.W., et al.: Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff. **33**(7), 1123–1131 (2014)
12. de la Torre Díez, I., et al.: Big data in health: a literature review from the year 2005. J. Med. Syst. **40**(9), 209 (2016)
13. Ng, K., et al.: PARAMO: a PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records. J. Biomed. Inform. **48**, 160–170 (2014)
14. Suresh, S.: Big data and predictive analytics: applications in the care of children. Pediatr. Clin. North Am. **63**(2), 357–366 (2016)
15. Patient: Care.data - sharing your information (2014)
16. Sterckx, S., et al.: “You hoped we would sleep walk into accepting the collection of our data”: controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. Med. Health Care Philos. **19**(2), 177–190 (2016)
17. Cooper, G.F., et al.: The center for causal discovery of biomedical knowledge from big data. J. Am. Med. Inform. Assoc. **22**(6), 1132–1136 (2015). https://doi.org/10.1093/jamia/ocv059
18. Ahumada, L.M., et al.: Medication alert fatigue: the design and use of a medication alert dashboard as part of a comprehensive approach to drug-drug interaction alerts. In: Anesthesia and Analgesia. Lippincott Williams & Wilkins, Philadelphia (2013)
19. Aggarwal, M., Madhukar, M.: IBM’s Watson analytics for health care: a miracle made true. In: Cloud Computing Systems and Applications in Healthcare, pp. 117–134 (2016)
20. IBM: IBM Watson Health: A New Partnership Between Humanity and Technology (2016). http://www.ibm.com/watson/health/. Accessed 12 June 2016
21. Househ, M., Hasman, A., Mantas, J.: Enabling Health Informatics Applications. Studies in Health Technology and Informatics. IOS Press, Amsterdam (2015)
22. UCI: UC Irvine Machine Learning Repository, Chronic_Kidney_Disease Data Set (2015). http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease. Accessed 10 Dec 2016
23. Marston, S., et al.: Cloud computing—the business perspective. Decis. Support Syst. **51**(1), 176–189 (2011)
24. Patil, T.R., Sherekar, S.: Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int. J. Comput. Sci. Appl. **6**(2), 256–261 (2013)

Clinical Practice for Diagnostic Causes for Obstructive Sleep Apnea Using Artificial Intelligent Neural Networks

Mashail Alsalamah[1,2(✉)], Saad Amin[1,2], and Vasile Palade[1,2]

[1] Coventry University, Priory Street, Coventry CV1 5FB, UK
mashail.alsalamah@gmail.com
[2] Qassim University, Buraydah, Al-Mulida, Saudi Arabia

Abstract. Sleep apnea is a serious sleep disorder phenomena which happens when a person's breathing is paused during sleep. The most common diagnostic technique that is used to deal with sleep apnea is Polysomnography (PSG) which is done at special sleeping labs. This technique is expensive and uncomfortable. New automated methods have been developed for sleep apnea detection using artificial intelligence algorithms, which are more convenient and comfortable for patients. This paper proposes a novel scheme based on deep learning for sleep apnea detection and quantification using statistical features of ECG signals. The proposed approach is experimented with three phases: (1) minute-based apnea classification, (2) class identification and minute-by-minute detection for each ECG recording unlike state-of-the-art methods which either identify apnea class or detect its presence at each minute and (3) comparison of the proposed scheme with the well-known methods that have been proposed in the literature, which may have not used the same features and/or the same dataset. The obtained results demonstrate that the proposed approach provides significant performance improvements when compared to state-of-the-art methods. The outcome of this study can be used as an assistant tool by cardiologists to help them make more consistent diagnosis of sleep apnea disorder.

Keywords: Obstructive sleep apnea · Deep learning · Neural networks
Sleep disorder · Big data

1 Introduction

Sleep apnea is a potentially common sleep disorder in which a person's breathing may have one or more pauses during sleep. These pauses may continue from a few seconds to several minutes, and may occur hundreds of times during the night. If the obstruction to breathing is total and continues for ten or more seconds, then this case is called apnea [1]. Sleep apnea may have long-term effect on the cardiovascular system which makes it a risk factor for increasing mortality rate [2]. Sleep apnea typically is classified into three types; obstructive sleep apnea (OSA), central sleep apnea (CSA) and mixed sleep apnea (MIX). OSA is the more common form of apnea; it is caused by a blockage of the airway and is generally associated with a reduction in blood oxygen saturation [3].

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 259–272, 2018.
https://doi.org/10.1007/978-3-319-95450-9_22

Traditionally, sleep-related breathing disorders are diagnosed by visual observation of Polysomnography (PSG) signals. PSG is a sleep test that is performed at special laboratories [4]. Even though PSG become the standard diagnostic tool for sleep disorder cases, there are some problems related to its implementation which make it expensive and time consuming. Therefore, the need for a simpler alternative detection method has been arising. Automated methods that use the artificial intelligence algorithms can solve PSG problems. In this light, there have been many algorithms and schemes proposed to address the problem of automatic OSA detection. such solutions used ECG patterns obtained during PSG studies and machine learning techniques to tackle OSA detection.

The dataset of ECG recordings used in this work is obtained from the PhysioNet web site. These recordings were arranged into three classes, as follows: (1) Class A (apnea): the learning and test sets each contain 20 class A recording files, each file contains at least 100 min with apnea; (2) Class B (borderline): recordings in this class contain between 5 and 99 apnea minutes. Each of the learning and test sets contain 5 class B files; and (3) Class C (control): recordings in this class contain fewer than 5 apnea minutes. The learning and test sets each contain 10 files [5].

Deep learning is currently one of the most important active research areas in machine learning. It has attracted extreme attention from researchers due to its potential in wide range of active applications such as object recognition [6]; [7], speech recognition [8]; [9], natural language processing [10], medical science [11]; [12], and other vital fields. Inspired by the biological nature of human brain mechanisms for natural signals processing, deep neural networks are representation learning methods with multiple levels of representation [13]; [14]. The expression “deep” is used because the depth of the network is greater when compared to the more conventional neural networks, which are sometimes called shallow networks. In most conventional learning methods, a simple network with one hidden layer may achieve acceptable performance for performing a specific task but, by applying a deep architecture with more hidden layers; higher efficiency can be achieved. This is because each hidden layer extracts more features from the previous layer and creates its own abstract representation. Therefore, to resolve more complicated features, we have to add more hidden layers, which make deep learning capable of learning latent information [15].

Most of the recent experimental results with deep architecture are obtained with models that can be turned into deep supervised neural networks, but with initialization or training schemes different from the classical feed-forward neural networks [16].

This work proposes a new contribution to the trend of training deep neural network based on topological concepts like weights initialization and activation functions. This is in part inspired by observations obtained from the work proposed by [17] which approved that for those neural networks with the same number of hidden units; deep architectures, with arctangent and polynomial activation functions, can realize maps with a higher complexity with respect to shallow ones. Also the work proposed by [18] approved empirically that deep supervised neural networks can reach their best performance without requiring any unsupervised pre-training. This finding was an attempt to close the performance gap between neural networks learnt with and without unsupervised pre-training.

The aim of this study is to propose a novel scheme for OSA detection based on features of ECG signals. This scheme is a hybrid algorithm that combines the Deep Neural Network (DNN) with the Decision Tree. The classification process in this proposed scheme consists of two phases; the first phase uses DNN for minute-based classification, then the output of this phase is fed into a decision tree model in order to perform the second phase; class identification. In addition to the proposed scheme, a comparative study of the most used classification methods, that have not been used with the same features and dataset, adopted in the literature is done.

The rest of this paper is organized as follows. Section 2 summarizes the related work in the literature. Section 3 contains an overview of the system and details the paper methodology. In Sect. 4, the experimentations and the obtained results are presented. Section 5 concludes the proposed study and lists possible extensions to this work.

2 Literature Review

Since PhysioNet/CinC challenge, many methods using the ECG signal to diagnose OSA have been proposed. The algorithms in this research area were divided into two types; some for apnea classes' identification and others for the minute-by-minute apnea classification.

Regarding the first challenge, several algorithms using different methods were developed to identify apnea class. For example, in [19–21] authors made use of spectral analysis of heart rate variability (HRV) to identify apnea class and achieved 30 correct score out of 30 (without class B consideration). While authors in [22, 23] used the Hilbert transform to extract frequency information from the heart rate signal and achieved a score of (28/30). Authors in [21, 24, 25] achieved the top three ranks in the PhysioNet's challenge on the subject of the minute-by-minute quantification. They reached an accuracy of 89.4%, 92.6%, 92.3%. In addition to HRV, authors made use of different features derived from ECG signals like ECG pulse energy [21], R-wave amplitude using power spectral density (PSD) [24] and T-wave amplitude using the discrete harmonic wavelet transform [25].

Since the challenge, different automated schemes have been proposed to detect OSA on the same PhysioNet Apnea-Ecg dataset. Khandoker et al. [26] employed wavelet based features and K-Nearest-Neighbour (KNN) classifier to achieve an accuracy of 83%. Xie and Minn [27] applied a number of classification algorithms as AdaBoost with Decision Stump and Bagging with REPTree to the extracted features from ECG and saturation of peripheral oxygen (SpO2) signals to obtain classification accuracy of 77.74%.

Different studies demonstrated that detection of obstructive sleep apnea can be achieved through HRV and the ECG signal. Quiceno-Manrique et al. [28] proposed a diagnostic scheme for OSA detection using time-frequency distributions and extracted features from ECG signal. This scheme was able to achieve up to 92.67% accuracy. In addition, authors in [29] proposed a technique that depends on features of the ECG signal and uses a bivariate auto regressive model to evaluate beat-by-beat power spectral density of HRV and R peak area and got classification accuracy higher than

85%. In 2012, Laiali Almazaydeh et al. proposed an automated classification scheme based on support vector machine (SVM) using statistical features extracted from ECG signals explicitly Heart Rate Variability (HRV) and obtained a classification accuracy of 96.5% [30].

Even though the aforementioned studies achieved relative satisfactory performance on apnea detection and quantification, there are some important aspects have to be highlighted. First, the proposed approaches either identify apnea class or detect the presence or absence of each minute of ECG data. To the best of our knowledge, only authors in [31, 32] addressed both apnea detection and quantification for each patient recording but both identify only two class not the three ones (class B is excluded). Second, different features are extracted from the RR intervals without any concern regarding its numbers and impact which implied that prediction of the classification models to be more excessive. Moreover, extracting and selecting features from such high-dimensional feature spaces require large computational resources, that is not reasonable for such wearable devices. As well, it is inadequate for home-based applications that have to aid physicians provide a quick pre-diagnosis for patient status. Therefore, to tackle these issues, this study proposes a novel OSA detection scheme to achieve acceptable performance using less features under limited capacities of wearable devices.

3 Methodology

This work is based on the ECG signal features to detect sleep apnea. Phases of the proposed methodology is discussed in the following subsections (see Fig. 1).

3.1 Data Collection Phase

The experimental data used in this study was obtained from the PhysioNet Apnea-ECG database [33]. The Apnea-ECG database contains ECG recordings for 70 different patients with OSA (classes a, b, c). Recordings vary in duration from slightly less than 7 h to nearly 10 h each. However, only 35 of these recordings contain minute-wise apnea annotations, which indicate the presence or absence of apnea during each minute of ECG data. ECG signals are sampled at 100 Hz with 12-bit resolution.

3.2 Data Pre-processing Phase

RR Intervals Extraction

The features used in our experimentation were all metrics based around RR intervals. An RR interval is defined as the time between two consecutive R peaks (see Fig. 2), which in turn are defined as the maximum amplitude of a given QRS complex. QRS is the combination of the Q wave, R wave and S wave and represents ventricular depolarization. The normal duration of the QRS complex is 0.08 and 0.10 s [34]. These metrics were chosen because RR intervals have been shown to be a telling indicator of HRV, which is a known byproduct of sleep apnea [35].

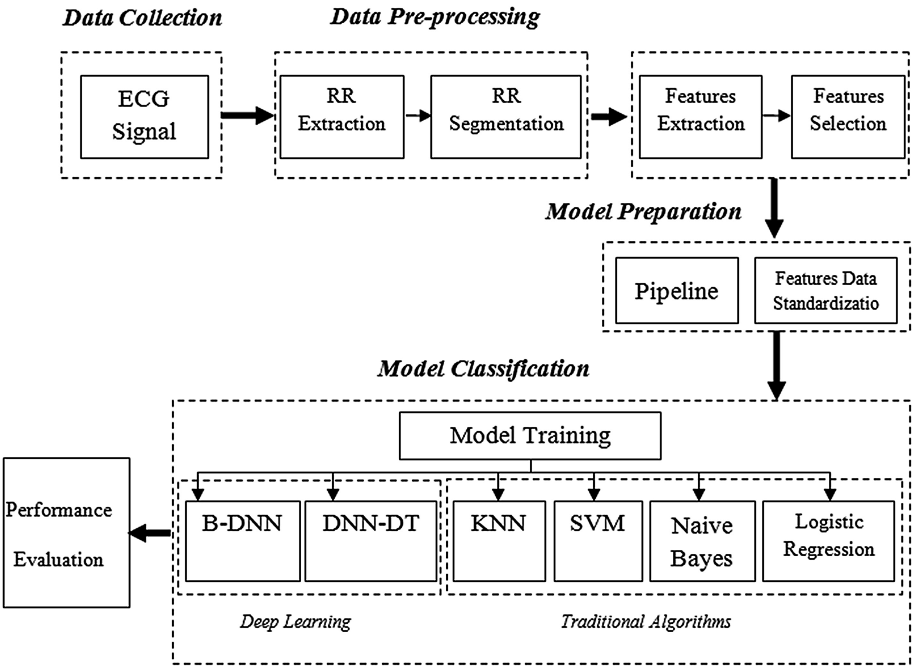


Fig. 1. Block diagram of the proposed methodology

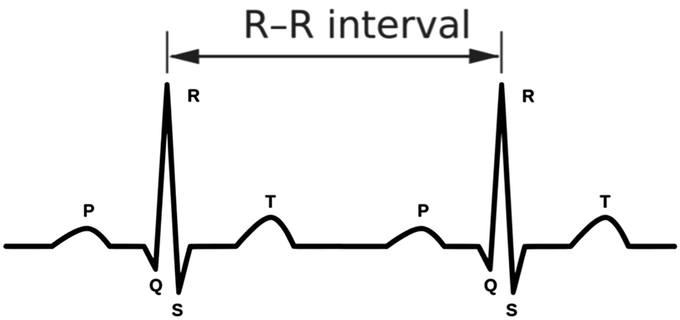


Fig. 2. Calculations of RR interval

RR Intervals Segmentation

As each ECG recording in the PhysioNet database was annotated per minute, the extracted RR intervals are segmented on a minute-by-minute basis according to the annotations. Therefore, RR intervals was calculated for each minute at each file; which implies that we have about 17003 RR records (35 file * file length (450–550 min)). Figure 3 presents the extracted RR values for one of the data set files (a01 file).

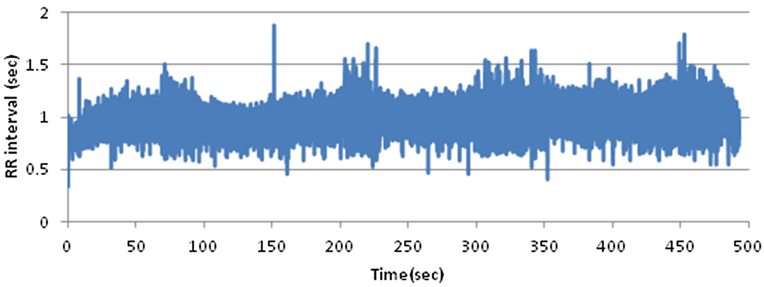


Fig. 3. Extracted RR intervals from (a01) file

3.3 Features Extraction Phase

Four each segment of the RR intervals obtained from the previous preprocessing phases, statistical features could be extracted and fed into the classification model for the possible classification of apnea events. Each feature vector was computed based on 60 s of ECG data; as each minute-wise annotation indicates the presence or absence of apnea at the beginning of the following minute. The following ECG features, which are the most common features used in the literature [4, 16] for apnea detection, are calculated:

1. Mean of the RR-interval.
2. Median of RR-intervals.
3. Standard deviation SD, of the RR-interval.
4. The NN50 measure (Variant 1), is the number of pairs of neighboring RR-intervals such that the first RR-interval exceeds the second RR-interval by more than 50 ms.
5. The NN50 measure (Variant 2), is the number of pairs of neighboring RR-intervals where the second RR-interval exceeds the first RR-interval by more than 50 ms.
6. The PNN50_1 measures, defined as NN50 (variant 1) measure divided by the total number of RR intervals.
7. The PNN50_2 measures, defined as NN50 (variant 2) measure divided by the total number of RR intervals.
8. The SDSD measures, defined as the standard deviation (SD) of the differences between neighboring RR-intervals.
9. The RMSSD measures, defined as the square root of the mean of the sum of the squares of differences between adjacent RR-intervals.
10. Inter-quartile range, defined as difference between 75th and 25th percentiles of the RR interval value distribution.
11. Mean absolute deviation values, defined as mean of absolute values by the subtraction of the mean RR-interval values from all the RR interval values in an epoch.

3.4 Features Selection Phase

In this phase, the features, that have the strongest effect on prediction, are selected. This stage scores the attributes according to their correlation with the classified apnea class.

It selects the most informative attributes. In total, 11 features were extracted from each ECG minute. In order to determine the discriminative power of each feature, ANOVA [36] statistical tests was adopted. After applying ANOVA test to features vector (see Fig. 4); it was induced that NN50_1, NN50_2, pNN50_1, pNN50_2 are the less relevant features and does not contribute highly in the classification results; so they are eliminated from the features set to have 7 features instead of 11.

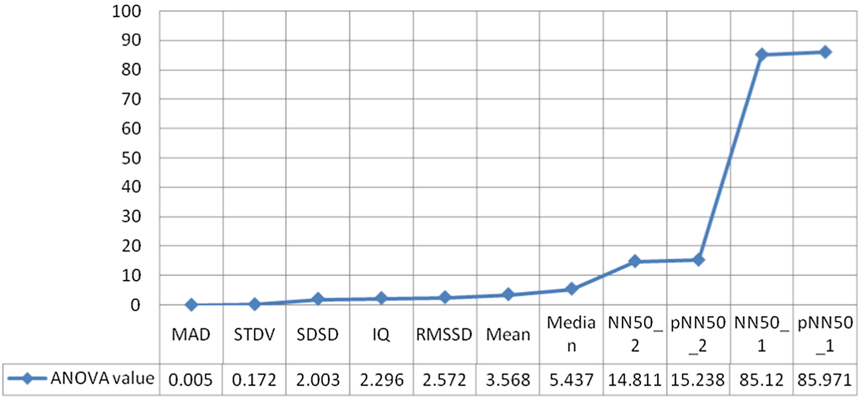


Fig. 4. Values of ANOVA test for features set

3.5 Model Classification

In the classification process, the extracted and selected features have to be fed into the training model to classify each minute of ECG data. In this work; two approaches were proposed for model training process. The first approach is based on the concept of deep learning and the other one is done using the traditional classification algorithms particularly Logistic Regression, KNN, SVM and Naïve Bayes models. In the rest of this section, detailed description of the proposed approaches is provided.

Deep Learning Approach

The proposed deep learning model passes two phases. In the first phase; a baseline deep neural network (B-DNN) model is proposed. This model is mainly used for the stage of minute-based classification. While in the second phase; a hybrid model is designed by the fusion of deep neural network model and decision tree model. This hybrid model is used for the stage of minute-class-based classification. Keras [37] library was used for building the proposed deep models.

Phase 1: Baseline Deep Neural Network Model (B-DNN)

The proposed Baseline Deep Neural Network (B-DNN) model uses feed-forward neural network architecture which is called a Multi-Layer Perceptron (MLP). It consists of 4 layers; the input layer, two hidden layers and output layer. A neural network topology with more hidden layers have the potential to extract better representations and features from the raw data to create much better classification models.

The first hidden layer has the same number of nodes as input parameters (7 neurons). While the second one was added to force a type of feature extraction by the network by restricting the representational space, since it take an input of 7 neurons (same number as of selected features) and reduce it to 5 (a new representation of the input features). This will enforce the model during training to select the most important representation of the input data.

Figure 5 presents the architecture of the proposed B-DNN model. It shows that the model consists of 4 layers; the first is the input layer with 7 neuron (same number as selected features), the second in the first hidden layer with 7 neuron which in turn passes the values to the second hidden layer that squeezes the representational space of the network to have 5 neurons, that is then fed to the output layer which have one neuron that presents the prediction result (Apnea or Non-apnea).

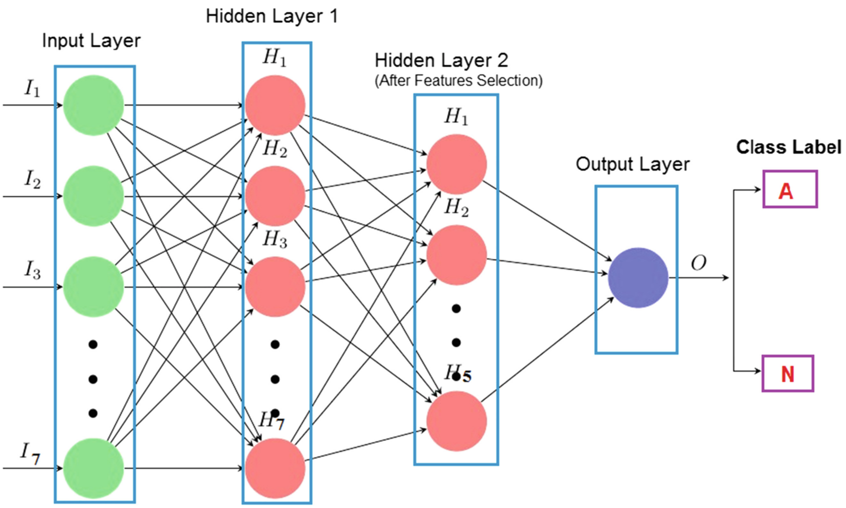


Fig. 5. DNN classification model architecture

As shown by Bengio and Glorot in [16]; units with more incoming connections should have relatively smaller weights; so that weights of the proposed DNN model are initialized using a small Gaussian random number. Rectified Linear Unit (RELU) activation is used as a transfer function for the weights to the net input value, which is then fed to the output layer that uses the sigmoid activation function to produce a probability output in the range of 0 to 1 which will then be converted to crisp class values. The used loss function during DNN training is Cross entropy, also referred to as logarithmic loss, is one of the most favourite loss functions that improve the performance of the DNN training. Once all the derivatives are computed, parameters are updated using the efficient Adaptive Moment Estimation (Adam) optimization algorithm [38] for gradient descent. Figure 6 shows the structure of the B-DNN model. This B-DNN model is used for achieving the first phase of the proposed scheme, which is minute-based classification.

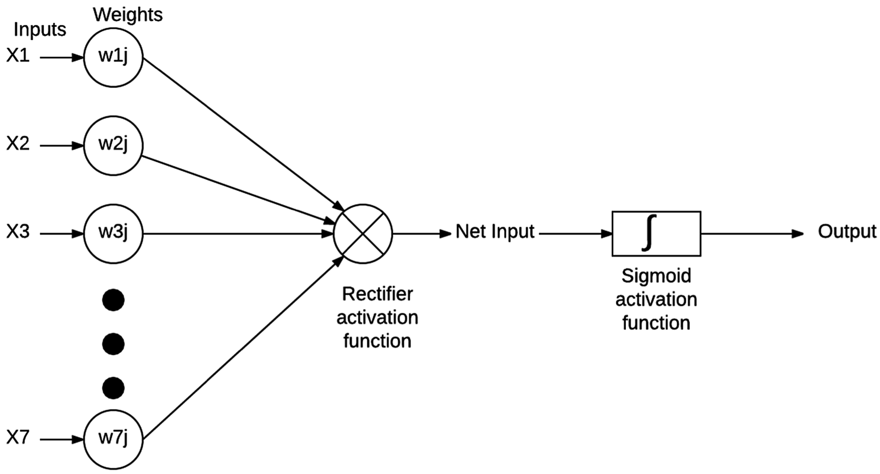


Fig. 6. DNN structure of B-DNN model

Phase 2: Hybrid Deep Neural Network and Decision Tree Model (DNN-DT)
This model is a hybrid algorithm that combines the Deep Neural Network (DNN) classifier with the Decision Tree classifier. The output of the first phase that is performed using B-DNN model (classified minutes as apnea on non-apnea) is fed into a decision tree model in order to perform class identification (Class A, B or C). Totally, the result is used for the fully minute-class-based classification phase. Figure 7 shows the architecture of the proposed DNN-DT.

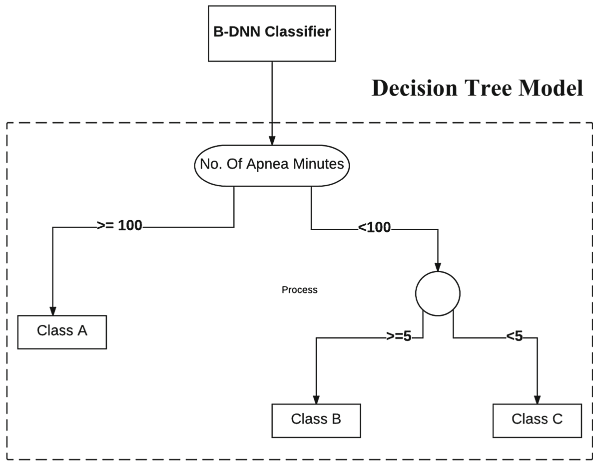


Fig. 7. DNN-DT classification model architecture

Traditional Classification Approaches

In addition to the proposed classification model, the extracted and selected features are applied to traditional classifiers being used previously in the literature for the same dataset and compared the results. The explored classifiers are: Logistic Regression, KNN, SVM and Naïve Bayes Classifier. Table 1 summarizes parameters for these classification models. Regarding the KNN model; Euclidean distance function is used to find the nearest neighbours. The model assigns uniform weights to each neighbour. In regards to the SVM model; the classification types is C-SVM, where C is chosen to be 1.0. The model uses the RBF Kernel with numerical tolerance equals to 0.001 and iterations can reach maximum of 100 iteration.

Table 1. Model parameters for the applied classifiers

Logistic Regression	KNN	SVM
Regularization: Ridge (L2), C = 1	Number of neighbours: 20 Metric: Euclidean Weight: Uniform	SVM type: C-SVM, C = 1.0 Kernel: RBF, $\exp(-1.0\|x - y\|^2)$ Numerical tolerance: 0.001 Iteration limit: 100

Orange Data Mining toolset [39] was used to simulate the traditional classifiers and compare results.

4 Experimental Results

4.1 Minute-Based Classification

Since only the training set (35 ECG recording) of PhysioNet Apnea-ECG database contain minute-wise apnea annotations; given the necessity of annotated test data to evaluate the classifier's performance, we were forced to use only these 35 recordings in the experiment. As aforementioned, we evaluated our approach using 10-fold cross validation technique.

The performance of the proposed classifier is compared to those of the state-of-the-art classifiers employing Apnea-ECG data-set at the same statistical features. This comparison is presented in Table 2. It is clear that the proposed B-DNN emerges as the classifier with the highest performance.

Table 2. Summary of various classifiers performance for minute-based classification.

Classifier	**# Features = 11**					**# Features = 7**				
	CA	Prec.	Rec.	Sens.	Spec.	CA	Prec.	Rec.	Sens.	Spec.
Logistic Regression	60.3	67.3	94.2	94.2	26.4	67.9	66.9	94.9	94.9	23.9
KNN	76.8	80.0	83.2	83.2	66.4	80.5	84.8	83.3	83.3	76.0
SVM	52.3	64.6	50.2	97.1	5.6	62.0	62.3	97.1	97.1	5.6
Naive Bays	63.2	66.0	57.8	57.8	71.8	67.2	76.0	61.5	61.5	68.7
Proposed B-DNN	79.0	80.0	79.0	79.7	77.7	92.7	95.3	92.8	92.8	92.6

4.2 Minute-Class-Based Classification

The second phase of the proposed approach is both detection of apnea class and quantification of apnea minutes. The number of the classified minutes for each recording is used to determine whether a patient recording belongs to class A, B or C unlike start-of-art methods that were able to classify only two classes instead of three. As mentioned before, the PhysioNet database for the training set contains 20 recording of class A, 5 of class B and 10 of class B.

In more details, the minutes of each files in classified to apnea or non-apnea minute using D-BNN classifier. At the same time, each recording file is classified to its corresponding class (using decision tree classifier) based on the number of classified apnea minutes.

The performance of the DNN-DT scheme for class-based classification is presented in Fig. 8. Hybrid scheme performed well in all performance metrics. Figure 9 views the confusion matrix of the scheme. It is also clear that the most misclassified classes are from class B which is a misleading class as induced from other schemes.

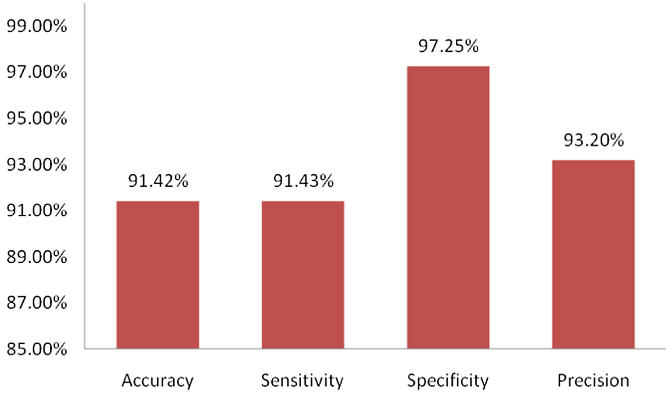


Fig. 8. Performance of the class-based classification using DNN-DT scheme

	A	B	C	Σ
A	19	1	0	20
B	0	3	2	5
C	0	0	10	10
Σ				35

Fig. 9. Confusion matrix for the class-based classification using DNN-DT scheme

5 Conclusion

In this work, a hybrid approach is proposed that includes deep neural networks and decision trees, for detection and quantification of sleep apnea using features of ECG signals. Statistical features were extracted from the RR interval and serves as training and testing data for the applied classifiers. Deep architectures have benefited much more from the pre-training stage in terms of training efficiency and test performance. Deep learning can adapt all the weights parameters in the network with respect to the tasks they solve. The success of this approach inspired further research to go in depth and perform an investigation into other scenarios of deep learning.

The proposed approach treated, with novelty, the following points: (i) identifies both apnea classes and detects minute-by-minute classification unlike state-of-the-art methods which either identify apnea class or detect its presence, (ii) identifies the three apnea classes (A, B and C) while other papers only identifies two classes (A and C), (iii) and makes a comparative study of the most used classification methods adopted in the literature but using the same features and the same dataset. The experimental results showed that this approach is robust and computationally efficient and clearly outperforms state-of-the-art methods.

References

1. Derrer, D.: WebMD Medical Reference, September 2014. http://www.webmd.com/
2. Caples, S.M., Garcia-Touchard, A., Somers, V.K.: Sleep-disordered breathing and cardiovascular risk. Sleep **30**(3), 291–303 (2007)
3. Morgenthaler, T., Kagramanov, V., Hanak, V., Decker, P.: Complex sleep apnea syndrome: is it a unique clinical syndrome? Sleep **29**(9), 1203–1209 (2006)
4. Chazal, P., Penzel, T., Heneghan, C.: Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. Physiol. Meas. **25**(4), 967–983 (2004)
5. PhysioNet: Detecting and quantifying apnea based on the ECG, January 2012. https://www.physionet.org/challenge/2000/
6. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, vol. abs/1409.1556 (2014)
8. Hinton, G.E., et al.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Sig. Process. Mag. **29**(6), 82–97 (2012)
9. Sainath, T.N., et al.: Deep convolutional neural networks for large-scale speech tasks. Neural Netw. **64**, 39–48 (2015)
10. Socher, R., Bengio, Y., Manning, C.D.: Deep learning for NLP (without magic). In: The 50th Annual Meeting of the Association for Computational Linguistics, Jeju Island, Korea (2012)
11. Brebisson, A.D., Montana, G.: Deep neural networks for anatomical brain segmentation. Computing Research Repository (CoRR), vol. abs/1502.02445 (2015)
12. Shin, H., Orton, M., Collins, D.J., Doran, S.J., Leach, M.O.: Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1930–1943 (2013)
13. LeCun, Y., Bengio, Y., Hinton, G.E.: Deep learning. Nature **521**, 436–444 (2015)
14. LeCun, Y.: Learning invariant feature hierarchies. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012. LNCS, vol. 7583, pp. 496–505. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33863-2_51

15. Soniya, S.P., Singh, L.: A review on advances in deep learning. In: IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI), Kanpur, India, pp. 1–6 (2015)
16. Bengio, Y., Glorot, X.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of AISTATS 2010, pp. 249–256 (2010)
17. Bianchini, M., Scarselli, F.: On the complexity of neural network classifiers: a comparison between shallow and deep architectures. IEEE Trans. Neural Netw. Learn. Syst. **25**(8), 1553–1565 (2014)
18. Seide, F., Li, G., Chen, X., Yu, D.: Feature engineering in Context-Dependent Deep Neural Networks for conversational speech transcription. In: Automatic Speech Recognition and Understanding (ASRU), Waikoloa, HI, USA (2011)
19. Kaguara, A., Myoung Nam, K., Reddy, S.: A deep neural network classifier for diagnosing sleep apnea from ECG data on smartphones and small embedded systems. Swarthmore College, Pennsylvania, United States, Thesis (2014). https://doi.org/10.13140/2.1.4174.5448
20. Jarvis, M., Mitra, P.: Apnea patients characterized by 0.02 Hz peak in the multitaper spectrogram of electrocardiogram signals. In: Computers in Cardiology, vol. 27, pp. 769–772 (2000)
21. McNames, J., Fraser, A.: Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram. In: Computers in Cardiology 2000, Cambridge, MA, USA, vol. 27, pp. 749–752 (2000)
22. Mietus, J., Peng, C., Ivanov, P., Goldberger, A.: Detection of obstructive sleep apnea from cardiac interbeat interval time series. In: Computers in Cardiology 2000, Cambridge, MA, USA, vol. 27, pp. 753–756 (2000)
23. Schrader, M., Zywietz, C., Einem, V., Widiger, B., Joseph, G.: Detection of sleep apnea in single channel ECGs from the PhysioNet data base. In: Computers in Cardiology, vol. 27, pp. 263–266 (2000)
24. De Chazal, P., et al.: Automatic classification of sleep apnea epochs using the electrocardiogram. In: Computers in Cardiology 2000, Cambridge, MA, USA, vol. 27, pp. 745–748 (2000)
25. Raymond, B., Cayton, R., Bates, R., Chappell, M.: Screening for obstructive sleep apnoea based on the electrocardiogram – the Computers in Cardiology Challenge. In: Computers in Cardiology 2000, Cambridge, MA, USA, vol. 27, pp. 267–270 (2000)
26. Khandoker, A., Karmakar, C., Palaniswami, M.: Automated recognition of patients with obstructive sleep apnoea using wavelet based features of electrocardiogram recordings. Comput. Biol. Med. **39**(3), 88–96 (2009)
27. Xie, B., Minn, H.: Real-time sleep apnea detection by classifier combination. Inf. Technol. Biomed. **16**(3), 469–477 (2012)
28. Manrique, Q., Hernandez, A., Gonzalez, T., Pallester, F., Dominquez, C.: Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features. In: IEEE International Conference on Engineering in Medicine and Biology Society (EMBS 2009), Minneapolis, MN, USA, pp. 5559–5562 (2009)
29. Mendez, M., et al.: Detection of sleep spnea from surface ECG based on features extracted by an autoregressive model. In: IEEE International Conference on Engineering in Medicine and Biology Society (EMBS 2007), Lyon, France, pp. 6105–6108 (2007)
30. Almazaydeh, L., Elleithy, K.H., Faezipour, M.: Obstructive sleep apnea detection using SVM-based classification of ECG signal features. In: IEEE International Conference on Engineering in Medicine and Biology Society (EMBS 2012), San Diego, CA, USA, pp. 4938–4941 (2012)
31. Babaeizadeh, S., White, D., Pittman, S., Zhou, S.: Automatic detection and quantification of sleep apnea using heart rate variability. J. Electrocardiol. **43**, 535–541 (2010)

32. Rachim, V., Li, G., Chung, W.: Sleep apnea classification using ECG-signal wavelet-PCA features. Bio-Med. Mater. Eng. **24**(6), 2875–2882 (2014)
33. Goldberger, A., et al.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation **101**(23), E215–E220 (2000)
34. MedicineNet: Definition of QRS complex, September 2016. http://www.medicinenet.com/script/main/art.asp?articlekey=5160
35. Thuraisingham, R.: Preprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals. Comput. Methods Programs Biomed. **83**(1), 78–82 (2006)
36. Statistics Solutions: ANOVA (2013). http://www.statisticssolutions.com/manova-analysis-anova/
37. Keras Documentation. https://keras.io/
38. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
39. Orange Data Mining, August 2016. http://orange.biolab.si/

Morphological Analysis of Fenestrae in Arteries

Muhammad Moazzam Jawaid[3](✉), Francisco Ramirez-Perez[2], Antoine Plumerault[1], Flora Quilichini[1], Jose Alonso Solis-Lemus[1], Luis Martinez-Lemus[2], and Constantino Carlos Reyes-Aldasoro[1]

[1] School of Mathematics, Computer Science and Engineering, City, University of London, London, UK
[2] Dalton Cardiovascular Research Center, University of Missouri, Columbia, MO, USA
[3] Mehran University of Engineering and Technology, Jamshoro, Pakistan
moazzam.jawaid@faculty.muet.edu.pk

Abstract. Cardiovascular disease (CVD) is worldwide cause of death. The morphological structure of one of the regions of arteries called the internal elastic lamina (IEL) is associated with the stiffness of arteries, especially the presence and characteristics of small holes called *fenestrae*. Structural analysis of the IEL as observed with multiphoton or confocal fluorescent microscopy is complicated, primarily due to is three-dimensional distribution along a series of z-stack slices. In addition, whilst the top slices of an artery cross long lamina sections, the bottom slices only cross short sections of the lamina and would be better observed in a different plane than that of the acquired image.

In this work, we describe a framework to analyse 3D stacks of arteries as observed with fluorescent multiphoton microscopy. We apply this framework to the study of arterial stiffness as inferred from the presence/absence of fenestra the IEL and their morphological characteristics. The framework assumes the arteries as cylinders, then performs a series of geometrical transformations to align the data in three dimensions. The IEL is segmented from the external elastic lamina (EEL) and then each lamina is projected to a plane from which the fenestrae of all slices can be analysed in a single 2D image. The fenestrae are segmented by filtering with a Laplacian of a Gaussian kernel with optimal parameters. The results were compared against manually segmented fenestrae and provided a classification accuracy of [98–99%] and Jaccard Index values in the range of [38–54%], which are comparable the variations of 3 independent manual segmentations. The code is available from the website https://github.com/reyesaldasoro/Arterial-Fenestrae.

1 Introduction

Cardiovascular disease (CVD) is worldwide cause of death and sometimes considered *world's top killer* [1]. Consumption of excess fat and carbohydrate, in some cases called *Western diet*, (WD) is associated with alterations in the structural

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 273–282, 2018.
https://doi.org/10.1007/978-3-319-95450-9_23

characteristics of blood vessels and could lead to CVD. This WD is prevalent in some countries like the USA and has led to alarming increases in obesity and type II diabetes, risk factors to CVD [2]. Development of CVD and hypertension is even associated with exposure to adverse maternal environments e.g. Gestational diabetes mellitus (GDM) [3,4]. Vascular remodelling may contribute to the pathophysiology of vascular diseases and circulatory disorders [5,6]. One of the regions of particular interest in the structure of the terminal vascular bed is the internal elastic lamina (IEL) [7]. Remodelling is often associated with morphological changes in the IEL such as increased vascular elastin [8] or flow-induced arterial enlargement [9]. The IEL presents a characteristic formation of windows commonly called *fenestra* and its characteristics are important as in cases enlarged fenestra have been assumed to represent a weakness of the IEL and may contribute to microaneurysms [10].

Structural analysis of the IEL as observed with multiphoton or confocal fluorescent microscopy is complicated, primarily due to is three-dimensional distribution along a series of z-stack slices. It is common practice to analyse the IEL visually [5], and only a few regions of interest (ROI) in selected slices, and thus a large amount of information from the arteries, especially the region of the artery where the IEL is nearly perpendicular to the field of view, is not considered for analysis.

In this work, a framework for the analysis of IEL of arteries as observed with a multiphoton microscope is proposed. The main step of the framework is the projection of the artery, from a cylindrical structure to a 2D plane. In this way, all the regions of the artery can be observed in a single plane. The presence of the fenestrae in the two-dimensional IEL is analysed through intensity-based algorithms and validated against a manually-segmented ground truth. Finally, we apply the framework to compare two populations of different rats which provided statistical differences.

An additional advantage of the projection approach is that the lamina that resides outside the IEL, called the External Elastic Lamina (EEL) can also be projected and further analysed. Analysis of the EEL is beyond the scope of this work.

2 Materials and Methods

2.1 Data Sets

Details of acquisition and animals have been described previously [4] but briefly, mesenteric resistance arteries (MRA, 186–301 µm internal diameter) were harvested from spontaneously hypertensive rats (SHR, n = 4) and MRA segments from Wistar Kyoto rats (WKY, n = 4) following NIH and local Guidelines and Ethical Approvals. For illustration purposes, one femoral artery was also acquired (Set D in Fig. 2). Vessels were imaged using a Leica SP5 confocal/multiphoton microscope with a 63x/1.2 numerical aperture water objective. Alexa Fluor 633, to image elastin, was excited with a 633 nm HeNe laser. Alexa Fluor 546 phalloidin, to image F-actin components, was excited with a 543 nm HeNe laser.

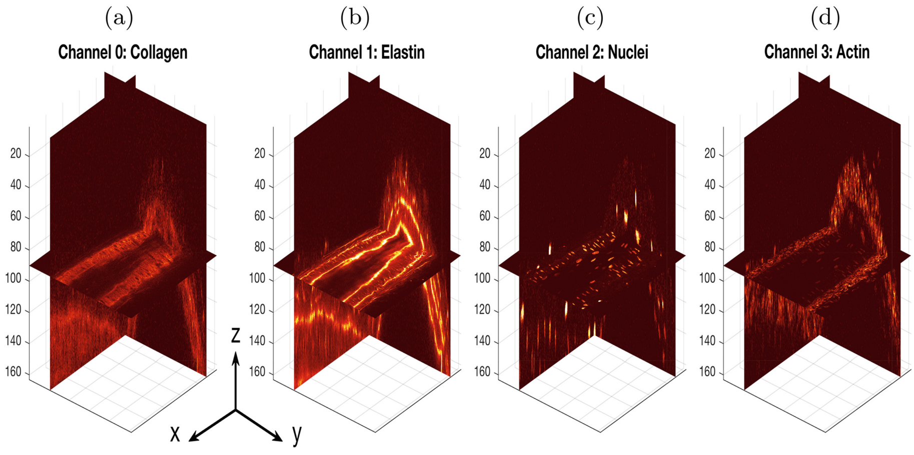


Fig. 1. Illustration of one representative artery and the four channels acquired under the following conditions: (a) subject to second harmonics to acquire collagen, (b) stained with Alexa633 to image elastin, (c) stained with DAPI to for cell nuclei, (d) stained with Phalloidin to for actin. Data is shown in 3 orthogonal planes, the horizontal xy is the acquired in the microscope and vertical planes (traversal yz to the artery on right and parallel xz on the left) are reconstructed. It can be noticed in (b) the presence of two laminae expressing Elastin and how these laminae locate as concentric regions. Another relevant observation is that the lines corresponding to the lamina in xy are getting closer towards the right side and closer to the yz plane, this indicates that the artery is not aligned with the plane. If the plane and artery were perfectly aligned, the lines should be parallel.

DAPI, to image nuclei, was excited with a multi-photon laser at 720 nm. Collagen was imaged via second-harmonic image generation using a multi-photon laser at 850 nm as previously described [5] (Fig. 1).

2.2 Methods

Alignment of the Arteries. Whilst care is taken at the position of the samples in the microscope, variations in alignment occur, especially in the *elevation* which is manifest in Fig. 2 in sets B and D. The data sets were thus aligned by rotating the data along the z axis (rotation of original images) and then a rotation along the y axis (axis perpendicular to the direction of the vessel). Once the vessel was oriented as close as possible to a cylinder aligned with the x axis, the maximum intensity projection over this axis was calculated. This projection (Fig. 3a) has a similar appearance to one yz plane as that of Fig. 2, but it compensates for any small variation along the artery.

Segmentation of the Laminae. To determine the edges of the internal and external laminae of the artery, a multiscale calculation of the Canny edges [11] was performed by increasing the variance of the Gaussian until an asymptotic

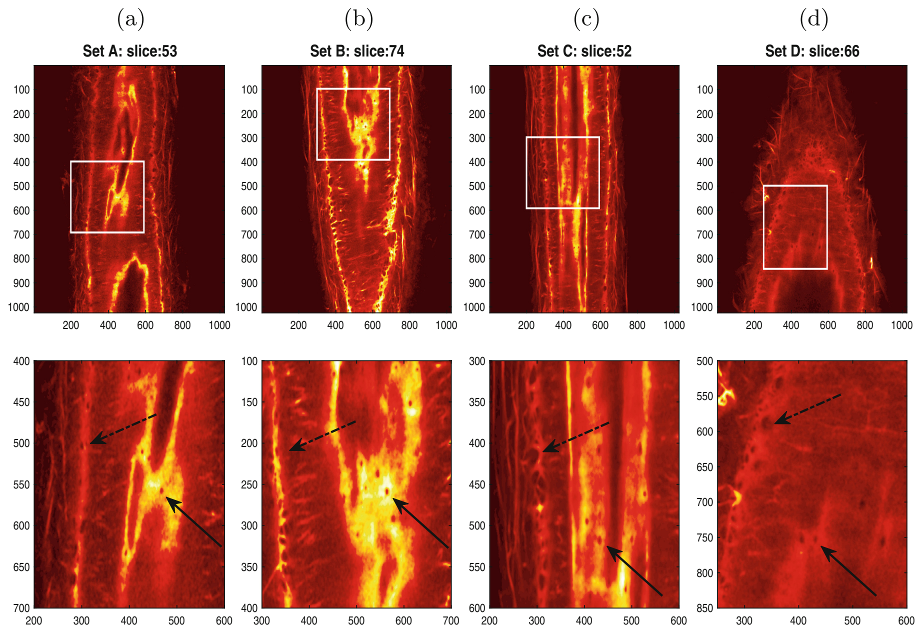


Fig. 2. Representative images of the IEL stained with Alexa633 that show the presence and variability of fenestrae in four different data sets. Regions of interest are denoted with a white rectangle and shown in the second row. Some fenestrae of the IEL are illustrated with solid arrows whilst structures of the EEL are indicated with dashed arrows. It is important to notice several points: Fenestrae appear as dark circular regions as compared with their surroundings, fenestrae-like structures can also be present in the EEL, but in general the fenestrae of the IEL are smaller in size, arteries are not always well aligned as shown in B and D where the arteries are tilted in opposite directions.

number of edges, ideally one for each side of the artery, was reached (Fig. 3a). When the top of the artery was too close to the edge of the field of view, the asymptotic number was 3 as the edge over the top was broken into two. The external edge presented considerable variability due to the nature of the EEL, therefore the segmentation of IEL was obtained from the dilation of the internal edge. To select the EEL, the dilated region was used to remove the intensity of the IEL. The regions corresponding to each of the laminae were further removed when the projections for each were calculated.

Projection to a Plane. The previous step identified the regions for the EEL and the IEL, but these were still distributed in a cylindrical shape over several dimensions, thus the next step was to project the intensities towards a single 2D plane. This was performed by a ray tracing from the central line of the assumed cylinder. Figures 3b, c illustrate both laminae in position with respect to the yz plane for one set, and Fig. 4 shows the maximum intensity projection of the IEL and EEL for the four data sets.

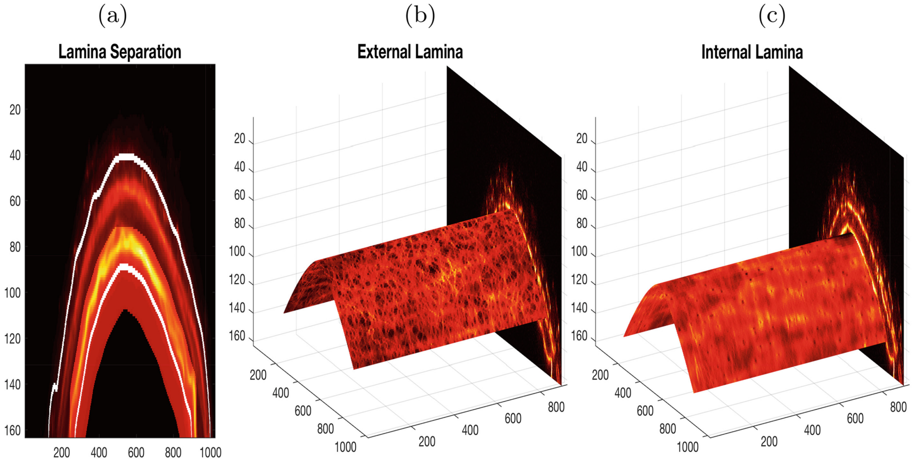


Fig. 3. Segmentation of the laminae of data set C. The maximum intensity projection along axis x towards the traversal plane are illustrated with white lines denoting the boundaries of the artery and the ROI of the internal lamina is highlighted with a brighter shade. On the right the EEL and IEL are illustrated on their position over the artery.

Segmentation of the Fenestrae. The fenestrae could be described as small, dark round regions surrounded by brighter pixels. Taking this characteristic into account, six segmentation algorithms were tested: (1) thresholding after a median filter was applied, (2) thresholding after median filter and morphological closing, (3) thresholding after a convolution with a Gaussian Kernel, (4) thresholding after finding peaks in the image, (5) detection of the fenestrae from the peaks of a Hough transform, (6) thresholding after convolution with a Laplacian of a Gaussian (L.O.G.).

Ground Truth. To validate the segmentation methods previously mentioned, three independent observers performed a manual delineation of the fenestra. Results are shown in the top row of Fig. 5 as colours (red, green, blue) assigned to each observer. The results illustrate the common intra-observer variability; regions selected by all 3 appear in white, with other colours as pairs or single selections.

Evaluation Metrics. Pixels were classified as true positives (TP Fenestrae segmented as Fenestrae), true negatives (TN Background segmented as Background), false positives (FP Background segmented as Fenestrae) and false negatives (FN Fenestrae segmented as Background). Then the accuracy was calculated as $(\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})$ and the Jaccard Similarity index [12] was calculated as $(\mathrm{TP})/(\mathrm{TP}+\mathrm{FP}+\mathrm{FN})$.

Additionally, the number of fenestrae detected automatically and the numbers of the ground truth were compared.

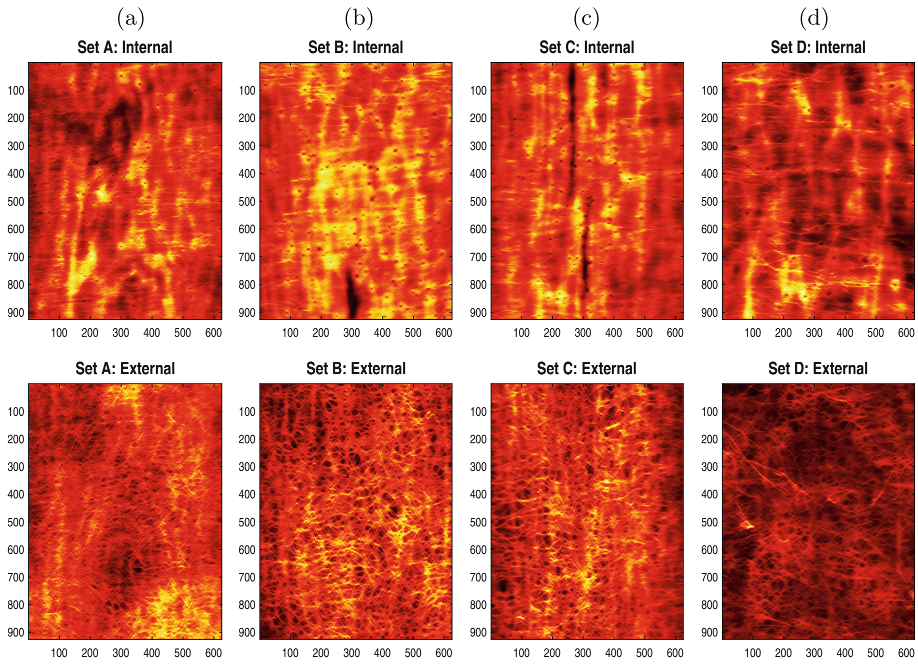


Fig. 4. Projected Internal and External laminae for the four representative cases of Fig. 2. Notice how the projections contain a considerable larger number of fenestrae to be analysed as compared with single slices. This is particularly important on the lower slices of the set (corresponding to the left and right edges of the images) as those fenestrae would be difficult to observe in horizontal planes. The EEL presents interesting structures that will be analysed in the future as it is beyond the scope of this paper.

3 Results and Discussion

For purposes of comparison, valid fenestrae were considered those regions selected by 2 or more observers, i.e. the three manual segmentations were added and the ground truth were all pixels with values greater than 1. These correspond to the yellow, magenta, cyan and white regions of Fig. 5. It should be noticed that this criterion is far from perfect as there were several regions that were detected by one single observer, particularly red regions in sets A and C, and green regions in set D. This procedure assumes that all the manual selections are equally likely and if only one observer is more careful or has a better understanding of the fenestrae, his or her observations could be removed by two less experienced ones. Any future work should have a more careful manual selection of fenestrae.

Six segmentation algorithms were tested with the four data sets displayed previously. Since there were large areas of background, the majority of pixels were allocated as TN and thus the accuracy measurements were rather high and

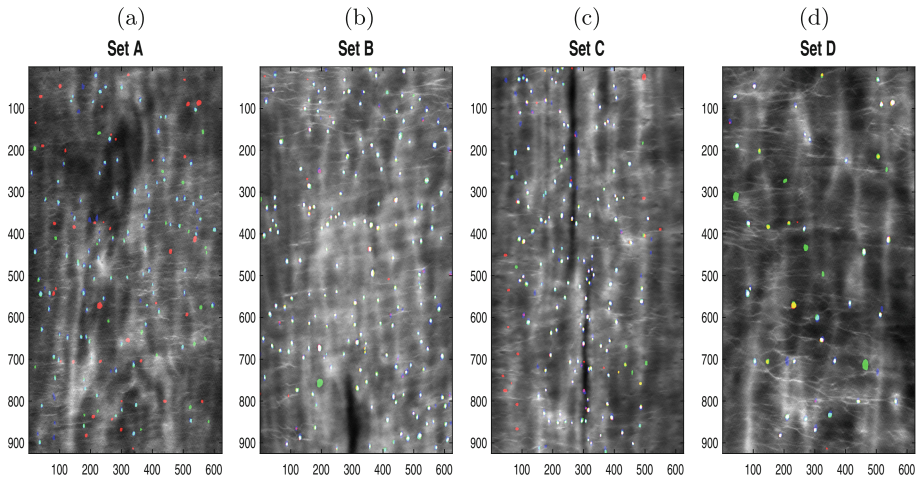


Fig. 5. Comparison of manual segmentation of 3 independent observers, red regions correspond to observer 1, green to 2, blue to 3, common selections correspond to 1 + 2 yellow, 1+3, magenta, 2+3 cyan, 1+2+3 white. It should be noticed the large variation of the observers. Regions where 2 or more observers coincided were considered as ground truth for classification. (Color figure online)

it was considered that the Jaccard would be a better indication of the algorithms. The results obtained were the following: (1) thresholding after a median filter was applied **37.2%**, (2) thresholding after median filter and morphological closing **19.4%**, (3) thresholding after a convolution with a Gaussian Kernel **34.3%**, (4) thresholding after finding peaks in the image **34.1%**, (5) detection of the fenestrae from the peaks of a Hough transform **34.0%**, (6) thresholding after convolution with a Laplacian of a Gaussian (L.O.G.) **45.58%**.

Overall, the classification accuracy obtained (pixelwise foreground/ background) was in the range [98–99%] and Jaccard Index values (pixelwise intersection over union) in the range of [38–54%] for data sets A–D (Fig. 6), which are comparable to the variation of the manual classifications.

The algorithm that provided best results was performing a convolution with a Laplacian of a Gaussian kernel and then applying a threshold. The parameters of this last segmentation algorithm were subject to a sensitivity analysis by systematically varying the parameters of Laplacian of a Gaussian and the threshold. Optimal values corresponded to size = 13, variance = 4, threshold of 0.25%. It was noted that the algorithm was sensitive to the size of the kernel only for small kernels and for larger ones it remains fairly stable. The variability with the threshold displayed a shape similar to a Poisson distribution (Fig. 7a). A comparison of the number of fenestrae detected showed that the optimal results were obtained with the same range of parameters, with results of around 15% variation in the number of fenestra detected (Fig. 7b).

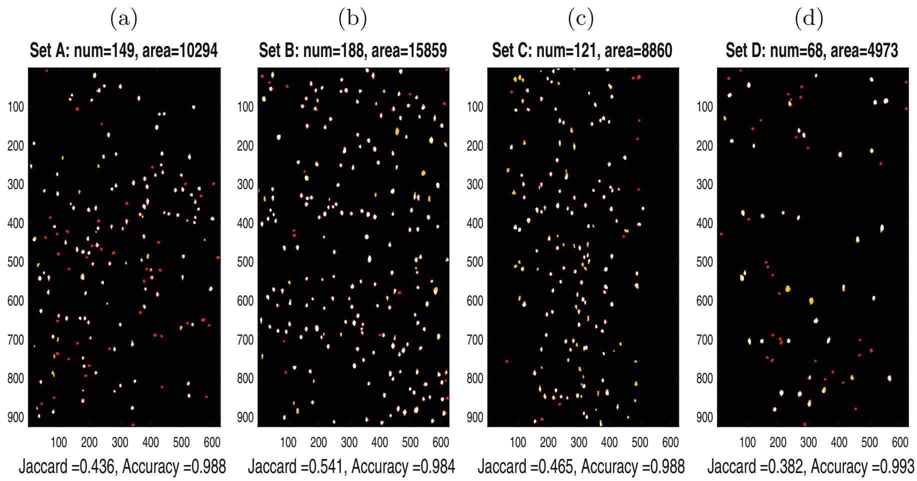


Fig. 6. Automatic segmentation: black = background, white = True positives, red = false positives, yellow = false negatives. Notice the difference between mesenteric arteries (A, B, C) and a femoral artery (D). (Color figure online)

The algorithms were then applied to analyse the arteries of the two groups: spontaneously hypertensive rat (SHR) and normotensive control (WKY) rats. WKY rats reported a higher number of fenestrae and fenestrae of larger areas (Fig. 8). These results are consistent in the number of fenestrae with previously published data [13] and our own previous experimental analyses made by visualisation and manual counting. The sizes correspond to range of sizes reported by [10] within the normal range (mean diameter $2.13 \pm 0.13\,\mu$m), and much smaller that the enlarged fenestrae (mean diameter $7.0 \pm 0.34\,\mu$m).

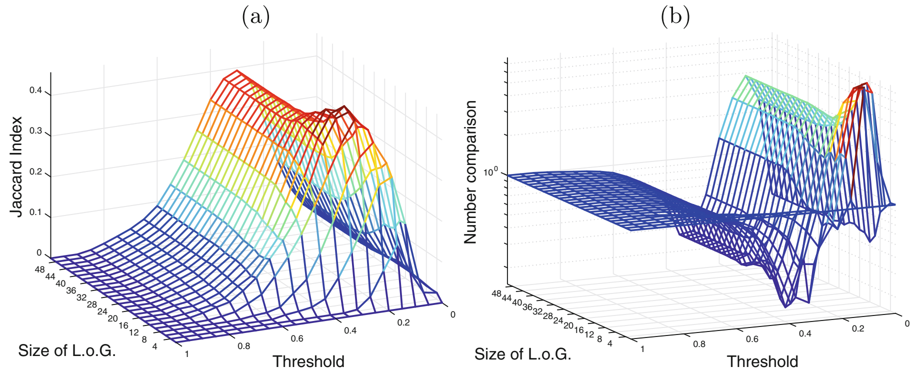


Fig. 7. Number and size of fenestrae showed statistical difference between WKY and SHR rats.

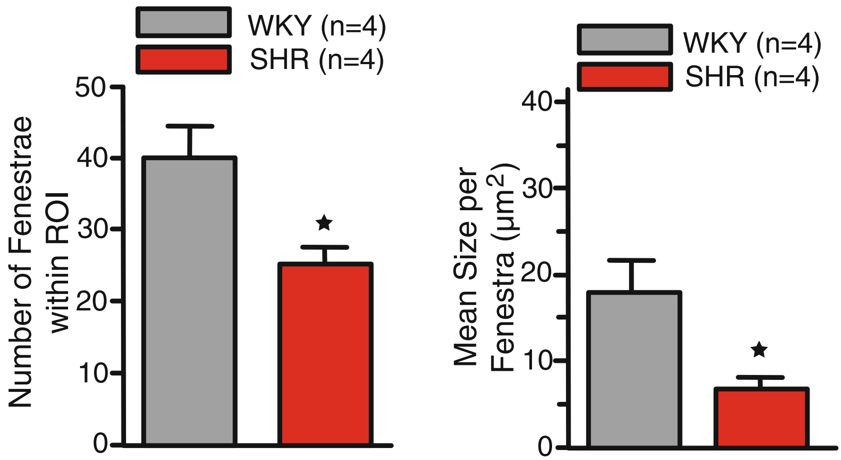


Fig. 8. Number and size of fenestrae showed statistical difference between WKY and SHR rats.

However, the calculations on the mean size of fenestrae were discordant between the two methods. Further analyses need to be performed to determine the source of variation in the measurement of fenestrae size.

Further analysis will focus also on the mean density of fenestrae, mean percentage area and shapes of the fenestrae. Another possible extension would be to compare fenestrae characteristics with age, as it has been reported in rabbits that fenestra increases dramatically with postnatal development [14].

References

1. Gerszten, R.E., Wang, T.J.: The search for new cardiovascular biomarkers. Nature **451**, 949–952 (2008)
2. Hedley, A.A., Ogden, C.L., Johnson, C.L., Carroll, M.D., Curtin, L.R., Flegal, K.M.: Prevalence of overweight and obesity among us children, adolescents, and adults, 1999–2002. JAMA **291**, 2847–2850 (2004)
3. Ramirez-Perez, F.I., et al.: Maternal hyperleptinemia increases arterial stiffening and alters vasodilatoy responses to insulin in adult male mice offspring. FASEB J. **30**(Suppl. 1), 721.8 (2016)
4. Pennington, K.A., et al.: Maternal hyperleptinemia is associated with male offsprings altered vascular function and structure in mice. PLOS One **11**(5), e0155377 (2016)
5. Bender, S.B., et al.: Regional variation in arterial stiffening and dysfunction in Western diet-induced obesity. Am. J. Physiol. Heart Circ. Physiol. **309**(4), H574–582 (2015)
6. Dzau, V.J., Gibbons, G.H.: Vascular remodeling: mechanisms and implications. J. Cardiovasc. Pharmacol. **21**(Suppl 1), S1–5 (1993)
7. Movat, H.Z., Fernando, N.V.P.: The fine structure of the terminal vascular bed I. Small arteries with an internal elastic lamina. Exp. Mol. Pathol. **2**(6), 549563 (1963)

8. Foote, C.A., et al.: Arterial stiffening in western diet-fed mice is associated with increased vascular elastin, transforming growth factor-β and plasma neuraminidase. Front. Physiol. **7**, 285 (2016)
9. Masuda, H., Zhuang, Y.-J., Singh, T.M., Kawamura, K., Murakami, M., Zarins, C.K., Glagov, S.: Adaptive remodeling of internal elastic lamina and endothelial lining during flow-induced arterial enlargement. Arterioscler. Thromb. Vasc. Biol. **19**(10), 22982307 (1999)
10. Campbell, G.J., Roach, M.R.: Fenestrations in the internal elastic lamina at bifurcations of human cerebral arteries. Stroke **12**(4), 489496 (1981)
11. Canny, J.: A computational approach to edge detection. IEEE Trans. PAMI **8**(6), 679–698 (1986)
12. Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull. Soc. Vaud. Sci. Nat. **37**, 547–579 (1901)
13. Briones, A.M., et al.: Role of elastin in spontaneously hypertensive rat small mesenteric artery remodelling. J. Physiol. (Lond.) **552**(Pt 1), 185–195 (2003)
14. Wong, L.C.Y., Langille, B.L.: Developmental remodeling of the internal elastic lamina of rabbit arteries: effect of blood flow. Circ. Res. **78**(5), 799805 (1996)

mLearning and eLearning

Game Based Social Skills Apps to Enhance Collaboration Among Young Children: A Case Study

Najmeh Behnamnia(✉), Amirrudin Kamsin, Maizatul Akmar Binti Ismail, and A. Hayati

Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
n.behnamnia@siswa.um.edu.my, {amir,maizatul}@um.edu.my, a.hayati@gmail.com

Abstract. One of the latest trends in mobile phones is the wave of smart phone apps, which include game based social skills and collaboration apps for young children. Despite the massive development and invention on the game based apps, it is also noted that there are still limited studies that discuss on the right guideline in developing the game based social skill apps or to consolidate them into real life learning process. This study discusses about the efficacy of game based social skills apps about Japanese culture and etiquette of collaboration. The aims of this paper included: (1) To investigate and identify features for fostering collaboration through game based social skills apps. (2) To design and develop a model that integrates content, behavioural and applying components of collaboration, roles and relationship with learning Japanese culture and etiquette in game based social skills apps for pre-schooler's level (3–6 years old). To avoid selective attention bias and assess medium term retention, evaluation methodology uses the extended version of instruments of prior studies by assessing the effects of the game.

Keywords: Educational games · Game-based learning · Pre-schoolers
Case study · Culture intervention · Peer collaboration
Game based social skills apps

1 Introduction

Recently, the awareness on the good impacts in the usage of games apps for education is increasing. One of the latest trends in mobile phones is the wave of smart phone apps, which include game based social skills and collaboration apps for young children. In spite of that, the research conducted on methodologies and proper guidelines in developing game based social skills apps and how to consolidate the game apps in the real life scenario learning processes is still limited. In order to integrate them, the process depends highly on acquiring the evidence of game based social skills apps' effectiveness [1].

Children who experience difficulties with social communication and language skills are known to have problematic peer relationships [2]. They are less accepted than

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 285–291, 2018.
https://doi.org/10.1007/978-3-319-95450-9_24

typically-developing children [3] and can be at greater risk of being bullied [4]. Emerging evidence also suggests that they are less likely to be able to benefit from the collaborative activities with peers that are a common feature of educational settings [5]. Cultural content varies in different definition and perspectives. There is physical cultural heritage that influence a community's evolution and customs such as old buildings, historical monuments, archive documentations, artwork, machineries or even historical sites that provide evidence of previous human settlements. These are called tangible cultural heritage. Apart from that, there is also intangible heritage such as social belief, customary practises, philosophical and religious values, literatures, folk tales, behavioural rules and impacts from historical events on that particular society. It is evidenced that it is hard to preserve intangible heritage as the society evolved over the years. However, the game apps scenario is believed to be maintained and communicated effectively. As a matter of fact, the game apps are able to be recreated in physical settings and also stipulate a complete exposure in spoken language in a certain society, appreciation of traditional music and aesthetics matters.

Prior works on educational media's role in early childhood learning mainly paid attention on various types of topics such as early literacy prosocial skill acquisition and adoption of healthy behaviours. However, only few researches have investigated the role of learning culture in early childhood in order to promote the collaboration in the future [6].

Results of recent studies on collaboration and game based social skills apps show that the application of the components of collaboration, roles and relationship with culture in game based social skills apps for pre-schooler's level is still low [31, 32]. Conceptual design model needs to be developed in designing the features based on the intervention strategies to match between applying culture in game based social skills apps among young children and to improve their collaboration [31–33].

2 Research Questions

The research problem described in last Section has led to a formulation of four research questions for this paper. Each of these questions can be divided into a set of objectives, as follows:

1. What are the problems of existing models and frameworks of game based social skills apps for fostering collaboration for pre-schooler to better learning Japanese culture, etiquette and heritage?
2. How to develop a model of game based social skills apps that can enhance collaboration among pre-schooler to better learning Japanese culture, etiquette and heritage?
3. How can a prototype of Japanese culture game based social skills app for pre-schooler level based on the proposed conceptual model of GBSS apps enhancing collaboration and achieve better Japanese cultural awareness and heritage among young children?

4. What is the effect on collaboration and learning Japanese culture, etiquette and heritage when the proposed model of game based social skills apps is evaluated among pre-schoolers?

3 A Comparison Between Recent Research Projects

Recently, the usage of game based learning apps through touch screen devices is increasing for young children. Most of pre-schoolers have access to a device, such as tablet and smartphone (touch screen) at home or school. These devices (tablet and smartphone) have become essential part of many young children daily routine [7, 8]. According to the survey of 1028 children (3–5) years that was carried out by the National Literacy Trust, it demonstrated that more than 70% have access to a device such as tablet or smartphone (touch screen) at home and school [9]. In addition, the number of game based learning apps in the market is increasing. The attention of parents and teachers to this kind of apps (educational game) is rising. The lack of knowledge about these game based apps on education and collaboration is a cause for concern. Schuler 2012 reported that more than 80% of the top-selling apps in the Apple Apps Store are about pre-school, education, collaboration and culture. So, to see how this program have been selected and used, research and analysis are needed [10, 11].

In addition, teachers and parents have repeatedly asked researchers to investigate the effects of educational programs among young children. Due to the large number of these requests, an essential demand for study on the use of media and technology in pre-school age were observed [12–14]. A wide range of prior studies on the use of educational games for pre-school is largely focused on the training in the basic settings and not expressly on game analysis, social skills, collaboration and culture [15–17]. Also, according to the increasing use of technology among children, play and foster collaboration and learning culture in the virtual environment has become a challenge to investigate [18–20]. Few studies have pointed out that young children can use an extensive area of technologies in order to elevate collaboration and culture [21]. However, further research on the variety, collaboration and nurture it during games training is needed [21, 22] (Fig. 1).

4 Finding the Features of Collaborative Learning and Culture

To identify and compare existing models and frameworks in order to find features of collaborative learning and culture and provide an initial conceptual model of game based social skills to foster collaboration for pre-school level. Online survey of data searching is: IEEE Explore, Science Direct, Web of Science, Springer and Scopus to gather data from historical data, pilot study, experience survey and case study to approach these sub-objectives:

1. To identify problems of existing models and frameworks of game based social for fostering collaboration approach pre-schooler to better learning cultures

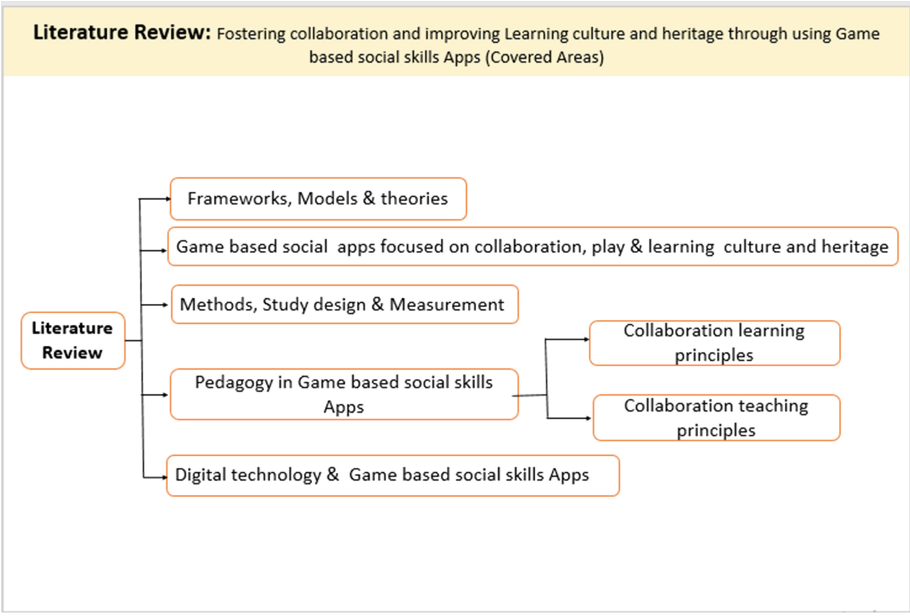


Fig. 1. Literature review: fostering collaboration and improving learning culture and heritage through game based social skills Apps (Covered Areas)

2. To find strengths of existing frameworks and models regarding learning and fostering collaboration
3. To search regarding outcomes of fostering collaboration and learning culture on existing studies about game based social skills apps
4. To search and select best measurement tools and methods used by these studies to evaluate their ideas.

5 Experiment

To determine the domain and features of exploratory research techniques, of existing methods and models of more effective game based social skills apps in order to find features of fostering collaboration. Top existing game based social skills apps that are mentioned as more effective on collaboration by researches, will be selected and evaluated. This phase is going to approach these sub-objectives:

1. To identify support/features of fostering collaboration of top existing game based social skills apps
2. To investigate overall design features: game genre, game element, theory of learning are used by these studies to foster collaboration and learning culture
3. To identify supporting (scaffolding) of use
4. To investigate the problems of pedagogy of top existing game based social skills apps.

6 Participation

Montessori education for pre-schooler is fundamentally a method of learning through playing game and collaboration. Through case study methods, a researcher is able to go beyond the quantitative statistical results and understand the behavioural conditions through the actor's perspective. By including both quantitative and qualitative data, case study helps explain both the process and outcome of a phenomenon through complete observation, reconstruction and analysis of the cases under investigation (Tellis 1997).

7 Observations and Interviews

'Go Pro' chestcam is a camera that is strapped to the child's chest and allows the recording of action as the child moves and interacts with other people and objects, including tablets without their care. In an isolated location, a well-trained researcher will conduct the experiment together with each individual child respondent. To start the session, the researcher will engage the child participant in some assessments of the child's ability to speak. Parents of participating children will have to complete a questionnaire via online on their child's media habits, behaviour, and family information. All experiment sessions will be recorded with "chestcam" video and audio.

8 Conclusion

From this study, it is evidenced that people are more aware of the positive impacts of game apps in learning process of various disciplines. One of the latest trends in mobile phones is the wave of smart phone apps, which include game based social skills and collaboration apps for young children. Unfortunately, there is still lack of studies concerning methodologies and proper guidelines in developing game apps that is able to integrate game based social skills in real life learning settings. To integrate such matters, the evidence of game based social skill effectiveness is definitely needed. This proposal purposed to report evaluation in the effectiveness of game based social skills apps about Japanese culture and etiquette of collaboration. The aims of this proposal included: (1) To investigate and identify features for fostering collaboration through games based social skills apps. (2) To design and develop a model that integrates content, behavioural and applying components of collaboration, roles and relationship with learning Japanese culture and etiquette in game based social skills apps for pre-schooler's level (3–6 years old). (3) To design and implement a novel prototype of game based social skills of Japanese culture app for pre-schooler level based on the proposed model of game based social skills apps. (4) To evaluate the proposed model of Japanese culture game based social skills and collaboration apps. The research methodology in this study used the extended version of instruments in previous studies by assessing the consequences of the game on raising awareness by avoiding selective attention bias and medium term retention. This research focuses to gain empirical evidences of assessing the advantages of game based social skills in Japanese cultural and heritage recognition among young children.

References

1. Mortara, M., Bellotti, F., Berta, R., Catalano, C.E., Fiucci, G., Houry-Panchetti, M., Petridis, P.: Serious games for cultural heritage: the GaLA activities. In: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage, VAST 2011, Prato (FI), Short and Project Papers, pp. 69–72, October 2011
2. Ellis Weismer, S.: Specific language impairment. In: Cummings, L. (ed.) Communication disorders, pp. 73–87. Cambridge University Press, New York (2013)
3. Laws, G., Bates, G., Feuerstein, M., Mason-Apps, E., White, C.: Peer acceptance of children with language and communication impairments in a mainstream primary school: associations with type of language difficulty, problem behaviours and a change in placement organization. Child Lang. Teach. Ther. **28**, 73–86 (2012)
4. Conti-Ramsden, G., Botting, N.: Social difficulties and victimisation in children with SLI at 11 years of age. J. Speech Lang. Hearing Res. **47**, 145–161 (2004)
5. Brinton, B., Fujiki, M., Montague, E.C., Hanton, J.L.: Children with language impairment in cooperative work groups: a pilot study. Lang. Speech Hearing Serv. Schools **31**, 252–264 (2000)
6. Mortara, M., Catalano, C.E., Bellotti, F., Fiucci, G., Houry-Panchetti, M., Petridis, P.: Learning cultural heritage by serious games. Elsevier J. Cult. Heritage (in press). https://doi.org/10.1016/j.culher.2013.04.004
7. Ofcom: Children and parents: media use and attitudes report
8. Plowman, L., Stephen, C.: Children, play and computers in pre-school education. Br. J. Educ. Technol. **36**(2), 145–157 (2005)
9. National Literacy Trust: Parents' Perspectives: Children's Use of Technology in the Early Years (2014). http://www.literacytrust.org.uk/assets/0002/1140/Early_years_parent_report.pdf
10. Shuler, C.: iLearn II: An analysis of the education category of Apple's app store (2012). http://www.joanganzcooneycenter.org/wpcontent/uploads/2012/01/ilearnii.pdf
11. Buckingham, D.: The Media Literacy of Children and Young People. Ofcom report, London (2005)
12. Gillen, J., Cameron, C.A. (eds.): International Perspectives on Early Childhood Research: A Day in the Life. Palgrave Macmillan, Basingstoke (2012)
13. Holloway, D., Green, L., Livingstone, S.: Zero to eight. Young children and their internet use. LSE, London. EU Kids Online. http://eprints.lse.ac.uk/52630/
14. Kucirkova, N.: Children interacting with books on iPads: research chapters still to be written. Front. Psychol. Dev. Psychol. **4**, 1–3 (2013)
15. Merchant, G.: Keep taking the tablets: iPads, story apps and early literacy. Aust. J. Lang. Lit. **38**(1), 3–11 (2014)
16. Lynch, J., Redpath, T.: 'Smart' technologies in early years literacy education: a meta-narrative of paradigmatic tensions in iPad use in an Australian preparatory classroom. J. Early Childhood Lit. (2012). https://doi.org/10.1177/1468798412453150
17. Burke, A., Marsh, J. (eds.): Children's Virtual Play Worlds: Culture, Learning and Participation. Peter Lang, New York (2013)
18. Marsh, J.: Young children's play in online virtual worlds. J. Early Childhood Res. **8**(1), 23–39 (2010)

19. Marsh, J., Bishop, J.C.: Changing Play: Play, Media and Commercial Culture from the 1950s to the Present Day. Open University Press/McGrawHill, Berkshire (2014)
20. Verenikina, I., Kervin, L.: iPads Digital Play and Preschoolers. He Kupu **2**(5), 4–19 (2011)
21. Sutton-Smith, B.: The Ambiguity of Play. Harvard University Press, Cambridge (1997)
22. Playing History website. http://www.playinghistory.eu/front
23. Frochauer, J., Seidel, I., Gartner, M., Berger, H., Merkl, D.: Design and evaluation of a serious game for immersive cultural training. In: 16th International Conference on Virtual Systems and Multimedia (VSMM), pp. 253–260 (2010)
24. Bellotti, F., Berta, R., De Gloria, A., D'Ursi, A., Fiore, V.: A serious game model for cultural heritage. ACM J. Comput. Cult. Herit. **5**, 4 (2012)
25. Froschauer, J., Merkl, D., Arends, M., Goldfarb, D.: Art history concepts at play with ThIATRO. ACM J. Comput. Cult. Herit. **6**(2) (2013)
26. Froschauer, J., Arends, M., Goldfarb, D., Merkl, D.: Towards an online multiplayer serious game providing a joyful experience in learning art history. In: Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES) (2011)
27. Seidl, M., Judmaier, P., Baker, F., Egger, U., Jax, N., Weis, C., Grubinger, M., Seid, G.: Multi-touch rocks: playing with tangible virtual heritage in the museum - first user tests. In: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage, VAST 2011, Prato (FI), Short and Project Papers, pp. 73–76, October 2011
28. Froschauer, J.: Serious heritage games: playful approaches to address cultural heritage. Ph.D. Dissertation, Faculty of Informatics, Wien University of Technology, May 2012
29. Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R.: Taxonomy of educational objectives: the classification of educational goals. In: Handbook I. Cognitive Domain. Longmans, Green, New York (1956)
30. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., Nunge, E.: Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In: 3rd Symposium on Usable Privacy and Security (SOUPS 2007), pp. 88–99. ACM, New York (2007)
31. Hainey, T., Connolly, T.M., Boyle, E.A., Wilson, A., Razak, A.: A systematic literature review of games-based learning empirical evidence in primary education. J. Comput. Educ. University of the West of Scotland, United Kingdom (2016)
32. Mortara, M., Catalano, C.E., Fiucci, G., Derntl, M.: Evaluating the Effectiveness of Serious Games for Cultural Awareness: The Icura User Study (2014)
33. Shabalina, O.: Game-Based Learning as a Catalyst for Creative Learning, Web of Science (2014)

The Main Components of Creativity in Educational Game: A Case Study

Najmeh Behnamnia[1(✉)], Amirrudin Kamsin[1], Maizatul Akmar Binti Ismail[1], and A. Hayati[2]

[1] Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
n.behnamnia@siswa.um.edu.my, {amir,maizatul}@um.edu.my
[2] Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia
a.hayati@gmail.com

Abstract. Recently, due to the increase use of technology among children, fostering creativity in the virtual environment has become a challenge to investigate. Few studies have pointed out that young children can use an extensive area of technologies in order to elevate creativity and learning. The aim of the study is to identify pre-school children's usage of tablet apps and the response in terms of the impact on their learning, playing and creativity. In this study, researcher used a sample of apps that were preloaded onto one tablet. Participants were seven pre-school children (3–6 years old) in Foundation Stage 1 and 2 in a Montessori school. The video recording took place for ten days. The result from the study showed that apps can promote learning and creativity in a wide range of ways, subjected to the design of the apps and the child's individual preferences.

Keywords: Educational games · Game-based learning · Preschoolers Case study · Creativity · Creative thinking

1 Introduction

Recently, the trend of using game based learning apps through touch screen devices among young children has been increasing. Most of preschoolers have access to devices, such as tablet and smartphone which has touch screen at home or school. These devices (tablet and smartphone) have become essential in daily routine of many young children [1]. According to the survey that was conducted to 1028 children around 3 to 5 years of age that was carried out by the National Literacy Trust demonstrated that more than 70% of the respondents have access to device such as tablet and smartphone (touch screen) either at home or school [2]. In addition to that, the number of games based learning apps in the market is also increasing. This also brings a rise in the attention of parents and teachers to this kind of educational games apps. Schuler (2012) reported that more than 80% of the top-selling apps in the Apple Apps Store are targeted to those in preschool and education purposes.

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 292–296, 2018.
https://doi.org/10.1007/978-3-319-95450-9_25

So, in order to see how this app has been selected and used, research and analysis is needed [3, 4].

Apart from that, teachers and parents have raised their concerns to the researchers to investigate the effects of educational programs apps among young children. Due to the large number of these requests, a huge demand for study on the use of media and technology in preschool age were observed [4–6]. A wide range of prior studies on the use educational games for preschool, were largely focused on the training in the basic settings but not much were done on game analysis and creativity [7–9]. Also, due to the increasing use of technology among children, playing and fostering creativity in the virtual environment has become a challenge to be investigated [10–12]. Few studies have pointed out that young children can use an extensive area of technologies in order to elevate creativity [13]. However, further research on how to nurture the variety and creativity during games training is still needed [13, 14].

2 Literature Review

Creativity is a skill that can be used both in teaching and learning. Digital games are used as a means of teaching/learning that can improve student's learning motivation [15–17]. The games can also improve students' academic achievements [18], nurturing and enhancing thinking skills such as creativity, problem solving, collaboration, and critical thinking. According to the presence of elements in game such as; fantasy, curiosity and challenge, digital games will rise interest as well as motivate the children. The game-based learning have interacted with children, resulting in positive learning attitudes namely satisfaction in learning and motivation. These games can stimulate curiosity in children as well as. Other than that, they also encourage children to find new solution in problem solving for specifics circumstances.

3 Method

3.1 Interview Procedure

In this study, researchers used a sample of apps that were preloaded onto one tablet for eight children aged 3–5 years old in Foundation Stage 1 and 2 in a selected Montessori school. The apps used were the ones identified as the top six preschool children's apps commonly used by 3–5 year-olds based on research by Marsh et al. (2015–2016) in United Kingdom. In addition, six augmented reality apps were identified by the research team as suitable for this age group. All children in Foundation Stage classes 1 and 2 were invited to participate in the research. The children's age are outlined in Table 1. Prior to the video recording session, all the children had experience using tablets, although not all of the children have access of tablets at their homes.

Table 1. Demographic profiles of the case study children

Children No.	Name (pseudonym)	Gender	Age at start of study	Class
1	Olivia	Female	6 years	Key stage 1
2	Yu Xin	Male	5 years 8 months	Key stage 1
3	Mani	Male	5 years	Foundation stage 3 (Reception)
4	Michael	Male	4 years 3 months	Foundation stage 3 (Reception)
5	Max	Male	4 years 6 months	Foundation stage 2 (Reception)
6	Ian	Female	3 years 8 months	Foundation stage 2 (Reception)
7	Lela	Female	3 years 2 months	Foundation stage 1 (Nursery)

3.2 Observation of Type of Creative Thinking and Apps Children Use

The video recording took place on two phases of ten separate days over a period of 3 months. The children were recorded using apps sometimes chosen by them but at certain times, they were directed to specific apps by the researcher and teachers. The camera focused on the child's interaction with the screen. Researcher recorded the video in total of 21 h per phase as illustrated in Fig. 1.

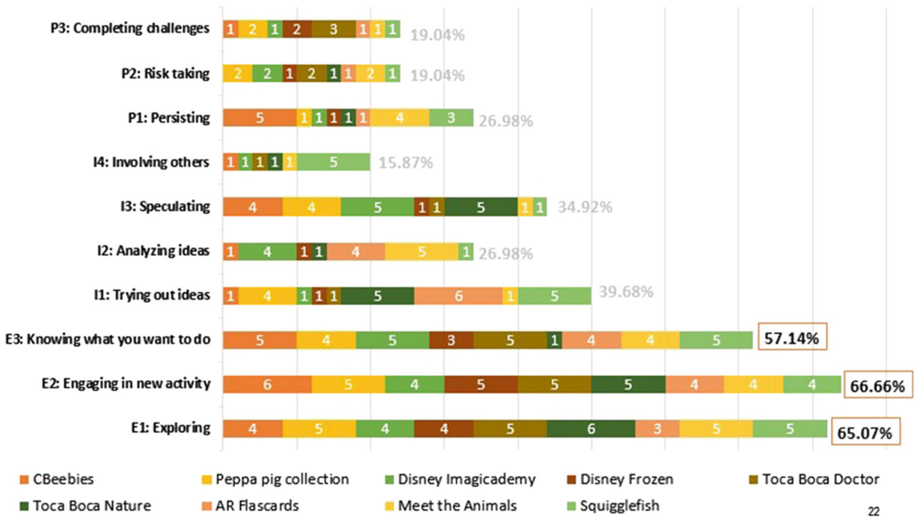


Fig. 1. Observation of the type of creative thinking and apps children use

4 Conclusion

One of the latest trends in mobile phones is the wave of smart phone apps, which includes game-based learning and creativity apps. As at current, research in this area is actively conducted, although its shape and lines are not yet clear. It also needs more insight and knowledge about what is happening in this emerging line at the current stage. This paper aims to divide the research into four categories: reviews and surveys, research studies on apps, developmental efforts and framework suggestions through the review and categorization of explanations. Thus, combining both the components of creativity and learning within a single model or framework could provide better performance for preschooler's level.

References

1. Ofcom. Children and parents: Media use and attitudes report (2014). Accessed at: Plowman, L., Stephen, C.: Children, play and computers in pre-school education. Br. J. Educ. Technol. **36**(2), 145–157 (2005)
2. National Literacy Trust. Parents' Perspectives: Children's Use of Technology in the Early Years (2014). http://www.literacytrust.org.uk/assets/0002/1140/Early_years_parent_report.pdf
3. Shuler, C.: iLearn II: An analysis of the education category of Apple's app store (2012). Accessed http://www.joanganzcooneycenter.org/wpcontent/uploads/2012/01/ilearnii.pdf
4. Buckingham, D.: The Media Literacy of Children and Young People. Ofcom report, London (2005)
5. Gillen, J., Cameron, C.A. (eds.): International Perspectives on Early Childhood Research: A Day in the Life. Palgrave Macmillan, Basingstoke (2012)
6. Holloway, D., Green, L., Livingstone, S.: Zero to Eight: Young Children and their Internet Use. LSE, London. (2013). EU Kids http://eprints.lse.ac.uk/52630/
7. Kucirkova, N.: Children interacting with books on iPads: research chapters still to be written. Front. Psychol. Dev. Psychol. **4**, 1–3 (2013)
8. Merchant, G.: Keep taking the tablets: iPads, story apps and early literacy. Aust. J. Lang. Literacy **38**(1), 3–11 (2014)
9. Lynch, J., Redpath, T.: 'Smart' technologies in early years literacy education: a meta-narrative of paradigmatic tensions in iPad use in an Australian preparatory classroom. J. Early Child. Literacy (2012). https://doi.org/10.1177/1468798412453150. Accessed 3 Aug 2012
10. Burke, A., Marsh, J. (eds.): Children's Virtual Play Worlds: Culture, Learning and Participation. Peter Lang, New York (2013)
11. Marsh, J.: Young children's play in online virtual worlds. J. Early Child. Res. **8**(1), 23–39 (2010)
12. Marsh, J., Bishop, J.C.: Changing Play: Play, Media and Commercial Culture from the 1950s to the Present Day. Open University Press/McGrawHill (2014)
13. Verenikina, I., Kervin, L.: iPads digital play and preschoolers. He Kupu **2**(5), 4–19 (2011)
14. Sutton-Smith, B.: The Ambiguity of Play. Harvard University Press, Cambridge (1997)

15. Bai, H., Pan, W., Hirumi, A., Kebritchi, M.: Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. Br. J. Educ. Technol. **43**(6), 993–1003 (2012). https://doi.org/10.1111/j.1467-8535.2011.01269.x
16. Yang, Y.-T.C.: Building virtual cities, inspiring intelligent citizens: digital games for developing students' problem solving and learning motivation. Comput. Educ. **59**(2), 365–377 (2012). https://doi.org/10.1016/j.compedu.2012.01.012
17. Hwang, G.-J., Wu, P.-H., Chen, C.-C.: An online game approach for improving students' learning performance in web-based problem-solving activities. Comput. Educ. **59**(4), 1246–1256 (2012). https://doi.org/10.1016/j.compedu.2012.05.009
18. Kim, S., Chang, M.: Computer games for the math achievement of diverse students. Educ. Technol. Soc. **13**(3), 224–232 (2010)

Role of Absorptive Capacity in Predicting Continuance Intention to Use Digital Libraries: An Empirical Study

Mohamed Emran Hossain[1], Touhid Bhuiyan[2], Imran Mahmud[2(✉)], T. Ramayah[3], and Brenda Scholtz[4]

[1] Universiti Sains Islam Malaysia, Nilai, Malaysia
emran@daffodilvarsity.edu.bd
[2] Daffodil International University, Dhaka, Bangladesh
{t.bhuiyan, imranmahmud}@daffodilvarsity.edu.bd
[3] Universiti Sains Malaysia, Penang, Malaysia
ramayah@usm.my
[4] Nelson Mandela University, Port Elizabeth, South Africa
Brenda.Scholtz@mandela.ac.za

Abstract. The purpose of this paper is to investigate the impact of absorptive capacity and the quality dimensions of technology on students' continuance intention to use the e-library system. To measure the continuance intention, an integrated research model was developed using expectation-confirmation theory (ECT) and absorptive capacity theory. This empirical study was undertaken at a university in Bangladesh with a sample size of 297. Data was collected via a survey questionnaire. The results reveal that the dimensions of absorptive capacity have a strong effect on confirmation of the system and a partial impact on perceived usefulness (PU). Confirmation of the system has a significant effect on the PU of and satisfaction with the system. Satisfaction was found to be a strong predictor for the continuance intention to use the e-library. Finally, ECT fully fits in this context and students' satisfaction has the largest effect on the continuance intention.

Keywords: Absorptive capacity · Bangladesh · Continuance intention e-library · Expectation-confirmation theory · Partial least squares

1 Introduction

In the current digital era, students and teachers across all educational levels are extensively using information technology (IT). Those who use IT for academic activities within schools, colleges or universities can be classified in different ways—as digital natives, the Google generation, or the Internet generation [1]. Among other technologies within the academic environment, there has been a global increase in the provision of online library services, which provide access to digitized traditional books, e-books, online journals, research papers and theses [2]. According to [3], 70% of students complete their tasks online. Additionally, 77% of educators consider online education to be similar to face-to-face teaching, and perceive an electronic library

M. H. Miraz et al. (Eds.): iCETiC 2018, LNICST 200, pp. 297–308, 2018.
https://doi.org/10.1007/978-3-319-95450-9_26

(e-library) as a powerful, reliable, and valid way for students to prepare high-quality reports [4]. The advantages of using an e-library include efficient services and remote access, improved ease of tracking digital resources and flexibility in searching [5]. Academic online library provides information-rich products, catalogues, digital collections and also allow access to students/researchers high quality scientific information through the web [6].

An e-library is a new concept that could enhance teaching and learning in developing countries like Bangladesh [2]. Despite the lack of interest from the University Grants Commission (UGC), two public universities currently provide partial support for an e-library. Meanwhile, private universities, such as the Daffodil International University (DIU), BRACU and ASAUB, offer full e-library support by collaborating with global libraries [2].

This research focuses on the e-library of DIU, a renowned university in Bangladesh. Despite providing a comprehensive web-based library [2] (see Table 1), DIU has found that student adoption of the e-library is still low. Among the students who have adopted the e-library, total usage is at 55.75% (see Table 2), which means that 44.25% of resources remain unused. As a result, attention is needed to measure the continuance intention which indicates students are influenced by the initial use of e-library to continue to use it further.

Table 1. Online resources of DIU library

Number of registered students in library	10,148
Online book transition report monthly and yearly	Monthly = 2,350 Yearly = 29,850
Total number of books, e-books, journals, articles, and other resources	Books = 26,250 E-books = 13,985 Articles and Journals = 13,877 Voice Library = 32 E-magazines = 286
Connection with foreign libraries, IEEE saga, etc.	30
Frequency of e-books, journals, articles, and others resources downloaded	E-books = 150 Articles and Journals = 167 Voice Library = 80 E-magazines = 100
Member information: 1. Number of male and female students 2. Number of students based on department	1. Male students = 6,780 2. Female students = 3,368 (i) FSIT = 4,124 (ii) Business and Economics = 3,100 (iii) Allied health Science = 685 (iv) Humanities and Social Science = 719

Table 2. Online resource usage of DIU library

	Items	Frequency	Usage
Books	26,250	29,850	113%
E-books	13,985	150	1.07%
Articles and Journals	13,877	167	1.2%
Voice Library	32	80	250%
E-magazines	286	100	34.96%

The purpose of this paper is to understand the phenomenon of continuance intention of students to use an e-library and the antecedents affecting this. This study aims to extend the expectation-confirmation theory (ECT) for information systems' (IS) continuance intention suggested by [7] by integrating the absorptive capacity theory identified by [8]. Thus, the research question of this study is:

RQ: Is there a relationship between a student's absorptive capacity and their continuance intention to use an e-library system?

Drawing on the ECT for IS continuous and absorptive capacity theories, it is proposed that absorptive capacity of understanding, assimilating and applying have a positive effect on perceived usefulness (PU) and the confirmation of using an e-library, which will lead to students' satisfaction and continuance intention of the e-library. The research question of this study was addressed by distributing a survey among users of the e-library provided by DIU and analyzing this data.

This paper is organized as follows: the first section presents a brief review of e-libraries in Bangladesh, the relevant literature, and highlights the unique contributions of the work, while the second section outlines the theoretical foundation of the research. The third section discusses the research model and identifies the hypotheses that were tested, which is followed by a summary of the research method, a description of the data analysis and presentation of the results. Finally, the findings, the implications thereof and future research directions are discussed.

2 Introduction Theoretical Foundation and Development of Hypotheses

2.1 Expectation-Confirmation Theory for IS Continuance

The concept of ECT is based on the satisfaction observed in users' behaviour after the purchase of an item. ECT is widely used in marketing, where [9] suggested that measuring consumer behaviour and customer satisfaction is essential, for they are key determinants of subsequent behaviour, such as repurchase intention. In the case of IT use, [7] suggested using ECT to measure the post-implementation behaviour of users. The model of ECT was thus renamed to the expectation-confirmation model (ECM).

The ECM suggests that the users' intention to reuse a product or continue service use is determined primarily by their satisfaction with prior use of that product or service (see Table 3). Satisfaction is viewed as the key to building and retaining a user's focus

Table 3. Definition of the constructs of ECT theory

Constructs	Definition
Perceived usefulness	Users' perception of the expected benefits of System use
Confirmation	Users' perception of the congruence between expectation of system use and its actual performance
Satisfaction	User satisfaction with prior use of that product or service
Continuance intention	Users' intention to continue using the system

on a system. Thus, investing in user satisfaction should result in the continued intention to use that system.

[10] declared PU, adapted from TAM model, to be a key determinant of user satisfaction because it reflects a long-term belief in the expected benefits of the system. In literature, ECT-IS continuance theory was used to measure the continuance intention of different technologies such as the virtual community [11], mobile messaging [12], information system continuance [13], and online shopping [14]. In the e-library context, a similar idea was emphasized by [15, 16]. As a result, the first formulated hypothesis of the paper is:

H1: PU has a positive effect on satisfaction with an e-library system.

User confirmation (CON) is assessed by users to determine their evaluative response of the system. CON is also a determinant of PU and satisfaction with the system in an educational setting [15]. So, the following hypotheses are:

H2: User confirmation has a positive effect on the PU of an e-library system.
H3: User confirmation has a positive effect on satisfaction with an e-library system.

The users' intention to continue using the IT system primarily relates to their level of satisfaction with using this system. [7] reported that, after the first stage relating to users' belief of the PU of a system, the users' confirmation of using that system will follow. Finally, if they are satisfied with the functionality of the system, they may continue to use it. Similar results were found in the e-library context from the research of [15, 16]. Thus, the fourth hypothesis of this study is:

H4: User satisfaction has a positive effect on his/her continuance intention of an e-library.

2.2 Absorptive Capacity and IS Continuance

[8] described absorptive capacity as a theory that investigates the capabilities of a firm to understand, assimilate, and apply new knowledge through its employees. The theory explained that an organisation could become innovative and achieve a higher level of employee performance when its employees absorbed new knowledge. Absorptive capacity has been applied in a diverse range of research streams, such as knowledge management, IT governance, IT innovation, IT business value and IT adoption [17].

Table 4. Definition of the constructs of absorptive capacity

Construct	Definition
Absorptive capacity for understanding (ACU)	Users' ability to understand the system and its exact value
Absorptive capacity for assimilating (ACA)	Users' beliefs about their ability to operate the system
Absorptive capacity for applying (ACC)	Users may use the system if they have relevant basic knowledge, confidence in their operational ability, and the ability to apply the knowledge gathered from the system to improve learning outcomes

The dimensions of absorptive capacity are understanding, assimilating and applying (see Fig. 2). Several definitions of the constructs of absorptive capacity are provided (see Table 4).

According to [18], the applicability of absorptive capacity is in the context of innovation, inter-organizational learning and new product development which can foster a firm's performance. To adopt knowledge, it is necessary to examine absorptive capacity as a skill of users [19]. [20] argued that, the dimensions of absorptive capacity improve the accuracy of the purpose of any system, accelerate the innovation and enhance learning within the organization.

Absorptive capacity has a positive impact on users' decision quality in deciding whether to use an IS system [21]. In educational settings, [22] explained that a new IS with innovative and useful features but poor understanding will lead to difficulties in using this system, which is also confirmed by the previous research of [23]. As a result, [22] emphasized that absorptive capacity had a positive relationship with the PU of a system related to student learning. So, three additional hypotheses of this paper are:

H5: Absorptive capacity for understanding has a positive effect on the PU of an e-library.
H6: Absorptive capacity for assimilating has a positive effect on the PU of an e-library.
H7: Absorptive capacity for applying has a positive effect on the PU of an e-library.

Absorptive capacity proved successful not only for existing knowledge but also to exploit newly acquired knowledge to perform regular tasks. Universities have invested resources to develop e-libraries, but if students fail to adopt these systems, the universities may decide not to adopt an e-library in the future [19]. This consequence implies that a lower level of absorptive capacity of assimilation will decrease the confirmation of an e-library. Meaningful learning is a non-stop process that involves linking different information with existing student knowledge and extends learning beyond an educational institution [23]. [22] suggested that if students feel confident in their ability to use e-library resources, it is possible that the e-library will a have high potential to make their learning as comfortable as possible. These considerations lead to the following three hypotheses:

H8: Absorptive capacity for understanding has a positive effect on the confirmation of using an e-library.
H9: Absorptive capacity for assimilating has a positive effect on the confirmation of using an e-library.
H10: Absorptive capacity for applying has a positive effect on the confirmation of using an e-library.

Finally, the integrated research model is shown in Fig. 1

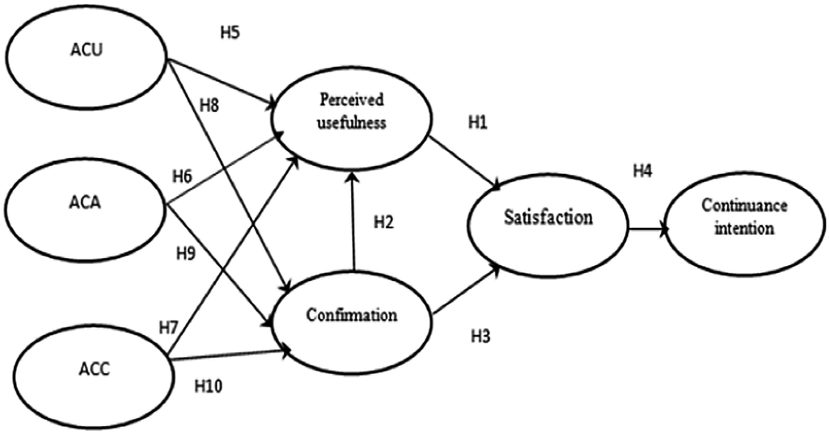


Fig. 1. Proposed research model

3 Research Method

3.1 Data Collection Procedure

The target participants of the study were undergraduate students of DIU from three different campuses. Data was obtained from students who were aware of the e-library provided by DIU and who were registered users of this e-library. Data for this study was collected through a self-administered questionnaire. Research assistants from the project team handed the questionnaires to lecturers who were willing to distribute these questionnaires to their students. Four hundred questionnaires were distributed and 297 were returned with full responses (response rate of 74.25%). The questionnaire consisted of two sections. The first section captured demographic data of the students, while the second section captured information about their absorptive capacity, their level of satisfaction, and their intention to continue using the e-library. The sampling method used in this research was purposive sampling due to two criteria: (1) students must be aware of the e-library provided by the university; (2) Students must visit the system at least once during the course of their studies.

3.2 Measurement

The evaluation metrics in the study were all adapted from published literature. The evaluation metrics for absorptive capacity were adapted from research by [22] PU was adapted from research by from [15], while satisfaction, confirmation and continuance intention were adapted from research by [23].

3.3 Profile of Participants

Demographic information including gender, age and department were collected for the 297 participants of the study (see Table 5). Among the participants, 80.5% were male and 19.5% were female. Most participants (77.7%) were aged between 21 and 24 years old, 21.8% of participants were below 21 years old and 1.5% were aged 25 years old or older.

Table 5. Demographic information

Demographic	Frequency	Percentage
Gender		
Male	239	80.5%
Female	58	19.5%
Age group		
18–20 years	65	21.8%
21–24 years	231	77.7%
25 years and above	1	0.5%

4 Data Analysis and Result

4.1 Measurement Model

To assess the measurement model, two types of validity were examined: firstly the convergent validity and secondly the discriminant validity. The convergent validity of the measurement is usually ascertained by examining the loadings, average variance extracted (AVE), and composite reliability (CR) [24]. The loadings were all higher than 0.7, the composite reliabilities were all higher than 0.7, and the AVE were higher than 0.5, as suggested by [24] (see Table 6). The result of item loadings, AVE and CR are provided in Table 6.

The discriminant validity of the measures, i.e. the degree to which items differentiate among constructs or measure distinct concepts, was examined by following the [25] approach of comparing the correlations between constructs and the square root of the AVE for that construct (see Table 7). All the values on the diagonals were greater than the corresponding row and column values, thus indicating that the measures were discriminant.

Table 6. AVE and CR

Constructs	AVE	CR
ACA	0.763	0.906
ACC	0.673	0.861
ACU	0.584	0.847
CI	0.667	0.889
CON	0.726	0.888
PU	0.649	0.880
SA	0.693	0.900

Table 7. Discriminant validity

	ACA	ACC	ACU	CI	CON	PU	SAT
ACA	**0.873**						
ACC	0.492	**0.821**					
ACU	0.536	0.579	**0.764**				
CI	0.260	0.491	0.455	**0.817**			
CON	0.686	0.626	0.586	0.425	**0.852**		
PU	0.372	0.444	0.443	0.614	0.510	**0.806**	
SAT	0.390	0.409	0.471	0.575	0.526	0.690	**0.832**

4.2 Structural Model

From the final result of the structural model, it was identified that among the ten hypotheses proposed for this study, only H_6 was not supported (see Table 8). The original ECT-IS continuance theory was very strongly supported. PU had a strongly positive relation and large effect size with SAT ($\beta = 0.570$, $f^2 = 0.496$), where confirmation had a positive relation and small effect with both SAT ($\beta = 0.234$, $f^2 = 0.084$) and PU ($\beta = 0.331$, $f^2 = 0.063$). Finally, significant evidence was found for the relation between SAT and continuance intention (CI) of the e-library ($\beta = 0.575$, $f^2 = 0.493$).

Table 8. Hypothesis test result

Hypothesis	Relationship	Beta value	T value	Decision
H1	**PU- > SAT**	0.570	12.179	Supported
H2	**CON- > PU**	0.331	4.935	Supported
H3	**CON- > SAT**	0.234	4.371	Supported
H4	**SAT- > CI**	0.575	14.585	Supported
H5	**ACU- > PU**	0.177	2.873	Supported
H6	**ACA- > PU**	–0.023	0.337	Not Supported
H7	**ACC- > PU**	0.146	2.279	Supported
H8	**ACU- > CON**	0.170	2.922	Supported
H9	**ACA- > CON**	0.443	7.925	Supported
H10	**ACC- > CON**	0.309	4.918	Supported

Absorptive capacity of understanding (ACU) had a positive significant relation with both PU (β = 0.177, f^2 = 0.025) and Confirmation with a small effect size (β = 0.170, f^2 = 0.042). From this result, it was also identified that the relation of absorptive capacity of assimilating (ACA) with PU (β = –0.023, f^2 = 0.000) was not supported, but supported with CON with a medium effect size (β = 0.443, f^2 = 0.323). The relation between absorptive capacity of applying (ACC) and PU was supported, but with a very weak relation and no effect, but with CON, ACC had a strong positive relation and small effect size (β = 0.309, f^2 = 0.147). See final model in Fig. 2.

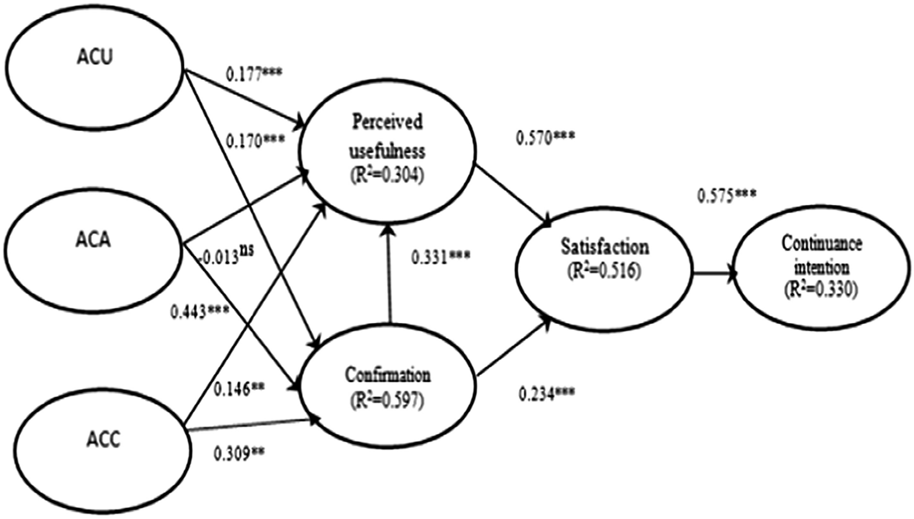


Fig. 2. Final research model with result

5 Discussion

5.1 Relationship Among the Variables Within ECT

This research aimed to understand the role of absorptive capacity in order to measure the continuance intention of students to use an e-library. The resulting research model was conceptualized based on the integration of ECT for IS continuance theory or ECM together with absorptive capacity theory. This research examined the impact of absorptive capacity (understanding, assimilating, and applying) on the PU and confirmation of using an e-library, both of which lead to user satisfaction and continuance intention. The results indicated that PU had a large effect, while confirmation had a small effect on students' satisfaction and satisfaction had a very high impact on students' continuance intention of the e-library. It can be deduced that satisfaction is the key to explaining students' continuance intention of the e-library.

This result also reveals that the ECT-IS continuance theory fits within this context, which is consistent with previous research in educational environments by [15, 16]. Since the variable satisfaction has the highest impact on students' continuance intention, it implies a need for further research to identify the antecedents that have the highest variance and effect on user satisfaction.

5.2 Impact of Absorptive Capacity on ECT

In the case of absorptive capacity, ACU was a significant predictor of the PU and confirmation in using an e-library. These results confirm the research of [17, 22]. ACC is also a significant predictor for confirmation and PU, which is similar to previous research of [17, 22]. ACC led to a medium effect on confirmation and a very weak effect on PU, which implies that some students believe in the importance of the e-library in reaching their academic goals, but not all students agree. It is assumed that since the participants were pursuing an undergraduate bachelor degree, they may possibly not know how to use e-library resources and depend largely on textbooks. Surprisingly, in this study, the ACA, which focused on students' ability to operate the e-library, was not significant with the PU but was significant with the confirmation of the e-library. This finding contradicts the findings of previous researchers. Absorptive capacity improves users' knowledge, which is critical in the IS environment. The university administration and faculty should provide more scope to learn and operate the e-library. For example, providing a manual on how to use the e-library and/or conducting a demonstration on how to use online resources for classroom studies could also assist in becoming familiar with the e-library [22]. In addition, to improve student perceptions of absorptive capacity in an e-library, appropriate training should be provided to the students, thus increasing their familiarity with the e-library environment.

[26, 27] explained that increased participation in the system by top management leads to a higher assimilation of knowledge within the organization. This notion implies that if teachers and employees of institutions use the e-library to upload resource materials or provide students tasks that require the use of the e-library, then assimilation will increase.

6 Conclusion

The first theoretical contribution of this research was the extension of the ECT-IS continuance theory. The majority of previous research focused on e-library adoption and limited research had aimed to integrate absorptive capacity within their research models. The second theoretical contribution was to evaluate the impact of absorptive capacity on users' perception of the usefulness and confirmation of an e-library. Previous literature mostly investigated absorptive capacity in a mandated environment, where users were required to use the system regardless whether they understood the system or not— for example an enterprise system [17, 26]. The proposed research model will provide scientific insights into using absorptive capacity in a voluntary context.

Further, the final research model will encourage the stakeholders of an e-library (developers, librarians, university administration) to focus on the significant variables and analyze the impact on students' continuance intention, which is vital for the success of any university's e-library. This research examined the motivation for and understanding of the factors that drive students towards using an e-library. The research also offers a further research area for academics to measure students' intention before investing resources to expand and improve the e-library associated with their organization.

This study provides insight into the factors influencing students' continuance intention of an e-library in relation to the students' absorptive capacity dimensions: understanding the e-library, belief in using the system and belief of applying the e-library to assist in furthering their education. The significant absorptive capacity of understanding, applying, confirmation and satisfaction to continuance intention was outlined. The study contributes to a better understanding of the ECT and how to improve the usage of an e-library. Factors related to the ECT should be emphasized for students to use the e-library. Further investigation and analysis of this research will assist system developers, librarians and university administration to integrate enjoyment factors within the e-library. Research was undertaken to measure students' continuance intention of an e-library at DIU, which used existing library software, Koha, and was linked with rich resources from various academic and research organizations, such as the British Council, Bangladesh Bureau of Statistics (BBS), The Asia Foundation, Bangladesh Bank, and the Center for Policy Dialogue Bangladesh (CPD). Due to the students' lack of use of the e-library resources, this research sought to measure the students' continuance intention of the e-library. The results of this research contribute to existing literature on e-library implementation projects as well as on ECT-IS continuance intention theory in the field of information systems.

References

1. Gibbons, S.L.: The Academic Library and the Net Gen Student: Making the Connections. American Library Association, Chicago (2007)
2. Ahmed, A.A.A., Siddique, M.N.E. A., Al Masum, A.: Online library adoption in Bangladesh: an empirical study. In: Fourth International Conference on e-Learning Best Practices in Management, Design and Development of e-Courses: Standards of Excellence and Creativity 2013, pp. 216–219. IEEE (2013)
3. Allen, I.E., Seaman, J.: Changing Course: Ten Years of Tracking Online Education in the United States. Newburyport, Sloan Consortium (2013)
4. Silk, K.J., Perrault, E.K., Ladenson, S., Nazione, S.A.: The effectiveness of online versus in-person library instruction on finding empirical communication research. J. Acad. Librarianship **41**(2), 149–154 (2015)
5. Jeong, H.: An investigation of user perceptions and behavioral intentions towards the e-library. Libr. Collections Acquisitions Tech. Serv. **35**(2), 45–60 (2011)
6. Torres-Pérez, P., Méndez-Rodríguez, E., Orduna-Malea, E.: Mobile Web adoption in top ranked university libraries, a preliminary study. J. Acad. Librarianship **42**(4), 329–339 (2016)
7. Bhattacherjee, A.: Understanding information systems continuance: an expectation–confirmation model. MIS Q. **25**(3), 351–370 (2001)
8. Cohen, W.M., Levinthal, D.A.: Absorptive capacity: A new perspective on learning and innovation. Adm. Sci. Q. **35**(1), 128–152 (1990)

9. Oliver, R.L.: Value as excellence in the consumption experience. Consumer Value: A Framework for Analysis and Research. Routledge, New York (1999)
10. Bhattacherjee, A., Perols, J., Sanford, C.: Information technology continuance: a theoretic extension and empirical test. J. Comput. Inf. Syst. **49**(1), 17–26 (2008)
11. Chen, M., Qi, X.: Members' satisfaction and continuance intention: a socio-technical perspective. Ind. Manag. Data Syst. **115**(6), 1132–1150 (2015)
12. Oghuma, A.P., Libaque-Saenz, C.F., Wong, S.F., Chang, Y.: An expectation-confirmation model of continuance intention to use mobile instant messaging. Telematics Inform. **33**(1), 34–47 (2016)
13. Fleischmann, M., Amirpur, M., Grupp, T., Benlian, A., Hess, T.: The role of software updates in information systems continuance—an experimental study from a user perspective. Decis. Support Syst. **83**, 83–96 (2016)
14. Mohamed, N., Hussein, R., Hidayah Ahmad Zamzuri, N., Haghshenas, H.: Insights into individuals online shopping continuance intention. Ind. Manag. Data Syst. **114**(9), 1453–1476 (2014)
15. Cheng, Y.M.: Extending the expectation-confirmation model with quality and flow to explore nurses' continued blended e-learning intention. Inf. Technol. People **27**(3), 230–258 (2014)
16. Baker-Eveleth, L., Stone, R.W.: Usability, expectation, confirmation, and continuance intentions to use electronic textbooks. Behav. Inf. Technol. **34**(10), 992–1004 (2015)
17. Mayeh, M., Ramayah, T., Popa, S.: The role of absorptive capacity in the usage of a complex information system: the case of the enterprise information system. J. Univ. Comput. Sci. **20**(6), 826–841 (2014)
18. Lane, P.J., Koka, B.R., Pathak, S.: The reification of absorptive capacity: a critical review and rejuvenation of the construct. Acad. Manage. Rev. **31**(4), 833–863 (2006)
19. Park, S.Y.: An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. Educ. Technol. Soc. **12**(3), 150–162 (2009)
20. Deng, X., Doll, W.J., Cao, M.: Exploring the absorptive capacity to innovation/productivity link for individual engineers engaged in IT enabled work. Inf. Manage. **45**(2), 75–87 (2008)
21. Seo, Y., Chang Lee, K., Sung Lee, D.: The impact of ubiquitous decision support systems on decision quality through individual absorptive capacity and perceived usefulness. Online Inf. Rev. **37**(1), 101–113 (2013)
22. Lin, H.F.: The effect of absorptive capacity perceptions on the context-aware ubiquitous learning acceptance. Campus-Wide Inf. Syst. **30**(4), 249–265 (2013)
23. Harrington, S.J., Guimaraes, T.: Corporate culture, absorptive capacity and IT success. Inf. Organ. **15**(1), 39–63 (2005)
24. Mahmud, I., Ramayah, T., Kurnia, S.: To use or not to use: modelling end user grumbling as user resistance in pre-implementation stage of enterprise resource planning system. Inf. Syst. **69**, 164–179 (2017)
25. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. **18**(1), 39–50 (1981)
26. Saraf, N., Liang, H., Xue, Y., Hu, Q.: How does organisational absorptive capacity matter in the assimilation of enterprise information systems? Inf. Syst. J. **23**(3), 245–267 (2013)
27. Alzahrani, A.I., Mahmud, I., Ramayah, T., Alfarraj, O., Alalwan, N.: Modelling digital library success using the DeLone and McLean information system success model. J. Librarianship Inf. Sci. (2017). https://doi.org/10.1177/0961000617726123

Author Index