Lecture Notes in Electrical Engineering 782

Hyuncheol Kim
Kuinam J. Kim *Editors*

IT Convergence and Security

Proceedings of ICITCS 2021

Springer

Lecture Notes in Electrical Engineering

Volume 782

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases. ****

More information about this series at http://www.springer.com/series/7818

Hyuncheol Kim · Kuinam J. Kim
Editors

IT Convergence and Security

Proceedings of ICITCS 2021

Springer

Editors
Hyuncheol Kim
Namseoul University
Cheonan, Korea (Republic of)

Kuinam J. Kim
Institute of Creative Advanced Technologies
Science and Engineering
Suwon, Korea (Republic of)

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-16-4117-6 ISBN 978-981-16-4118-3 (eBook)
https://doi.org/10.1007/978-981-16-4118-3

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

Organizing Committee

General Chairs

Hyuncheol Kim, Namseoul University, South Korea

Steering Committee

Nikolai Joukov, New York University and modelizeIT Inc., USA
Borko Furht, Florida Atlantic University, USA
Bezalel Gavish, Southern Methodist University, USA
Kin Fun Li, University of Victoria, Canada
Kuinam J. Kim, Kyonggi University, Korea
Naruemon Wattanapongsakorn, King Mongkut's University of Technology Thonburi, Thailand
Xiaoxia Huang, University of Science and Technology Beijing, China

Publicity Chair

Minsu Kim, Kyonggi University, Republic of Korea
Suresh Thanakodi, National Defence University of Malaysia, Malaysia

Financial Chair

Jongmin Kim, Kyonggi University, Republic of Korea

Publication Chair

Hyeunchul Kim, Namseoul University, Republic of Korea

Program Chair

Kuinam J. Kim, Institute of Creative Advanced Technologies, Science and Engineering, Korea
Nakhoon Baek, Kyungpook National University, Republic of Korea

Organizers and Supporters

Institute of Creative Advanced Technologies, Science and Engineering (iCatse)
University of Science and Technology Beijing, China
Korean Industry Security Forum (KISF)
Korea Convergence Security Association (KCSA)
Hongik University, Republic of Korea
Kyonggi University, Republic of Korea
University of Canterbury, New Zealand
Kyungpook National University, Republic of Korea
Korea Institute of Science and Technology Information (KISTI)

Program Committee

William Emmanuel Yu, Ateneo de Manila University, Philippines
Dr. Nagarajan Velmurugan, APEC, Anna University, Chennai, Tamilnadu, India
Vasco N. G. J. Soares, Instituto de Telecomunicações/Instituto Politécnico de Castelo Branco, Portugal
Pascal Lorenz, University of Haute Alsace, France
Jose Manuel Ribeiro da Fonseca, NOVA University of Lisbon, Portugal
Heming Cui, The University of Hong Kong, Hong Kong
Robert S. Laramee, University of Nottingham, UK
Marco Listanti, DIET Department, University Sapienza of Roma, Italy
Sangseo Park, The University of Melbourne, Australia
Wun-She Yap, Universiti Tunku Abdul Rahman, Malaysia
Siti Rahayu Selamat, Universiti Teknikal Malaysia Melaka, Malaysia
Stelvio Cimato, Università degli Studi di Milano, Italy
Wolfgang A. Halang, Qingdao University of Science and Technology, China

Hyunsung Kim, Kyungil University, South Korea
Zuriati Ahmad Zukarnain, University Putra Malaysia, Malaysia
Terje Jensen, Telenor
Baojun Ma, Shanghai International Studies University, China
Mohd. Saifuzzaman, Daffodil International University, Bangladesh
Mohd Helmy Abd Wahab, Universiti Tun Hussein Onn Malaysia, Malaysia
Prof. Dr.-Ing. Sandro Leuchter, Mannheim University of Applied Sciences, Germany
Pi-Chung Hsu, Shu-Te University, Taiwan ROC.
Yanling Wei, KU Leuven, Belgium
Mainguenaud Michel, INSA-Rouen/LITIS Normandy University, France
Pao-Ann Hsiung, National Chung Cheng University, Taiwan
Dr. Kittisak Jermsittiparsert, Henan University of Economics and Law, China
Bikram Pal Singh, Global Group of Institutes, India
Harikumar Rajaguru, Bannari Amman Institute of Technology Sathyamangalam-India
Ong Thian Song, Multimedia University Malaysia, Malaysia
Dr. Ahmad Kamran Malik, COMSATS University Islamabad (CUI), Pakistan
Dr. Shitala Prasad, I2R, A*Star Singapore, Singapore
Longzhi Yang, Northumbria University, UK
Ad Lalit Prakash Saxena, Combo Legal Consultancy Obra India, India
James Braman, Community College of Baltimore County, USA
Ljiljana Trajkovic, Simon Fraser University, Canada
Maurantonio Caprolu, Hamad Bin Khalifa University, Doha, Qatar
Pavel Loskot, Swansea University, UK
Dr. Chittaranjan Pradhan, Kalinga Institute of Industrial Technology (KIIT) Deemed to be University, Bhubaneswar, India
Fu-Hau Hsu, National Central University, Taiwan
Xiaochun Cheng, Middlesex University, London, UK
Mauro Gaggero, National Research Council of Italy, Italy
Ir. Prof. Dr. Goi Bok Min, Universiti Tunku Abdul Rahman (UTAR), Malaysia
Hardeep Singh, IKG Punjab Technical University, India
Shimpei Matsumoto, Hiroshima Institute of Technology, Japan
Maicon Stihler, Centro Federal de Educação Tecnológica de Minas Gerais—CEFETMG
Dr. Ng Hui Fuang, Universiti Tunku Abdul Rahman, Malaysia
Dennis Pfisterer, University of Luebeck, Germany
Daniel B.-W. Chen, Monash University, Australia
Nasir Uddin, Senior Security Architect at Microsoft, USA
Dr. Ramadan Elaiess, University of Benghazi, Libya
Suresh Thanakodi, Universiti Pertahanan Nasional Malaysia, Malaysia
Siti Sarah, Asia Pacific University (APU), Malaysia

Preface

This LNEE volume contains the papers presented at the iCatse International Conference on IT Convergence and Security (ICITCS2021) which was held on May 15–17, 2021.

ICITCS2021 will provide an excellent international conference for sharing knowledge and results in IT Convergence and Security. The aim of the conference is to provide a platform to the researchers and practitioners from both academia as well as industry to meet the share cutting-edge development in the field.

The primary goal of the conference is to exchange, share and distribute the latest research and theories from our international community. The conference will be held every year to make it an ideal platform for people to share views and experiences in IT Convergence and Security-related fields.

On behalf of the Organizing Committee, we would like to thank Springer for publishing the proceedings of ICITCS2021. We also would like to express our gratitude to the 'Program Committee and Reviewers' for providing extra help in the review process. The quality of a refereed volume depends mainly on the expertise and dedication of the reviewers. We are indebted to the Program Committee members for their guidance and coordination in organizing the review process and to the authors for contributing their research results to the conference.

Our sincere thanks to the Institute of Creative Advanced Technology, Engineering and Science for designing the conference web page and also spending countless days in preparing the final programme in time for printing. We would also like to thank our organization committee for their hard work in sorting our manuscripts from our authors.

We look forward to seeing all of you at the next year's conference.

Editors of ICITCS2021
Kuinam J. Kim
Kyounggi University
Suwon, Korea (Republic of)

Hyuncheol Kim
Namseoul University
Cheonan, Korea (Republic of)

Contents

Machine Learning and Deep Learning

Security and Privacy

Other Related Topics

Future Network Technology

Intelligent Vehicular Networking and Applications

Blockchain and Cryptocurrency

Machine Learning and Deep Learning

An NMT-Based Approach to Translate Natural Language Questions to SPARQL Queries

Jia-Huei Lin and Eric Jui-Lin Lu

Abstract SPARQL is a powerful query language which has already been widely used in various natural language question answering (QA) systems. As the advances of deep neural networks, Neural Machine Translation (NMT) models are employed to directly translate natural language questions into SPARQL queries in recent years. In this paper, we propose an NMT-based approach with Transformer model to generate SPARQL queries. Transformer model is chosen due to its relatively high efficiency and effectiveness. We design a method to encode a SPARQL query into a simple sequence with only RDF triples reserved. The main purpose of this step is to shorten the sequences and reduce the complexity of the target language. Moreover, we employ entity type tags to help modify the translation result due to the vocabulary mismatch problem. The proposed approach is evaluated with an open-domain question answering dataset (QALD-7) on BLEU score and accuracy which result in an outstanding 80.83% on BLEU score and 78.07% on accuracy.

Keywords SPARQL generation · Neural machine translation · Question answering · Transformer

1 Introduction

SPARQL is the W3C recommended query language for retrieving data from knowledge bases (KB) stored in the Resource Description Framework (RDF), a directed and labeled graph data format [1]. Searching for information with SPARQL is significantly more accurate and efficient than using a regular search engine since SPARQL requires an understanding of the syntax and semantics of a natural language and constructs structured queries [2]. Due to its highly powerful capability, SPARQL has already been widely used in various information systems, making it an important and urgent task to automatically generate SPARQL queries.

J.-H. Lin · E. J.-L. Lu (✉)
National Chung Hsing University, Taichung, Taiwan (R.O.C.)
e-mail: jllu@nchu.edu.tw

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_1

Traditionally, there are two main approaches to generate SPARQL. The first one is to turn natural language questions into an intermediary format (ex. entity type tags), generally with some multiclass classifiers, followed by templates or dependency structure to generate the final SPARQL queries [3]. The second one is to rely on the structure of knowledge graph, applying algorithms for subgraph search to find all possible RDF triples [4, 5]. These approaches are capable of handling some complex SPARQL queries; however, the cost of querying stays at a relatively high level.

As the advances of deep neural networks, Neural Machine Translation (NMT) models are employed to directly translate natural language questions into SPARQL queries in recent years [2, 6, 7], which have already reached good results in some closed-domain question answering datasets [2, 8]. However, the proposed approaches failed to solve open-domain question answering datasets like LC-QuAD [9], mainly because of the complexity of queries as well as the vocabulary size [2]. These difficulties make it challenging to practically apply NMT models into question answering systems by far.

In this paper, we propose an NMT-based approach with the Transformer model [10] to automatically translate natural language questions into SPARQL queries. A Transformer model is chosen due to its relatively high efficiency than RNN-based models as well as the effectiveness in various natural language processing tasks. To overcome the complexity problem, we design a method to encode a SPARQL query into a simple sequence with only RDF triples reserved to be the target language for the NMT model. Moreover, pre-trained Word2Vec embedding [11] is used to minimize the unknown word problem. Finally, we modify the translation result by entity type tags [5] to reduce vocabulary mismatch in the target language (the encoded SPARQL).

Our approach combines benefits of NMT models and approaches involve an intermediary format, with resolutions to remedy their main drawbacks. Results are evaluated with an open-domain question answering dataset (QALD-7) on BLEU score and accuracy. The experimental results showed that our approach achieved 80.83% and 78.07% on BLEU score and accuracy; respectively.

2 Related Works

2.1 Neural Machine Translation

Deep neural networks under the encoder-decoder architecture have shown to be powerful in not only machine translation but any other sequence to sequence tasks [12]. Most of the prevalent approaches were based on recurrent neural networks (RNN) that benefit sequential learning until the Convolutional Sequence to Sequence Model (ConvS2S) [13] and Transformer Model [10] were launched.

Google's Neural Machine Translation system (GNMT) [14] is constructed by a deep LSTM network which consists of 8 encoder and 8 decoder layers with attention mechanism between the bottom layer of the decoder to the top layer of the encoder. ConvS2S [13] is a fully convolutional architecture uses convolutions and gated linear units in its encoder and decoder layers, including attention mechanism as well to capture the relation between input and target language. Transformer model [10] is, on the other hand, the state-of-the-art machine translation model which based solely on attention mechanism, with its encoder and decoder layers are all constructed by self-attention. Both ConvS2S and Transformer models are way more efficient and better exploit the hardware than RNN-based methods [10, 13].

2.2 SPARQL Generation Using NMT Models

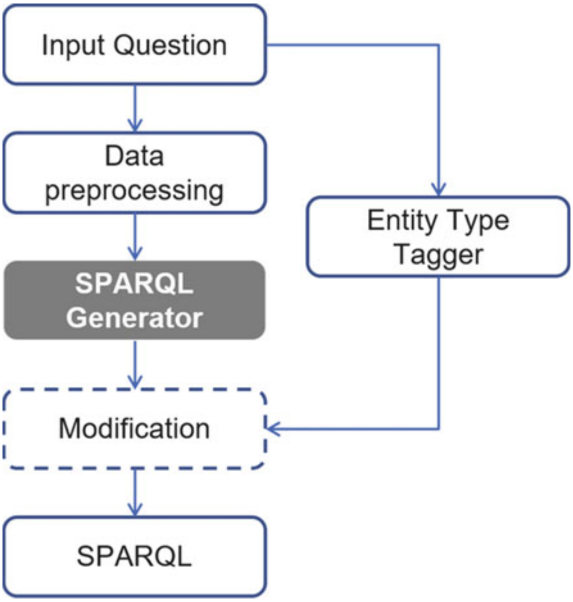
Using NMT models in SPARQL generation has been a popular field in recent years. Soru et al. [6] proposed an end-to-end architecture to translate natural language questions into SPARQL with a basic LSTM-based NMT model. To generate the target language of the NMT model, SPARQL is encoded into a sequential format that operators, brackets, and URIs are replaced by string-look variables. For instance, a SPARQL query 'SELECT DISTINCT ?uri WHERE {dbr:Fort_Knox dbo:location ?uri.}' would be encoded into 'select distinct var_uri where brack_open dbr_Fort_Knox dbo_location var_uri sep_dot brack_close'. Results of their approach were evaluated with the Monument Dataset, a closed-domain dataset. Further research also done by Soru et al. [7] showed that shortening SPARQL sequences and adding direct entity translations are proved to be better ways to improve the translation results.

Yin et al. [2] utilize eight NMT models, including 6 LSTM-based models, ConvS2S model, and Transformer model, as well as three question answering datasets (The Monument Dataset [6], DBNQA [15], and LC-QuAD [9]) to investigate suitable models for SPARQL generation. ConvS2S outperformed the other 7 models in this task, reaching a 97.12% in the Monument Dataset and a 59.54% in LC-QuAD on BLEU score. As for the Transformer model, it was in the second place and performed well in smaller datasets but failed to handle DBNQA which contains about 900,000 question pairs. In addition, none of the eight models produced one fully correct query in LC-QuAD, where the highest accuracy was only 8% by the ConvS2S model, meaning that there are still defects on these approaches to solve an open-domain and complex dataset.

3 Our Approach

The architecture of our approach relies on three main phases, including data preprocessing, the SPARQL generator trained with a Transformer model, and the modification phase considering entity type tags (see Fig. 1).

Fig. 1 Architecture of the presented approach



3.1 Data Preprocessing

To encode SPARQL into sequential format is an essential step when adopting an NMT model for SPARQL generation [7]. In our data pre-processing phase, question-query pairs in the dataset are artificially turned into a newly designed format. We encode a SPARQL query into a simple sequence. Moreover, only RDF triples in a SPARQL query are reserved for presenting the query intention, while all the other elements are removed to shorten the sequences and simultaneously reduce the complexity of the target language. The process is described in detail below and accompanied by a schematic diagram as shown in Fig. 2.

First, name entity tags are extracted with the NLTK NER tagger. If any phrase in the input natural language question is recognized as a name entity, we replace the phrase with a variable ‘NER’. At the same time, if the recognized name entity is used in the target SPARQL query, the entity will also be replaced with variable ‘NER’. As shown in Fig. 2, for the question ‘Where is Fort Knox located?’, it is converted to ‘Where is NER located?’, because Fort Knox was tagged as an NER. Also, dbr:Fort_Knox in the target SPARQL query is converted to NER.

As for property or class entities in the target SPARQL query, we look up every single entity in the input question and find a most suitable word to replace the entity. For example, ‘dbo:location’ is replaced by ‘located’.

Finally, less important elements in the SPARQL queries are all removed and only the RDF triples are reserved, while commas are added between each entity to separate subject, predicate, and object in an RDF triple. After all, only ‘NER, located, ?ans.’ is

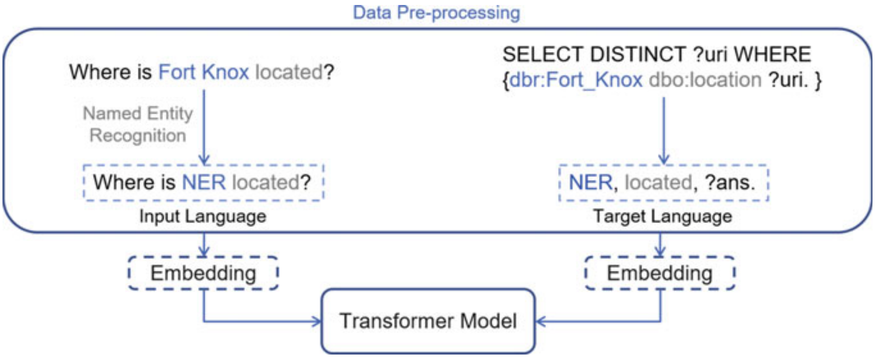


Fig. 2 The data pre-processing and training phase with an example

Table 1 Comparison of hyperparameters between the original version of Transformer model and the presented one in this paper

Model	Layer number	Multi-head number	DFF units	Embedding size	Embedding corpus
Original transformer	6	8	1024	512	Training data
The presented transformer	2	4	1024	256	Word2Vec

reserved in our approach, which is way more uncomplicated for a machine translation model to learn than the prior methods [2, 6].

3.2 *The SPARQL Generator*

The SPARQL generator is then trained by the pre-processed data using a Transformer model. In this paper, we focus only on Transformer model due to its relatively high efficiency. The hyperparameter experimenting and tuning of the presented Transformer model will be described in detail in Sect. 4.2. After experiments, Table 1 lists the best hyperparameters of the presented model, while pre-trained Word2Vec embedding is employed.

3.3 *The Modification Phase*

The modification phase will only be done in the testing phase as a remedy for unsolvable vocabulary problems. Since vocabulary mismatch is profusely found in the translation result at first, we employ entity type tags from Ou's study [5] to help modify the result. Thus, the final SPARQL queries are decided jointly by both the result of Transformer model and the entity type tags. V, N, E, R, and C are five main entity type tags, where V presents the interrogatives, N presents the unimportant or stop words, and E, R, C, respectively, present the name entities, relations and classes in the knowledge base. For question 'Where is Fort_Knox located?', the entity type tagger will give an output 'V-B N E-B R-B'.

For instance, for the same question 'Where is Fort Knox located?', the correct translation result will be'NER, located, ?ans.'. If the question is incorrectly translated into 'NER, married, ?ans.', where the word 'married' does not exist in the question, it is defined as a vocabulary-mismatched question. Once a vocabulary-mismatched question is found, the mismatched words will be replaced by words with E, R, or C tags, depending on their position.

4 Experiments

4.1 Dataset and Evaluation

As one of the main purpose of this paper is to present an approach to automatically generate SPARQL and can be used in an open-domain question answering system, we utilize the 7th Open Challenge on Question Answering over Linked Data (QALD-7) [16] as our training and testing dataset, which is one of the most popular question answering competition that provides up-to-date benchmark and datasets annually. In the current state of our research, we select only 'simple question' in QALD-7, which contains only queries without filter, including 186 question-query pairs in training set and 23 pairs in testing set.

A combination of evaluation metrics is chosen to present and evaluate the output result in this paper, including the BLEU score, accuracy, and F1 score. BLEU score is commonly used in machine translation and proved to be highly correlated to artificial evaluations [17], while it presents only word or phrase correctness but lack considering about the order. Therefore, the exact string matching accuracy and F1 score are employed to evaluate if every single element in the output sequence is in the right position.

4.2 Hyperparameter Experimenting and Tuning

Since generating SPARQL queries with an NMT model is to some degree different from a regular translation task between two natural languages, which has a less complex target language, experiments are done to find out the most suitable model for our task.

Table 2 BLEU score between different multi-head numbers and DFF units

Multi-head number	DFF units = 512 (%)	DFF units = 1024 (%)
4	75.81	**80.83**
8	78.10	72.56
16	72.31	76.18
32	71.98	74.91

Table 3 BLEU score between different layer numbers

Layer number	2	4	6
BLEU score	**80.83%**	70.33%	73.86%

We first fix the layer number of the encoder and decoder on 2 to observe the relationship between the multi-head number and unit of the deep feed forward neural network (see Table 2). While a deeper feed forward neural network seems to be necessary for extracting more information from the data, 4 or 8 heads, which means to project the data to 4 or 8 subspaces, seem to be enough for our model because of the lower complexity of our target language.

Then, we focus on the layer number of the encoder and decoder. As can be seen in Table 3, a smaller layer number seem to better fit our goal since layer number is usually related to the size of dataset, where QALD-7 is obviously a smaller one, with only hundreds of data.

4.3 Results

The final evaluation results of our approach are shown in Table 4. It is shown that the translation result of our Transformer model reaches 61.89% on accuracy, which is already an acceptable result in an open-domain question answering dataset. After modifying the result by entity type tags, the final accuracy improved about 16%, reaching an outstanding result, 78.07% on accuracy. Roughly compare our results

Table 4 Results with and without the Modification Phase

	Transformer only (%)	Entity type tags modified
BLEU Score	56.90	80.83% (+23.93%)
Accuracy	61.89	78.07% (+16.18%)
F1 Score	55.26	78.05% (+22.79%)

with Yin et al.'s study [2], a significantly improvement can be observed on accuracy as their highest accuracy evaluated with LC-QuAD [9] was only 8%.

5 Conclusion

In this work, we propose an NMT-based approach with the Transformer model to automatically translate natural language questions into SPARQL queries, with some remedy and modification. Our approach is evaluated with QALD-7 and reached an outstanding result among open-domain question answering datasets. For future works, we plan to employ a larger dataset with higher complexity to strengthen our current model, where the LC-QuAD might be the first choice by far.

Furthermore, there are few drawbacks that can be improved in the future. Firstly, the current approach heavily relies on name entity recognition and entity type tagging, which might be a waste of training cost. Moreover, QALD-7 dataset is relatively small than other question answering datasets, which provides limited vocabulary size. These problems are probable to be solved by generating a pre-train embedding under DBpedia entities, which is however a hardware-intensive and time-consuming work.

References

1. Prud'hommeaux E, Seaborne A (2021) SPARQL Query Language for RDF. https://www.w3.org/TR/rdf-sparql-query/. Acccessed 01 Aug 2021
2. Yin X, Gromann D, Rudolph S Neural machine translating from natural language to SPARQL. https://arxiv.org/pdf/1906.09302.pdf. Accessed 12 Feb 2020
3. Diefenbach D, Both A, Singh K, Maret P (2018) Towards a question answering system over the semantic web. Semantic Web 11(3) https://doi.org/10.3233/SW-190343
4. Hu S, Zou L, Xu YJ, Wang H, Zhao D (2018) Answering natural language questions by subgraph matching over knowledge graphs. IEEE Trans Knowl Data Eng 30(5):824–837
5. Ou T-A (2018) A natural language query system base on two steps' machine learning (Master' s thesis). National Chung Hsing University, Taichung City, Taiwan
6. Soru T, Marx E, Moussallem D, Publio G, Valdestilhas A, Esteves D, Neto CB (2017) SPARQL as a foreign language. In: SEMANTiCS CEUR workshop proceedings 2044. Amsterdam, The Netherlands
7. Soru T, Marx E, Valdestilhas A, Esteves D, Moussallem D, Publio G (2018) Neural machine translation for query construction and composition. In: ICML workshop on neural abstract machines & program induction v2. Stockholm, Sweden
8. Bordes A, Usunier N, Chopra S, Weston J (2015) Large-scale simple question answering with memory networks. http://arxiv.org/abs/1506.02075. Accessed 12 Feb 2020
9. Trivedi P, Maheshwari G, Dubey M, Lehmann J (2017) LC-QuAD: a corpus for complex question answering over knowledge graphs. Lecture Notes in Computer Science, Springer, Cham, pp 210–218
10. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: 31st conference on neural information processing systems. Long Beach, CA, USA

11. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. https://arxiv.org/pdf/1301.3781.pdf. Accessed 01 Aug 2021
12. Firat O, Cho K, Sankaran B, Yarman Vural FT, Bengio Y (2017) Multi-way, multilingual neural machine translation. Comput Speech Lang 236–252
13. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN (2016) Convolutional sequence to sequence learning. In: 34th international conference on machine learning. Sydney, Australia
14. Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K, Klingner J, Shah A, Johnson M, Liu X, Kaiser L, Gouws S, Kato Y, Kudo T, Kazawa H, Stevens K, Kurian G, Patil NW (2016) Google's neural machine translation system: bridging the gap between human and machine translation. https://arxiv.org/pdf/1609.08144.pdf. Accessed 01 Jan 2021
15. Hartmann A-K, Soru T, Marx E (2018) Generating a large dataset for neural question answering over the DBpedia knowledge base. https://www.researchgate.net/publication/324482598_Generating_a_Large_Dataset_for_Neural_Question_Answering_over_the_DBpedia_Knowledge_Base. Accessed 12 Feb 2020
16. Usbeck R, Ngomo A-CN, Haarmann B, Krithara A, Röder M, Napolitano G (2017) 7th open challenge on question answering over linked data (QALD-7). In: SemWebEval, 2017 Semantic Web Challenges. Springer, Cham
17. Papineni K, Roukos S, Ward T, Zhu W-J (2002) BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting of the association for computational linguistics. Philadelphia, Pennsylvania, USA

Efficient Training Convolutional Neural Networks on Edge Devices with Gradient-Pruned Sign-Symmetric Feedback Alignment

Ziyang Hong and C. Patrick Yue

Abstract With the prosperity of mobile devices, the distributed learning approach enabling model training with decentralized data has attracted wide research. However, the lack of training capability for edge devices significantly limits the energy efficiency of distributed learning in real life. This paper describes a novel approach of training DNNs exploiting the redundancy and the weight asymmetry potential of conventional back propagation. We demonstrate that with negligible classification accuracy loss, the proposed approach outperforms the prior arts by 5x in terms of energy efficiency.

Keywords CNNs · Training · Feedback alignment · Gradient pruning · Edge devices

1 Introduction

The rapid development of deep learning (DL) algorithms brings the Artificial Intelligence (AI) within the bounds of possibility. The prosperity of deep neural networks (DNNs) applications requires a huge amount of feature data. However, there are limitations of centralizing the data for model training. Such huge amount of local data makes it harder to upload due to the limitation of communication bandwidth. Also, people tend to be conservative more and more when it comes to sharing their private data. For instance, for the facial recognition, the data cannot be uploaded to the cloud for training purpose. These issues give rise to the federated learning [20], which investigates the collaborative model training and inference attacks and constructs robust DNNs models. In the federated learning, rather than the local data, the client nodes such as edge devices always send the updated models or gradients to

Z. Hong (✉) · C. P. Yue
Department ECE, HKUST-Qualcomm Joint Innovation and Research Laboratory,
Hong Kong University of Science and Technology, Hong Kong, China
e-mail: zhongad@connect.ust.hk

C. P. Yue
e-mail: eepatrick@ust.hk

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_2

the central server, which requires the client nodes to be equipped with the capability of model training/re-training. Most of the studies [2, 16, 20] on federated learning work on the encryption and model sharing strategy primarily while the lack of local model re-training capability for underlying hardware remains an issue. With current model training approaches, both the throughput and the power offered by edge devices are not enough for federated learning scenario, due to its constraints of power and inferior energy efficiency shown in the Fig. 1.

What is worse, as the computational complexity of the myriad DNNs is increasing, the demand for higher throughput rises drastically. Besides, the slow-down of the Moore's Law makes it harder to meet such a requirement solely depending on technology scaling. The pressure to improve energy efficiency by elaborately designing the specific accelerator architecture on architects and circuit designers becomes higher.

One of the significant factors of limiting the energy efficiency of the accelerators is the excessive external memory access. Containing many epochs until convergence, the model training process of DNNs consumes a great amount of energy by accessing off-chip DRAM for each training sample. Horowitz [9] shows the energy consumption of basic arithmetic and memory operations in a 45 nm CMOS process is dominated by the DRAM access, which is more than 200x larger the average of other operations. In such case, the power consumption solely brought by DRAM access is always beyond the power envelope of typical edge devices such as mobile phones, even in inference process. It poses great challenges on the edge devices for training tasks.

With these considerations in mind, the aim of this work is to design an efficient training algorithm that leverages the redundancy of conventional model training such that the energy efficiency of the accelerators during training is improved. We propose a novel approach called EfficientGrad, which can reduce DRAM access while maintaining high throughput by utilizing the sign-symmetric feedback. As we discussed in more detail in Sect. 2, the symmetric modulatory signal used in conventional back-propagation-based training algorithm is replaced by the sign-symmetric feedback, for both convolutional layers and fully-connected layers. To eliminate the overhead brought by minor gradients calculation in the backward phase while preserving the original training accuracy, the error gradients generated by sign-symmetric feedback is further pruned in a stochastic fashion. The main contributions of the paper are summarized as follows:

- **EfficientGrad algorithm** which imposes the sign-symmetric fixed feedback as the modulatory signals for error gradients and prunes the error gradients that follow in a stochastic approach. To ease the classification accuracy drop of CNNs, the stochastic error gradient pruning is described to maintain the learning capability by reaching low angles between the modulatory signals prescribed by itself and back propagation.
- **Specific data reused architecture** to utilize the sparsity and memory access reduction brought by EfficientGrad. By eliminating the transposed weight matrix fetching/storing and minor gradients being involved in backward phase, the energy efficiency is dramatically increased.

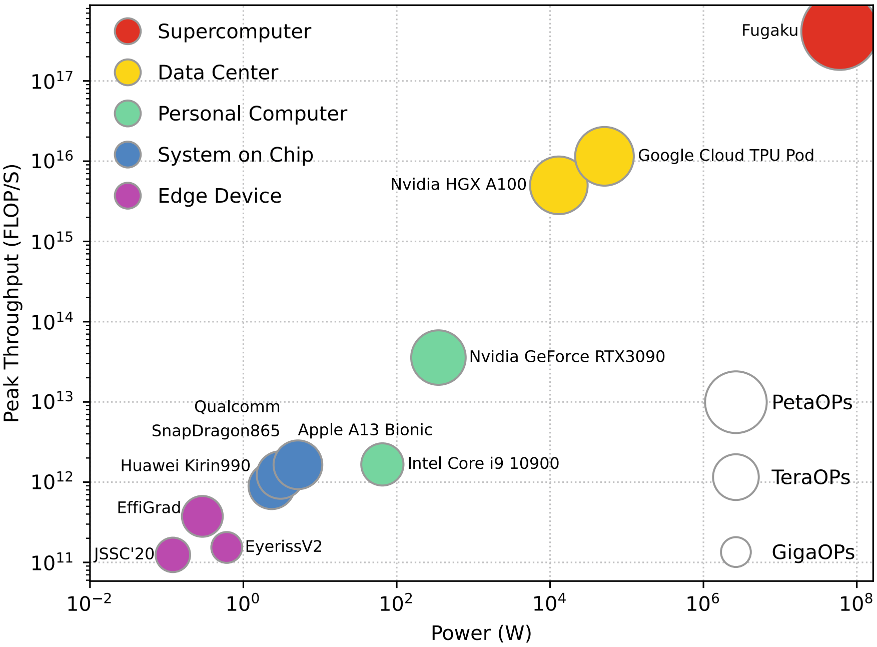


Fig. 1 Throughput versus power comparison of the hardware hierarchy

Our PyTorch [18] implementation used to train the benchmark models with gradient-pruned sign-symmetric feedback alignment is available at https://github.com/HaFred/EfficientGrad.

2 Related Work

There has been substantial studies focusing on accelerating the inference of popular DNNs [4, 7], but no too much efforts are invested on training process for dedicated edge devices. Even though DaDianNao [3] proposed a prototype accelerator for convolutional training, the power consumption is with 14 W therefore it is too large for edge devices, which is normally around hundreds of mW as shown in the Fig. 1. Reference [6] proposes a low-power training processor utilizing feedback alignment, which was first presented by [15], to reduce the energy consumption brought by external DRAM access . However, [6] limits the training capability within fully-connected layers by disabling convolutional training. This is because as [15] analyzes, the angle of the error gradients between back propagation and feedback alignment will be stuck with 90° therefore no learning happens. To deal with this problem, [5, 14, 17] work on the variants of feedback alignment. In section. 4, we show our design solve the aforementioned issues and achieve the learning capability on convolutional layers while minimizing the computations needed in the training process.

3 Generic Back Propagation

Algorithm 1: Generic Convolutional Neural Networks Training

```
Input: [$Img_1$, $Img_2$, ..., $Img_N$]: Input batch with the size of $N$ images, [$W_1$,
        $W_2$, ..., $W_L$]: $L$ layers of trainable weights
Output: Trained network for inference
// Phase 1: Forward
for $l \leftarrow 0$ to $L - 1$ do
    $a_{l+1} = \sigma(W_{l+1} * a_l)$ ;
    if $l = L - 1$ then
        $Loss = C(a_{l+1}, y)$
    end
end
// Phase 2: Backward
for $l \leftarrow L$ to $1$ do
    if $l = L$ then
        $e = \frac{\partial Loss}{\partial a_l} = (a_l - y) \odot \sigma'(a_l)$
    else
        $\delta_l = W_{l+1}^T * \delta_{l+1} \odot \sigma'(a_l)$
    end
end
// Phase 3: Weight gradients update
for $l \leftarrow L$ to $1$ do
    $\Delta W_l = \frac{\partial Loss}{\partial W_l} = a_{l-1} * \delta_l$;
    $W_l = SGD(W_l, \Delta W_l, lr = \gamma, momentum = \mu)$
end
```

Back Propagation (BP) algorithm [12] and the Stochastic Gradient Descent (SGD) algorithm [13] are the canonical approaches for DNNs training. They remains powerful and effective and being used invarious nowadays AI systems. Generally, the BP algorithm is an efficient use of the Chain Rule for generating gradients, and the SGD algorithm takes the average of gradients of a mini-batch input to update the weights. The conventional way of performing back propagation is illustrated in Fig. 2.

In Algorithm 1, a given learning rate γ will be used as the coefficient of the weight gradients. Additionally, the previous updating value is also taken into account by a factor of the momentum μ to help accelerate SGD in the target direction and mitigates oscillations. N denotes the batch size of each training iteration. Practically, we will neither take the whole training data set as Batch Gradient Descent (BGD) nor a single training sample for every weight update. Consequently, mini-batch SGD, which updates parameters by randomly select a mini-batch N of training sample from the training set, is utilized. It helps the optimization to jump out of local minimum comparing to BGD. The training algorithms for CNNs consists of three phases as in Algorithm 1.

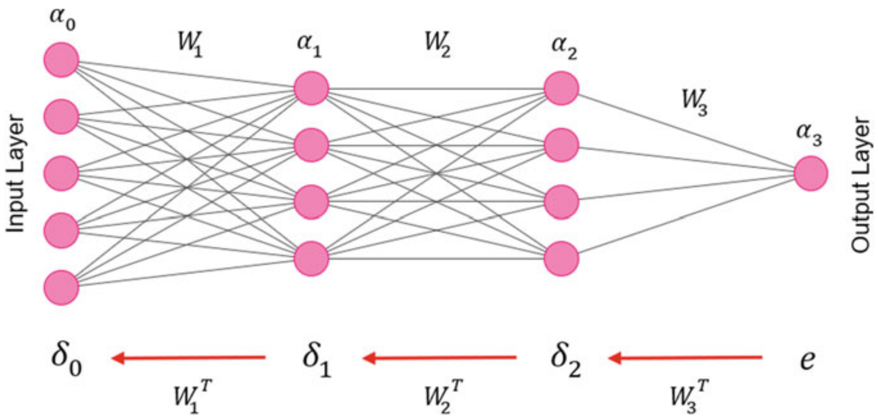


Fig. 2 The forward phase and backward phase in back propagation

4 EfficientGrad

4.1 Algorithm

Consider the feedback alignment proposed by [15], the modulatory signals in the phase 2 of Algorithm 1 is a random feedback matrix B, which eliminates the usage of transposed or 180 degree rotated weight matrix W^T as shown in Fig. 2.

$$\delta_l = B_{l+1} * \delta_{l+1} \odot \sigma'(a_l) \tag{1}$$

Nevertheless, as mentioned in [15], the defect of feedback alignment is that the fixed feedback cannot be directly imposed on convolutional layers. It is because all the neurons with in a convolutional layer share precisely the same receptive field, and such weight sharing aggravates the regularization effect of feedback alignment and leads to over-regularization. Beyond that, the activation function σ in [15] compromises into hyperbolic tangent. From our experiments, We observe that in the early training stages, the regularization effect of feedback alignment will often improperly impel the activation into negative region, which will lead to dead neurons if ReLU is applied.

To address these limitation for feedback alignment on CNNs, we mitigate the over-regularization issues by assigning the fixed random feedback with the symmetric signs of its corresponding weights. Moreover, to restore the improper killed neurons in the hidden layers, we append batch normalization [10] layers in between wherever the neurons tend to be killed. The sign-symmetric feedback alignment in the phase 2 of Algorithm 1 could be obtained:

$$\delta_l = sign(W_{l+1}) \odot |B_{l+1}| * \delta_{l+1} \odot \sigma'(a_l) \tag{2}$$

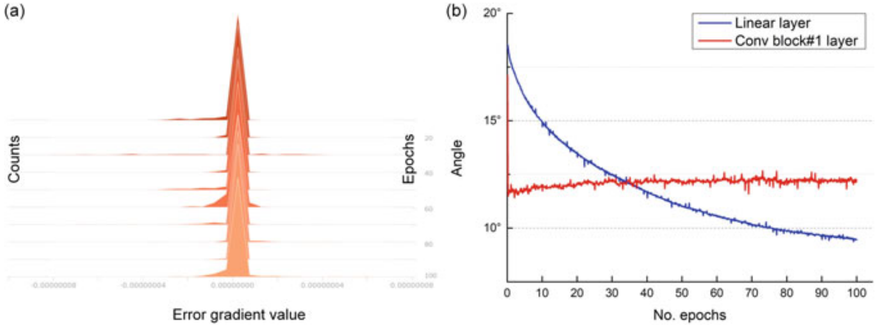


Fig. 3 **a** Error gradients δ distribution of ResNet-18 over 100 epochs, **b** The angles between error gradients prescribed by BP and those prescribed by EfficientGrad of two representative layers

Besides, the resulting error gradients in (2) turns out to be small in magnitude. We observe that the error gradients of adopting sign-symmetric feedback alignment in the phase 2 of Algorithm 1 is distributed in a long tailed normal distribution, which is shown in the Fig. 3a. It means that the computation brought by (2) can be abandoned as long as its expectation remains. Inspired by [21], we propose a stochastic gradient pruning algorithm on (2) to reduce these redundant gradient computations. The basic idea is to prune the error gradients prescribed by sign-symmetric feedback while maintaining their mathematical expectation. To make expectation remain unchanged, it is natural to compensate the pruned gradients back to the pruned threshold:

$$\hat{\delta_{l_i}} = \begin{cases} \delta_{l_i} & \text{if} |\delta_{l_i}| > \tau \\ \tau \odot sign(\delta_{l_i}) & \text{if} \tau \geq |\delta_{l_i}| \geq r\tau \ , r \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where r is a uniform random number ranging from 0 to 1. Note that (3) is applied on top of (2), we need to ensure that the angles of error gradients with EfficientGrad is well under 90°. Since the error gradients are pruned with expectation remains, the sign-symmetric feedback stays unchanged, thus it is still affecting the weight to be aligned with the random fixed feedback as analyzed in [15]. Comparing to the original feedback alignment, the sign-symmetric one with stochastic gradient pruning could be even reach lower angle under 45°, the angles over 100 epochs training on ResNet-18 [8] is shown in the Fig. 3. As discussed in Sect. 2, the lower angle between error gradients the better learning capability. The linear (fully-connected) classifier layers keeps align with the random feedback because of the over-regularization is suppressed in fully-connected layers, whereas the convolutional layers drops fast but tend to be stable. It makes sense because the batch normalization layer liberates the neuron turn-off problems mentioned above and restores the internal covariate shift layer-wisely.

One of the critical parts of EfficientGrad algorithm is to determine a dynamic pruning threshold τ that will preserve the original accuracy that a given DNNs model could reach. Consider the cumulative density function (CDF) Φ of a given δ_l, if we use a pruning rate P to control the gradient sparsity, then (4) holds.

$$P = 1 - \{[1 - \Phi(\frac{\tau}{\sigma})] \times 2\} = 2\Phi(\tau) - 1 \tag{4}$$

$$\tau = \Phi^{-1}(\frac{1+P}{2}) \cdot \sigma \tag{5}$$

With (5), the ratio of δ_l which will be stochastic pruned in (3) is set as P. The expectation of δ_l in EfficientGrad is almost unchanged, thus it leads to a negligible classification accuracy loss.

4.2 Hardware

To enhance the energy efficiency of edge devices for DNNs training task, we design the architecture that leverages the sparsity and memory access reduction brought by EfficientGrad. The DNNs mapping dataflow is based on the row stationary proposed in EyerissV2 [4]. Our DNNs training accelerator (in Fig. 4) consists of 6 Processing Clusters (PCs). Each of the cluster is made up of 12 Processing Elements (PEs) which is in charge of multiply-accumulations (MACs). Within a PC, the weight matrix row is preloaded and shared across each row of the PC. That is to maximize the

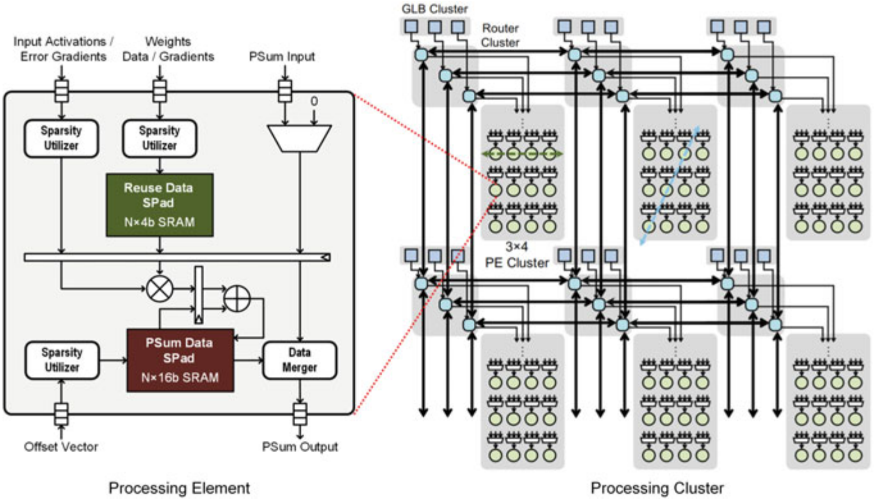


Fig. 4 Overall architecture of the proposed EfficientGrad DNNs Accelerator

convolutional kernel reuse especially when the activation size gets larger. Besides, the activation row is shared anti-diagonally across each PC. In such case, the activation row needs to be operated by certain weight row is manipulated in a systolic array fashion without stall. There is a router cluster and global buffer cluster (GLB) for each PC to communicate PC-wisely and storing the data that is yet to push back to the external DRAM.

Unlike EyerissV2 that only supports CNNs inference, each PE of EfficientGrad accelerator can proceed all three phases of training in Algorithm 1. The weight in phase 1 and fixed feedback in phase 2 are both stored in the reuse data scratchpad inside the PE. It maximizes the convolutional reuse for all three phases and minimizes the external memory access in phase 2.

5 Evaluation

We compare the classification accuracy of ResNet-18 on CIFAR-10 [11] of EfficientGrad with other feedback alignment variants [6, 14] in Fig. 5a. For deep CNNs training, binary random feedback in [6] and sign-symmetric only feedback in [14] degrade in terms of accuracy. EfficientGrad compromises negligible accuracy loss over sign-symmetric random magnitude feedback, to achieve less backward phase calculation.

The hardware part of EfficientGrad is designed in Chisel [1] with the behavioral simulation with ChiselTest [19] and synthesized using the Synopsys Design Compiler with SMIC 14 nm tt process. Utilizing the fork-join property of ChiselTest, we are able to build a simulation-based timing model in scala, to handle the intra-PE pipeline. The power and clock rate of the SRAM/RegFile is generated by the 14 nm off-the-shelf memory compiler. EfficientGrad can achieve peak throughput at 121GOP/S with the clock frequency of 500MHz and the power of 790 mW theoretically. In

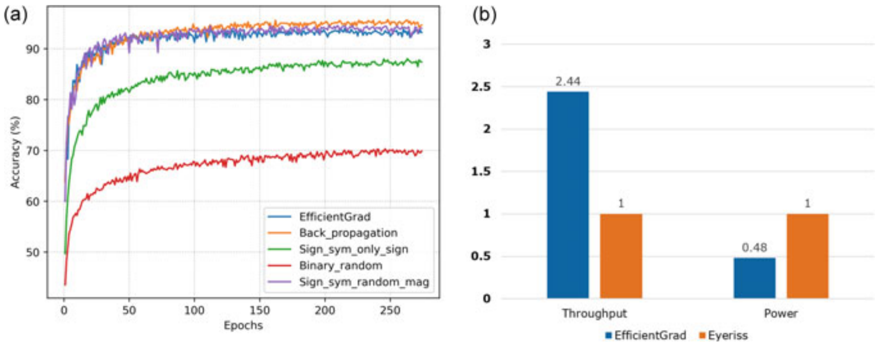


Fig. 5 **a** Classification accuracy convergence of ResNet-18 for training over 270 epochs, **b** The performance of our proposed EffcientGrad comparing to EyerissV2

such case, for one patch forward phase of training on ResNet-18, EfficientGrad can finish within 0.69ms.

We normalize the throughput of EfficientGrad with reference to the unpruned back propagation version of EyerissV2 [4]. It is significantly faster than EyerissV2 with the 2.44x throughput improvement and 0.48x power reduction as shown in the Fig. 5b. We also include EfficientGrad in Fig. 1 to compare with other popular computational devices. It appears that EfficientGrad reaches higher energy efficiency and is well-suited to edge devices.

6 Conclusion

In this paper, we present EfficientGrad, an efficient back-propagation-based DNNs training algorithm that enables one to make full use of both the elasticity of weight symmetry problem, and the redundancy residing in the conventional back propagation algorithm. As demonstrated in the paper, our proposed EfficientGrad can increase the throughput by approximately 2.5x and decrease the power consumption to a half, which leads to superior energy efficiency as 5x against prior accelerators. s

Acknowledgements This work is supported by the Hong Kong Research Grants Council under General Research Fund (GRF) project no. 16215620 and the HKUST-Qualcomm Joint Innovation and Research Laboratory.

References

1. Bachrach J, Vo H, Richards B, Lee Y, Waterman A, Avižienis R, Wawrzynek J, Asanović K (2012) Chisel: constructing hardware in a scala embedded language. In: Proceedings of the 49th annual design automation conference, pp 1216–1225. DAC ’12, Association for Computing Machinery, New York, NY, USA
2. Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, Kiddon C, Konečný J, Mazzocchi S, McMahan HB, Overveldt TV, Petrou D, Ramage D, Roselander J (2019) Towards federated learning at scale: system design. arXiv preprint arXiv:1902.01046
3. Chen Y, Luo T, Liu S, Zhang S, He L, Wang J, Li L, Chen T, Xu Z, Sun N, Temam O (2014) Dadiannao: a machine-learning supercomputer. In: 2014 47th Annual IEEE/ACM international symposium on microarchitecture, pp 609–622
4. Chen Y, Yang T, Emer J, Sze V (2019) Eyeriss v2: a flexible accelerator for emerging deep neural networks on mobile devices. IEEE J Emerg Sel Topics Circuits Syst 9(2), 292–308
5. Frenkel C, Lefebvre M, Bol D (2019) Learning without feedback: fixed random learning signals allow for feedforward training of deep neural networks. Frontiers Neurosci. arXiv:1909.01311
6. Han D, Lee J, Lee J, Yoo H (2019) A low-power deep neural network online learning processor for real-time object tracking application. IEEE Trans Circuits Syst I: Regular Papers 66(5), 1794–1804. doi: 10.1109/TCSI.2018.2880363
7. Han S, Mao H, Dally WJ (2106) Deep compression: compressing deep neural network with pruning, trained quantization and huffman coding. In: Bengio Y, LeCun Y (eds) 4th international conference on learning representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings. arXiv:1510.00149

8. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
9. Horowitz M (2014) 1.1 computing's energy problem (and what we can do about it). In: 2014 IEEE international solid-state circuits conference digest of technical papers (ISSCC), pp 10–14
10. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. PMLR (2015). arXiv:1502.03167
11. Krizhevsky A, Hinton G (20019) Learning multiple layers of features from tiny images
12. LeCun Y (1988) A theoretical framework for back-propagation. In: Touretzky D, Hinton G, Sejnowski T (eds) Proceedings of the 1988 connectionist models summer school, CMU, Pittsburg, PA, pp 21–28. Morgan Kaufmann
13. LeCun Y, Bottou L, Orr GB, Müller KR (1998) Efficient backprop. Neural Networks: Tricks of the Trade. This Book is an Outgrowth of a 1996 NIPS Workshop. Springer, Berlin, Heidelberg, pp 9–50
14. Liao Q, Leibo JZ, Poggio T (2016) How important is weight symmetry in backpropagation? In: Proceedings of the Thirtieth AAAI conference on artificial intelligence, pp 1837–1844. AAAI'16, AAAI Press
15. Lillicrap TP, Cownden D, Tweed DB, Akerman CJ (2016) Random feedback weights support learning in deep neural networks. Nat Commun
16. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA (2017) Communication-efficient learning of deep networks from decentralized data. In: Singh A, Zhu XJ (eds) Proceedings of the 20th international conference on artificial intelligence and statistics. Proceedings of Machine Learning Research, vol 54, pp 1273–1282. PMLR
17. Nø kland A (2016) Direct feedback alignment provides learning in deep neural networks. In: Advances in neural information processing systems 29, pp 1037–1045. Curran Associates, Inc
18. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017) Automatic differentiation in pytorch. In: 31st conference on neural information processing systems (NIPS 2017), Long Beach, CA, USA
19. UCB-BAR: Chiseltest, a test harness for chisel-based rtl designs (2018). https://github.com/ucb-bar/chisel-testers2
20. Yang Q, Liu Y, Chen T, Tong Y (2019) Federated machine learning: concept and applications. ACM Trans Intell Syst Technol **10**(2). https://doi.org/10.1145/3298981
21. Ye X, Dai P, Luo J, Guo X, Qi Y, Yang J, Chen Y (2020) Accelerating cnn training by pruning activation gradients. ECCV. arXiv:1908.00173

Design of the Prediction Model for Adolescents’ Stress Using Deep Learning

Hyejoo Lee and Euihyun Jung

Abstract Adolescents are exposed to various types of stress from parents, study, or friends in their school life. Though these stresses cannot be avoided, the proper monitoring of affecting variables can give educators a chance to help their youths to ease the stresses. Deep Learning is superior to other machine learning methods in terms of prediction performance, but it has a weakness to explain the effects of variables due to its black box characteristic. In addition to this, it requires all input variables to use the trained neural network, but it is frequently impractical to gather all variables in the actual education field. To resolve this issue, the authors suggest the design approach combining Deep Learning and kNN. The authors use feature importance with kNN and reduce variables into one third, but the result of performance evaluation shows that the approach can save the advantage of prediction performance of Deep Learning while reducing the number of variables.

Keywords Deep learning · Prediction · kNN · Adolescents’ stress

1 Introduction

Recently, researchers have been getting surprised by the ability and the possibility of Deep Learning for years [1]. It has changed the rules of the game in every academic and industry domain. Deep Learning can find criminals from surveillance cameras [2], detect cancers in medical imaging [3], or play Go [4]. In some fields, it already surpasses human abilities and starts to replace human roles [5]. For these reasons, researchers in every field have been eagerly trying to find a way to adopt it into their domain and the first movers already tasted the fruits of success.

H. Lee
Department of Education, Chung-Ang University, Seoul, Korea
e-mail: ladyzen@cau.ac.kr

E. Jung (✉)
Department of Convergence Software, Anyang University, Anyang City, Kyunggi-do, Korea
e-mail: jung@anyang.ac.kr

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_3

However, in the educational domain, Deep Learning has not been widely used than expected. There are two reasons for educational researchers to hesitate to adopt it into their studies. Firstly, Deep Learning is weak in explaining the effect of causal variables because of its black box model [6]. It is excellent in prediction or classification, but it fails to present how variables work for the result. However, in the educational domain, finding causal variables and evaluating the variables' effects are especially important, because educators can help students by monitoring the variables. Secondly, in Deep Learning, there is no process of the selection of the most impacting variables. In the training process, all variables are treated equally, and the weights of the variable's importance are automatically calculated in a neural network. It seems very convenient to researchers, but they should collect all input variables to use the trained neural network in return for convenience. Frequently, it is impractical or even impossible to get all the variables of a specific individual in the educational domain.

Nevertheless, Deep Learning is too valuable to be thrown away. Most of all, the method shows much higher prediction results than other machine learning methods. For this reason, the authors decided to adopt Deep Learning to predict the stress level of adolescents. However, it is still difficult to figure out the inner process of the black box and it is beyond the scope of the research, so the authors concluded it was a more practical way to find a method reducing the number of input variables without compromising the performance of Deep Learning. By combining the reducing method, we can design a Deep Learning model which takes advantage of higher prediction performance with a small set of variables.

Youths are constantly exposed to a lot of stress from parents' expectations of study, peer pressures, or comparing appearance. The stresses frequently associate with a variety of misbehaviors in school life such as violence, bullying, or even youth crime. To find the way reducing youth stress, researchers in the educational domain have tried to find the relevant variables of youth stress and have suggested solutions to ease it [7, 8]. If educators or parents predict the degree of stress with high accuracy, they can help their youths to ease the stress by eliminating causes. Especially, in detecting stress, the number of monitoring variables should be as small as possible because it should be detected easily and rapidly in real school life.

In the paper, the authors combined Deep Neural Network (DNN) [9] and k-Nearest Neighbors (kNN) [10] to predict the stress with a small number of variables. In the first step, DNN was trained with total of 28 variables selected from the Korea Youth Panel Survey (KYPS) data [11]. In the second step, the authors reduced input variables to nine variables by calculating feature importance with kNN. In the last step, we trained DNN again with the selected nine variables and compared the performances. The authors expected the performance after reducing variables will be poorer than the original one, but surprisingly it is a little bit better than the original one. It reveals the possibility of Deep Learning model adequate for the educational domain.

The rest of this paper is organized as follows. Section 2 describes a dataset for this study and explains how to train DNN with the dataset. In Sect. 3, the authors show how to reduce input variables to some important ones and compare the performances after training DNN again with the selected ones. Section 4 concludes the paper.

2 Dataset and Basic Model

2.1 Overview of Dataset

The dataset for this research is originated from the Korea Youth Panel Survey (KYPS), a longitudinal study of a nationally representative sample of South Korean children and adolescents collected by the National Youth Policy Institute (NYPI) of South Korea [11]. The KYPS conducted six follow-up surveys from 2003 (Wave 1) to 2008 (Wave 6). For this study, data from the first wave were analyzed and the adolescents were second-graders in middle school.

To make a qualified dataset, data preprocessing has been done. Most of the variables have a 5-point Likert scale. A variable of *chinman* (=meeting friends) has a 3-point Likert scale and a variable of *buhark* (=level of father's education) has an 8-point Likert scale. The higher the score of the variables, the stronger their characteristics. Consequently, a total of 3,147 subjects were included after dropping out those which had missing fields and 27 variables were selected as input variables. The selected variables are summarized in Table 1. The target variable, *stress* was categorized into two values of high and low depending on the degree of stress.

2.2 Deep Learning Model

The dataset is randomly split into a training and a test set in a ratio of 8:2. The training dataset contains a known output, and the model is trained with this data for generalization. The test dataset is used to evaluate the trained model's prediction. The designed neural network model is shown in Table 2. The authors deployed four linear layers to predict the level of adolescents' stress. The output is categorized into two values of high or low.

In the model, Cross Entropy Loss is used as the loss function and Adam optimizer is deployed as the optimizer. The model has trained 700 epochs and the accuracy of the model is 0.69. Figure 1 shows the training curves of the original and the reduced variables models.

3 Reduction of Variables

3.1 The Decision of Feature Importance with kNN

Feature Importance refers to techniques that assign a score to input variables based on how useful they are at predicting a target variable [10]. There are many methods getting feature importance such as decision trees, permutation, and correlation scores.

Table 1 Overview of input variables

Variable	Description	Mean	Standard deviation	Min	Max
buae	Parent attachment	20.04	4.67	6	30
bugam	Parent's monitoring	12.89	3.47	4	20
bupo	Spouse abuse	3.70	1.85	2	10
bumopo	Parent abuse	3.46	1.78	2	10
jatong	Self-control	16.05	4.04	6	30
kyoae	Teacher attachment	7.38	2.47	3	15
chinman	Meeting friends	2.83	0.46	1	3
thoae	Peer attachment	16.22	2.62	4	20
comoh	Preference of computer game	3.45	1.24	1	5
jajon	Self-esteem	18.56	3.79	6	30
jain	Self-awareness of bullying	3.82	1.61	2	10
jumoon	Surround-awareness of bullying	3.39	1.60	2	10
jubee	Concern about critics from surround	6.09	2.10	2	10
jasin	Self-belief	10.39	2.21	3	15
gong	Aggressiveness	16.53	2.25	6	30
knock	Optimistic propensity	10.19	2.35	3	15
saengman	Life satisfaction	3.48	0.87	1	5
buhark	Level of father's education	4.74	1.31	1	8
bugong	Parent's expectation of studying	3.12	1.08	1	5
sunglem	Grades of Korean, English, and math	9.38	2.47	3	15
sulem	Class participation	9.76	2.46	3	15
gunsang	Health	4.00	1.03	1	5
zipkyung	Family economic power	4.09	0.99	1	5
harkjerk	School adaptation	4.14	1.03	1	5
harkgong	Interest in school study	3.83	1.01	1	5
yeoga	Leisure satisfaction	3.27	0.92	1	5
gongbu	Worry about studying	3.58	1.02	1	5

Table 2 DNN model summary

Layer (type)	Output shape	Param #
Linear-1	[−1, 1, 128]	1,280
Linear-2	[−1, 1, 256]	33,024
Linear-3	[−1, 1, 64]	16,448
Linear-4	[−1, 1, 2]	130

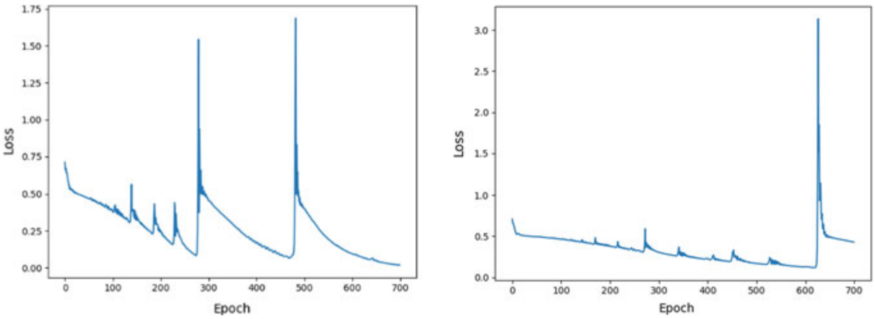


Fig. 1 The training curves of the original (left) and the reduced variables (right) models

The authors select permutation feature importance from k-Nearest Neighbor to decide the feature importance. After running kNN, the values of importance are generated as shown in Fig. 2. The authors cut off the variables below 0.0015 and the remained variables are *buae*, *bumopo*, *jatong*, *chinman*, *gong*, *saengman*, *bugong*, *yeoga*, and *gongbu.*

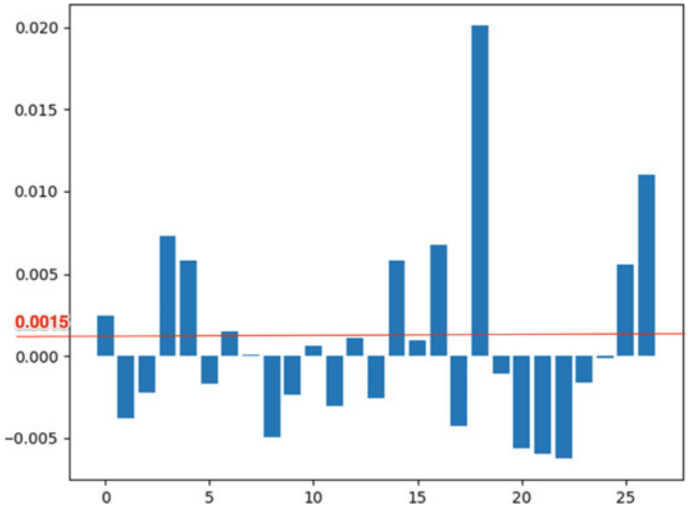


Fig. 2 A feature importance value from running kNN analysis

Table 3 The performances of two models

Metric	Original model	Reduced model
AUC	0.6905	0.7255
Accuracy	0.702	0.738
Precision	0.644	0.691

3.2 Performance Evaluation

After reducing variables, the same DNN model was trained with the reduced nine variables. The following Table 3 gives the comparison of the results from the evaluation of two models. Although the number of variables is reduced to one-third, the performance is even a little bit better. That means educators can use Deep Learning to monitor their students' stress with only a small set of variables and Deep Learning can be a practical way in the educational domain especially.

4 Conclusion

Although Deep Learning is a popular and amazing machine learning method, it has been rarely used in the educational domain. There can be a lot of reasons, the main reason is originated from the black box characteristic of Deep Learning. Researchers cannot figure out which variables are important and how much the variables affect when they use Deep Learning. Additionally, since all the variables are equally required to use the trained model in Deep Learning, every variable should be collected in the educational field to use the trained model. These two issues are critical in the educational domain because educators need to know the effect of variables and they usually get only a small number of variables in the actual educational field.

To resolve them, the authors suggest the design approach combining Deep Learning and kNN. After training a DNN model in a normal way, the authors reduce the input variables into one-third with kNN. Then, the original DNN model is trained again with the reduced variables. The result of performance evaluation shows that the approach can save the advantage of prediction performance of Deep Learning while reducing the number of variables. This implies that the proposed approach enables educators to predict accurately youths' stresses with small monitorable variables which are available in the actual educational field.

Acknowledgements This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5B5A07092270).

References

1. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyum M, Chen S, Iyengar SS (2018) A survey on deep learning: algorithms, techniques, and applications. ACM Comput Surv 51(5):1–36
2. Sreenu G, Durai MS (2019) Intelligent video surveillance: a review through deep learning techniques for crowd analysis. J Big Data 6(1):1–27
3. Ker J, Wang L, Rao J, Lim T (2017) Deep learning applications in medical image analysis. IEEE Access 6:9375–9389
4. Granter SR, Beck AH, Papke DJ Jr (2017) AlphaGo, deep learning, and the future of the human microscopist. Arch Pathol Lab Med 141(5):619–621
5. Sejnowski TJ (2018) The deep learning revolution. MIT Press
6. Samek W, Wiegand T, Müller KR (2017) Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296
7. Hampel P, Petermann F (2006) Perceived stress, coping, and adjustment in adolescents. J Adolesc Health 38(4):409–415
8. Misra R, Crist M, Burant CJ (2003) Relationships among life stress, social support, academic stressors, and reactions to stressors of international students in the United States. Int J Stress Manag 10(2):137–157
9. Aggarwal CC (2018) Neural networks and deep learning, vol 10. Springer, pp 978–983
10. Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. IEEE Trans Pattern Anal Mach Intell 19(2):153–158
11. National Youth Policy Institute (2009) Korea youth panel survey (KYPS) VII: survey summary report for years. Seoul, Korea, pp 1–6

Security and Privacy

The Effect of Sampling Methods on the CICIDS2017 Network Intrusion Data Set

Yan-Bing Ho, Wun-She Yap, and Kok-Chin Khor

Abstract Handling unbalanced intrusion detection data sets are difficult as minority intrusion classes may not be easy to detect. One of the possible causes of the problem is the characteristic of learning algorithms that usually favour majority classes in data sets. The contribution of this study is to improve the detection rate for intrusions in the unbalanced CICIDS2017 data set by using sampling techniques. We evaluated Random Under-Sampling (RUS), Synthetic Minority Over-sampling Technique (SMOTE) and the combination of RUS and SMOTE. After applying the sampling techniques, we performed intrusion detection and used the accuracy plus True Positive Rate (TPR) as the evaluation metrics for the detection results. The results showed that RUS gave the best detection performance overall. Besides, 12 out of the 15 classes, including some hard-to-detect minority classes, were detected with result improvement.

Keywords Samplings · Intrusion detection · CICIDS2017

1 Introduction

Intrusion Detection Systems (IDS) are used to detect malicious activities in the networks and information systems. Due to the increasing network scale and traffic, large network data are generated almost every seconds. However, intrusion activities are relatively rare compared to the overall traffic amount causing the network data to be unbalanced.

Y.-B. Ho (✉) · W.-S. Yap · K.-C. Khor
Lee Kong Chian Faculty of Engineering Science, Universiti Tunku Abdul Rahman, Kampar, Malaysia
e-mail: yanbing114@1utar.my

W.-S. Yap
e-mail: yapws@utar.edu.my

K.-C. Khor
e-mail: kckhor@utar.edu.my

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_4

Using such data to train detection models for IDS is difficult because the learning algorithms usually favour large classes for maximising accuracy and may have difficulties detecting the minority intrusions. Further, minority intrusions may not be able to form actual decision boundaries for the learning algorithms. Decision boundaries are important as they are the regions in a feature space that separates classes of a data set so that the learning algorithms can learn the classes effectively.

In this study, we attempted to improve the detection rates for minority intrusions by balancing the data set involved. The data set in this study used was CICIDS 2017 [1]. We firstly attempted under-sampling for the large class. Secondly, we attempted over-sampling, Synthetic Minority Over-sampling Technique (SMOTE) [2] for the seven weak intrusion classes that usually give weak detection rates. Finally, we combined both sampling techniques to seek better improvement in intrusion detection.

2 Literature Review

2.1 CICIDS2017 Data Set Overview

The CICIDS2017 data set [1] contains eight different files, and each of them contains network activities collected over five days. Table 1 shows the class distribution of the CICIDS2017 data set after combining the eight files. It comprises 2,830,743 instances, 78 features and 15 classes with no duplicated data. The data set is highly unbalanced as the BENIGN class takes 80.3% of the data set.

2.2 Sampling Techniques

Unbalanced class distribution is a common problem for real-world data sets such as network intrusions detection [1] and credit card fraud detection [3]. The rare classes are often the primary interests of classification [4]. Researchers have proposed several sampling techniques to tackle the unbalanced class distribution and improve classification performance, i.e., over-sampling, under-sampling, and combining sampling [5].

2.2.1 Over-Sampling

Over-sampling duplicates instances of minority classes or generates the duplicates based on the characteristic of the minority classes. This shall decrease the rareness of minority classes, thereby decreasing the overall level of class imbalance [4]. A basic over-sampling method is Random Over-Sampling (ROS) that duplicates the minority instances randomly [6]. Increasing the size of a minority class using ROS

Table 1 The class distribution of the CICIDS2017 data set

No	Normal/attack label	Number of instances	% of the total instances
1	BENIGN	2,273,097	80.3004
2	DoS hulk	231,073	8.1630
3	PortScan	158,930	5.6144
4	DDoS	128,027	4.5227
5	DoS goldeneye	10,293	0.3636
6	FTP-patator	7938	0.2804
7	SSH-patator	5897	0.2083
8	DoS slowloris	5796	0.2048
9	DoS slowhttptest	5499	0.1943
10	Bot	1966	0.0695
11	Web attack–Brute force	1507	0.0532
12	Web attack–XSS	652	0.0230
13	Infiltration	36	0.0013
14	Web attack–Sql injection	21	0.0007
15	Heartbleed	11	0.0004
	Total	2,830,743	100.0000

can increase the time taken to build a model and may lead to an overfitting problem [7]. Further, the lack of minority information may persist even after duplicating existing instances using ROS. Studies [7] show that ROS is less effective at improving the detection of minority classes. Therefore, Chawla et al. [2] proposed an advanced over-sampling method, Synthetic Minority Over-sampling Technique (SMOTE), to create new minority instances rather than duplicating existing instances. This technique creates synthetic instances using the nearest neighbour rule in the feature space. However, SMOTE considers only minority classes without taking care of majority classes. Therefore, increasing the size of minority classes may increase the chances of overlapping among classes [8].

2.2.2 Under-Sampling

Under-sampling removes the existing majority instances to balance a data set. A basic under-sampling technique is Random Under-Sampling (RUS) that removes the majority instances randomly. However, this may cause the removal of potentially useful information from a data set and the performance degradation in classification [4, 9].

2.2.3 Combining Sampling

Combining sampling is to apply a combination of sampling techniques on an unbalanced data set to improve the classification performance [10]. One example of combining sampling is to combine under-sampling and over-sampling. Das et al. [6] stated that under-sampling should be applied before over-sampling as a data cleaning method because it helps reduce the overlapping classes' effect.

3 Methodology

We transformed the CICIDS2017 data set into a format understandable by data mining algorithms used with data pre-processing. We replaced the missing values in the CICIDS2017 data set with the mean values of the features. Infinity values were then replaced by values that were ten times the maximum feature value. We also used Z-score normalisation to standardise all the features because the original range of their values is varied widely.

The unbalanced class distribution has caused the learning algorithms to bias majority classes and may produce low detection rates for minority classes. We used three sampling methods to address the problem, namely, over-sampling, under-sampling and hybrid sampling.

Four learning algorithms were used for intrusion detection, i.e., Gaussian Naïve Bayes (GNB) [11], C4.5 [11], Neural Network (NN-MLP) [12], K-Nearest Neighbour (KNN) [13], and Logistic Regression (LR) [14]. We used tenfold cross-validation to evaluate the performance of the learning algorithms. The data set was split into ten groups for both training and testing purposes.

The CICIDS2017 data set is unbalanced. Therefore, accuracy is a less suitable metric to evaluate learning algorithms. If the majority class is correctly classified, then the accuracy shall be high even though the rare classes are wrongly classified. Complementing accuracy with True Positive Rate (TPR) to examine learning algorithms' performance is a better option. This is because TPR can examine the detection performance for each of the classes in the data set.

4 Results and Discussion

Table 2 shows the detection result using the learning algorithms, i.e., Gaussian Naïve Bayes (GNB), C4.5, Neural Network (NN-MLP), K-Nearest Neighbour (KNN), and Logistics Regression (LR). By comparing the average TPR, C4.5 was the best performer among the single classifiers, with an accuracy (average TPR) of 0.9927. We noticed seven weak intrusion classes (bold classes in Table 2) that were hard to detect; below-average TPRs (less than 0.8.) were obtained using some of the learning algorithms. They were Bot, DoS Slowloris, Heartbleed, Infiltration, and

Table 2 The intrusion detection result for the full CICIDS 2017 data set using the learning algorithms. The classes in bold are the weak intrusion classes that give below average TPR

	GNB	C4.5	NN-MLP	KNN	LR
BENIGN	0.6500	0.9939	0.9955	0.9930	0.9709
Bot	0.9980	0.7872	0.3481	0.5607	0.0092
DDoS	0.9573	0.9996	0.9984	0.9977	0.9648
DoS goldeneye	0.9320	0.9173	0.9484	0.9610	0.8060
DoS hulk	0.7123	0.9898	0.9874	0.9874	0.9210
DoS slowhttptest	0.6767	0.9073	0.8440	0.8658	0.8056
DoS slowloris	0.6290	0.9154	0.8848	0.8681	0.4756
FTP-patator	0.9956	0.9961	0.9880	0.9948	0.5491
Heartbleed	0.8000	0.9000	0.0000	1.0000	0.0000
Infiltration	0.8417	0.7583	0.0000	0.2000	0.0750
PortScan	0.9885	0.9913	0.9809	0.9845	0.9939
SSH-patator	0.9944	0.9969	0.9646	0.9866	0.0270
WA–Brute force	0.0916	0.7306	0.1128	0.7658	0.0000
WA–Sql injection	1.0000	0.5667	0.0000	0.1500	0.0000
WA–XSS	0.9232	0.4125	0.0169	0.2990	0.0000
Average TPR	0.6907	**0.9927**	0.9922	0.9911	0.9614

three Web Attacks (WAs)—Brute Force, Sql Injection and XSS. Such performance could be caused by the unbalanced class distribution of the data set as the BENIGN is the immense majority in the data set. The learning algorithms' characteristic that favours the majority class (BENIGN) also contribute to such performance.

To improve the detection rate overall and for these weak intrusion classes, we attempted under-sampling, over-sampling and a combination of them to balance the data set.

Firstly, we attempted random under-sampling (RUS) on the majority class, BENIGN to balance the data set and reduce the effect of the majority BENIGN. Table 3 shows the RUS results using C4.5. C4.5 was used since it gave the best performance among the single classifiers, as shown in Table 1. The best overall accuracy (average TPR of 0.9985) was achieved by reducing BENIGN between 30 and 90% of its original size. The result was not much different from the full data set. However, the TPR for 12 of the classes were improved, including four of the weak intrusion classes.

We then attempted the over-sampling technique, Synthetic Minority Over-sampling (SMOTE), to increase the size of the seven weak intrusion classes. Table 4 shows the results of the over-sampling. The best average TPR (0.9900) was achieved by increasing the size of these minority classes to 250% of the full data set, and the result was slightly weak compared with the full data set. Improvements were noticed for some of the classes, including only three of the weak intrusion classes. Overall, the detection performance was slightly weak as compared to RUS.

Table 3 The intrusion detection result for the CICIDS 2017 data set resampled using RUS on BENIGN. The numbers in bold shows better TPRs than the results obtained using the full data set

Label	30%	40%	50%	60%	70%	80%	90%	Full data set
BENIGN	**0.9989**	**0.9990**	**0.9991**	**0.9991**	**0.9992**	**0.9992**	**0.9992**	0.9939
Bot	**0.8861**	**0.8861**	**0.8698**	**0.8596**	**0.8474**	**0.8210**	**0.8128**	0.7872
DDoS	**0.9998**	**0.9998**	**0.9998**	**0.9998**	**0.9998**	**0.9998**	**0.9998**	0.9996
DoS goldeneye	**0.9971**	**0.9965**	**0.9969**	**0.9963**	**0.9965**	**0.9963**	**0.9959**	0.9173
DoS hulk	**0.9995**	**0.9994**	**0.9992**	**0.9993**	**0.9992**	**0.9992**	**0.9992**	0.9898
DoS slowhttptest	**0.9891**	**0.9891**	**0.9862**	**0.9869**	**0.9898**	**0.9840**	**0.9876**	0.9073
DoS slowloris	**0.9959**	**0.9959**	**0.9962**	**0.9959**	**0.9959**	**0.9955**	**0.9962**	0.9154
FTP-Patator	**0.9992**	**0.9995**	**0.9992**	**0.9992**	**0.9992**	**0.9992**	**0.9992**	0.9961
Heartbleed	0.8000	0.8000	0.6000	0.6000	0.8000	0.8000	0.8000	0.9000
infiltration	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.7583
PortScan	**0.9978**	**0.9968**	**0.9961**	**0.9954**	**0.9948**	**0.9942**	**0.9931**	0.9913
SSH-Patator	**0.9983**	**0.9993**	**0.9986**	**0.9983**	**0.9980**	**0.9980**	**0.9980**	0.9969
WA–Brute force	0.7145	0.7092	0.7118	0.7145	0.7131	0.7118	0.7158	0.7306
WA–Sql Injection	**0.6000**	**0.6000**	0.1000	0.3000	0.2000	0.2000	0.5000	0.5667
WA–XSS	**0.4141**	0.4049	0.3988	0.4049	0.4049	**0.4141**	**0.4202**	0.4125
average TPR	**0.9985**	**0.9985**	**0.9985**	**0.9985**	**0.9985**	**0.9985**	**0.9984**	0.9926

Finally, we combined RUS and SMOTE to seek improvement in detection. Figure 1 shows the TPRs achieved using the combination of these two sampling techniques. The x-axis represents the percentage of the remaining majority class samples, BENIGN, after under-sampling. On the other hand, the y-axis represents the TPRs. There are four line-plots that represent the percentage of over-sampling on the seven minority classes. The best result achieved was 30% under-sampling on BENIGN and 300% over-sampling on the seven weak intrusion classes. The average TPR obtained was 0.9934.

Table 5 shows the result comparison of the full data set and the resampled data sets. The RUS (30%) achieved the best average TPR (0.9985) among the sampling techniques. Using RUS, we achieved the best TP rates for 11 out of 15 classes as compared with the full data set, and the data sets resulted using SMOTE and RUS (30%) + SMOTE (300%). To conclude, the sampling technique RUS gave a slight improvement in detection overall and most of the CICIDS 2017 data set classes.

Table 4 The results for the data set yielded using SMOTE on the five minority classes (bolded font). The numbers in bold shows better TPRs than the results obtained using the full data set

Label	100% (Full data set)	150%	200%	250%	300%
BENIGN	0.9939	0.9929	0.9929	0.9929	0.9926
Bot	0.7872	**0.7968**	**0.7968**	**0.8075**	**0.8121**
DDoS	0.9996	0.9994	0.9996	0.9994	0.9995
DoS goldeneye	0.9173	0.9110	0.9109	**0.9178**	0.9177
DoS hulk	0.9898	**0.9899**	**0.9901**	**0.9907**	**0.9907**
DoS slohttptest	0.9073	0.8564	0.8762	**0.9269**	**0.9254**
DoS slowloris	0.9154	0.8826	0.8824	0.8817	0.8948
FTP-patator	0.9961	0.9958	**0.9966**	**0.9972**	**0.9962**
Heartbleed	0.9000	0.8000	0.9000	0.8000	0.9000
Infiltration	0.7583	**0.8417**	**0.7833**	**0.7917**	**0.8500**
PortScan	0.9913	0.9912	0.9912	0.9912	0.9911
SSH-patator	0.9969	0.9969	0.9969	**0.9973**	**0.9971**
WA–Brute Force	0.7306	**0.7399**	**0.7472**	**0.7446**	**0.7339**
WA–Sql Injection	0.5667	0.5667	0.4167	**0.6333**	0.4167
WA–XSS	0.4125	0.3678	0.3649	0.3894	0.3990
Average TPR	0.9927	0.9899	0.9899	0.9900	0.9898

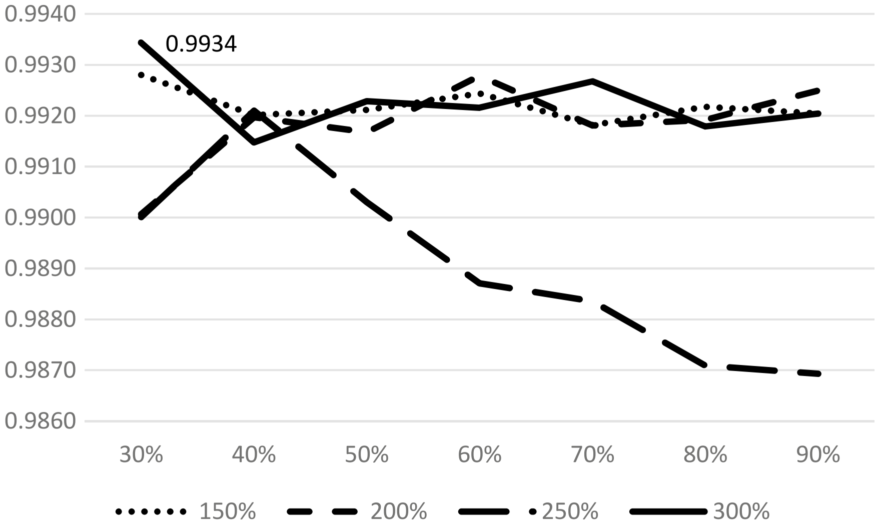


Fig. 1 The detection results achieved using the combination of under-sampling and oversampling. RUS (30%) + SMOTE (300%) gives the best result—a TPR of 0.9934

Table 5 The result comparison using the full and resampled data sets. The numbers in bold shows the best TPRs obtained by using RUS

	Full dataset	SMOTE	RUS (30%)	RUS (30%) + SMOTE (300%)
BENIGN	0.9939	0.9929	**0.9989**	0.9945
Bot	0.7872	0.7842	0.8861	0.8918
DDoS	0.9996	0.9995	**0.9998**	0.9994
DoS goldeneye	0.9173	0.9172	**0.9971**	0.9743
DoS hulk	0.9898	0.9896	**0.9995**	0.9880
DoS slowhttptest	0.9073	0.8520	**0.9891**	0.8735
DoS slowloris	0.9154	0.9173	**0.9959**	0.8577
FTP-patator	0.9961	0.9961	**0.9992**	0.9976
Heartbleed	0.9000	0.9000	0.8000	1.0000
Infiltration	0.7583	0.8667	0.6667	0.7500
PortScan	0.9913	0.9912	**0.9978**	0.9971
SSH-patator	0.9969	0.9969	**0.9983**	0.9968
WA–Brute Force	0.7306	0.7438	0.7145	0.7439
WA–Sql Injection	0.5667	0.5833	**0.6000**	0.4500
WA–XSS	0.4125	0.4141	**0.4141**	0.3834
Average TPR	0.9927	0.9918	**0.9985**	0.9934

5 Conclusion and Future Work

This paper aims to use sampling techniques to improve the intrusion detection rate using the CICIDS 2017 data set. We attempted under-sampling, over-sampling, and a hybrid of them to balance the data set, as the learning algorithms work by assuming data sets involved are balanced in class distribution. The RUS gave the best overall accuracy, measured using average TPR. The average TPR obtained was 0.9985, a slight improvement compared to the full and resampled data sets using SMOTE and RUS + SMOTE. We also noticed an improvement in detecting most of the classes, including some weak intrusion classes.

Acknowledgements This work was supported by the UTAR Research Grant (IPSR/RMC/UTARRF/2019-C2/K02) provided by the Universiti Tunku Abdul Rahman, Malaysia.

References

1. Sharafaldin I, Lashkari AH, Ghorbani AA (2018) Toward generating a new intrusion detection dataset and intrusion traffic characterisation. In: ICISSP 2018–Proc. 4th Int. Conf. Inf. Syst. Secur. Priv. 2018-Janua. pp 108–116

2. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
3. Kalid SN, Ng K, Tong G, Khor K (2020) A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes. IEEE Access 8:28210–28221
4. Weiss G (2004) Mining with rarity: a unifying framework. SIGKDD Explor 6:7–19
5. Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 5:221–232
6. Das B, Krishnan NC, Cook DJ (2014) Handling imbalanced and overlapping classes in smart environments prompting dataset
7. Drummond C, Holte RC (2003) C4.5, class imbalance, and cost sensitivity : why under-sampling beats over-sampling
8. Sáez JA, Luengo J, Stefanowski J, Herrera F (2014) Managing borderline and noisy examples in imbalanced classification by combining smote with ensemble filtering. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 8669 LNCS. pp 61–68
9. Dal Pozzolo A, Caelen O, Bontempi G (2010) Comparison of balancing techniques for unbalanced datasets. Mach Learn Gr Univ Libr Bruxelles Belgium 16:732–735
10. Seiffert C, Khoshgoftaar TM, Van Hulse J (2009) Hybrid sampling for imbalanced data. Integr Comput Aided Eng 16:193–210
11. Abdulrahman AA, Ibrahem MK (2019) Evaluation of DDoS attacks detection in a new intrusion dataset based on classification algorithms. Iraqi J Inform Commun Technol 1:49–55
12. Toupas P, Chamou D, Giannoutakis KM, Drosou A, Tzovaras D (2019) An intrusion detection system for multi-class classification based on deep neural networks. In: Proc.–18th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2019. pp 1253–1258
13. Yong Y (2012) The research of imbalanced data set of sample sampling method based on K-Means cluster and genetic algorithm. Energy Procedia 17:164–170
14. Zhang Y, Chen XU, Jin LEI, Wang X, Guo DA (2019) Network intrusion detection: based on deep hierarchical network and original flow data. IEEE Access 7:37004–37016

The Gaps of Identity Management in Fulfilling Personal Data Protection Regulations' Requirements and Research Opportunities

Drishty Rai Ratti and Hui Na Chua

Abstract The utilization of social media and online services affirms the benefits that are being relished by consumers. Organizations that provide the services require consumer personal data to be integrated into the service applications. Consequently, personal information is progressively made available over the internet which promotes the risk of the data being misused. Therefore, data privacy and security have taken prominence worldwide which resulted in data protection regulations being enforced around the globe. Technology is one of the main aspects in protecting data besides these regulations. There are various approaches when it comes to the technologies of protecting data particularly in protecting personal data through user profiles, data management, encryption, and access policies are among the main approaches. Identity Management (IdM) technologies are developed surrounding these main approaches as it represents the amalgamation of the authentication and authorization process that manages data associated to an individual's identity. Yet, there is a lack of investigation on how these identity-protecting technologies can complement the functional requirements of personal data protection regulations. A functional requirement in this research is denoted as a specification that may involve "calculations", "technical details", "data manipulation" and "processing" which a system is supposed to accomplish. Thus, we performed a study to identify gaps of IdM technologies in fulfilling the functional requirements to discover the potential opportunities for researchers to conduct further studies to bridge the identified gaps while providing a mapping for organizations to recognize the required IdM for compliance.

Keywords Identity management · Personal data protection · Information privacy · Information security · Systematic literature review methodology

D. R. Ratti · H. N. Chua (✉)
Department of Computing and Information Systems, Sunway University, Selangor, Malaysia
e-mail: huinac@sunway.edu.my

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_5

1 Introduction

The cascading utilization of social media and internet services affirm the benefits that are being relished by consumers. Organizations that provide the previously mentioned services require consumer personal data to be integrated into their day-to-day operations. As a result, personal information is progressively made available over the internet which promotes the risk of the data being lost or stolen [1]. Therefore, data privacy and security have taken prominence worldwide which resulted in data protection regulations being enforced in various countries around the globe. These regulations are implemented to ensure that personal data is protected from any malicious intent and allows citizens of their respective countries to obtain more control over their data.

Technology is one of the main aspects in protecting data besides these regulations to protect personal data [2]. There are various approaches when it comes to the technologies of protecting data. Identity Management (IdM) technologies are developed surrounding these main approaches, particularly in protecting personal data through user profiles, data management, encryption, and access policies [3]. IdM represents the amalgamation of the authentication and authorization process and technologies that implement this process [4] associated with the data management of an individual's identity. Nevertheless, there is a lack of investigation on how these identity-protecting technologies can fulfill the functional requirements of personal data protection regulations. A functional requirement in this research is labeled as a specification that may involve "calculations", "technical details", "data manipulation" and "processing" which a system is supposed to accomplish [5].

Thus, this research paper is aimed to first classify the regulations' functional requirements and then, identify the technological capabilities of IdM and how they can fulfill the identified functional requirements. Subsequently, the gaps between IdM and the functional requirements are identified as well. Providing that there is a gap between them would propose the opportunity for scholars to conduct further research to invent a technology or systematic approach that will be able to do so. The findings of this research project could be used as a reference to ensure that organizations utilize the appropriate identity management technologies according to the functional requirements. To meet the research aim, we propose the following questions:

RQ1: what are the functional requirements of the personal data protection regulations?

RQ2: what are the existing technological capabilities of IdM that can support the functional requirements?

RQ3: what are the gaps of IdM in fulfilling the regulations' functional requirements that lead to potential research opportunities?

Table 1 Comparison table of personal data protection regulations

Concepts	GDPR	PDPA	PA1988	FIA	PIPA
Applicable to whom	Organizations within/beyond the EU who provide products/services to customers/ organizations in the EU	Individuals processing personal data related to commercial transactions	Organizations and private sectors that have an annual turnover > $3 million	Federal executive branch agency records and government agencies	Organizations within/beyond Korea who provide products/services to customers/ organizations in Korea
Applicable to what	Personal data that reveals an individual's distinctive and confidential identity	Sensitive data/an opinion that defines an individual	Personal data/ opinion that explains an identified individual	Records from federal agencies that allow the public to access these records	Any form of data that easily defines and identifies an individual

2 Literature Review

2.1 *Personal Data Protection Regulations*

Personal data protection regulations inaugurate law mainly about data protection that accommodates rules legalizing the obtainment, usage, and revelation of personal data. These laws acknowledge an individual's right in protecting their data, the privilege of accessing and amending, and require of organizations to accumulate, utilize, and reveal the personal data which is only eligible for constitutional and authorized purposes.

Numerous personal data protection regulations are enforced around the globe such as the EU General Data Protection Regulations (EU GDPR), Australian Privacy Act 1988 (PA1989), Freedom of Information Act (FIA, USA), Personal Data Protection (PDPA, Malaysia), and the Personal Information Protection Act (PIPA, Korea). Table 1 provides a summarized comparison of each of the personal data protection regulations that are included in this paper.

2.2 *Identity Management (IdM)*

IdM is composed of three areas; (i) data security—protecting data from unauthorized access, (ii) provisioning—the ability to grant and manage access to a specific identity whilst preserving the confidentiality, integrity, and availability of the information, (iii) compliance—the acquiescence towards the policies that revolve around IdM [6].

IdM can perform certain functions such as administration, discovery, maintenance, enforcement of policies, management, the exchange of information, and primarily, authentication [7]. It is utilized for authenticating users, devices, and services and to grant or deny access to any form of information, application, or system. IdM has a substantial organizational domain that handles identifying cloud objects, entities and controlling access to resources [8]. Many organizations that utilize IdM began by conducting a risk assessment which is mainly to establish the demand for the controls of IdM to secure information appropriately [9]. This is to create and perpetuate an identity, to substantiate and authenticate an identity, to accord permits and authorities, to perform auditing and appraisal towards the process of identity management.

We consolidated IdM capabilities based on prior studies' interpretation as follows:

(1) Identity Lifecycle Management. This capability enables the entire data process to be transparent to the data subject, which provides log functionalities on the user's side [9, 10] and ensures that the entire data process is transparent to the data subject [11]. Besides, it also creates identity-based on certain conditions and keeping it up to date and is deleted once the conditions are no longer met, and under no circumstance [12] will it be resurrected or recycled.

(2) Consent Management. These technologies enable the data subject to provide their consent for the controller to collect, process, or transfer their data [13].

(3) Access Control Management. Authentication and authorization go hand in hand in which both make up a portion of Access Control Management [12]. Authentication is used for validating the identity of the user. As soon as the authentication completes, the authorization process begins by determining what exactly the user can access. These technologies enable the data subject to provide their authorization for the controller to collect, process, or transfer their data [13] and functions that enable users to exercise their rights [9]. These are used to create, manage, disseminate, implement, store, and invalidate or retract digital certificates and public keys [14].

(4) Policy Management. Policy management is composed of identity policy, provisioning policy, and password policy because these capabilities define the policies of the accounts and access that are assigned to each user [15]. This capability is considered the undertaking of creating, transferring, and maintaining policies that have been established by the authorities for the organization to comply with [16].

3 Research Methodology

In this research, a systematic literature review methodology was chosen. This method is used to synthesize research discovery in a systematic, explicit, and reliable way. It is a process of identifying, analyzing, and interpreting available research that is relevant to a research question or area of interest and identifying gaps in a certain research area to later provide suggestion areas that require further investigation [17].

There were two data sources collected from the identity management journals and the personal data protection regulations. Mapping was carried out between the identity management capabilities and the personal data protection regulation

Table 2 Inclusion and exclusion criteria for identity management journals

Inclusion criteria	Available as full text and relevant to the research questions Within the domain of Identity Management when explaining (in detail): definitions, theories, capabilities or functionalities, evolution, different types of models, approach or concepts, trends, stakeholders involved
Exclusion criteria	Not available in full text or not relevant to the research question Not within the computing definition of Identity Management Not within the scope of Identity Management

requirements; hence it was important to identify the mapping criteria that are implemented.

3.1 Data Collection

3.1.1 Search Strategy, Inclusion, and Exclusion Criteria

Identity Management Journals. Google Scholar is selected as the main domain of collecting the journals within the past 15 years (i.e., 2004–2019) as it is the provider of relevant information concerning the research question of this study. Predefined keywords are used to narrow the scope of the search. Due to the term 'identity management' being too broad, more specific keywords were needed to extract the relevant journals. Table 2 presents the inclusion and exclusion criteria.

Personal Data Protection Regulations. The personal data protection regulations selected for the analysis are originated from each continent around the globe. These regulations are the Australian Privacy Act (PA1988), US Freedom of Information Act (FOIA), EU General Data Protection Regulations (GDPR), Malaysian Personal Data Protection Act (PDPA), and Korean Personal Information Protection Act (PIPA). The inclusion and exclusion criteria used are as follows:

- Inclusion: articles that contain the requirements that can be supported by IdM including automated or systematically operated functions, processing or handling of data, and using appropriate safety measures.
- Exclusion: articles that contain the requirements that cannot be supported by IdM technology including penalties and compliance or relevance to other regulations.

3.1.2 Study Selection and Data Aggregation Processes

Personal Data Protection Regulations. After selecting the functional requirements that are relevant to the research questions based on the inclusion and exclusion criteria, they were further aggregated. The selected functional requirements are aggregated in a way when the regulations offer similar meaning and provide identical requirements. This step is crucial as it aids in avoiding any duplicates or repeated functional requirements which will help in preventing redundancy when the analysis is conducted. Figure 1 shows the process flow concerning how the identity management capabilities are handled before conducting the mapping process.

Identity Management Journals. Figure 2 provides a clearer visualization of the process in this study selection. Google Scholar produced 325 research papers and journals. Out of all the collected journals, 5 were removed due to the title not being related to the research questions of this research. 8 more papers out of the remaining 320 papers were removed due to the irrelevance of its abstract to the area of study of this research. The 312 journals left to undergo the inclusion and exclusion criteria which were pre-written to narrow the search. During this process, 40 journals were

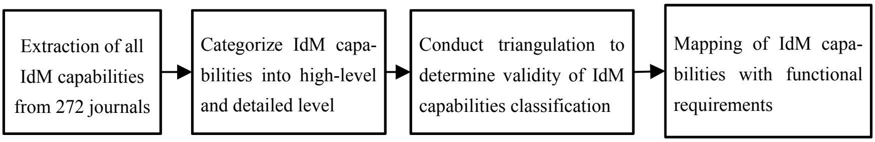

Extraction of all IdM capabilities from 272 journals
Categorize IdM capabilities into high-level and detailed level
Conduct triangulation to determine validity of IdM capabilities classification
Mapping of IdM capabilities with functional requirements

Fig. 1 Process flow of the handling of the extracted identity management capabilities

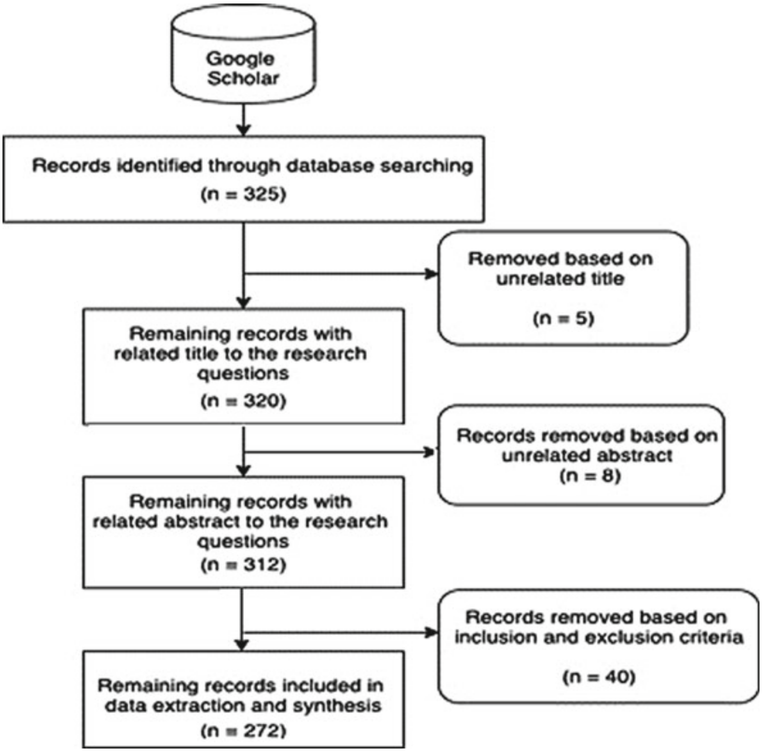

Google Scholar
Records identified through database searching (n = 325)
Removed based on unrelated title (n = 5)
Remaining records with related title to the research questions (n = 320)
Records removed based on unrelated abstract (n = 8)
Remaining records with related abstract to the research questions (n = 312)
Records removed based on inclusion and exclusion criteria (n = 40)
Remaining records included in data extraction and synthesis (n = 272)

Fig. 2 Study selection process for the identity management journals

excluded. Finally, the remaining 272 journals that fulfilled the inclusion criteria were used in the data extraction and synthesis of this research.

3.2 Mapping Identity Management Capabilities with Functional Requirements Criteria

A predefined list of criteria was set to map the identity management capabilities and the functional requirements. Keywords within the technological requirement are important and finding similar keywords within the IdM capabilities is one way to map the two. Many functional requirements can fall under the same IdM capability if deemed appropriate e. g., some articles within one section represent the same meaning so it is possible to map them to one identity management capability. A functional requirement may have multiple IdM capabilities mapped to it.

3.3 Reliability and Validity

To ensure the validity and reliability of the findings of this research paper, two types of reliability tests were carried out. Cohen's κ was run to determine if there was an agreement between two researchers' judgment on whether the 325 articles are qualified according to the selection criteria for analysis. The definition of κ can be found below [18], where p_o is the relative observed agreement among researchers, and p_e is the hypothetical probability of chance agreement:

$$k = \frac{\rho_o - \rho_\rceil}{1 - \rho_\rceil} \tag{1}$$

The result of this reliability test is $\kappa = 0.837$ (95% CI, 0.300 to 0.886), $p < 0.0005$. This shows that there was a strong agreement with 95% confidence interval between the two researchers' judgments, along with indicating the result is statistically significant due to the p-value <0.0005.

Another test implemented was triangulation. This was conducted during numerous occasions which are the establishment of the high and detailed level of the identity management capabilities, the aggregation of the functional requirements, and finally, the mapping of the capabilities of identity management to the functional requirements.

4 Results

4.1 What are the Functional Requirements of the Personal Data Protection Regulations?

We identified and consolidated functional requirements that can be fulfilled for data processing and controlling [5]. As shown in Table 3, GDPR consists of all the functional requirements that have been listed, closely followed by PIPA regulations.

4.2 What are the Existing Technological Capabilities of IdM that Can Support the Functional Requirements?

According to Table 4, We observed the “Data Protection Impact Assessment” is not fulfilled by any existing IdM capability. Although most of the requirements are mapped, certain underlying gaps remain which the mapped identity management technologies are unable to fully fulfill. We further discuss the gaps in Sect. 4.3.

4.3 What are the Gaps of IdM in Fulfilling the Regulations’ Functional Requirements that Lead to Potential Research Opportunities?

Based on the mapping results presented in Table 4, further investigation was conducted to identify the implicit gaps as follows:

Processing of Personal Data. The compliance process assessment system is missing. The regulation requirements specify that it should be the responsibility of organizations to have a standardization to predetermine the quality of the technology used to process data. However, there is no mechanism to assess the quality of the technology used.

Research opportunity. A metric system and mechanisms for assessing the compliance flow are needed which may be as part of an Identity Life Cycle Management system. Organizations can use this assessment system to benchmark their compliance from the aspect of technology used to process personal data.

Purpose of Data Collection. Missing a system that can perform classification to categorize different types of personal data. Based on the technologies that have been mapped onto this requirement, it is uncertain to identify what is considered as “adequate”, “relevant” and “necessary” personal data as there are various forms of existing personal data.

Research opportunity. A classification system for categorizing different personal data types is required. The classification enables greater transparency for individuals

Table 3 Functional requirements of personal data protection regulations

Functional requirements	GDPR	PDPA	PA1988	FIA	PIPA
Processing of Data: Processed in a way that ensures appropriate security of personal data and is processed lawfully, fairly, transparently, encrypted pseudonymized, confidentiality, integrity, availability, and resilience when processing	√	√	√	√	√
Purpose of Data Collection: Should only be for legitimate purposes, relevant, limit to necessity, accuracy, stored for no longer than necessary	√	√	√	√	√
Consent: The data subject is provided with portability rights, restriction rights, objection rights, erasure rights, access rights, and rectification rights	√	√		√	√
Data Shall Be Processed for Special Categories of Data: Protect vital interests of data subjects, to perform a task carried out in public interest, for the legitimate interests pursued by the controller, and for compliance with legal obligations	√				√
Information to Be Provided to Data Subject: Should be concise, transparent, and easily accessible. The information provided should be the contact details of the controller and the data protection officer, purpose of processing, and a copy of the personal data undergoing processing	√	√	√		
Data Protection Impact Assessment: Must be carried out and a consultation with the supervisory authority is necessary if the assessment results reveal a high risk in the absence of measures taken by the controller	√				√
Data Protection Officer: Needed when processing is carried out by public authority, when requires regular systematic monitoring of data subject on a large scale, to advice controllers on the regulations when processing, cooperate with the supervisory authority, act as the contact point for the supervisory authority	√	√			√
Transferring of Personal Data: Transferring to a third party should acquire appropriate safeguards or a legally binding and enforceable instrument between public authorities and a standard data protection clause adopted by the Commission	√		√		√
Codes of Conduct and Corporate Rules: Supervisory authorities should formulate codes of conduct intended for the proper application of the regulations and corporate rules should be applied, enforced, and complied with by those involved with the data processing stage	√	√			

(continued)

Table 3 (continued)

Functional requirements	GDPR	PDPA	PA1988	FIA	PIPA
Obligations of Secrecy: Member State shall set more specific rules that deemed necessary towards the authority of the supervisory authorities, controllers, and processors with the obligation of secrecy	√				

Table 4 Mapping of identity management capabilities with the functional requirements

IdM capabilities Functional requirements	Identity lifecycle management [9–12]	Consent management [13]	Access control management [9, 12–14]	Policy management [15, 16]
Processing of personal data	√	√	√	√
Purpose of data collection	√			
Consent	√	√	√	
Processed for special categories of data		√		
Information provided to data subject	√		√	
Data protection impact assessment				
Data protection officer			√	√
Transferring of personal data			√	
Codes of conduct and corporate rules				√
Obligations of secrecy			√	

to understand what category of their data is relevant for the service provided; this could allow them to exercise their right to choose not to disclose irrelevant data categories instead of being forced to provide unnecessary data.

Consent. Lack of effective technologies for ensuring awareness. In the requirement, it is stated that individual (i.e., data subject) has the right to restrict, rectify, object the processing of their data or withdraw their consent. However, for the data subject to carry these out, they must first be made aware of their rights as a data subject. The technologies that have been mapped to the requirement do not play a role in allowing the data subject to familiarize their rights. Although the regulations require organizations to provide notice about their practice in protecting personal data, the common approach for organizations to tackle the requirement is realized through the publishing of online privacy policy which does not guarantee the awareness of their rights [19].

Research opportunity. Technologies for prompting or increasing awareness of data rights are needed. These technologies can be embedded within web browsers' or application functions that collect personal data, to alert users of their data rights.

Data Protection Impact Assessment (DPIA). Based on the requirements, it is mentioned that the organization is required to conduct DPIA for data processing that is likely to result in a high risk. However, there is a lack of a risk assessment metric system embedded in IdM to detect data that is "likely to result in high risk".

Research opportunity. An automated risk assessment system that can estimate the likeliness of the data processing to be of "high risk" is required. Risk and impact analysis on different types of personal data is necessary for developing this risk assessment system. With this assessment system, organizations would be able to evaluate their data processing to enhance the level of security on personal data. Access control to different personal data types can be imposed with different restriction levels.

5 Conclusion

The area of interest that is addressed lies in the gaps that have been identified through our findings. These identified gaps establish the potential research opportunities to further be exploited by researchers, additionally, providing a mapping for organizations to identify the required identity management technologies to remain compliant with the regulations. This research paper may also contribute towards organizations in which the analysis may aid in developing identity management technologies that are more personal data protection regulation centric.

Acknowledgments Funding: This research was supported by the Malaysian government FRGS grant [FRGS/1/2019/ICT04/SYUC/02/2].

References

1. Statista (2014) Statista. The statistics portal. https://www.statista.com/. Accessed 18 June 2020
2. Sim WL, Chua HN and Tahir M (2019) Blockchain for identity management: The implications to personal data protection. In 2019 IEEE Conference on Application, Information and Network Security (AINS). IEEE. pp. 30–35
3. Blazic AJ, Dolinar K and Porekar J (2007) Enabling privacy in pervasive computing using fusion of privacy negotiation, identity management and trust management techniques. In First International Conference on the Digital Society (ICDS'07). IEEE. pp. 30–30

4. Yang Y, Chen X, Wang G and Cao L (2014) An identity and access management architecture in cloud. In 2014 Seventh International Symposium on Computational Intelligence and Design (Vol. 2):200--203
5. Lightsey B (2001) Systems engineering fundamentals. DEFENSE ACQUISITION UNIV FT BELVOIR VA
6. Kunz M, Hummer M, Fuchs L, Netter M and Pernul G (2014) Analyzing recent trends in enterprise identity management. In 2014 25th international workshop on database and expert systems applications. IEEE. pp. 273–277
7. Balasubramaniam S, Lewis GA, Morris E, Simanta S and Smith DB (2009) Identity management and its impact on federation in a system-of-systems context. In 2009 3rd annual IEEE systems conference. IEEE. pp. 179–182
8. Chehab MI and Abdallah AE (2010) Assurance in identity management systems. In 2010 Sixth International Conference on Information Assurance and Security. IEEE. pp. 216–221
9. Dhungana RD, Mohammad A, Sharma A and Schoen I (2013) Identity management framework for cloud networking infrastructure. In 2013 9th International Conference on Innovations in Information Technology (IIT). IEEE. pp 13–17
10. Kumar V and Bhardwaj A (2018) Identity management systems: A Comparative Analysis. Int J Strat Decis Sci (IJSDS) 9(1):63–78
11. Zhao G, Hu X, Li Y and Du L (2009) Implementation and testing of an identity-based authentication system. In 2009 ISECS International Colloquium on Computing, Communication, Control, and Management (Vol. 4). IEEE. pp. 424–427
12. Bhardwaj A and Kumar V (2011) Cloud security assessment and identity management. In 14th International Conference on Computer and Information Technology (ICCIT 2011). IEEE. pp. 387–392
13. Galpin R and Flowerday SV (2011) Online social networks: Enhancing user trust through effective controls and identity management. In 2011 Information Security for South Africa. IEEE. pp. 1–8
14. Lee H, Jeun I, Jung H (2009) Criteria for evaluating the privacy protection level of identity management services. In 2009 Third International Conference on Emerging Security Information, Systems and Technologies. IEEE. pp. 155–160
15. Thakur MA and Gaikwad R (2015) User identity & lifecycle management using LDAP directory server on distributed network. In 2015 International Conference on Pervasive Computing (ICPC). IEEE. pp. 1–3
16. Dsouza, Clinton, Gail-Joon Ahn, and Marthony Taguinod. "Policy-driven security management for fog computing: Preliminary framework and a case study." In Proceedings of the 2014 IEEE 15th international conference on information reuse and integration (IEEE IRI 2014), pp. 16-23. IEEE, 2014.
17. Torres-Carrión PV, González-González CS, Aciar S and Rodríguez-Morales G (2018) Methodology for systematic literature review applied to engineering and education. In 2018 IEEE Global engineering education conference (EDUCON). IEEE. pp. 1364–1373
18. Umesh UN, Peterson RA and Sauber MH (1989) Interjudge agreement and the maximum value of kappa. Educ Psychol Measure 49(4):835–850
19. Chua HN, Herbland A, Wong SF and Chang Y (2017) Compliance to personal data protection principles: A study of how organizations frame privacy policy notices. Telematics Inform 34(4):157–170

Verifying MQV-Based Protocols Using ProVerif

Ernest-YongYi Yap, Ji-Jian Chin, and Alwyn Goh

Abstract ProVerif is an automatic protocol verifier that is usually used to find symbolic attacks in a protocol as described in the Dolev-Yao Security Model [7]. But according to its manual [2], it can also be used to verify some computation attacks such as those described in the Bellare-Rogaway (BR) or Canetti-Krawczyk (CK) Security Model [5]. This cryptographic tool does not recognize the laws of mathematics and the laws needed to be applied manually. This paper shows the security verification of authenticated MQV-based key exchange (AKE) protocols. We show the proof of correctness using this protocol verifier tool as well as some of the known computational attacks done by others such as Unknown-Key-Share attack using it. Included in our results are two MQV-based protocol variants: an identity based key agreement (FG IB-KA) and a certificateless identity authenticated based key agreement (CLAKA).

Keywords ProVerif · Protocol · MQV · IBKA · CLAKA · UKS · KCI

This manuscript was written during the second author's visit to Information Security Lab, MIMOS Berhad. The authors appreciate the financial assistance from the Ministry of Education of Malaysia in supporting this work with the Fundamental Research Grant Scheme (FRGS/1/2019/ICT04/MMU/02/5). The authors would also like to thank Jason Chia for assisting in discussions on cryptography proof techniques.

E.-Y. Yap (✉) · J.-J. Chin
Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia

J.-J. Chin · A. Goh
Information Security Lab, MIMOS Berhad, Cyberjaya, Malaysia

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_6

1 Introduction

Cryptographic protocols have existed in the computing world for a long time and methods to create one have been proposed throughout the years. To start an encrypted conversation between two parties, they have to first decide on a secret key for encrypting their messages. The Diffie-Hellman (DH) key exchange [6] is one of the methods to allow two parties to share a symmetric key, but since the DH key exchange does not have entity authentication, it can be broken easily with a man-in-the-middle attack. Since then, cryptographers have modified the DH key exchange with additional entity authentication properties using methods such as digital signatures and certificates [3, 4, 13]. More secure and efficient mathematical methods such as the elliptic curves have been proposed as well.

In this paper, the automatic security analyzer tool for cryptographic protocol used is ProVerif [1]. This cryptographic protocol verifier is designed based on the Dolev-Yao model which means that it is primarily used to detect symbolic attacks [15–17]. It supports both symmetric and asymmetric cryptographic protocols, hash functions, digital signatures and key exchanges based on DH mechanics. It can also allow multiple sessions of the protocol to be run at once and providing an unlimited message capacity. Sometimes ProVerif may results in a false positive attack, but if some property is stated to be satisfied, then the property is confirmed to be satisfied. Some of the properties that ProVerif can verify is the secrecy of a message, entity authentication and strong secrecy, which means the adversary can't detect the value change of a secret.

To the best of our knowledge, it seems that there is no existing work that provides a ProVerif verification for MQV-based protocols. Thus, we fill in that research gap with these results. We use ProVerif to demonstrate the security properties of the main MQV protocol. We also demonstrate the Unknown Key Share (UKS) attack on MQV using it. As corollary results, we utilize ProVerif to verify the security of two related protocols: and ID-Based Key Agreement (FG IB-KA) protocol by Fiore and Gennaro [9] which is of an identity-based construction, as well as the Certificateless Authenticated Key Agreement (CLAKA) Protocol by He et al. [11]that uses a certificateless construction.

2 Protocols

2.1 MQV Protocol

Security Feature. The MQV protocol [14] is considered one of the most efficient DH-based protocols that uses public key cryptography to provide entity authentication. This protocol does not require a third-party key provider such as a KGC and it only allows two parties in a session. The protocol designers specifically designed this protocol to resist key compromise impersonation (KCI) attacks, known key attacks

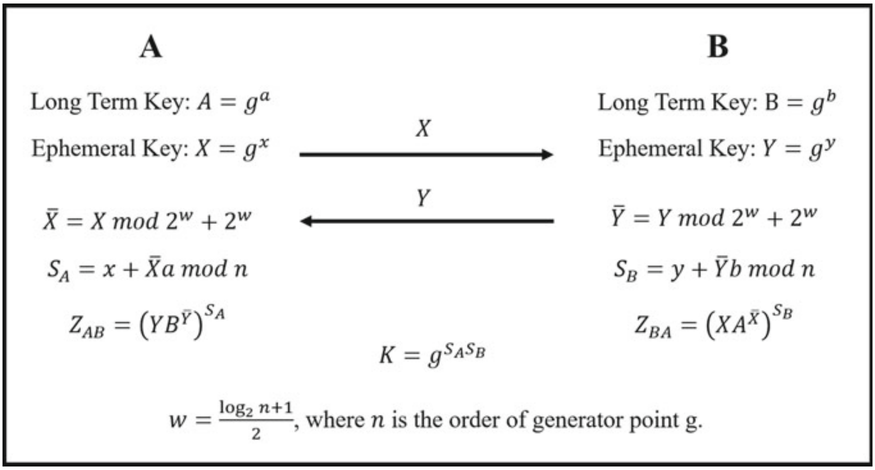

A
B
Long Term Key: $A = g^a$
Long Term Key: $B = g^b$
Ephemeral Key: $X = g^x$
Ephemeral Key: $Y = g^y$
X
Y
$\bar{X} = X \bmod 2^w + 2^w$
$\bar{Y} = Y \bmod 2^w + 2^w$
$S_A = x + \bar{X}a \bmod n$
$S_B = y + \bar{Y}b \bmod n$
$Z_{AB} = \left(YB^{\bar{Y}}\right)^{S_A}$
$Z_{BA} = \left(XA^{\bar{X}}\right)^{S_B}$
$K = g^{S_A S_B}$
$w = \frac{\log_2 n+1}{2}$, where n is the order of generator point g.

Fig. 1 MQV protocol

and provide perfect forward secrecy which means the adversary can't obtain every used session key despite having the long-term key. This protocol is proven to be secure in the BR model. However, an UKS attack was discovered by Kaliski [12] when the entity authentication is done implicitly.

Protocol Outline. The first MQV protocol proposed only uses implicit entity authentication, which mean both users only exchange the ephemeral key. Both users create their own long-term keys and derive a partial private key S_i as shown in Fig. 1, The shared key is then calculated using each other's long-term public key and self partial private key forming K.

2.2 FG IB-KA Protocol

Security Feature. The Fiore-Gennaro ID-Based Key Agreement (FG IB-KA) protocol [9] is an identity-based protocol derived from the MQV protocol. Identity-based cryptography is a method to remove certification of public keys by allowing principals to compute the public key of another principal based on the identity's information. The FG IB-KA protocol is modelled under the CK model which shows that the adversary can't distinguish between an actual session key and a random generated key with the same length. This protocol also provides forward secrecy, but it is considered weak forward secrecy which means the past used session keys are all safe but not the future ones [10]. Besides that, it resists most symbolic and computational attacks such as reflection attacks, KCI attacks and impersonation attacks.

Protocol Outline. Unlike the original MQV protocol, the FG IB-KA uses explicit entity authentication which means that the user identity is also sent by each of the

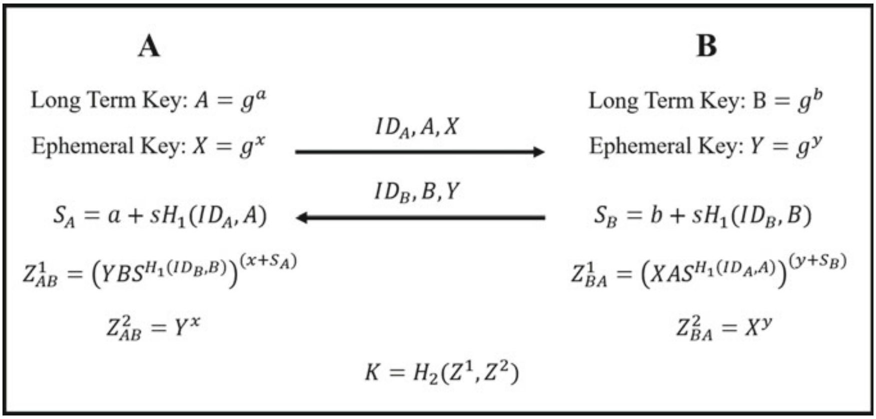


Fig. 2 FG IB-KA protocol

user in a session with it's long-term and ephemeral key. A Key Generation Center (KGC) takes in an user identity and derive a private key using Schnorr's Signature and also provide longterm public keys to both users. Using those KGC keys, Z^1 and Z^2 is calculated and the share key K is formed as shown in Fig. 2.

2.3 He et al. CLAKA Protocol

Security Feature. The Certificateless Authenticated Key Agreement (CLAKA) Protocol also known as the He-Padhye-Chen Protocol [11] is the certificateless version of the MQV-based protocol, where there is no key escrow. The CLAKA protocol is proved to be secure in the eCK model under the Gap-Diffie-Hellman (GDH) assumption. If the GDH problem has been broken with negligible probability, the advantage of the adversary in this protocol is said to be still negligible. In the security analysis done by Farouk [8], it is proven that this protocol is secure against an eCK model adversary, which means it resist known key attacks, KCI, UKS and provides forward secrecy.

Protocol Outline. The He-Padhye-Chen CLAKA uses elliptic curve cryptography, but it is converted to Diffie-Hellman notation for easier understanding. Since this is a certificate-less protocol, it does not rely completely on KGC but it has some similarity to FG IB-KA. Different from FG IB-KA, the KGC in CLAKA provides a random key and a similar Schnorr's Signature as private key to both principal. Both users generate their own long-term key and will exchange their identity, random key and ephemeral key. The long-term key is used to calculate Z^2, Z^1 and Z^3 is calculated in a similar way with FG IB-KA. At last, the share key is K as shown in Fig. 3.

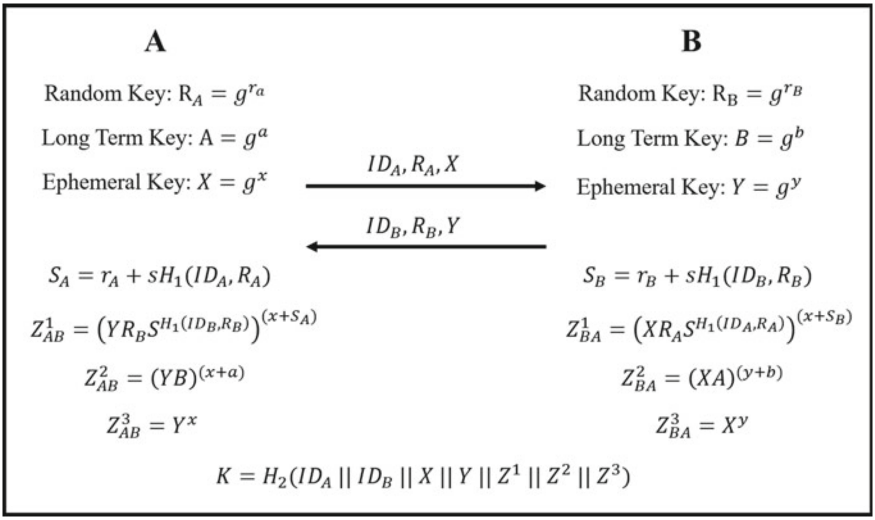

A
B
Random Key: $R_A = g^{r_a}$
Random Key: $R_B = g^{r_B}$
Long Term Key: $A = g^a$
Long Term Key: $B = g^b$
Ephemeral Key: $X = g^x$
ID_A, R_A, X
Ephemeral Key: $Y = g^y$
ID_B, R_B, Y
$S_A = r_A + sH_1(ID_A, R_A)$
$S_B = r_B + sH_1(ID_B, R_B)$
$Z^1_{AB} = \left(YR_BS^{H_1(ID_B,R_B)}\right)^{(x+S_A)}$
$Z^1_{BA} = \left(XR_AS^{H_1(ID_A,R_A)}\right)^{(x+S_B)}$
$Z^2_{AB} = (YB)^{(x+a)}$
$Z^2_{BA} = (XA)^{(y+b)}$
$Z^3_{AB} = Y^x$
$Z^3_{BA} = X^y$
$K = H_2(ID_A \,||\, ID_B \,||\, X \,||\, Y \,||\, Z^1 \,||\, Z^2 \,||\, Z^3)$

Fig. 3 CLAKA protocol

3 ProVerif

Since ProVerif does not recognize mathematical properties, the programmer needs to define the properties such as commutative, associative and distributive using the equation and reduc syntax. ProVerif is mostly used in protocols that do not have complicated mathematical expressions such as pairings, for example the Needham-Schroeder Public Key protocol [2]. The ProVerif uses queries to detect the vulnerabilities of the protocol. Each query can be programmed differently to simulate different types of attacks. A false query usually means that an attack is detected.

Proof of Correctness. To prove that the protocol works as intended, a proof of correctness is needed. Alice and Bob start a session with each other and exchange the secret key, Alice and Bob will then send each other a message encrypted using the secret key they exchanged. When Alice or Bob received the message, they decrypt it with their own key and check whether the message holds. If the message holds, the event will be executed, and the query will be false as the adversary only acts as a wire and attacks passively.

Secrecy of Messages. The secrecy of the message is defined as the adversary can't obtain the secret message that Alice sends to Bob or vice versa. The secrecy of a message is the most fundamental security property of a protocol: if an adversary can obtain the secret message of a session easily without the compromise of any keys, it means that the protocol is vulnerable to other attacks such as man-in-the-middle attacks or replay attacks. ProVerif can easily validate this security property using $query\ attacker(secretMessage)$.

KCI. A KCI attack is when the adversary possesses the long-term key of Alice or Bob, it can impersonate as the intended principal of Alice or Bob. To detect KCI attack in ProVerif, the long-term private key of Alice or Bob will be leaked out to the adversary via a public channel. The initiator will try to send out a secret message then execute the $eventSend(A, B, M)$. When the responder receives the message she will execute the event $Recv(A, B, M)$. If the responder executed $Recv(A, B, M)$ without the event $Send(A, B, M)$ executed, KCI attack is successful. The declared query for KCI attack is $event(Recv(A, B, M)) ==> event(Send(A, B, M))$.

Implicit Entity Authentication. Entity authentication is where one principal knows the identity of another principal that is in the same session, and is used to avoid MITM attacks or impersonation attacks. In ProVerif, Alice will execute $acceptsA(A, B, M, K)$ where A is her own public key, B is Bob's public key, M is the message she sent out and K is the shared key. When Bob received Alice's message, he will also compute the shared key and end the protocol by executing $termB(A, B, M, K)$. The declare query is $event(termB(A, B, M, K)) ==> event(acceptsA(A, B, M, K))$.

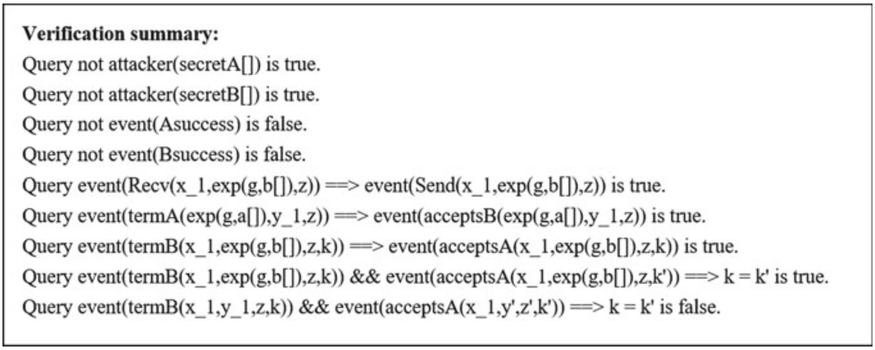
Key Indistinguishability. To fit in the model of BR or CK, an adversary must not have the ability to distinguish a real session key from a random key with negligible probability. ProVerif shows an example for key indistinguishability query in it's manual. This query shows the secrecy of the keys established by Alice when it starts a session with an honest principal Bob, with the sense that these keys are indistinguishable from independent random numbers.

UKS. UKS attack is an attack that allows an adversary to cause one principal to believe it is sharing a key with the adversary, but the principal actually shares it with another different principal that is not the adversary. In ProVerif, the query is $event(termB(A, B, M, K)\&\& \, event(acceptsA(A, B', M', K')) ==> K = K'$. This query allows the adversary to send different messages to Alice and Bob whereby the message can be different, but the key must the same.

4 Results and Discussions

Since these are quite complicated protocols, ProVerif requires some time to validate all queries. The more secure the protocol is, the longer time it takes for ProVerif to verify it. The time used to verify all the queries in every protocol is shown in Fig. 4. The specification of the computer used to verify these computer is i5-4460 core processor, 16GB RAM and operating system is Windows 10 Home with a solid state drive. All the queries in MQV finish processing under 1.5 h; FG IB-KA took under 10 h; and CLAKA completed under 47 h because of the long hash function in the end.

Although all the protocols have the same output from ProVerif, the trace graphs are different for every protocol. The adversary is passive in $event(Asuccess)$ and



```
Verification summary:
Query not attacker(secretA[]) is true.
Query not attacker(secretB[]) is true.
Query not event(Asuccess) is false.
Query not event(Bsuccess) is false.
Query event(Recv(x_1,exp(g,b[]),z)) ==> event(Send(x_1,exp(g,b[]),z)) is true.
Query event(termA(exp(g,a[]),y_1,z)) ==> event(acceptsB(exp(g,a[]),y_1,z)) is true.
Query event(termB(x_1,exp(g,b[]),z,k)) ==> event(acceptsA(x_1,exp(g,b[]),z,k)) is true.
Query event(termB(x_1,exp(g,b[]),z,k)) && event(acceptsA(x_1,exp(g,b[]),z,k')) ==> k = k' is true.
Query event(termB(x_1,y_1,z,k)) && event(acceptsA(x_1,y',z',k')) ==> k = k' is false.
```

Fig. 4 ProVerif result

$event(Bsuccess)$ as it is acting as a wire giving the proof of correctness in the protocols. Secrecy, implicit entity authentication, KCI resistance and key indistinguishability are all output true for all three protocols.

However, the UKS query is outputted as false on every protocol. Since ProVerif has the possibility of giving a false attack, the trace graph of the attack can be checked to verify the attack. The trace graph will be shown below in Fig. 5. The trace graph indeed shows the proof of correctness and the adversary is simply acting as a wire. For Fig. 5, an UKS attack is found on MQV as Alice thinks that she shares her key with the adversary, but Bob thinks that he is sharing his key with Alice, hence this is a positive attack. The UKS attack shown in FG IB-KA and CLAKA protocol is a false attack because both Alice and Bob is sharing the key with each other instead of the adversary, hence the UKS attack does not hold. This paper only shows one of the graph for explanation purpose, the rest of the trace graphs and codes can be found on GitHub on https://github.com/ernestyyy0306/ProVerif-MQV-Based.

5 Conclusion

In this work, the security properties of MQV, FG IB-KA and CLAKA protocol is verified using an automatic cryptographic verifier tools called ProVerif. This tool successfully detected an UKS attack on MQV protocol but gives a false UKS attacks on FG IB-KA and CLAKA protocol. These three protocols are proven to be secure in term of secrecy, key indistinguishability, KCI resistance and most symbolic attacks such as replay attacks and MITM attacks. This work shows that ProVerif is not only able to show symbolic attack as described in Dolev-Yao Model, but is also able to verify computational attacks such as KCI and UKS as described in BR or CK model. Besides, it also proves that complicated protocols that uses complicated mathematical properties such as point addition and point multiplication can be verified in ProVerif.

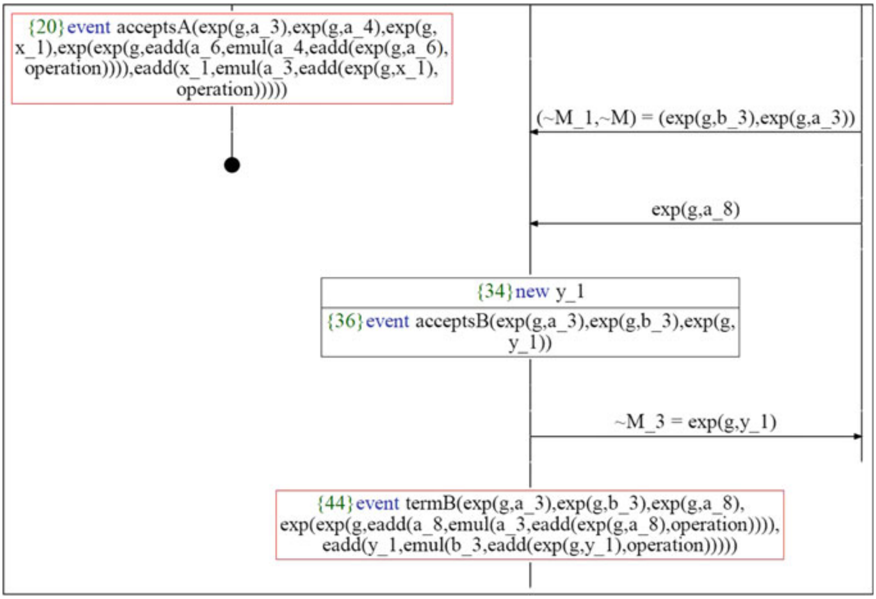

{20}event acceptsA(exp(g,a_3),exp(g,a_4),exp(g,x_1),exp(exp(g,eadd(a_6,emul(a_4,eadd(exp(g,a_6),operation)))),eadd(x_1,emul(a_3,eadd(exp(g,x_1),operation)))))
(~M_1,~M) = (exp(g,b_3),exp(g,a_3))
exp(g,a_8)
{34}new y_1
{36}event acceptsB(exp(g,a_3),exp(g,b_3),exp(g,y_1))
~M_3 = exp(g,y_1)
{44}event termB(exp(g,a_3),exp(g,b_3),exp(g,a_8),exp(exp(g,eadd(a_8,emul(a_3,eadd(exp(g,a_8),operation)))),eadd(y_1,emul(b_3,eadd(exp(g,y_1),operation)))))

Fig. 5 MQV UKS attack trace graph

References

1. Blanchet B (2016) Modeling and verifying security protocols with the applied pi calculus and proverif
2. Blanchet B, Smyth B, Cheval V, Sylvestre M (2020) Proverif 2.02 pl1: automatic cryptographic protocol verifier, user manual and tutorial
3. Boyd C, Mathuria A, Stebila D (2003) Protocols for authentication and key establishment, vol 1. Springer
4. Choo KKR (2006) Key establishment: proofs and refutations. PhD thesis, Queensland University of Technology
5. Choo KKR, Boyd C, Hitchcock Y (2005) Examining indistinguishability-based proof models for key establishment protocols. In: International conference on the theory and application of cryptology and information security, pp 585–604. Springer
6. Diffie W, Hellman M (1976) New directions in cryptography. IEEE Trans Inform Theory 22(6), 644–654
7. Dolev D, Yao A (1983) On the security of public key protocols. IEEE Trans Inform Theory 29(2), 198–208
8. Farouk A, Miri A, Fouad MM, Abdelhafez AA (2014) Efficient pairing-free, certificateless two-party authenticated key agreement protocol for grid computing. In: 2014 fourth international conference on digital information and communication technology and its applications (DICTAP), pp 279–284. IEEE
9. Fiore D, Gennaro R (2010) Making the diffie-hellman protocol identity-based. In: Cryptographers' track at the RSA conference, pp 165–178. Springer
10. Fiore D, Gennaro R, Smart NP (2010) Constructing certificateless encryption and id-based encryption from id-based key agreement. In: International conference on pairing-based cryptography, pp 167–186. Springer

11. He D, Padhye S, Chen J (2012) An efficient certificateless two-party authenticated key agreement protocol. Comput Math Appl 64(6), 1914–1926
12. Kaliski BS Jr (2001) An unknown key-share attack on the mqv key agreement protocol. ACM Trans Inform Syst Secur (TISSEC) 4(3):275–288
13. Katz J, Lindell Y (2020) Introduction to modern cryptography. CRC Press
14. Menezes A (1997) Some new key agreement protocols providing implicit authentication. In: Workshop on selected areas in cryptography. CRC Press
15. Shashidhara R, Nayak SK, Das AK, Park Y (2021) On the design of lightweight and secure mutual authentication system for global roaming in resource-limited mobility networks. IEEE Access 9:12879–12895
16. Wu TY, Yang L, Lee Z, Chen CM, Pan JS, Islam S (2021) Improved ecc-based three-factor multiserver authentication scheme. Secur Commun Netw
17. Zhang J, Yang L, Gao X, Tang G, Zhang J, Wang Q (2021) Formal analysis of quic handshake protocol using symbolic model checking. IEEE Access (2021)

A Study on Using Emojis in a Shoulder Surfing Resistant Authentication Method

Mohamed Mahrous Amer, Yvonne Hwei-Syn Kam, and Vik Tor Goh

Abstract While images or emojis offer good memorability when used in an authentication method, inherently graphical data are highly susceptible to shoulder surfing attacks. An authentication system incorporating emojis was proposed and designed, which offers resistance to shoulder surfing attacks. The proposed system implements emojis in place of numerics in the reference method, DragPIN, and adds cue questions. The methods are compared in terms of performance and memorability, through user testing. The proposed authentication system was found to be successful at resisting shoulder surfing attacks. After 4–6 weeks, memorability was also higher in the proposed method compared to the reference method.

Keywords Graphical authentication system · PIN · Password · Emoji · Shoulder surfing

1 Introduction

User authentication systems must consider human factors such as ease of use and accessibility. Traditionally, alphanumeric passwords are perceived to have memorability predicaments. Graphical passwords use pictures instead of texts and are easier to remember than text-based passwords [1]. They provide a mechanism offering more user-friendly passwords. An advantage of graphical passwords is that they are easy to recall compared to alphanumeric passwords. However, while images or emojis might offer great memorability, inherently graphical data are highly susceptible to shoulder surfing attacks. Designing an effective graphical user authentication challenge scheme has been an ongoing subject of research for more than a decade. The

M. M. Amer · Y. H.-S. Kam (✉) · V. T. Goh
Multimedia University (MMU), 63000 Cyberjaya, Selangor, Malaysia
e-mail: hskam@mmu.edu.my

M. M. Amer
e-mail: 1151102010@student.mmu.edu.my

V. T. Goh
e-mail: vtgoh@mmu.edu.my

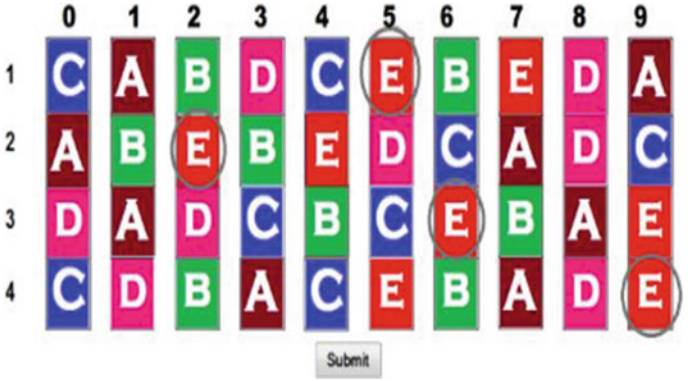
H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_7

challenge is managing the tradeoffs between the competing requirements of usability, memorability, and security.

2 Literature Review

Graphical passwords can offer advantages with regards to memorability and safety. The memorability can be explained by a hypothesis called the pictorial-superiority effect. The pictorial-superiority effect is a hypothesis that claims that a person can more easily recall graphical data than a text [1]. In an attempt to mitigate the risks associated with the vulnerabilities of Graphical authentication systems towards shoulder surfing attacks, Srinivasan proposed DragPIN [2]. Two schemes were created; manual and auto sliding. The proposed system revolves around a four-by-ten grid Personal Identification Number (PIN) entry scheme with characters that move automatically or manually. This system uses a four-digit PIN.

Figure 1 illustrates the DragPIN scheme. It has a manual and auto-sliding variant. The login process of the manual variant is initially commenced by mentally choosing an alphabet letter. Users must then use either the left or right arrow keys to move the alphabets in the left or right direction by remembering the digits of the PIN so that the chosen alphabet on the Nth row is located on a column number equal to the Nth PIN. For example, let the PIN be 5269. In Fig. 1, 'E' is the chosen letter which the user moves until it is aligned with each of the PIN digits. The automatic variant auto slides the rows and the user has to click at the right time to select the PIN digit. The advantage of DragPIN is that it is shoulder surfing resistant, especially its auto-sliding variant. The display in its auto-sliding variant does not directly correspond to the fixed password making it more difficult to reconstruct or deduce the fixed PIN from the result. In contrast, the password candidates of some shoulder-surfing resistant schemes can reduce radically after one observation [3].



	0	1	2	3	4	5	6	7	8	9
1	C	A	B	D	C	[✓ E]	B	E	D	A
2	A	B	[✓ E]	B	E	D	C	A	D	C
3	D	A	D	C	B	C	[✓ E]	B	A	E
4	C	D	B	A	C	E	B	A	D	[✓ E]

Fig. 1 Screenshot of DragPIN scheme [2]

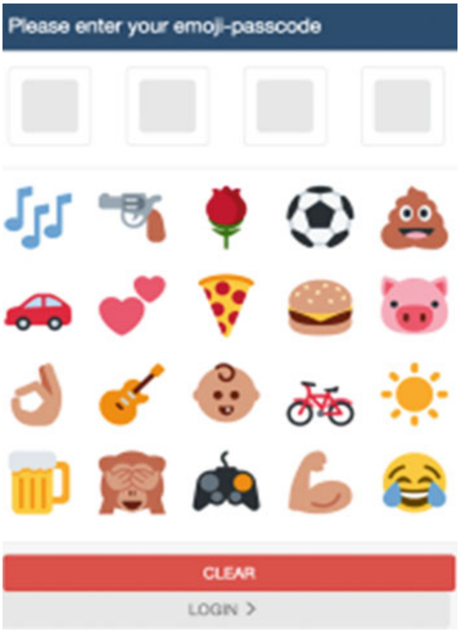


Fig. 2 Screenshot of EmojiAuth system [4]

While DragPIN achieved sufficient results in ensuring the resistance to shoulder surfing attacks, it may lack the ease of memorability, as common with methods that use alphanumeric passwords. This could be enhanced by using images rather than digits. Emojis are said to be familiar to users and can be expressed into memorable stories [4]. EmojiAuth [4] is an emoji-based authentication method. Figure 2 below demonstrates the login interface of the system. It is vulnerable to shoulder surfing because a potential shoulder surfer could watch the user as he/she selects the icons to enter his/her emoji password.

A highly effective approach to resisting shoulder surfing is by creating a secret channel between the system and the user to send information, which can cue the user in entering their session password. These channels could be, for example, audio [5], tactile [6], etc. A downside is that additional hardware or capabilities are needed to achieve these secret channels. Binbeshr et al. [7], suggested that future research should focus on methods that do not require an additional channel and/or hardware to ensure its acceptance and adoption.

Table 1 lists the advantages and disadvantages of related works. The methods [5] and [6] both require additional channels and thus additional hardware. Both DragPIN [2] and the methods in [3] and [8] are resistant to shoulder surfing without requiring additional hardware. The auto sliding variant implemented by DragPIN has the advantage where the displayed state does not directly correspond to the fixed password, which makes it more shoulder-surfing resistant. In [3], the fixed password can be deduced after multiple observations. In [8], intersection attacks on multiple observations can also narrow down candidates.

Table 1 Advantages and disadvantages of reviewed systems

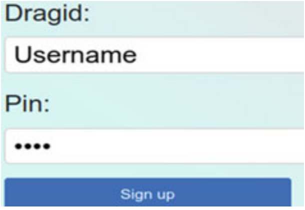
Paper	Advantages	Disadvantages
Srinivasan [2]	Resistant to shoulder surfing attacks The display does not directly correspond to the fixed password (auto-sliding variant)	Numbers are not as memorable as pictures
Salman et al. [3]	Resistant to shoulder surfing attacks	Vulnerable to intersection attack, with multiple observations
Kasat and Bhadade [8]	Resistant to shoulder surfing attacks	Vulnerable to intersection attack, with multiple observations
Rajarajan et al. [5]	Resistant to shoulder surfing attacks	Additional hardware/capabilities needed (earphones)
Ku and Xu [6]	Resistant to shoulder surfing attacks	Additional hardware/capabilities needed (vibration)
Golla et al. [4]	Memorable and easy to use	Not resistant to shoulder surfing attacks

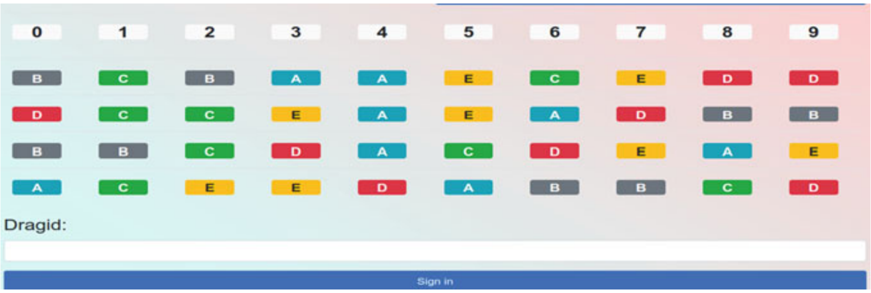
To summarise, DragPIN proposed a graphical authentication system for resisting shoulder surfing attacks. However, it uses numbers, which are not as memorable as pictures. EmojiAuth proposed a very easily memorable and useable system. But it is not resistant to shoulder surfing attacks. Both these methods came with their associated disadvantages. Therefore a proposed system using emojis instead of digits in DragPIN can achieve memorability and resistance to shoulder surfing attacks while solving both systems' disadvantages, while not losing their particular advantages.

3 Methodology

Both DragPIN and EmojiAuth methods were implemented to better understand their workings. A prototype of DragPIN was created as a proof of concept. As illustrated in Fig. 3, a signup page that allows a user to register a 4-digit pin and a username were created. The login scheme proposed in DragPIN was also implemented. The interface allowed users to sign in using the manual and automatic variants shown in Fig. 4.

Fig. 3 DragPIN prototype Sign up page



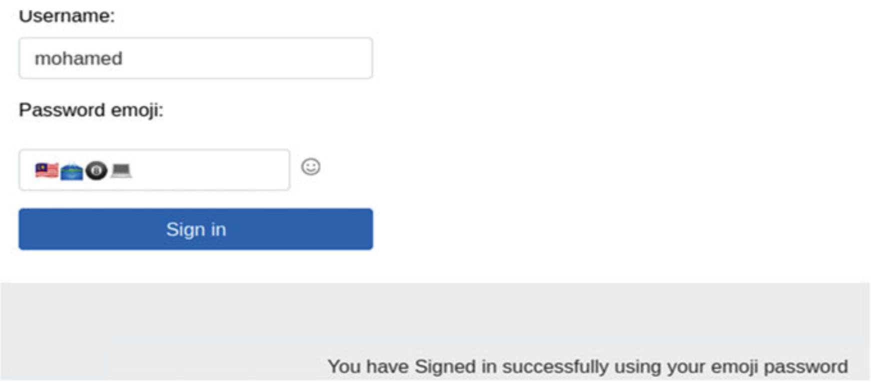


0	1	2	3	4	5	6	7	8	9
B	C	B	A	A	E	C	E	D	D
D	C	C	E	A	E	A	D	B	B
B	B	C	D	A	C	D	E	A	E
A	C	E	E	D	A	B	B	C	D

Dragid:

Fig. 4 DragPIN prototype (DragPIN auto sliding variant shown)

Figure 4 shows our implementation of the DragPIN interface. The login process of the manual variant was as described in Sect. 2. The auto sliding (or automatic) variant automatically slides the rows and the user has to click at the right time to select the PIN digit. The login process of the automatic scheme is commenced similarly by mentally choosing an alphabet letter then pressing the start button. This is followed by clicking the Spacebar once when they see their preferred alphabet appearing at the same column as their first PIN digit. The first row continues to slide, however. The user needs to press the Enter key to stop the sliding of the first row and to begin the sliding of the second row. The slipping of the second row then begins. The process is repeated for the four digits of the PIN. Then a prototype for EmojiAuth was created. The signup page allowed users to register a username and an emoji password. A login page (Fig. 5) allowed the user to input his/her previously registered username and emoji password using an online rendered emoji keyboard (shown in Fig. 6) or the emoji keyboard available on most smartphones.



Username: mohamed

Password emoji: 🇲🇾🏟🎱💻

You have Signed in successfully using your emoji password

Fig. 5 EmojiAuth prototype sign in

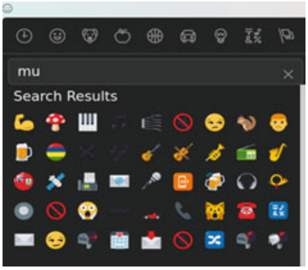
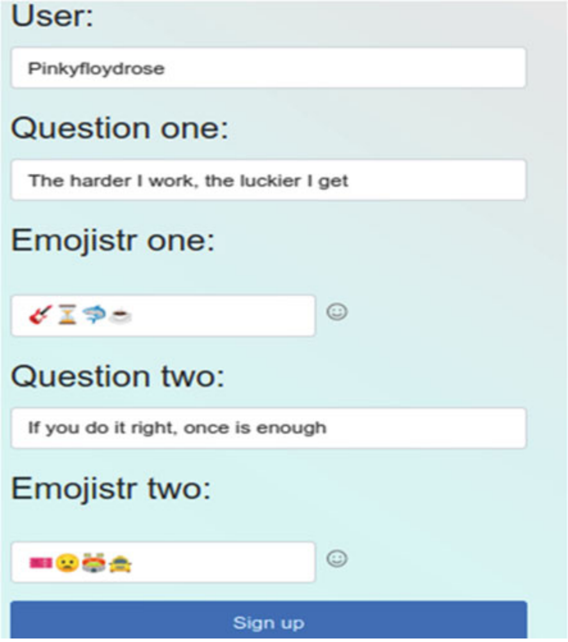


Fig. 6 Emoji Keyboard

3.1 Proposed Method Design

The proposed system was designed as a web application. To achieve better memorability, emojis are implemented rather than the PIN digits in DragPIN. The user only submits four emojis for each of the two passwords needed to register. Therefore, the system randomly generates six more decoy emojis from Unicode 'version six' to allow for $4 + 6 = 10$ emojis to be placed as table column indexes which are only generated by the system once and stored. For a particular user, the same set of emojis is shown at every login. To achieve memorability and usability, cue questions were added to the method, which was not in DragPIN. Cue questions are used to provide a cue for the particular password expected during authentication. Shoulder surfing resistance is enhanced as there are two cue questions, where the attacker may not be challenged with the same cue question as the one observed. Figure 7 below further illustrates the proposed usage of cue questions. The cue question is created by the user and ideally ultimately creates a virtual mental channel in which the system would seem to communicate details about the password to the user. Each user is required to register two cue questions associated with two passwords consisting of four emojis each, as shown in Fig. 7 below. For each login session, the column indexes are rendered by the system along with a random cue question allowing users to login using either the first or second emoji password.

Users are then expected to submit their username in the form shown in Fig. 8 below as the first part of a two-part login process. During this phase, Cross-site request forgery (CSRF) tokens are generated and passed with the form to the user. Those tokens get checked as well as the requested username. A CSRF token is a unique secret value generated by the server-side application. This token is sent to the client so that it would be checked after being sent back in a subsequent HTTP request made by the client. An example of the usage of CSRF tokens in the proposed system is shown in Fig. 9.



User:

Pinkyfloydrose

Question one:

The harder I work, the luckier I get

Emojistr one:

🎸⏳🌪☕ ☺

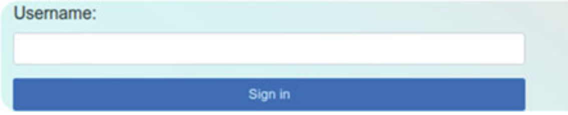
Question two:

If you do it right, once is enough

Emojistr two:

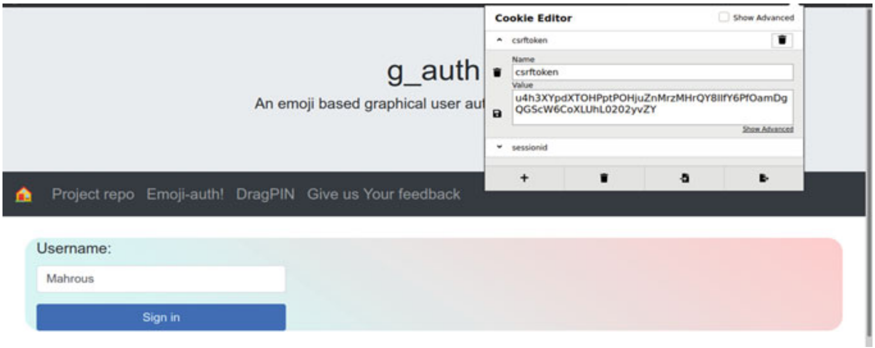
🎟😦🎂🚖 ☺

Fig.7 Proposed method Sign up page



Username:

Fig. 8 Proposed method Sign-in page



g_auth

An emoji based graphical user aut

Cookie Editor □ Show Advanced

csrftoken

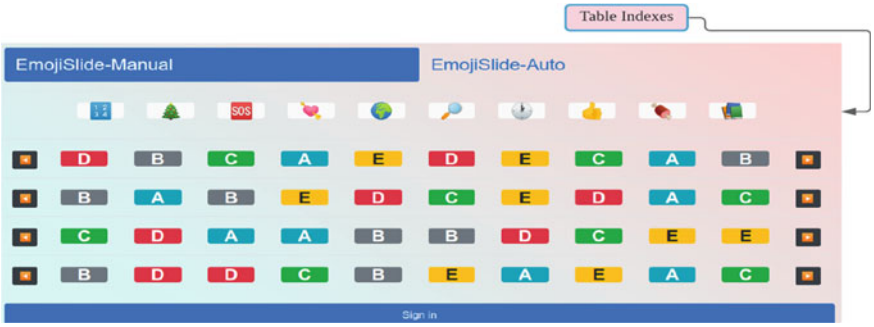
Name: csrftoken

Value: u4h3XYpdXTOHPptPOHjuZnMrzMHrQY8llfY6PfOamDgQGScW6CoXLUhL0202yvZY

Username:

Mahrous

Fig. 9 CSRF token



EmojiSlide-Manual EmojiSlide-Auto

	🔢	🎄	🆘	💘	🌍	🔎	🕛	👍	🍖	📚	
◀	D	B	C	A	E	D	E	C	A	B	▶
◀	B	A	B	E	D	C	E	D	A	C	▶
◀	C	D	A	A	B	B	D	C	E	E	▶
◀	B	D	D	C	B	E	A	E	A	C	▶

Sign in

Fig. 10 Proposed manual implementation

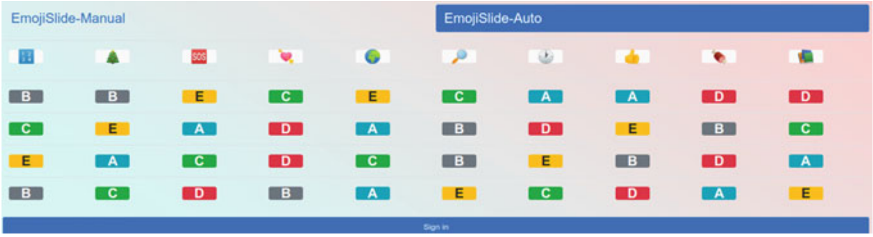
3.2 Authentication

During authentication, a user may choose to authenticate via the manual or the automatic scheme. If the manual scheme shown in Fig. 10 is chosen, the user needs to mentally pick a color or a character to represent his marker. The user's four emoji password could be for example 🔢, 👇, 🕐, 🆘. The icons may look slightly different on different platforms. The arrows located at both ends of each row move the chosen color/letter under the first emoji in the password. The process is repeated for the remaining three emojis in the password. In Fig. 10, the chosen letter was 'D' and this letter was aligned to the emojis in the password.

Otherwise, if the automatic scheme (Fig. 11) is chosen, similar to the manual variant, firstly the user needs to pick a color or a character to represent his marker. Secondly, the user's four emoji password (e.g. 🔢, 👇, 🕐, 🆘) is then mapped by allowing the first row to shift until the chosen color/letter is in the required position. The "space" key is pressed to record that entry and the process can be commenced in the second row by pressing the "enter" key to stop the existing row from shifting and begin shifting the second row. The same technique is constant throughout the rest of the rows thus the process is repeated for the remaining emojis in the password. In Fig. 11, the chosen letter was 'B' and the user pressed the "enter" key after the row had slid past the password emoji. That is the reason the letter 'B' is no longer under the corresponding emojis. This misalignment resists shoulder surfing.

3.3 User Test Study

In the first phase, we invited participants to a google meet that was being recorded for later evaluation on the ability of this scheme to resist shoulder surfing. Initially, the motives behind the system were briefly explained. A test user was then created while explaining each field on the sign-up page. The participant (or user) then attempted to log in using the test user account, using each of the implemented schemes in both



EmojiSlide-Manual

EmojiSlide-Auto

B	B	E	C	E	C	A	A	D	D
C	E	A	D	A	B	D	E	B	C
E	A	C	D	C	B	E	B	D	A
B	C	D	B	A	E	C	D	A	E

Sign in

Fig. 11 Proposed automatic implementation

DragPIN and EmojiSlide (our proposed method). This enabled the users to fully understand the login process associated with both schemes of the authentication systems. Next, users were instructed to register. The users had to create two cue questions and their corresponding emoji passwords. Some users resorted to creating an emoji password that resembled a story. They then attempted to authenticate. Participants were given only 3 consecutive attempts to log in. None of the users failed 3 consecutive attempts. Users registered and logged in using both schemes. In both, the time needed for the users to successfully authenticate was recorded. An understanding of what users thought of the system was reached by allowing the users to answer a quick survey upon completion of the experiment. To test for shoulder surfing resistance, in the first phase, shoulder surfing was carried out on the user login video recordings. Four “shoulder surfers” were first subjected to a demo round in which each created a user and tried logging in to gain an understanding of each of the schemes. Each shoulder surfer was then given the video recordings of the logins performed by the users, to attempt a shoulder surfing attack. In the second phase, users from the first phase were tasked to log in again, to test for memorability of the emoji passwords vs PINs. This phase was held from four to six weeks after the initial tests.

4 Results

There were 30 participants in the study. The age groups of the participants who participated are shown in Fig. 12. Most of them (76.7%) were between 20 and 30 years old. Table 2 below summarises the different aspects of the participants that contributed to this study. A note about the 26.7% of participants who claimed to be computer savvy. They were faster and tried to protect their identity by using a variation of less commonly used emojis that did not relate to their cue question at the first glance.

Each of the participants attempted to login to two methods with two variants each (EmojiSlide Manual, EmojiSlide Automatic, DragPIN Manual, and DragPIN Automatic). Table 3 shows the average time needed for the participants to login using each of the schemes, with the two variants each. When comparing the login time

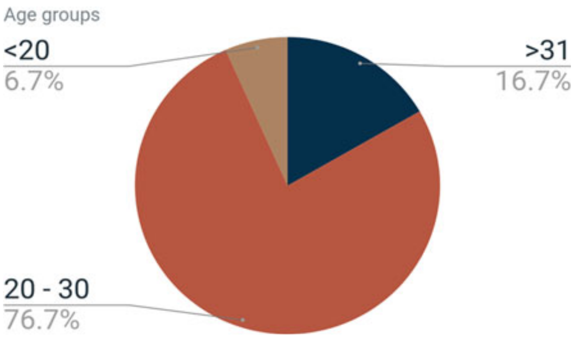


Fig. 12 Age groups of users

needed for DragPIN vs EmojiSlide, users took less time to login using EmojiSlide. Results also showed that the automatic variants required more time than the manual variants. Users were requested to rate their trust in the system as being shoulder-surfing resistant by the question shown in Fig. 13. The figure shows that 76.7% of participants felt that they can trust this system to resist shoulder surfing. 23.3% of participants were unsure and none (0%) of the participants answered no to this question.

The shoulder surfers described trying to observe the users' passwords as an "excruciating task". They also commented that reversing the recorded videos or slowing the playback did not help much in identifying the password. This was especially true for the automatic variants. None of them were able to get any full PIN or emoji password. They were only able to get a maximum of two emojis right, from three users, which was due to those users pointing their cursor at their desired emoji. None of the participants took more than three tries to log in to both the proposed system and DragPIN implementations. Most were in the DragPIN auto variant during phase 1, where three participants took three tries to log in. The proposed system has a higher average partial success rate and higher memorability as apparent in Fig. 14. The partial success rate is calculated as such: for each participant, if a successful login is made in one attempt, the success rate is 100%, if in two attempts, the success rate is 1/2 or 50% and if in three attempts, the success rate is 1/3 or 33.33%. The average partial success rate is taken as the average across all participants. After 4–6 weeks, during the phase 2 evaluation, the partial success rate of the proposed system was still higher compared to DragPIN. The percentage of decline was also smaller. User testing of the proposed scheme showed that users were able to enter their graphical passwords with high accuracy. Enhanced memorability was also achieved by implementing emojis in DragPIN schemes.

Table 2 Summary of the participants

	Males	Females	Graduates/Employed	Undergraduates	Computer savvy users	Average computer users	Non-frequent computer users
Percentage (%)	66.66	33.33	26.67	73.33	26.66	53.33	20

Table 3 Average login time

Parameters	DragPIN manual	DragPIN automatic	EmojiSlide manual	EmojiSlide automatic
Average login time (s) seconds	19.3	30.1	16.7	29.5

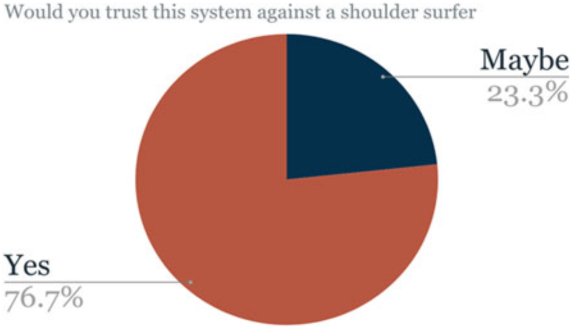


Fig. 13 Users' trust in the proposed system to resist shoulder surfing

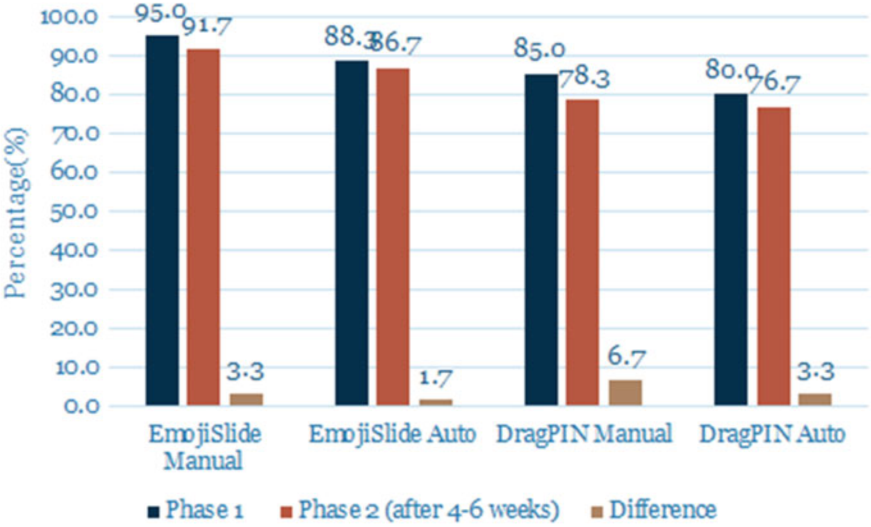


Fig. 14 Average success rate (including partial fails) in EmojiSlide and DragPIN during phase 1 and phase 2

5 Conclusion

In this paper, a graphical authentication method was proposed where it was shown to be shoulder surfing resistant as well as memorable. The proposed system implements emojis in place of numerics in the reference method, DragPIN, and in addition, implements cue questions that aid memorability while improving theoretical shoulder surfing resistance. A user study was carried out and participants demonstrated the features of high usability and security in the proposed method. Memorability was tested by having the same participants log in using the methods, four to six weeks after the initial phase. The result showed a higher partial success rate for the proposed method compared to the DragPIN implementation. The proposed method (both manual and auto-sliding variants) showed resistance to shoulder surfing in a recording-based shoulder surfing experiment, where none of the shoulder surfers managed to obtain the passwords.

Acknowledgements This work was supported by the Malaysian Ministry of Higher Education, Fundamental Research Grant Scheme (FRGS) [grant number FRGS/1/2015/ICT04/MMU/03/6]; and Multimedia University IR Fund [grant number MMUI/210071-IR Fund].

References

1. Paivio A, Csapo K (1973) Picture superiority in free recall: imagery or dual coding?. Cogn Psychol 5(2):176–206
2. Srinivasan R (2018) DragPIN: a secured PIN entry scheme to avert attacks. Int Arab J Inf Technol 15(2):213–223
3. Salman M, Li Y, Wang J (2019) A graphical PIN entry system with shoulder surfing resistance. In: 2019 IEEE 4th international conference on signal and image processing (ICSIP). IEEE, pp 203–207
4. Golla M, Detering D, Dürmuth M (2017) EmojiAuth: quantifying the security of emoji-based authentication. In: Proceedings of the usable security mini conference (USEC)
5. Rajarajan S, Kalita R, Gayatri T, Priyadarsini PLK (2018) Spinpad: a secured pin number based user authentication scheme. In 2018 international conference on recent trends in advance computing (ICRTAC). IEEE, pp 53–59
6. Ku W, Xu H (2019) Efficient shoulder surfing resistant PIN authentication scheme based on localized tactile feedback. In: 2019 6th IEEE international conference on cyber security and cloud computing (CSCloud)/ 2019 5th IEEE international conference on edge computing and scalable cloud (EdgeCom), pp 151–156
7. Binbeshr F, Kiah MM, Por LY, Zaidan AA (2020) Systematic review of PIN-entry methods resistant to shoulder-surfing attacks. Comput Secur 102116
8. Kasat OK, Bhadade US (2018) Revolving flywheel pin entry method to prevent shoulder surfing attacks. In: 2018 3rd international conference for convergence in technology (I2CT). IEEE, pp 1–5

A Framework System Using Word Mover's Distance Text Similarity Algorithm for Assessing Privacy Policy Compliance

Chun Yee Chaw and Hui Na Chua

Abstract Privacy policies are important as they outline how organizations manage the personal data of consumers who use their services. However, a key issue with privacy policies is that they are lengthy and verbose, hindering the public from fully understanding the contents stated in the privacy policy. While there have been existing research works on assessing privacy policies, most of them are manually done by humans. Besides lacking automated solutions for assessing privacy policies' compliance with data protection regulations, there has been no usage of semantic text analytics approaches in the study of privacy policy compliance. As such, we researched and implemented a framework system embedded with a data protection requirements dictionary where privacy policies are assessed automatically based on its coverage with the dictionary. We selected the General Data Protection Regulation (GDPR) as the primary source of our experiment for its broader requirements compared to other regulations. The assessment by the framework is realized through the Word Mover's Distance (WMD) text similarity algorithm which calculates the similarity distance of how close the meaning of a privacy policy and the data protection regulation requirements in the dictionary. Our framework system is a novel implementation of the WMD text similarity algorithm in assessing privacy policies semantically and it contributes to an automated assessment on privacy policy compliance with personal data protection requirements.

Keywords Word Mover's distance · Document similarity · Text analytics · Personal data protection · Privacy policy compliance

1 Introduction

Privacy policies are important in ensuring transparency and accountability for an organization's use of personal data and helping consumers understand their rights to data protection. However, a key issue with privacy policies is that they are

C. Y. Chaw · H. N. Chua (✉)
Department of Computing and Information Systems, Sunway University, Selangor, Malaysia
e-mail: huinac@sunway.edu.my

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_8

often lengthy and hard to be understood by the average consumer [1]. Due to this, consumers are more likely to accept the terms in the privacy policy without truly understanding its contents [2]. As such, it is importance that a privacy policy is examined to determine what the terms in the policy provide for consumers. However, existing research in privacy policies analysis is mostly manually by humans reading through the policies, comparing them with already established legislations, and using questionnaires and experiments with human subjects to assess the scope and coverage of the policies [3–5].

Cradock et al. [6] proposed a novel solution using Jaccard Similarity [7], a document similarity algorithm to compare privacy policies both against each other and an external list of recommendations from existing legislation. Document similarity is the process of finding how similar two pieces of text are based on the words used and how the words are used. Unfortunately, this solution is disadvantageous when semantics need to be considered as it only compares similarities based on words used, not the context of their usage.

Unlike Jaccard Similarity, Word Mover's Distance (WMD) [8] compares the similarity between two documents by calculating the minimum amount of distance that the embedded words of one document need to "travel" to reach the embedded words of another document and takes semantics into account by considering the similarities in each word embedding. Therefore, we consider WMD as a more effective algorithm than Jaccard Similarity from the aspect of analyzing semantics. This consideration shapes our motivation and objective to investigate the realization of automated assessment systems using WMD algorithm for analyzing privacy policies compliance. This investigation leads to an implemented framework system using the WMD algorithm to assess privacy policies by identifying the closest matching regulation in the dictionary based on their semantics. The data protection requirements (DPR) dictionary we developed is used to capture the legislations' requirements based on GDPR.

2 Literature Review

2.1 Personal Data Protection Regulations and Privacy Policy

Personal data protection regulations form rules legalizing how personal data is obtained, used, and disclosed. Among these regulations, GDPR is considered the tightest personal data protection regulation based on the requirements it covers and its global impact [9]. Most of these regulations consider a similar conception of consent, purpose, reasonableness. To comply with these principles, regulations require companies to inform customers how an organization collects and processes personal data [10]. Despite the importance of privacy policies, most of them are difficult to read due to their length and verbosity [1]. Because of this, users turn out to agree to the terms without truly understanding their meaning [11]. Consequently, many customers rarely understand the full scope of the privacy policies they are agreeing to [2]. As such, companies respond differently and vary their privacy notice' contents [12] for various reasons such as organizational strategies and customers' perspectives of information privacy [13, 14].

2.2 Text Similarity and Word Embedding Techniques

Text similarity falls within the scope of Natural Language Processing (NLP). There are two distinct concepts of "similarity" to note: lexical similarity is comparing text based on similar words used and the real meaning in the text is unaccounted for, whereas semantic similarity focuses on the context where the text is used, and a piece of text is broken into relevant groups of related words before computing their similarity. There are many types of text similarity algorithms in NLP; some algorithms take semantics into account while others do not. As of to-date, only Cosine Similarity [15] and WMD [8] consider semantics when comparing the similarity between sentences. On the other hand, algorithms like Latent Dirichlet Allocation (LDA) [16] with Jensen-Shannon distance compute similarity in topic distribution, whereas Siamese Manhattan Long Short-Term Memory (MaLSTM) [17] computes the similarity between networks.

Cosine similarity and WMD are similar as they require pre-trained word embedding and both consider semantics when comparing text sentences. However, their mechanisms for text comparison are different. Cosine similarity, on its own, only measures the orientation in which two sentences are apart, and this depends on what words are used in the sentences. If two sentences have entirely different words yet convey a similar meaning, the cosine similarity becomes zero and inaccurate. To account for semantics, cosine similarity must be paired with another mechanism called Smooth Inverse Frequency (SIF) that provides more weight to words for semantic contribution. While SIF does help cosine similarity consider semantics, it adds another step to the overall process of using the cosine similarity algorithm. By contrast, the WMD algorithm only requires a word embedding model as a reference to calculate the distance score, with no other mechanisms required, and results in less complexity during implementation.

In word embedding techniques, one word is mapped to one vector and the vector values are learned such that it resembles a network. These vectors form a learned representation of the words which are called word embeddings. Word embeddings embody a group of methods in which words are represented as real-valued vectors in a vector space. As such, a densely distributed representation for each word is the key to word embedding techniques. This distributed representation is learned depending on how the words are used, allowing words that are used similarly to have similar representations and obtaining their meaning at the same time. There are many different techniques to obtain word embeddings, such as Bag of Words (BoW) [18], Term Frequency—Inverse Document Frequency (TF-IDF) [19], Continuous BoW (CBoW) model [20], Skip-gram model [21], and Pre-trained word embedding model [22]. Among the word embeddings techniques, pre-trained word embeddings are considered the most useful in doing NLP-related tasks as the word embeddings are trained on a large corpus of text documents and the vectors for a word can easily be retrieved from the word embedding model.

2.3 Prior Research on Privacy Policy Analysis

Paul et al. [5] analyzed the privacy policies of Internet of Things services by creating a framework based on GDPR which contains parameters to assess the coverage of the privacy policy being tested framed as yes-or-no questions. Then Paul et al. read through manually each privacy policy and marked a table containing the parameters with yes-or-no depending on the coverage of the privacy policy. Zapata et al. [4] utilized a similar measure where privacy policies of mobile personal health records are assessed by comparing the articles in the privacy policy to existing regulations and creating a questionnaire that assesses the privacy policy based on specific parameters. Schmidt et al. [3] conducted an empirical study to test the ability of requirements engineers to extract compliance requirements from privacy policies using analysis approaches including CPR (commitment, privilege, and right) analysis, goal-based analysis, and non-method assisted (control) analysis.

Literatures in [3–5] focus on the process of manually reading through each privacy policy for assessment and using an established data privacy documentation to compare and assess privacy policies. None of the above-mentioned studies used an automated text analysis approach to assess privacy policies and at the same time consider semantics. Research which has used text similarity algorithms to assessing privacy policies is rare, except the literature found in Cradock et al. [6] that analyzed each privacy policy by manually reading through the policy. These prior research works then used thematic analysis and cross-document structure theory to identify themes in each policy. The research conducted in [6] is important as it builds up the motivation of using text similarity algorithms for assessing privacy policies. However, the use of Jaccard Similarity is disadvantageous when semantics need to be considered since it only checks for the similarity of the words used and not how it is used.

3 Framework System and Methodology

Figure 1 shows the high-level architecture view of our framework system. A data protection requirements (DPR) Dictionary based on the GDPR is generated for comparisons with privacy policies to be assessed.

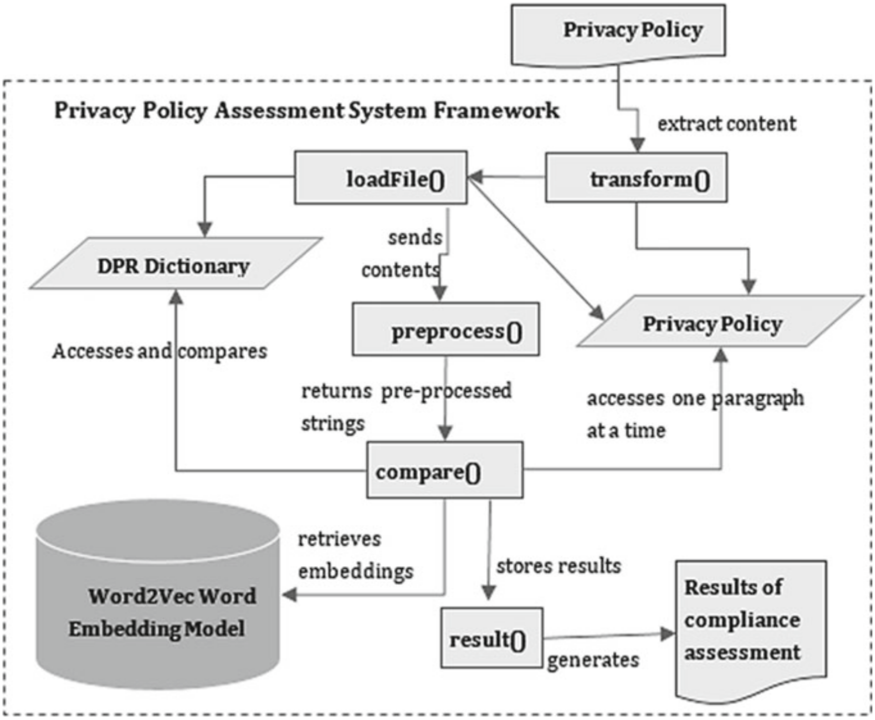
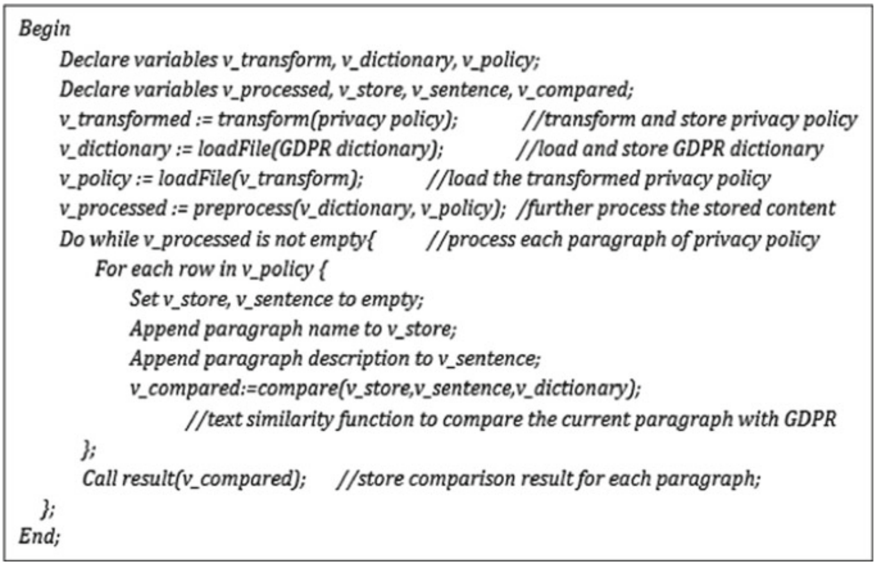


Fig. 1 Framework for assessing privacy policy compliance

3.1 Assessment System Main Procedure

The framework system has several essential sub-functions to assess privacy policy compliance:

- "transform()"—This function extracts and transforms the privacy policy from an organization into the format structure that can be systematically interpreted by the subsequent function for further data preparation.
- "preprocess()"—Pre-processes a text string provided as input to ensure only the words in the text strings which contribute meaning are used for comparison. It performs the tasks of splitting the string into individual words; converting each word to lowercase; removing punctuation from each word and any non-alphabetic words; and, filtering out stop-words. This sub-function outputs an array of words that form the pre-processed string.
- "loadFile()"—Loads the GDPR dictionary and privacy policy documents. This sub-function converts the contents from both sets of documents into the output structure for later comparison.
- "compare()"—Conducts the actual comparison using WMD. When a privacy policy is being compared to this dictionary, the function will loop through the



```
Begin
    Declare variables v_transform, v_dictionary, v_policy;
    Declare variables v_processed, v_store, v_sentence, v_compared;
    v_transformed := transform(privacy policy);          //transform and store privacy policy
    v_dictionary := loadFile(GDPR dictionary);           //load and store GDPR dictionary
    v_policy := loadFile(v_transform);         //load the transformed privacy policy
    v_processed := preprocess(v_dictionary, v_policy);  /further process the stored content
    Do while v_processed is not empty{         //process each paragraph of privacy policy
        For each row in v_policy {
            Set v_store, v_sentence to empty;
            Append paragraph name to v_store;
            Append paragraph description to v_sentence;
            v_compared:=compare(v_store,v_sentence,v_dictionary);
                    //text similarity function to compare the current paragraph with GDPR
       };
       Call result(v_compared);      //store comparison result for each paragraph;
  };
End;
```

Fig. 2 The high-level pseudocode illustrating the main procedure for the system

entire privacy policy, taking each paragraph from the policy and finding which requirement in the dictionary is the closest match to the privacy policy paragraph.

- "result()"—Outputs the results of comparison for privacy policy assessment.

The high-level privacy policy compliance assessment system procedure is logically illustrated through the pseudocode presented in Fig. 2.

3.2 Data Protection Requirements (DPR) Dictionary

There are some articles in the GDPR do not directly relate to both organizations (denoted as the data controller or processor in regulations) and individuals (denoted as the data subject). These articles only describe the process of enforcing the GDPR by the EU and it is unlikely that data controllers will include these types of articles in their privacy policy. As such, their inclusion in the GDPR dictionary will be irrelevant to the purpose of the assessment. Therefore, we applied a filtering method in which an article is only selected if the article directly implicates the key entities in the GDPR which are Data Subject, Personal Data, Controller, Processor, Recipient, and Third-party.

The DPR dictionary is constructed through manual reading and selecting the relevant articles. This process is done only once, and the dictionary is unchanged throughout the comparison process. The manual process for article selection is considered the most accurate approach as it avoids any bias generated by the system due to the possibility of misinterpreting the meaning of data protection regulations.

3.3 Privacy Policy

The privacy policy to be assessed can be any online policy that is sourced from an organization's website. A privacy policy transformer (i.e. "transform()" in Fig. 1) is required to convert the structure of the online privacy policy so that it is in a compatible format as the GDPR dictionary to allow efficient matching by the assessment system. The information about the transformed privacy policy is paragraph name and the paragraph description. In the comparison algorithm, the transform() function presented in Fig. 1 is a logical component that has not been implemented physically for our experiment. Therefore, like how the DPR dictionary is created, articles in the privacy policy are manually analyzed and selected for comparison.

3.4 Text Similarity Algorithm and Word Embedding Model

In deciding the text similarity algorithm to compare the content of DPR Dictionary and privacy policy, semantics must be considered. For the comparison algorithm, a privacy policy is compared to only one reference document, which is the DPR Dictionary.

Most privacy policies are often narrow in scope, only focusing on topics regarding usage and protection of personal data. As such the topic distribution in both the DPR Dictionary and privacy policy are likely to be relatively narrow as well. For this reasoning, algorithms that consider similarity in topic distributions are not considered necessary for our research aim. As mentioned in Sect. 2.2, WMD has lesser steps needed to consider semantics for calculating the similarity of text strings when compared to Cosine Similarity. Therefore, the WMD algorithm is chosen for this paper.

Pre-trained word embeddings are used in the comparison algorithm as they are trained on large datasets, which ensures that the words in both the GDPR dictionary and privacy policy have a higher chance of having their embeddings in the model. Besides, this avoids the necessity of having to manually train a word embedding model which is time-consuming and does not guarantee that there are sufficient words that can cover the language used in both the GDPR dictionary and privacy policy.

3.5 Text Similarity Accuracy Evaluation

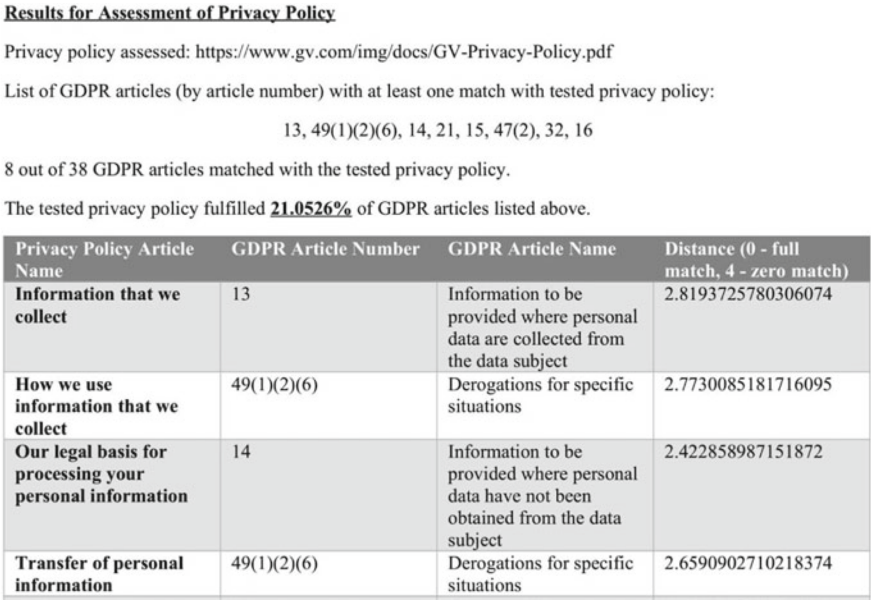
WMD works by comparing the pre-processed text strings of both the privacy policy paragraph and DPR Dictionary requirements. The algorithm calculates the overall WMD distance score of both text strings by measuring the distance of each word in the string, then aggregating all the distances as a distribution to obtain an average distance score. Applying this idea for the framework, the reference is the word embeddings in vectors, and the text sentences to be compared are the privacy policy paragraph and the DPR requirement respectively. Each word in the pre-processed text string is compared with the word embeddings in the model to find the word embedding with the lowest distance. The WMD distance score is then calculated by aggregating the distances of the individual words with their embeddings. When comparing the pre-processed text strings of a privacy policy paragraph and a DPR requirement, the WMD distance between the two strings is greatly influenced by how many individual words are in the pre-processed text strings. The more the number of words in the privacy policy paragraph and DPR requirement that matches with those in the model, the lower the distance score yielded with the comparison.

4 Experimental Results

The view of one of the results after assessing a privacy policy is presented in Fig. 3.

Figure 3 shows that the assessed privacy policy fulfilled 21.0526% of the overall requirements in the DPR Dictionary. The figure also shows the result of matched privacy policy paragraphs with a more detailed description of similarity distance (the close the smaller distance number). The unit for the WMD score is a floating-point number ranging from zero to infinity, where zero means an exact match for both sentences while infinity means no match. The higher the distance score, the greater the difference between the two sentences. Since the WMD score can go all the way to infinite, a sensible benchmark for the maximum score on the WMD score is necessary to provide more context to the WMD scores calculated in the results. To derive a reasonable maximum limit for interpreting the context, several text strings are tested with the DPR dictionary to obtain the highest WMD score possible. These text strings do not have words that are in the DPR Dictionarys so there is a higher chance of each of them having a large WMD distance from the DPR articles. With this experiment, we derived a maximum score to interpret the highest distance, which is the score of 4.

We performed a random accuracy evaluation of our results using a qualitative approach. We observed in the privacy policy of GV company, the article “Right to object” is correctly matched with the “Right to object” requirement in the DPR dictionary. Likewise, the article “Information security” in the same privacy policy is also correctly matched with the “Security of processing” requirement in the DPR dictionary. The WMD algorithm identifies the closest matching articles correctly



Results for Assessment of Privacy Policy

Privacy policy assessed: https://www.gv.com/img/docs/GV-Privacy-Policy.pdf

List of GDPR articles (by article number) with at least one match with tested privacy policy:

13, 49(1)(2)(6), 14, 21, 15, 47(2), 32, 16

8 out of 38 GDPR articles matched with the tested privacy policy.

The tested privacy policy fulfilled **21.0526%** of GDPR articles listed above.

Privacy Policy Article Name	GDPR Article Number	GDPR Article Name	Distance (0 - full match, 4 - zero match)
Information that we collect	13	Information to be provided where personal data are collected from the data subject	2.8193725780306074
How we use information that we collect	49(1)(2)(6)	Derogations for specific situations	2.7730085181716095
Our legal basis for processing your personal information	14	Information to be provided where personal data have not been obtained from the data subject	2.422858987151872
Transfer of personal information	49(1)(2)(6)	Derogations for specific situations	2.6590902710218374

Fig. 3 The partial view of one of the privacy policies assessment results

even without knowing the article titles. This further validates that the contents of both the privacy policy and DPR dictionary affect which requirement-paragraph pairs are the closest match with one another.

5 Analysis of Results and Limitations

In analyzing the efficacy of our proposed algorithm, we have identified several key limitations that need to be highlighted. Firstly, This experiment does not include a mechanism to evaluate the accuracy of the comparison results, as it falls out of the scope of this paper. Our main purpose is primarily to demonstrate how the WMD algorithm may be implemented to compare privacy policies with data protection legislation. However, based on the random accuracy evaluation mentioned in Sect. 4, we can conclude that the WMD algorithm in our proposed framework successfully measured the similarity of privacy policy articles with those in the DPR dictionary to a reasonable degree of accuracy. This proves that WMD is applicable as our framework text similarity algorithm.

Another limitation is that the comparison results can be affected by the contents of the DPR dictionary. As the number of requirements in the DPR dictionary is increased, each article in the privacy policy needs to be compared with a greater number of requirements in the dictionary. This will affect the final percentage of overall compliance in the results. Future work is required to tackle this challenge.

The WMD algorithm in our proposed framework can be helpful for end-users to gauge compliance with a privacy policy by quantifying the comparison results. However, the proposed framework only considers the technical aspect of measuring compliance. End-users may interpret compliance with the privacy policy differently as language is subjective. As such, we can only conclude that our study has laid down a proven framework using WMD to measure compliance of privacy policies, one that considers not only quantitative results but also language and context.

6 Conclusion

The results of our implemented framework system show a novel implementation of the use of the WMD text similarity algorithm in assessing a privacy policy by comparing the contents of the privacy policy and a data protection regulation dictionary developed. As such, there exists an opportunity to further improve the inner workings of this system. Different types of word embeddings and text similarity algorithms other than those utilized in this system can be tested and analyzed to ascertain their efficiency in the comparison algorithm. Also, further work can be done to measure the accuracy of our proposed framework in terms of the comparison results as well as other existing frameworks in evaluating privacy policies. Another improvement to our framework is a structured methodology in deciding what requirements are needed for the DPR dictionary.

Acknowledgments Funding: This research was supported by the Malaysian government FRGS grant [FRGS/1/2019/ICT04/SYUC/02/2].

References

1. Litman-Navarro K (2019) We read 150 privacy policies. They were an incomprehensible disaster. The New York Times 12
2. Hart K (2019) Privacy policies are read by an aging few. Tech Rep
3. Schmidt JY, Antón AI, Earp JB (2012) September. Assessing identification of compliance requirements from privacy policies. In: 2012 Fifth IEEE International Workshop on Requirements Engineering and Law (RELAW), IEEE, pp 52–61
4. Zapata BC, Niñirola AH, Fernández-Alemán JL, Toval A (2014, August) Assessing the privacy policies in mobile personal health records. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp 4956–4959
5. Paul N, Tesfay WB, Kipker DK, Stelter M, Pape S (2018 September) Assessing privacy policies of Internet of Things services. In: IFIP International Conference on ICT Systems Security and Privacy Protection, Springer, Cham, pp 156–169

6. Cradock E, Millard D, Stalla-Bourdillon S (2015 May) Investigating similarity between privacy policies of social networking sites as a precursor for standardization. In: Proceedings of the 24th International Conference on World Wide Web, pp 283–289
7. Glen S (2016) Jaccard index/similarity coefficient. Statistics how to. https://www.statisticshowto.com/jaccard-index. Accessed 22 Oct 2020
8. Kusner M, Sun Y, Kolkin N, Weinberger K (2015 June) From word embeddings to document distances. In: International Conference on Machine Learning, PMLR, pp 957–966
9. Sim WL, Chua HN, Tahir M (2019 November) Blockchain for identity management: the implications to personal data protection. In: 2019 IEEE Conference on Application, Information and Network Security (AINS), IEEE, pp 30–35
10. PDPA (2010) Laws of Malaysia, Act 709, Personal Data Protection 2010
11. Chua HN, Wong SF, Low YC, Chang Y (2018) Impact of employees' demographic characteristics on the awareness and compliance of information security policy in organizations. Telematics Inform 35(6):1770–1780
12. Chua HN, Herbland A, Wong SF, Chang Y (2017) Compliance to personal data protection principles: A study of how organizations frame privacy policy notices. Telematics Inform 34(4):157–170
13. Greenaway KE, Chan YE (2013) Designing a customer information privacy program aligned with organizational priorities. MIS Quart Execut 12(3)
14. Parks RF, Wigand RT (2014) Organizational privacy strategy: Four quadrants of strategic responses to information privacy and security threats. J Inf Privacy Secur 10(4):203–224
15. Prabhakaran S (2018) Cosine similarity - understanding the math and how it works? (with Python). Mach Learn Plus. https://www.machinelearningplus.com/nlp/cosine-similarity/. Accessed 22 Jan 2020
16. Sieg A (2018) Text similarities: estimate the degree of similarity between two texts. Medium. https://medium.com/%40adriensieg/text-similarities-da019229c894. Accessed 22 Jan 2020
17. Cohen E (2017) How to predict quora question pairs using siamese manhattan LSTM. https://medium.com/mlreview/implementing-malstm-on-kaggles-quora-question-pairs-competition-8b31b0b16a07. Accessed 22 Jan 2020
18. Brownlee J (2017) A gentle introduction to the bag-of-words model. Machine learning mastery. https://machinelearningmastery.com/gentle-introductionbag-words-model/. Accessed 22 Jan 2020
19. Góralewicz B (2018) The TF*IDF algorithm explained. Onely. https://www.onely.com/blog/what-is-tf-idf/. Accessed 22 Jan 2020
20. Sarkar D (2018) Implementing deep learning methods and feature engineering for text data: the continuous bag of words (CBOW). KDnuggets. https://www.kdnuggets.com/2018/04/implementing-deep-learning-methods-feature-engineering-text-data-cbow.html. Accessed 22 Jan 2020
21. Huang K (2018) Word2Vec and FastText word embedding with Gensim. Medium. https://towardsdatascience.com/word-embedding-with-word2vec-and-fasttext-a209c1d3e12c. Accessed 22 Jan 2020
22. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781

A Novel Approach to Classify Vulnerabilities Based on Authenticated Measurements

Georges El-Hajal, Roy Abi Zeid Daou, and Yves Ducq

Abstract Cyber incidents occur on a regular basis. Moreover the COVID-19 pandemic pushed most enterprises on the path of digital transformation thus expanding the attack surface for cyber threats. The severity of the risks posed by cyber threats makes cyber security a key element of their transformation plans. Senior leadership of these enterprises doesn't want to be the next cyber incident, yet it has to run their work with limited resources. They want to do just the necessary actions to mitigate threats but not be excessive in the implementation. To choose the best plan of action, the risk must be identified and prioritized. This paper develops a procedure to calculate the Risk Score (RS) of detected cyber threats and to produce a Priority Score (PS) to rank the threats that will clarify the course of action needed to achieve the integrity, confidentiality and availability of data inside the enterprise.

Keywords Cyber threat · Cyber classification · Cyber risk management · Automation in cyber security · Risk assessment · CVSS

1 Introduction

The vast reliance on connected devices made them a critical asset. Data and its availability has become the world most valuable resource [1]. Based on this fact, threat actors are creating more and more malware to infiltrate and steal data from devices. In case of failure to acquire data, attackers are trying to make it unavailable by Distributed Denial-of-Service (DDoS) attacks [2]. Security experts try to deter

G. El-Hajal · Y. Ducq
IMS Laboratory, UMR 5218 CNRS – Tssalence, University of Bordeaux, Bordeaux, France
e-mail: georges.el-hajal@u-bordeaux.fr

Y. Ducq
e-mail: yves.ducq@u-bordeaux.fr

R. Abi Zeid Daou (✉)
Biomedical Technologies Department, Lebanese German University, Jounieh, Lebanon
e-mail: roydaou@mart-ler.org; r.abizeiddaou@lgu.edu.lb

Education and Research Center, MART Learning, Chananir, Lebanon

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_9

these threats by using existing software like antivirus and firewalls, and by proposing new methods that will be detailed later on. A lot of work has been made in this field, yet the number of security incidents is on the rise day after day [3].

Along the years, malware (Malicious Software) acquired different names based on their purpose and behavior such as adware, spyware, virus, backdoor, Command and Control (C&C) bot, worm, Trojan, rootkit and ransomware [4]. Cyber-attacks do not discriminate among governments and companies. Data breaches happen at any time and anywhere, and data can be sold or bought in the DarkNet [5].

Based on these studies, it is becoming more and more difficult for a network administrator to cope with the enormous number of cyber threats. The need for an effective and simple methodology is necessary to face an active security risk facing the enterprise. Therefore, vulnerability risk assessment is performed to select the one with the highest corresponding risk as a priority for network security reinforcement [6]. Traditionally the FIRST's Common Vulnerability Scoring Systems (CVSS) [7] is primarily the basis for vulnerability risk assessments [8]. The CVSS is an open framework for communicating the principle characteristics and the severity of software vulnerabilities. The CVSS metric ignores the impact of the vulnerability in a specific network, which leads to identical vulnerability values for different network environments [9]. In fact, a little survey of the CVSSv2 scores will show that until the end of 2020, there are 8330 CVEs with a score of 10 (HIGH) and 193 CVEs have a score of 10 (CRITICAL) in CVSSv3 scores. So the scores cannot be used alone as a scale for risk priority.

To improve the security of the Information Technology environment, the novelty of this paper is to propose a novel procedure to calculate the Risk Score (RS) of detected cyber threats and to produce a Priority Score (PS) to rank the threats. CVSS scores are used in conjunction with other metrics to produce a new score that can be more significant to enhance cyber threat classification.

Thus, this paper will be divided as follows: Sect. 2 will present the different metrics that will be used in the proposed procedure and the applied methodology to acquire them. Section 3 will show the implementation of the procedure and the analysis of the obtained results will be proposed in Sect. 4. At the end, a conclusion and some future ideas will be proposed in order to enrich this work.

2 Proposed Procedure

The objective is to classify existing risks in such a way to remediate the ones that can be solved and to embrace those whose consequences are acceptable. Our proposed solution aims at providing a simple and reliable procedure that can help administrators estimate a scoring probability of vulnerabilities. This scoring is being calculated referring to several databases that are being updated continuously by their owners. Here below are the different resources that are used to calculate the vulnerability score:

1. Common Vulnerability Scoring Systems (CVSS): they have been in wide use in vulnerability management programs for more than a decade. CVSS attempts to assign severity scores to vulnerabilities, allowing responders to prioritize responses and resources according to threat. The MITRE corporation [10] gives every publicly known information security vulnerabilities a Common Vulnerabilities and Exposures (CVE) identifier which make it easier to share data across separate network security databases and tools.

While CVSS scores can and should be an important part of your vulnerability management program, it is important to keep in mind that widely published CVSS scores for a vulnerability can be misleading, as these typically represents the base score only. Thus, the base score (V2/V3) will constitute a part of the total score of the vulnerability classification introduced in this paper. Data is available through an API for online access or by downloading the entire National Vulnerability Database (NVD) for offline use which is provided by the National Institute of Standards and Technology (NIST) data feeds [11]. NVD contains all the CVEs and their corresponding metrics and offers the vulnerability data feed using the JSON format. This paper used the offline files in order to have more flexibility in extracting the values and producing the new score. We extracted the base scores of both CVSSs, CVE published date and counted the number of citations associated to the CVE. Two important parameters can be also extracted: the Patch tag that indicates if the vulnerability has a patch or not, and the Vendor Advisory tag which indicates if the vendor has declared an advisory for this vulnerability or not. Table 1 represents a sample of the final extracted values for CVE-2018–19,458.

2. The Exploit Database (EDB): It is an archive of public exploits and corresponding vulnerable software [12]. It is maintained by Offensive Security, an information security training company that provides various Information Security Certifications as well as high end penetration testing services. The Exploit Database is a non-profit project that is provided as a public service by Offensive Security. It is used by penetration testers, vulnerability researchers, and security fanatics. It reports vulnerability for which there is a proof-of-concept exploit. EDB is considered as the white market for exploits. According to surveys, it is also preferred by many contemporary exploit developers [13]. It provides an updated CSV file in its GitHub repository of all its exploits to be used offline [14]; however, this file contains the ExploitId and the exploit date but not the CVE related to the exploit. So to get the CVE connected to the corresponding exploit, we had to scrape the website in order to correlate the ExploitID with the corresponding CVE then calculate the number of exploits for each CVE and the

Table 1 Data extracted from NVD for CVE-2018–19,458

CVE	Published	BaseScore V2	BaseScore V3	Nbr citations	Patch	Vendor advisory
CVE-2018–19,458	2018–11-22T20:29Z	5	7.5	2	FALSE	FALSE

Table 2 Measures extracted from EDB for CVE-2018–19,458

CVE	Number of exploits	Minimum date of all exploits
CVE-2018–19,458	1	05/11/2018

Table 3 Measure extracted from CIRCL for CVE-2020–17,518

CVE	Count of CAPEC
CVE-2020–17,518	10

date of the first published exploit in case of multiple exploits. Table 2 represents a sample of the final extracted values.

3. Computer Incident Response Center Luxembourg (CIRCL): It is a government-driven initiative designed to provide a systematic response facility to computer security threats and incidents [15]. CIRCL provides a contextual feed containing all software vulnerabilities including visibility ranking in Luxembourg. The data feed originates from the aggregated data-sources. It has an API to provide Online queries and it provides a daily CVE JSON file to be used Offline [16]. We extracted the number of Common Attack Pattern Enumeration and Classification (CAPEC) [17] associated to each CVE. Table 3 represents a sample of the final extracted values.

After describing all the resources, we will present the five parameters used to calculate the new vulnerability score and how they are created from the aggregation of different values extracted from the three already listed sources. Thus, Table 4 contains the detailed calculations, definitions and weights of these parameters.

Calculating the vulnerability score is only the first step to prioritize their risk. To improve the visibility, we relied on the presence of a patch or a vendor advisory for the CVE (as was discussed previously in the CVSS paragraph). Using the SVS score from Table 4, the priority score is obtained after adding a certain constant as shown in Table 5. The highest value is assigned when both tags (i.e., Patch and Vendor Advisory) are available.

Referring to Pareto Principle [18] or 80/20 rule, it considers that about 80% of effects are triggered by 20% of causes. For administrators, the main takeaway from this rule is that not all cyber security risks are equal. Therefore, security resources should be devoted to the risks that are likely to cause the most damage to the enterprise. So, based on the priority score obtained, if we have, for example, 20 vulnerabilities the remediation of top 4 will lead to the elimination of 80% of the total system risk.

Table 4 Classification of the CVE based on the proposed technique

Parameters	Description		Weight
Average Base Score (BS)	The average of the base scores from CVSSv2 and CVSSv3; if this latter doesn’t exist, BS will consist of CVSSv2 only		60%
nbrCitations (NC)	Internet links that provides additional information about the vulnerability		2.5%
Pub–dateExploit (PDE)	Difference between the Publication Date of the CVE and the date of the first available Exploit. According the value of the difference a value will be used:		5%
	Less than 7 days	4	
	Between 8 and 30 days	3	
	Between 31 and 365 days	2	
	More than 365 days	1	
ExploitExists (EE)	(TRUE = 1,FALSE = 0) if the CVE has at least one exploit in EDB		30%
CountCAPEC (CC)	The number of Common Attack Pattern Enumeration and Classification (CAPEC) ids associated with every CVE		2.5%
Vulnerability score (VS)	= 0.45*BS + 0.05*NC + 0.25*EE + 0.1*CC + 0.15*PDE		
Standardized vulnerability score (SVS)	= VS /Max(VS)		

Table 5 Priority score constants

Priority score (PS)	Value added to SVS
Patch and vendor advisory exists	3
Patch only exists	2
Vendor advisory only exists	1

3 Implementation of the Proposed Procedure

In this section, we will present in details the procedure used. Figure 1 represents the methodology used in order to implement the proposed procedure. As it shows, the first step is to acquire the files from the three sources cited previously in Sect. 2. Python scripts were used to automate the tasks. Script one downloaded the JSON files from the NVD website. Every file was parsed to extract the name of the CVE, its publish date, the CVSSv2 base score, the CVSSv3 base score, the number of citation, the presence of the tags Patch and Vendor Advisory. Script two downloaded the CSV files from the GITHUB of EDB. The files were parsed for the ExploitDB, date of exploit. Then, for every ExploitDb, the EDB page related to this ExploitID was scraped for the CVE linked to it and then calculate its number of exploits and the date of the first published exploit in case of multiple exploits. Added to that, Script

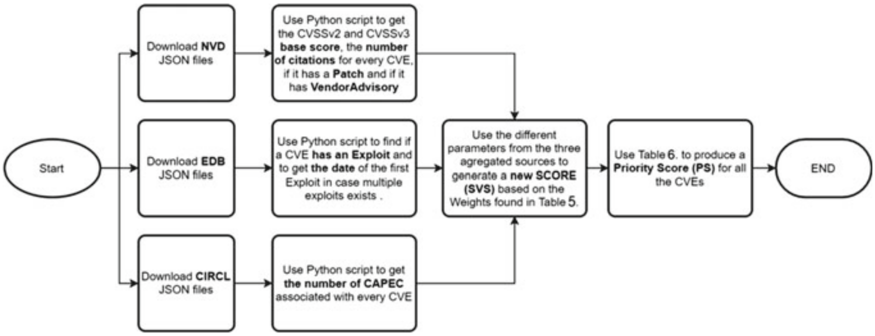

Start
Download NVD JSON files
Download EDB JSON files
Download CIRCL JSON files
Use Python script to get the CVSSv2 and CVSSv3 base score, the number of citations for every CVE, if it has a Patch and if it has VendorAdvisory
Use Python script to find if a CVE has an Exploit and to get the date of the first Exploit in case multiple exploits exists .
Use Python script to get the number of CAPEC associated with every CVE
Use the different parameters from the three agregated sources to generate a new SCORE (SVS) based on the Weights found in Table 5.
Use Table 6. to produce a Priority Score (PS) for all the CVEs
END

Fig. 1 Procedure used to calculate the proposed priority score

three, downloaded the files from CIRCL and extracted the count of CAPEC per each CVE. At this point, the three datasets were merged to create one MS Excel file.

In Microsoft Excel, we calculated the difference in days between the date the CVE was published and the date the first exploit of this CVE was announced. A constant value is assigned as shown in Table 4. Then we created a parameter indicating if the CVE has an exploit and gave it a value of 0 for no exploit found, and 1 if it has at least one exploit. We started the dataset with 148,789 entries, but before applying the formula, we removed 208 entries because they had no scores at all but we kept the 23 CVEs with CVSSv2 score equal to zero to maintain the integrity of the dataset. It is to note that out of the remaining CVEs, 73,688 don't have a CVSSv3 score which is due to the fact that it was released in 2015. After applying the formula to the CVEs, we obtain the VS. We standardize the values to obtain the SVS. Afterwards, based on the presence of the tags patch and/or Vendor Advisory, a new score will be acquired. Sorting, in descending order, the PS will give a priority list that can be used by the administrator. Of course, we applied the formula on all the CVEs as a proof-of-concept. However, for a real system, the procedure will be applied on the CVEs that the administrator detects in his environment to obtain a list with priority scores which can help in deterring the cyber risks in an organized manner.

4 Analysis of Results

Figure 1 represents the ten highest CVEs (we have shown only the first 10 because of space limitation) based on the proposed Priority Score while showing the CVSSv2, CVSSv3 and the SVS. The higher the SVS, the more serious the vulnerability is, and the higher the PS, the easier the vulnerability can be addressed. One can notice that the CVE-2014–0224, with a CVSSv2 score of 5.8 and a CVSSv3 score of 7.4, has the highest classification in both scores. This can be explained by the fact that this is a serious vulnerability in OpenSSL. It allows man-in-the-middle attackers to hijack sessions and obtain sensitive information, via a crafted TLS handshake. Yet, based on its Priority Score, we can conclude that it is easily remediable since both tags are available.

As for CVE-2008–1447, known as the Kaminsky Bug, it is a DNS vulnerability that allowed attackers to send users to malicious sites and impersonate any legitimate website and steal data. It is classified as number 8 on the list of the Top Ten Worst Vulnerabilities by infosecurity-magazine [19].

Concerning CVE-2014–3566, it is a vulnerability in the SSL protocol 3.0 which makes it easier for man-in-the-middle attackers to obtain clear-text data via a padding-oracle attack, aka the "POODLE" issue.

In addition to the fact that the classification of the CVEs is becoming more accurate, this new scoring technique eliminates the redundant scores which were widely present in the CVSSv2 and CVSSv3 (as already mentioned in the introduction part). Thus, one can notice that the first ten vulnerabilities have unique scores. which is not the case of the two above scores; these new values can facilitate the autonomous execution of remediation procedures to reduce cyber security risks.

As for the execution time to run all this procedure (as proposed in Sect. 3), it needs few minutes to classify and sort all vulnerabilities; thus, calculating the threats in a certain system will require no time.

5 Conclusions and Future Works

In this paper, a novel technique was proposed to classify and sort threats in order to propose a priority list of vulnerabilities that can be remediated for the sole purpose of increasing cyber security in enterprises in the easiest and fastest way. To do so, five parameters and two tags obtained from three authenticated datasets (NVD, EDB and CIRCL) were used. A new score, based on the cumulative weighted score of the previous parameters, was defined for each CVE. Then, a sorting was made in such a way to define the vulnerabilities that must be solved at a first stage. Results have shown that almost each CVE has a unique score, which is not the case of the today's used classification, i.e. the Base Score V2 and V3. Added to that, some vulnerabilities, that were not highly classified in these two methods, have a highest ranking in the new scoring technique as shown in the previous paragraph.

As for the future works, some other parameters may also be added to increase accuracy. Added to that, automation can be fully implemented in order to link the CVEs scores to the chatbots for immediate notification and to allow a fast course of action in case of serious vulnerabilities.

References

1. Economist T (2017) https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data
2. Wreski D (2012) https://linuxsecurity.com/news/intrusion-detection/ddos-attacks-against-us-banks-peaked-at-60-gbps
3. Galov N (2019) https://www.cyberdefensemagazine.com/cyber-security-statistics-for-2019/
4. Aycock J (2006) Computer viruses and malware. Springer, Boston
5. Biryukov A, Pustogarov I, Thill F , Weinmann R (2014) Content and popularity analysis of tor hidden services. In: 34th international conference on distributed computing systems workshops (ICDCSW), Madrid
6. López D, Pastor O, Villalba LG (2013) Dynamic risk assessment in information systems: state-of-the-art. In: Proceedings - 6th international conference on information and communication technology
7. https://www.first.org/cvss/
8. Houmb SH, Franqueira VNL, Engum EA (2010) Quantifying security risk level from CVSS estimates of frequency and impact. J Syst Softw 83(9):1622–1634
9. Wang W, Shi F, Zhang M, Xu C, Zheng J (2020) A Vulnerability risk assessment method based on heterogeneous information network. IEEE Access 8:148315–148330
10. https://www.mitre.org/
11. N. I. o. S. a. T. (NIST). https://nvd.nist.gov/vuln/data-feeds
12. https://www.exploit-db.com/
13. Fang M, Hafiz M (2014) Discovering buffer overflow vulnerabilities in the wild: an empirical study. In: Empirical software engineering and measurement (ESEM 2014), Torino
14. https://github.com/offensive-security/exploitdb
15. https://www.circl.lu/
16. https://www.circl.lu/opendata/
17. https://capec.mitre.org/
18. Gittens M, Kim Y, Godwin D (2005) The vital few versus the trivial many: examining the Pareto principle for software. In: 29th annual international computer software and applications conference (COMPSAC'05), Edinburgh
19. Raywood D (2020) https://www.infosecurity-magazine.com/magazine-features/top-worst-vulnerabilities/

Extension to CryptDB with Support for Arithmetic Expressions

Karthik Jagilinki and Ray Kresman

Abstract Growth of cloud infrastructure has made it attractive for enterprises to rethink how and where to position their resources. CryptDB (Raluca Ada Popa, et al.: CryptDB: Protecting Confidentiality with Encrypted Query Processing. MIT CSAIL, SOSP '11, October 23–26, 2011) allows storage of cloud data in encrypted form and helps mitigate privacy concerns. It performs encryption of data in layers without revealing plain-text data to the cloud vendor. While CryptDB supports simple queries, it does not appear to handle queries with arithmetic expressions. We discuss the pros and cons of a couple of schemes to address this issue and finally propose a component-based approach that provides support for queries with arithmetic expressions.

Keywords Encryption · Cloud · Service provider · Onion encryption

1 Introduction

With the advent of cloud computing, variety of services are offered by several cloud vendors in terms of infrastructure, platform, and software [1, 2]. This frees enterprises from the job of maintaining servers and associated resources. Applications routinely use databases and several new applications are emerging with databases as backend. Database as service is getting popular with service providers; they provide services to store, manage, maintain, and administer large amounts of business data in cloud databases [3, 4]. This comes with additional advantages such as availability, scalability, and usage-based cost models [5]. High availability can be ensured through replication in geographically distributed locations [6]. SQL Azure cloud database service, for example, allows partitioning of databases either horizontally or vertically using their in-house elastic tools. These and other features from service providers

K. Jagilinki · R. Kresman (✉)
Department of Computer Science, Bowling Green State University, Bowling Green, OH, USA
e-mail: Kresman@bgsu.edu

K. Jagilinki
e-mail: Jagilinki@bgsu.edu

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_10

make cloud databases more attractive and help entice customers to move even critical data to the cloud.

However, for customers to move their critical data to cloud, service providers must address the privacy concerns of their clients. Security of their sensitive data is a critical factor for customers in deciding whether to outsource data to these untrusted third-party cloud vendors. One approach is to store the data in encrypted form and not provide the encryption key to the cloud provider. The idea is to encrypt the data on client side before migration to the service provider and store the data in encrypted form on the untrusted server.

CryptDB employs such an approach [7, 8]. It allows storage of data in encrypted form on the cloud/untrusted server. Queries are performed on the encrypted data without fully decrypting the data on the server. It works by performing SQL-aware encryption schemes in layers of onion to encrypt the data. And the data is decrypted only on the client side with the server having no access to the key. This makes CryptDB a practical solution for querying encrypted data. While it supports simple queries, the version that we had access to could not handle queries with arithmetic expressions. This paper describes a generic extension to the CryptDB framework to address this issue.

2 Encrypted Cloud Storage

CryptDB [7] follows a novel approach in processing encrypted data. Their approach suggests use of (SQL) operator-specific encryption schemes. These SQL-aware encryption schemes are then used to encrypt a single data value in layers. It employs multiple encryption schemes stacked into layers (or onions) that are chosen based on the operators in SQL queries. This approach allows query processing on the server side without fully decrypting the cipher text. It eliminates the false positives that are common with other approaches [9]. In terms of efficiently executing queries over encrypted data, CryptDB has low overhead on the client side compared to other approaches. Use of onions of encryption appears to be novel and practical and provides adequate confidentiality.

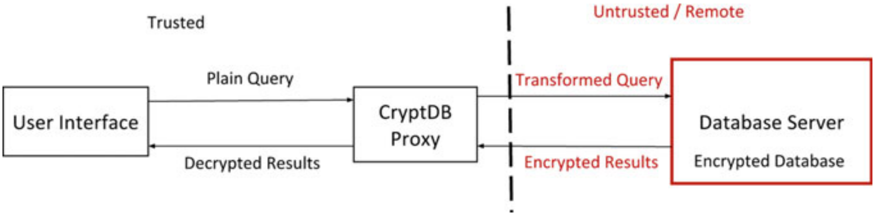


Fig. 1 CryptDB architecture

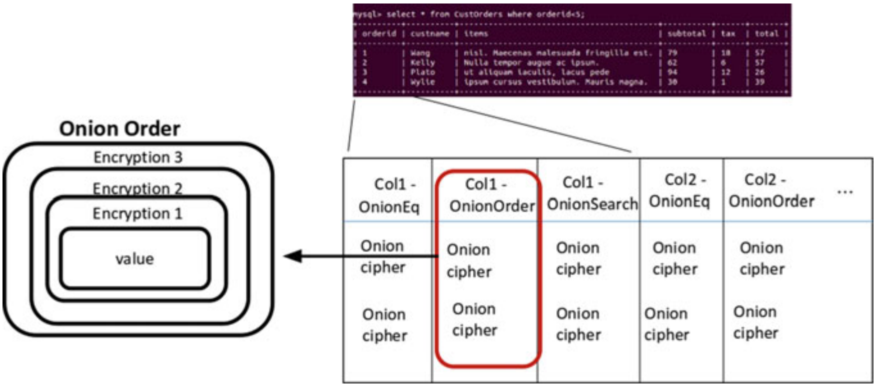
The model is depicted in Fig. 1, with a proxy between user interface and database server. The proxy is located on the client side and is trusted, while the database server is on the remote server in an untrusted location. Therefore, the data is always stored in encrypted form on the database server. The proxy located on the client side is lightweight and does not store any data. The only information that the proxy stores is schema and the master key.

Whenever the user issues a plaintext query from the interface to the database server, CryptDB proxy intercepts the query, analyzes it and then anonymizes the sensitive information in the query by performing certain encryption based on the operators in the query. The proxy performs two steps. First, it transforms the query by rewriting it with the anonymized names of columns and tables, and then sends it to the database server. Second, it strips off a layer of onion on the database server based on the intercepted query by invoking a user defined function. This way, based on the query, the proxy transforms the query and keeps the data on the server at the same level. The server processes the transformed query on the encrypted database and then returns the encrypted results back to the proxy. Note that the server works in the encrypted domain and cannot decrypt the underlying data. The proxy then decrypts them and sends the query result, in plaintext, to the user. The proxy does not perform any query execution, instead the query is fully executed on the database server. CryptDB utilizes server-side processing more and has less client-side overhead.

2.1 Encryption and Storage at the Backend

CryptDB has various encryption schemes based on the operators used in SQL queries. Figure 2 outlines the data encryption process on the database server.

As shown in Fig. 2, each column data item is encrypted using a different set of encryption schemes. Also, for the sample query, we see that a column orderID is



```
mysql> select * from CustOrders where orderid<5;
+---------+----------+-----------------------------------------+----------+-----+-------+
| orderid | custname | items                                   | subtotal | tax | total |
+---------+----------+-----------------------------------------+----------+-----+-------+
| 1       | Wang     | nisl. Maecenas malesuada fringilla est. | 79       | 18  | 57    |
| 2       | Kelly    | Nulla tempor augue ac ipsum.            | 62       | 6   | 57    |
| 3       | Plato    | ut aliquam iaculis, lacus pede          | 94       | 12  | 26    |
| 4       | Wylie    | ipsum cursus vestibulum. Mauris magna.  | 30       | 1   | 39    |
+---------+----------+-----------------------------------------+----------+-----+-------+
```

Col1 - OnionEq	Col1 - OnionOrder	Col1 - OnionSearch	Col2 - OnionEq	Col2 - OnionOrder ...
Onion cipher	Onion cipher	Onion cipher	Onion cipher	Onion cipher
Onion cipher	Onion cipher	Onion cipher	Onion cipher	Onion cipher

Fig. 2 Query processing

encrypted and stored in three different sets i.e. Col1-OnionEq, Col1-OnionOrder, and Col1-OnionSearch. Based on the operators present in the query, any of the encrypted dataset can be used to process the query. OnionEq, OnionOrder, and OnionSearch as the names suggest are used to perform query processing based on the different (SQL) operators equal, order, and search respectively. For example, in Fig. 2, the user issued the query: select * from CustOrders where orderID < 5. Here, we have an order operator < . Once the proxy sees this operator in the query, it will invoke the user defined function to map the user query to the OnionOrder encrypted dataset. As noted in Fig. 2, each of these data sets consists of an onion cipher that had encrypted the column value using multiple encryption schemes in layers i.e. Encryption 1, Encryption 2, and Encryption 3. Each onion cipher consists of data value being encrypted with these multiple encryption schemes in layers. Here, the functionality of lower layer is strictly higher than the functionality of higher layers to help ensure overall security of data.

2.2 Query Processing in the Encrypted Domain

Whenever the client (or the interface) issues a query, the proxy intercepts it and transforms the query based on the type of operators it has. For instance, if the query has equality operator then Deterministic encryption scheme is used. And if the query has order operator to check less than or greater than, the Order-Preserving scheme will be used to perform encryption to the query.

Invoke UDF to adjust the level of onion. Different queries coming from client may perform different computation on server data. Based on the user query, and proxy invokes a user defined function (UDF) to adjust the database encryption level on the backend. Whenever certain UDF is invoked based on the user query, proxy would give the key to the SQL function on database server for the specific layer of the onion. This would simply strip off certain layer of encryption. This way proxy processes queries over encrypted data without giving master key to the server. Thus, as noted earlier, the data on the server is always in encrypted form.

3 Proposed Framework

CryptDB does not appear to support queries with arithmetic expressions. This is because expressions are encrypted with different encryption schemes which may not be compatible with one another. To illustrate this problem, let us invoke a query with an arithmetic expression. The query and the error message/CryptDB response is this: *mysql> select subtotal + tax from CustOrders where total < 20;* **ERROR 1105 (0700): Current crypto schemes do not support this query**

The query, *Select subtotal + tax from CustOrders where total < 20* has an arithmetic expression *subtotal + tax*. Here subtotal and tax are two different columns. The data of these columns are stored in the backend encrypted at their highest secured level. When the user gives this query, CryptDB proxy intercepts it and sees that the query has an arithmetic expression, which is trying to perform an addition operation. Initially these columns are encrypted at layer 3. As addition operation is not possible at Encryption layer 3, proxy gives the partial key of layer 3 to the backend to decrypt a layer of onion for column subtotal from Encryption layer 3 to Encryption layer 2. At this layer 2, CryptDB uses homomorphic encryption scheme that is suitable to perform addition operation over encrypted data. However, the problem here is CryptDB proxy only decrypts the first column subtotal to Encryption layer 2, but it still keeps column tax at Encryption layer 3. Clearly, addition operation is incompatible between the two columns that belong to different encryption schemes. Our extension to correctly handle such queries is discussed next.

3.1 Tweak CryptDB UDFs

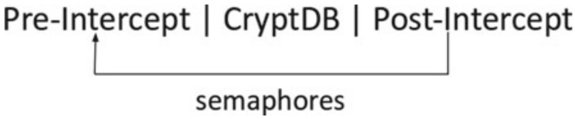
Given the complexity of systems such as CryptDB, we attempt to address the question of resolving queries with arithmetic expressions [10, 11] using a component-based approach. Our goal is to add a component that would interface with the underlying software architecture of CryptDB and yet respect its (CryptDBs) black-box nature.

One approach is to update CryptDB UDF to support queries with arithmetic expressions. When the UDF is triggered, if we could modify the UDF in a way to decrypt multiple columns that are present in the query at the same time, then that would help keep these columns on the server at the same layer allowing one to perform a valid operation, such as addition, over encrypted data. A second approach is to include additional encryption schemes to each of the onions, to help provide support for queries with arithmetic expressions. Unfortunately, neither approach respects the black-box nature of CryptDB components and so we set out to explore alternate mechanisms.

3.2 Piped Architecture

A third possibility is the piped architecture as shown in Fig. 3. The vertical bar means pipes between adjacent subsystems. The Pre-Intercept stage provides input to CryptDB and the Post-Intercept receives the results returned from CryptDB. The Pre-Intercept receives the user's query and does some pre-processing on it before

Fig. 3 Piped architecture



shipping it to CryptDB. If the query has no arithmetic expression, it is passed as it is.

However, for a query with arithmetic expressions, a simple approach is to split the user's query into multiple queries based on the arithmetic expression. Then, the queries are issued, one-by-one, to CryptDB. Post-Intercept can then combine the results of each of these split query results returned by CryptDB and generate the composite result (or response) to the original user's query. As shown in Fig. 3, we need some coordination between the Pre-Intercept and Post-Intercept subsystems. Semaphores can be used to synchronize such communication between the two stages and help ensure that the user's query is processed in a sequential manner. The advantage of the piped architecture is that it respects the black-box nature of CryptDB components.

3.3 Integrated Architecture

The piped architecture can be modified slightly to integrate the functionality of Pre-Intercept and Post-Intercept stages in a seamless manner (into one component) to process arithmetic expressions in the query. We call this component-based approach, 'Integrated Architecture.'

The proposed system will coordinate with CryptDB in an integrated manner to process queries with arithmetic expressions. As noted earlier, it also respects the black-box nature of CryptDB components.

Figure 4 is a high-level schematic of our Integrated Architecture. Whenever the user enters queries from the interface, they flow through EnhancedCryptDB to the CryptDB proxy. The middle step does some pre-processing on the query before giving to CryptDB proxy. As noted in Sect. 2, CryptDB follows a sequence of steps to interact with backend database server maintained by the cloud provider and finally returns the query response to EnhancedCryptDB. Then, the latter component does post-processing on these results before sending the results back to the user.

EnhancedCryptDB consists of two major pieces: Query Interceptor and Query Processor. Query Interceptor acts as a query manager that handles queries from the client (or interface). Query Processor provides the primary functionality in our approach, i.e. it receives the queries from the Query Interceptor and splits the ones with arithmetic expressions. Figure 5 shows the components of our architecture to illustrate the information flow in our framework.

Query Interceptor and Query Processor are two different processes that coordinate with each other in processing the user's query. As these are two different processes, we require interprocess communication between them. Query Interceptor reads user queries from the interface and handles them one at a time. For each query, it internally calls Query Processor to handle arithmetic expressions in the query. Query Processor then rewrites the query, as needed, before sending them to the CryptDB proxy. CryptDB proxy performs its functionality (see Sect. 2.1) and returns the results that are intercepted by Query Processor. Note that these results are in plaintext. Query

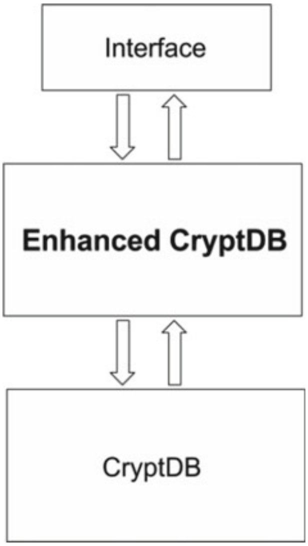


Fig. 4 Integrated architecture

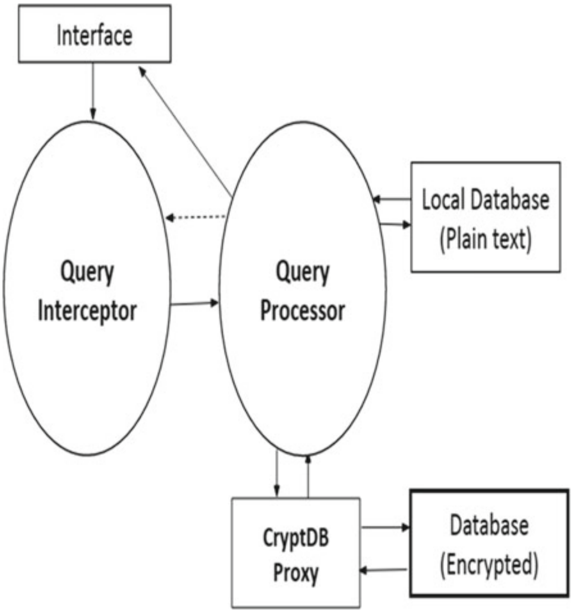


Fig. 5 EnhancedCryptDB srchitecture

Processor dumps these result tuples to a local database; it then queries this (local) database against the original user expression. The output is just the response to the original user's query and this output is returned to the user.

Our approach can handle any type of arithmetic expression in the query. Consider for example: select a + b / c from T1. Such a query will result in 3 calls to CryptDB: select a from T1; select b from T1; select c from T1. CryptDB response to these three queries is stored in a local database, say table localTable. At the end, the Query Processor does a query of the form, select a + b / c from localTable that yields the result to the original query, which is then returned to the user.

Clearly, one downside with our approach is that it is not as efficient compared to a scheme that handles the entire query in one indivisible unit. However, our approach is component-based in the sense that it integrates with CryptDB and at the same time respects the black box nature of the (CryptDB) software components. We feel that this trade-off between performance and simplicity is something for the client to consider. A second issue is security; since the local database handles data in plaintext, does it open any security holes? This work was done as a Master's project and we have not done any analysis on data leakage in our system, but we feel—assuming the proposed system sits next to CryptDB (perhaps in the same hardware as CryptDB)—that its security is perhaps comparable to CryptDB [12, 13]. In any event, these two issues are worthy of additional investigations.

4 Concluding Remarks

CryptDB lets users store cloud data in encrypted form. It supports several layers of encryption. Clients can issue SQL commands against the data; Queries are performed directly on the encrypted data and the cloud provider does not have access to the decryption key.

This paper proposed an extension framework to CryptDB to provide support for arithmetic expressions in SQL queries. The advantage of our approach is that it respects the black-box nature of CryptDB components - the proposed scheme does not affect, or is not even aware of, the internal details of various CryptDB onion layers. For brevity, software implementation details and performance evaluation results are omitted in our discussions. These details will be addressed in a future paper.

Acknowledgements We thank the reviewers for their insight into an earlier draft.

References

1. RDS (2016) Amazon relational database service (RDS) – AWS. Amazon Web Services, 2016. https://aws.amazon.com/rds/. Accessed 4 Aug 2016

2. Salesforce (2016) What is cloud computing technology? Cloud definition. Salesforce.com, 2000. http://www.salesforce.com/cloudcomputing/. Accessed 4 Aug 2016
3. Google Developers (2016) Cloud SQL—MySQL relational database. Google Developers. https://cloud.google.com/sql/. Accessed 4 Aug 2016
4. Oracle (2016) Database. https://cloud.oracle.com/database. Accessed 4 Aug 2016
5. Wikipedia (2016) SQL Azure. Wikimedia Foundation, 2016. https://en.wikipedia.org/wiki/SQL_Azure. Accessed 4 Aug 2016
6. MS Azure (2016) Scaling out with azure SQL database. 2016. https://azure.microsoft.com/en-us/documentation/articles/sql-database-elastic-scale-introduction/. Accessed 4 Aug 2018
7. Popa RA et al (2011) CryptDB: protecting confidentiality with encrypted query processing. MIT CSAIL, SOSP ‘11
8. CryptDB (2016) In: CryptDB. https://css.csail.mit.edu/cryptdb/#Software. Accessed 4 Aug 2016
9. Alwarsh M, Kresman R (2011) On querying encrypted databases. In: Proceedings of the 10th international conference on security and management, pp 256–262
10. Jagilinki K (2016) Enhanced query processing with CryptDB. Master’s project. Department of Computer Science, Bowling Green State University, Bowling Green, p 73
11. Jagilinki K, Kresman R (2020) An extension to CryptDB. IAET international conference on artificial intelligence, information systems, engineering, Budapest, Hungary. 2:1
12. Naveed M, Kamara S, Wright C (2015) Inference attacks on property-preserving encrypted databases. In: CCS ‘15: Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, pp 644–655. https://doi.org/10.1145/2810103.2813651
13. Almarwani M, Konev B, Lisitsa A (2019) Flexible access control and confidentiality over encrypted data for document-based database. In: Proceedings of the 5th international conference on information systems security and privacy - volume 1: ICISSP, pp 606–614. ISBN 978-989-758-359-9. https://doi.org/10.5220/0007582506060614

Automated Model-Based Test Case Generation Using Uml Activity Diagram: A Review

Rozi Nor Haizan Nor, Md Abdul Monim, Yusmadi Yah Jusoh, and Nur Ilyana Ismarau Tajuddin

Abstract Software or application testing is a process of executing a program with the goal of finding defect to make better system. In software testing phase, writing test cases is one of the important activities. Manually writing test cases approach is lengthy of time period and need more effort to accomplish the process. This paper describes test case, test case generation techniques, different types of software testing approaches and comparison of testing tool. Test cases usually writing at the beginning of testing process from the set of software requirements. Test cases are created by using two different approaches. One is manually written test cases and another is automatically generated test cases. Manually written test cases are a very lengthy process and need to give a lot of effort to make good quality test cases. On the other hand, automatically generated test cases are generated by automatically using some software tools and it saves a lot of time and effort.

Keywords Test case generation · Testing technique · Model based testing · Automated test case generation · UML diagram

1 Introduction

A software system or a web application had to go through into a development life cycle and testing is an important phase of this software/application development life cycle. Software testing can be done by manually or automatically. In manual

R. N. H. Nor (✉) · M. A. Monim · Y. Y. Jusoh
Department of Software Engineering and Information System, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia
e-mail: rozinor@upm.edu.my

Y. Y. Jusoh
e-mail: yusmadi@upm.edu.my

N. I. I. Tajuddin
Tamhidi Centre, Universiti Sains Islam Malaysia, Nilai, Malaysia
e-mail: nur_ilyana@usim.edu.my

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_11

testing, the testing requires doing all the process by manually includes input, 2 analysis, and writing and managing the test cases. Manual testing deals with human interaction with all the process from beginning to end of the process and it is a time-consuming process. A human can get tired of doing all the process continuously. On the other hand, in automated testing most of the testing processes are automated. Generating test cases, executing the test cases and producing the test result are done by automatically. Though, one of the software testing fundamental is hundred percent automation is not possible in the field of software testing. Some task still need the intervention of human. Based on software testing approaches there are mainly three type of testing techniques have been using in software testing, those are 8 Specification-based testing, Code-based testing, and Model-Based testing. In Specification-based testing techniques test cases are made by directly from the software specification or some other kind of document that may have a different type of flow or direction. Therefore, in Code-based testing techniques, all the statement or module of the code have to execute at least once during the testing process and that process are followed by test cases that have written based on the code [1]. Lastly, in Model-Based testing approaches, model are developed from requirements and that model is the main resource to generate the test cases by manually or automatically. Writing or generating good quality test cases is very important for the test execution process.

2 Literature Review

This section describes test case, test case generation techniques, different types of software testing approaches and comparison of testing tool. A test case is a set of conditions, test inputs or variables under which a tester will determine whether a system test satisfies requirements or works properly. Test cases usually writing at the beginning of testing process from the set of software requirements. Test case needs to be ready before starting the test execution. Testers are followed by the step as mentioned in a test case using test data and compare the output after the test execution. Test cases are created by using two different approaches. One is manually written test cases and another is automatically generated test cases. Manually written test cases are a very lengthy process and need to give a lot of effort to make good quality test cases. On the other hand, automatically generated test cases are generated by automatically using some software tools and it saves a lot of time and effort. Based on software testing approaches there are mainly three type of testing techniques have been using in software testing, those are 8 Specification-based testing, Code-based testing, and Model-Based testing. In Specification-based testing techniques test cases are made by directly from the software specification or some other kind of document that may have a different type of flow or direction. Therefore, in Code-based testing techniques, all the statement or module of the code have to execute at least once during the testing process and that process are followed by test cases that have written based on the code [1]. Lastly, in Model-Based testing approaches, model are developed from requirements and that model is the main resource to generate the test cases by manually or automatically. Writing or generating good quality test cases is very important for the test execution process.

2.1 Model-Based Testing

"Model-based Testing is a testing technique where the runtime behavior of an implementation under test is checked against predictions made by a formal specification or model" [2]. Jupudy et al. [3] presented in their paper, Model-based testing comes under the black-box testing approach in which the internal design of the system is not taken into account during the test case generation. The process for MBT starts with generating functional test models based on the software requirements then this model is used for generating the test cases. The final test execution can be done either by manual or automated test execution. For test case generation it is based on the information provided by the test models so that the models should need to include the system behavior that the tester wishes to test. Model-Based Testing is a very suitable way to represent any system. Models can be as simple as a graph, flowchart, or diagram. A model of a software or system is a portrait of its behavior where the behavior can be described in terms of the input sequences that accepted by the system, set of actions, conditions and the flow of data through the system's module as shown in Fig. 1. These models can be created using any modelling tools. There were several Model-Based Testing tool available in the market that can generate test cases using different types of input format. The list is shown in Table 1.

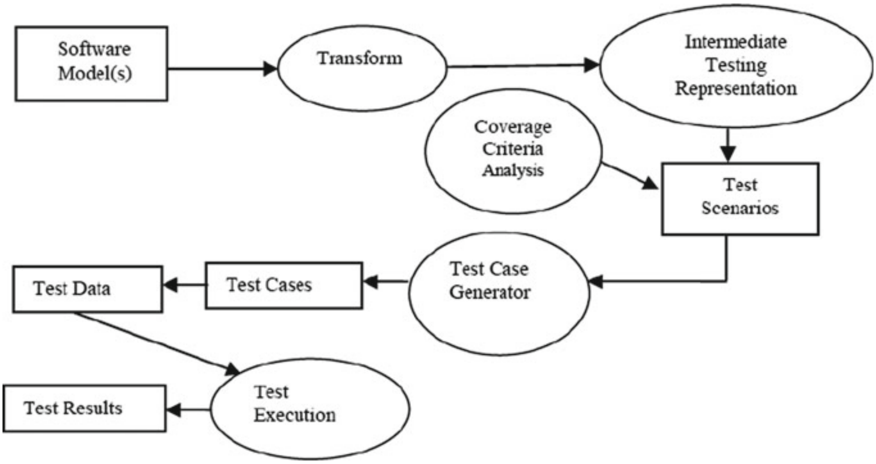


Fig. 1 A typical model-based testing (MBT) process Swain et al. [4]

Table 1 Different types of MBT tool for test case generation

No	Tool	Input format	Type	Description
1	BPM-Xchange	BPMN, UML	Commercial	BPM-Xchange creates test cases from business process models based on different criteria (statement, branch, path, condition). It can import models from several modelling tools, and can export test cases to Excel, HP Quality Center, etc
2	Conformiq designer	UML State Machines, QML	Commercial	Models can be created as UML State Machines and in Qtronic Modelling Language (QML). Test cases can be exported to test management tools or TTCN-3
3	fMBT	Custom (AAL)	Open source	fMBT (free Model-Based Testing) generates test cases from models written in the AAL/Python pre/post condition language using different heuristics (random, weighted random, look ahead)
4	Graphwalker	FSM	Open source	Test generation from Finite State Machines. Search algorithms: A* or random, with a limit for various coverage criteria (state, edge, requirement). Formerly called as MBT
5	JSXM	EFSM (Stream X-machines)	Academic	JSXM is model animation and test generation tool that uses a kind of EFSMs as its input. The generated tests can be transformed to JUnit test cases
6	MaTeLo	Markov chains	Commercial	MaTeLo (Markov Test Logic) is a commercial product to generate functional test cases. Strategies: random generation oriented by profiles, all transitions coverage. Can be connected to numerous test platforms
7	MISTA	Petri net	Academic	MISTA generates test cases from high-level Petri nets, and using a mapping it can generate executable test code for various platforms (JUnit, NUnit, Selenium). It can be used for functional or security testing
8	ModelJUnit	EFSM	Open source	Model J Unit allows to write simple finite state machine (FSM) models or extended finite state machine (EFSM) models as Java classes, then generate tests from those models and measure various model coverage metrics
9	MoMuT::UML	UML state machines, OOAS	Academic	MoMuT is a family of automated, model-based test case generation tools that can work off UML State Machines, Assume-Guarantee Contracts (REQS), and Object Oriented Action Systems (OOAS). The tools feature a fault-based test case generation strategy (using mutation operators)

(continued)

Table 1 (continued)

No	Tool	Input format	Type	Description
10	RT-Tester	UML/ SysML, Matlab	Commercial	RT-Tester starts from UML/SysML or Matlab models, transforms them to a internal representation based on Kripke structures, transform requirements to LTL formulae, and generates test cases using an SMT solver based on the goals from requirements and various model coverage criteria
11	SmartestingC ertifyIt	UML + OCL	Commercial	This tool is the successor of BZ Testing-Tools and Leiros Test Generator, and now is a commercial product from Smartesting. The SUT is given with UML models enriched with OCL constraints, the tool generates tests to satisfy various model coverage criteria
12	Spec explorer	Model programs in C#	Commercial	Spec Explorer is the successor of the AsmL. The Spec Explorer is now integrated into Visual Studio. The models can be written in C#, and test generation is directed with test 13 purposes written in the Cord language
13	Tcases	Custom	Open source	Tcases is a combinatorial testing tool where the inputs of thesystem could be specified in an XML file (with conditions, failure values, don’t cares,etc.). Test cases can generate n-wise or randomized test suites
14	Test cast	UML State Machines	Commercial	Test Cast MBT Edition generates TTCN-3 test cases from UML State Machines based on requirement and model structural coverage. The product is the successor of the MOTES prototype tool
15	Test optimal	(E) FSM	Commercial	Test Optimal supports FSM and EFSM modelling with several test case generation algorithms. It has various plug-ins for online testing web application, windows applications, database and web services, etc. It also supports data-driven testing and pair-wise algorithms right within the model and has the facility for performing load testing using the same models

2.2 Automated Test Case Generation

Rushby [5] elaborates in his paper, most of the process of software test execution and monitoring is now automated in modern software development life cycle. But for the test case generation has remained the existing labor-intensive manual task in some of the software development practice. But technology is changing by its own way, so this existing technique is also changed and now many methods or approach now become available to automate this process. If a tool can be able to generate the test cases automatically by using any appropriate input and any method then it can be called automated test case generation process. Then, Automated Model-Based test case generation is a technique that a visual model is used for generating the test cases based on some method or algorithm into that tool.

2.3 Related Works

Sumalatha and Raju [6] described in their paper, there are a lot of models and each describes different aspects of software or system behavior. Like, control flow, data flow and program dependency graph manifest how the implementation behaves by representing its source code structure. Decision table and state machines are used to describe external behavior. There are many models have been using in software testing those 14 are finite state machines, state charts, Unified Modeling Language (UML), Markov chains and grammars. Model-Based Testing (MBT) is the next step in the evolution of software testing. Model-Based testing has been proven to allow more effective work and increase attention on the substance of testing. Priya and Sheba [7] conducted a survey about test cases generation from different types of UML model and presented all the survey report in their paper. They choose five UML model that combination of UML structural and behavioral diagram those are, (i) Activity diagram, (ii) Sequence diagram, (iii) Class diagram, (iv) State-chart diagram and, (v) Collaboration diagram. They also described some test case generation techniques using those above UML diagram, some of the technique was based on single UML diagram and some of the technique was based on combination of two UML diagram. [8] presented a review study about different test case generation techniques, test case selection method, test case minimization techniques, test case prioritization techniques and some test case evaluation techniques. The principal techniques were critical path method, code based test generation, GUI based test case generation, and Dynamic path testing and evolutionary testing. And they also provided some algorithm for generating test case such as, Graph traversal algorithm, and Genetic algorithm. An Executable Test Generation from UML Activity Diagram Using Genetic Algorithm approach was proposed [9]. They described an automated test case generation techniques from UML activity diagram using Genetic algorithm. They used UML to XMI as a modelling interchange and also used an API named Robotium, based on Genetic Algorithm (Table 2).

Table 2 Summary table of some related works

No	Author	Title	Focus area/ Approach	Description
1	Kaur and Gupta [10]	Automated model–based test-path generation from UML diagrams via graph coverage techniques	-New approach for graph covering technique. -Automated tools. -Chinese postman algorithm. -Prefix based algorithm	They used a tool named Test Optimal, is an integrated next-generation test design and test automation toolset powered by Model-Based Testing (MBT) to test case generation and test automation
2	Priya and Sheba [7]	Test case generation from UML models—a survey	-Survey of five UML diagram. - Class diagram - State chart diagram - Sequence diagram - Activity diagram - Collaboration diagram.	Performed a literature review for test case generation from UML structural and behavioural diagram, test case generation by a combinational approach and, different type
3	Chouhan et al. [11]	Test case generation on the origin of activity diagram for navigational mobiles	- UML activity diagram - Model-basted testing - Mobile systems - Navigation systems	This work proposes a model for test case generation for navigation mobile application based on activity diagram. And the complexity calculated by Cyclomatic Complexity. The proposed mode introduces an algorithm that automatically creates a table called Activity Dependency Table (ADT) and then uses it to create a directed graph called Activity Dependency Graph (ADG). Finally, the ADG with the ADT is used to generate the final test cases

(continued)

Table 2 (continued)

No	Author	Title	Focus area/ Approach	Description
4	Suhag and Bhatia [12]	Model-based test cases generation for web applications	- MBT for web application - Using web diagram & sequence diagram	Test case generation technique has been applied to a web application which allows sharing of previous year papers, notes and other academic notices among each other
5	Hooda and Chhillar [8]	A review: study of test case generation techniques	Review study based on some test case generation techniques, - UML diagrams - Critical path method - Code based test generation - GUI based test case generation - Dynamic path testing and evolutionary testing - Graph traversal algorithm - Genetic algorithm	Presented a review study about different test case generation techniques, test case selection method, test case minimization techniques, test case prioritization techniques and some test case evaluation techniques

(continued)

Table 2 (continued)

No	Author	Title	Focus area/ Approach	Description
6	Jain et al. [13]	Automatic test case generation using UML models	- Automated test case generation - UML activity diagram - UML use-case diagram - Full predicate coverage criteria - XML standard for exchanging UML models	This paper describes an approach for test case generation from combination of UML Activity Diagram and Use Case Diagram. At first, AD are converted into activity Graph. Then extracting concurrent control flow path from activity graph. For CCFP they used depth first search (DFS) algorithm. Finally they use a combinational approach using use-case and activity diagram to generate test case
7	Shah et al. [14]	Automated test case generation using UML class & sequence diagram	- UML class diagram - UML sequence diagram - Model-based testing - Object oriented language - Test Automation	Visual Paradigm tool is used to create class diagram and again same tool is used to create sequence diagram. The developed diagrams then have exported into XML format. Total coding have done with C# and generated test cases saved into txt file

(continued)

Table 2 (continued)

No	Author	Title	Focus area/ Approach	Description
8	Anbunathan and Basu [9]	Executable test generation from UML activity diagram using genetic algorithm	- UML activity diagram - Genetic algorithm - Test automation - Pairwise testing	Activity Diagram is created to capture input scenarios. XMI file obtained from this AD is parsed to extract model information. A Control Flow Graph (CFG) is derived from edges by sorting the edges using Breadth First Search (BFS) algorithm. A recursive algorithm is developed to obtain path test cases from CFG. To generate test scripts, Robotium APIs are identified from Edges and Robotium API database using Genetic Algorithm
9	Patil and Jadhav [15]	Functional test case generation based on model driven testing using FSM and UML activity diagram	- UML activity diagram - Finite state machine - Model-based testing	Input FSM, Activity diagram in the form of XML. Then Process FSM ⊕ Activity Diagram. Activity Dependency Table for Activity Diagram and DFSM Graph Generator for FSM Output is test cases with all path, prioritization and removal of redundancy

(continued)

Table 2 (continued)

No	Author	Title	Focus area/ Approach	Description
10	Teixeira and Silva [16]	EasyTest: an approach for automatic test cases generation from UML activity diagrams	- UML activity diagram - Model-based testing - Test automation - EasyTest tool	Presents an automatic approach to generate test cases from UML activity diagrams using gray-box technique. The EasyTest approach comprises three phases, 1) importing activity diagrams in XMI; 2) test cases generation; and 3) applying test cases. Used Activity Dependency Graph and Activity Dependency Tree for sequencing the test path. Then applied into automated test case generation

3 Conclusion

This paper presented the background of different kind of test case generation techniques, methods, approaches, and tools were discussed. A summary table was provided based on existing similar research study. If the software testing stage can be finished early in software development life cycle then the total development process will be shortened and the software product is possible to deliver early. Manual test case writing approach is lengthy of time period and need more effort to accomplish the process. Manually have to write all test cases from the requirement and then execute the test cases are also done by manually. Writing test cases from the requirement is a very long process, boring and error-prone. Hence, automated test case generation is the way to solve this issue. Therefore, this paper contains an overview of test case generation and testing technique, Model-Based Testing, automated test case generation, comparison of different existing Model-Based Testing tools, related works that include a table of summary of some related previous works.

References

1. Kaushik S, Tyagi K (2016) Critical review on test case generation systems and techniques. Int J Comput Appl 133(7):24–29. https://www.ijcaonline.org/archives/volume133/number7/23798-2016907916
2. Sharma HK, Singh SK, Ahlawat P (2014) Model-based testing: the new revolution in software testing. Database Syst J 4(1):26–31
3. Jupudy I, Saraf N, Manjula R (2016) Comparative analysis of model based and formal based software testing methods. Int J Adv Res Comput Sci Softw Eng 6(3):49–58
4. Swain SK, Pani SK, Mohapatra DP (2010) Model based object-oriented software testing. J Theor Appl Inform Technol Vol-14. http://www.jatit.org/volumes/research-papers/Vol14No1/4Vol14No1.pdf
5. Rushby J (2007) Automated test generation and verified software. In Conference on verified software: theories, tools, and experiments (VSTTE). Zurich, Switzerland, pp 161–172
6. Sumalatha VM, Raju GSVP (2012) UML based automated testcase generation technique using activity-sequence diagram. Int J Comput Sci Appl (TIJCSA) 1(9):58–71
7. Priya SS, Sheba PD (2013) Test case generation from UML models–a survey. Int Conf Inform Syst Comput (ICISC) 3(1):449–459. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.414.104&rep=rep1&type=pdf
8. Hooda I, Chhillar R (2014) A review: study of test case generation techniques. Int J Comput Appl 107(16):33–37. https://www.ijcaonline.org/archives/volume107/number16/18839-0375
9. Anbunathan R, Basu A (2017) Executable test generation from uml activity diagram using genetic algorithm. Int J Comput Sci Inf Technol Secur (IJCSITS) 7(3):1–6
10. Kaur P, Gupta G (2013) Automated model-based test path generation from UML diagrams via graph coverage techniques. Int J Comput Sci Mob Comput 2(7):302–311. https://ijcsmc.com/docs/papers/July2013/V2I7201365.pdf
11. Chouhan C, Shrivastava V, Sodhi SP, Soni P (2013) Test case generation on the origin of activity diagram for navigational mobiles. Int J Adv Comput Eng Netw 1(1):32–36. http://iraj.in/journal/IJACEN/paper_detail.php?paper_id=21&name=Test_Case_Generation_on_The_Origin_of_Activity_Diagram_For_Navigational_Mobiles
12. Suhag V, Bhatia R (2014) Model based test cases generation for web applications. Int J Comput Appl 92(3):23–31. https://doi.org/10.1007/978-3-319-09153-2_19

13. Jain SP, Lalwani KS, Mahajan NK, Gadekar BJ (2014) Automatic test case generation using UML models. Int J Adv Comput Eng Netw 2(6):30–34. https://ieeexplore.ieee.org/abstract/document/4418295
14. Shah S, Shahzad R, Bukhari S, Humayun M (2016) Automated test case generation using UML class & sequence diagram. Br J Appl Sci Technol 15(3):1–12
15. Patil SS, Jadhav PA (2017) Functional test case generation based on model driven testing using FSM and UML activity diagram. Int J Adv Res Comput Sci 8(5):1527–1530. https://search.proquest.com/openview/cbc76318c598ff45d1a9fe6cb762447c/1?pq-origsite=gscholar&cbl=1606379
16. Teixeira FA, and Silva GB (2017) EasyTest: an approach for automatic test cases generation from UML activity diagrams. Adv Intell Syst Comput Inform Technol New Generat pp 411–417. https://doi.org/10.1007/978-3-319-54978-1_54

Advanced Android Malware Detection Utilizing API Calls and Permissions

Md Naseef-Ur-Rahman Chowdhury, Qudrat E. Alahy, and Hamdy Soliman

Abstract The Android operating system is a major presence in the proliferation of smartphones and IoT applications. Android apps can utilize security loopholes wherein developers may access user-critical data on the host device. A previously published Android malware-detection model analyzed 109,000 APKs, achieving better results than other peer models in accuracy, precision, recall, and F-Score metrics. In this paper, the model is expanded through the addition of API-call analysis, along with APK permissions. This expansion enabled more powerful and improved detection accuracy. Moreover, in an analysis of 158,000 APKs, the more recent model with newer settings achieved much better results than prior work on the same set of performance metrics. These results are an encouraging indication that further expansion of dynamic APK analysis will permit early detection of malware installation, allowing vulnerable Android systems to preempt damage from the malware application vector.

Keywords Malware detection · Machine learning · Android · Security · Android API · Android permissions

1 Introduction

Android is considered one of the most popular smartphones operating systems worldwide, as it is open-source and portable to any UNIX-based device [1]. Android applications are published by Google with minimal security screening [2]. Bad actors can take advantage of this and publish malicious Android apps to the Google Play Store. Thus, the need arises for an accurate Android Malware Detection Mechanism (AMDM). In general, Android applications utilize system resources via invocation of the system API [3], whose access is granted via permissions. Notably, malicious

M. N.-U.-R. Chowdhury (✉) · H. Soliman
New Mexico Tech, 801 Leroy Pl, Socorro, NM 87801, USA
e-mail: naseef.chowdhury@student.nmt.edu

Q. E. Alahy
Boise State University, 1910 W University Dr, Boise, ID 83725, USA

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_12

applications use sensitive APIs and permissions more frequently than benign apps [4]. Therefore, aside from Google system apps, the use of sensitive, high-risk APIs and permissions can be utilized in the process of detecting malicious apps. However, there are many such API calls for Android, so the challenge is to select specific calls as a reference for detecting malware [5].

Many AMDMs are mostly based on content signatures, which compare an app's signature to a database of known malware signature definitions [6]. Such a detection model can work to detect only well-known, traditional malware types, and not the latest, advanced versions [7]. This is a known problem for researchers, who indicate signature-based approaches cannot keep up with the pace of new malware development [8]. Hence, there is an urgent need to research and develop solutions for the alleviation of this problem. Examples of newly-developed, effective solutions include behavioral-based static or dynamic analyses [8]. One of the most popular static, behavioral-based methods of AMDM is based on analyzing the requested permissions list and resource usage, e.g., Location Services, Contact Information, WiFi, etc. [9]. A more powerful approach is dynamic, behavioral-based analysis, which observes the real-time behavior of running Android applications, i.e., analyzing on-demand API calls and capturing the live activity of the application [10]. However, due to the nature of dynamically monitoring a running program, the degree of automation and real-time processing required is relatively high. Moreover, the live analysis must ensure detection occurs prior to malicious code damaging the system. For these reasons, dynamic detection technology requires more resources than its static counterpart [11].

Our proposed methodology is based on tracing the type of system API calls invoked by potentially malicious applications. We developed a new Smart AMDM (SAMDM) which utilizes different machine learning (ML) classifiers for more intelligent detection of malicious apps. Android system API calls and associated permissions were extracted from more than 158,000 APKs collected from different, popular malware data sources, such as Drebin [12], MalGenome [13, 14], Marvin [15], and Virustotal [16].

In the current literature, researchers used either permissions-based or system API-based mechanisms. In this paper, we combine both mechanisms as a hybrid to establish SAMDM. The details of our technique, experimental approaches, and results are described later in this paper. The main goal was to accurately and quickly detect malware via the utilization of ensemble learning over the six involved ML models. The major contributions of this paper are listed below:

- We proposed an efficient Android Malware detection mechanism based on API calls and permissions utilizing ensemble learning algorithms.
- Experimental results showed that our SAMDM model achieved better performance than peer models while reducing the complexity of the detection process.

The remainder of the paper is structured as follows. Section 2 depicts related works. Our methodology is introduced in Sect. 3. Experimental results are reported in Sect. 4, and our conclusion is stated in Sect. 5.

2 Related Works

Much ongoing research has been performed in the field of Android malware detection utilizing different ML models [17]. Primarily, there are two main approaches to malware detection, namely Static and Dynamic Analysis Techniques (SAT & DAT).

2.1 SAT

SAT analyzes an Android APK without executing it. The two most common methods of SAT are signature & permission-based. Signature-based malware detection was introduced in the mid-90 s [18]. It was demonstrated using ML modeling on 525 malicious and 122 benign applications, achieving a detection accuracy of 86.56%, with non-disjoint testing and training samples on sets of Android applications. Yet, it was noticed that this method performs poorly when separating the training and testing samples [18]. In [19], and APK-Auditor [20] researchers obtained what seems to be impressive detection accuracy when carrying-out an API analysis-based method, utilizing around 10 K APKs to detect malware. However, its testing and training data were of limited sizes, and with no information regarding the most dangerous Android APIs. The above APK Auditor technique obtained an overall 88% accuracy when experimenting with 8,762 applications, of which 6,909 were malicious. Li et al. [21] developed a lightweight tool to run in the Android environment, extracting many run-time features from APKs, The reduced dimension data utilizing Principal Component Analysis is used to train a Support Vector Machine (SVM), obtaining an accuracy of approximately 79%. In [22] developed a platform-independent application instrumentation technique to trace sensitive APIs used at the run-time. This included hooking system functions in order to trace live API calls from the application. Experimentally, accuracy was not indicated but mentioned the weakness of their model and the need for developing next-generation anti-malware solutions for smartphones.

Some researchers proposed disassembling application binaries for static malware detection. Opcodes were extracted from the assembly code to describe the behavioral characteristics of a program. For example, statically generated N-gram opcode sequences were used as feature vectors and evaluated as a relation between opcodes their frequency of occurrence [23]. Yet, such an opcode sequence approach cannot reflect the run-time behavior of APKs accurately. In order to alleviate this deficiency, extracted opcode sequences and a control flow graph can implement malware detection based on a distance-matching map [24].

2.2 DAT

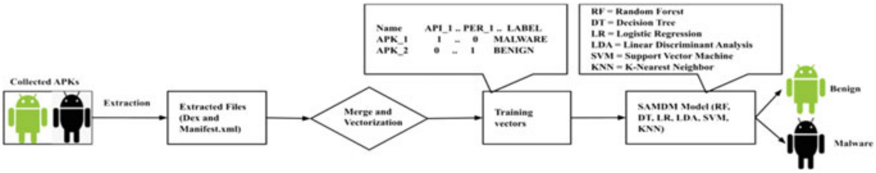
DAT involves executing an APK, then monitoring and analyzing its behavior. Once the Android app is installed and executed on the device, the app's log or trace is generated by Android OS, then recorded by the malware detection model. The next step is to generate a data-set that is extracted from the log file. Usually, this requires more time and resources towards analyzing an application's behavior than static code evaluation [25]. However dynamic analysis has a major advantage over the static approach because it can efficiently analyze packed malware. This stems from the fact that malware must be unpacked during run-time, where its original code is then loaded into memory. However, the price of this advantage is higher time and resource consumption, since the potential applications must be analyzed one-by-one. As a result, these shortcomings have limited its adoption in commercial analysis applications.

3 Our Approach

Our SAMDM model implementation is detailed in Fig. 1, which also illustrates our experiment workflow structure. In the beginning, Android APK files were collected from different popular malware families over the last decade. Then, AndroGuard APKTool [26] is used to analyze collected APK files to extract all of their invoked system API calls and permissions. The extracted APIs and permissions are used to build data-sets of vectors for training the different ML classifiers to be utilized in our SAMDM model. Vectors in any obtained training dataset are labeled as malware or benign, based on prior concrete knowledge. The following subsections depict the overall process details, phase-by-phase, as shown in Fig. 1.

3.1 Data Collection Phase

In this phase, the APKs are collected, to be used for training the ML models. This data is collected from different time periods and malware families. We utilized four



Name	API_1 .. PER_1 ..			LABEL
APK_1	1	..	0	MALWARE
APK_2	0	..	1	BENIGN

Fig. 1 Full work-flow of SAMDM

well-known data-set sources of APKs, combining them into an immensely large dataset of 158,000 APKs, with the goal of facilitating a fast and accurate SAMDM. Moreover, utilizing popular data-sets will enable more even results comparison to related peer work in literature. The datasets we utilized are **Marvin** (Contains over 133 K APKs), **Drebin** (Contains 5,560 Android Malwares), **VirusTotal** (Contains approximately 12 K APKs), and **Malgenome** (Contains 1200 APKs) [1, 26].

3.2 Data Preprocessing and Vectorization

The preparation for training our model involves two **sub-phases**.

- **Raw Data Extraction**–To extract the Permission list and API list, we utilized **APKTool**. For the permission list, we extracted the manifest file and fed it into APKTool to get the utilized permissions. For the API list, we used *isExternalMethod()* of APKTool to get utilized system APIs.
- **Vectorization**–The vectorization process is carried out by applying binary one-hot encoding (BOH) [27]. Each of the 20,500 (not including the classification label and APK name) fields of a vector will have their value set to 1 if the API or associated permission key exists in the JSON file; otherwise, they are set to zero.

3.3 Utilized ML Classifier Performance

SAMDM utilizes the following aforementioned ML classification models, with the expectation of varying performance levels: SVM, RF, KNN, LDA, LR, and DT. Although the utilization of large data-sets in modeling slows down the speed of training, it results in much better accuracy compared to small data-sets. Accuracy is more important than the speed of training because of SAMDM's frequency of utilization versus its solitary training phase.

4 Experiment Results

The data-set was trained with various size input-sets, randomly selected from different generations of malware and benign families. Each input set was split into training and testing sets while varying APK family and origin per set. In experimentation, the utilized ML models were observed to have varying performances for different data-set sizes; some excelled with larger or smaller set sizes. These variations in data selection allow for the enrichment of input sets to cover most of the malware types. In addition, our total separation between the testing and training data-sets increases trust in the obtained accuracy.

4.1 Evaluation Measures

There are three major performance metrics for ML modeling: Precision, Recall, and F-Score [28]. The above-mentioned metrics are the function of True Positive (TP), False Positive (FP), and False Negative. The formula is given as:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, F - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

Here, TP represents the number of benign apps correctly classified, FP is the number of benign apps classified as malware, and FN is the number of malicious apps classified as benign [29].

4.2 Tabulated Accuracy Results

For overall comparison with recent peer work in the literature, Table 1 contains experimental results from the previous model with randomly chosen testing and training data (25% for training and 75% for testing) [29] (Table 1). Table 2 demonstrates the

Table 1 Accuracy using 109,480 apps for analysis (Prior Work [29])

Classifiers	Train Acc.	Test Acc.	Precision	Recall	F-score
SVM	96.57	96.53	97	97	96
LR	96.38	96.38	98	97	98
RF	98.70	98.19	98	98	98
DT	98.75	97.85	98	98	98
LDA	95.81	95.62	96	96	96
KNN	98.02	97.79	98	98	98

Table 2 Accuracy using 158,460 apps for analysis

Classifiers	Train Acc.	Test Acc.	Precision	Recall	F-score
SVM	99.81	99.53	98	98	97
LR	99.53	99.27	98	97	98
RF	99.68	99.52	99	98	99
DT	99.58	99.09	99	98	99
LDA	96.28	95.79	96	96	96
KNN	99.30	99.15	97	97	98

Table 3 Accuracy using VirusTotal 2018 as a testing dataset (Prior Work [29])

Classifiers	Train Acc.	Test Acc.	Precision	Recall	F-score
SVM	97.35	77.45	100	77	97
LR	96.22	75.06	97	78	80
RF	99.18	61.02	100	61	76
DT	99.22	64.27	100	64	78
LDA	96.59	73.47	100	73	85
KNN	72.29	74.62	72	73	72

Table 4 Accuracy using VirusTotal 2018 as a testing dataset

Classifiers	Train Acc.	Test Acc.	Precision	Recall	F-score
SVM	98.45	88.07	100	89	89
LR	98.15	88.95	97	90	93
RF	99.87	90.79	100	91	94
DT	99.43	94.45	98	95	98
LDA	95.81	85.62	99	86	91
KNN	97.10	86.08	96	87	98

accuracy of the current, individually-utilized ML models, using the same data-set of 158,480 apps for training and testing.

To prove the validity of the new model in different environments with different types of malware, training, and testing data were separated for the ML modeling. In the first training phase of the model, 125,254 APKs collected from the Malgenome, Drebin, and Marvin data-sets were used Table 3. In the second phase, Virus-Total 2018 was used as a testing data-set with 33,298 APKs. Both of the training and testing data-sets are mentioned in Sect. 3. The result of this experiment is shown in Table 4 and results from our previous work in Table 3. From the result, it is evident that the newly introduced SAMDM performs much better than the previous model.

For further inter-domain validation, we utilized two separate input sets, one for testing and one for training. The first input set consisted of 157,254 APKs collected from the Virus Total 2018, Drebin, and Marvin datasets, which were used to train our Model Table 5. The second input set consisted of 1,250 APKs from the Malgenome dataset (collected between 2010 and 2011), which was used to test the new model. The result of this experiment is shown in Table 6 with the previous model's result in Table 5.

From Tables 3 and 4, it is apparent the new model performs well in varying domain data-sets, proving its validity and applicability. Experimental results obtained from the variously utilized ML models are very promising; about 20 ML models were tried. According to the result analysis, RF and DT models obtained slightly better results than the rest. This is likely due to the binary decision-making nature of the problem domain.

Table 5 Accuracy using Malgenome as a testing dataset (Prior Work [29])

Classifiers	Train Acc.	Test Acc.	Precision	Recall	F-score
SVM	97.35	76.53	100	77	87
LR	97.22	75.18	97	78	81
RF	99.18	96.27	100	96	98
DT	99.22	96.27	100	96	98
LDA	96.59	65.34	100	65	79
KNN	72.29	96.99	98	82	86

Table 6 Accuracy using Malgenome as a testing dataset

Classifiers	Train Acc	Test Acc	Precision	Recall	F-score
SVM	98.10	88.59	100	89	94
LR	97.78	86.45	98	87	86
RF	99.62	98.81	100	99	99
DT	99.67	99.37	100	99	100
LDA	95.81	84.62	84	86	85
KNN	97.25	94.06	99	95	95

4.3 High Accuracy Justification

During data extraction, the occurrence of API calls and Permissions for both malware and benign apps were tabulated. Based on that calculation, it was observed that the frequency of sensitive API and permission usage was much higher for malware than benign APKs. Figures 2 and 3 show the use ratio of sensitive APIs and permissions of benign versus malicious APKs.

In Fig. 2, permissions such as *READ_PHONE_STATE*, *SEND_SMS*, *WRITE_EXTERNAL_STORAGE*, *READ_SMS*, etc., were used significantly more by malware than benign APKs. In Fig. 3, API calls such as *startActivity*, *finish*, *getSystemService*, *getIntent*, *getApplicationContext*, etc. were used more frequently in malware than benign APKs.

Per analysis, malicious apps command some of the most sensitive APIs and permissions unnecessarily, which are not usually needed by an average application. Hence, when running an APK (on the Android OS) that excessively demands the aforementioned permissions and/or API calls, it is very likely said APK is malicious.

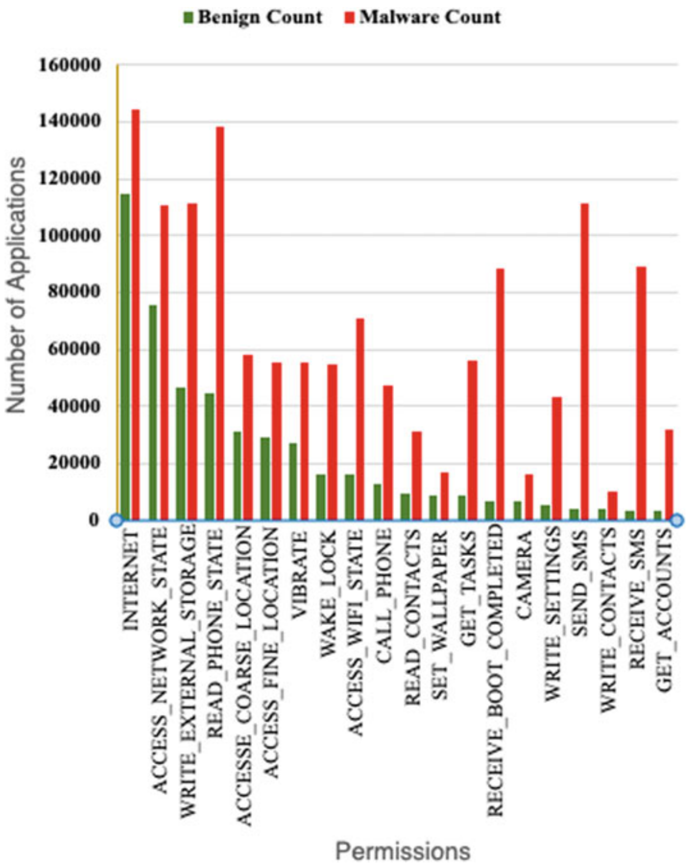


Fig. 2 Top 20 permissions used by malware compared to benign applications

5 Conclusions

In this paper, the detection of Android malware in large-scale data-sets was improved using new ML models. The new model detects more families of malware with greater accuracy. Moreover, SAMDM successfully distinguishes between malware and benign apps that were collected over a period of 12 years (between 2006 and 2018). The SAMDM utilizes the varying sizes, generations, and different families of data-sets to train the utilized ML models. In the new SAMDM design, the vector features were extended by adding a "utilized-API List" in addition to a "utilized Permission list", resulting in better performance. It is worth mentioning that extracting the utilized APIs from the data-set APKs using reverse engineering was difficult to accomplish. To address the problem, some of the code in the tools which carry out the reverse engineering was modified. For each of the three utilized data-sets, our new SAMDM model performs much better than the previously published

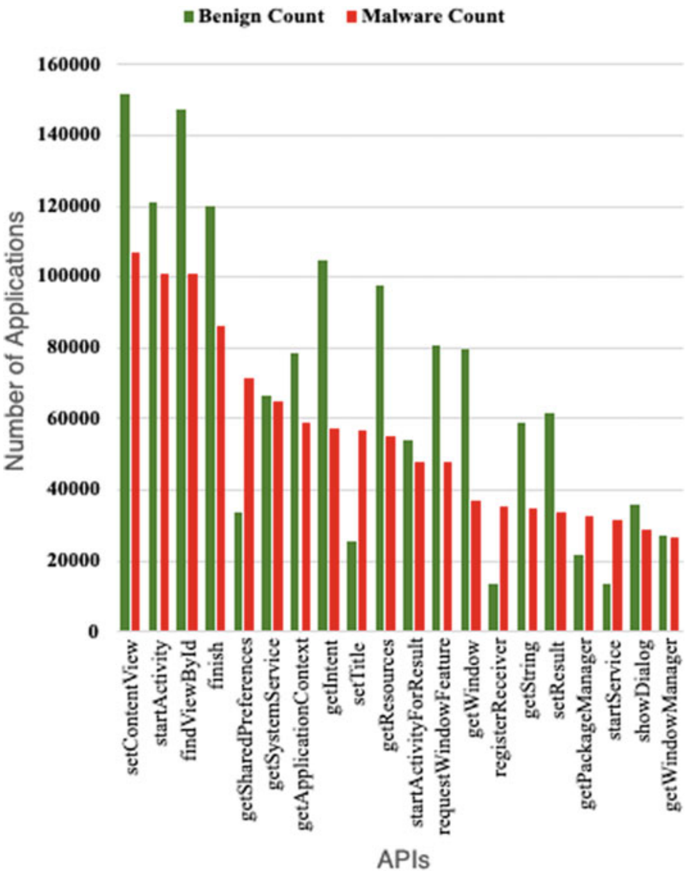


Fig. 3 Top 20 APIs used by malware compared to benign applications

work, with respect to the obtained Testing Accuracy, Precision, Recall, and F-score. In future work, the utilization of the dynamic analysis approach to predict, as early as possible at the system level, any malware attacks will be investigated. Such a capability would allow the refusal of app-based threats to Android devices.

References

1. Liu J, Yu J (2011) Research on development of android applications. In: 2011 4th international conference on intelligent networks and intelligent systems. Kunming, pp 69–72 https://doi.org/10.1109/ICINIS.2011.40
2. Oulehla M (2015) Investigation into google play security mechanisms via experimental botnet. In: 2015 IEEE international symposium on signal processing and information technology (ISSPIT). Abu Dhabi, pp 591–596. https://doi.org/10.1109/ISSPIT.2015.7394406

3. Wang W, Godfrey MW (2013) Detecting API usage obstacles: a study of iOS and Android developer questions. In: 2013 10th working conference on mining software repositories (MSR). San Francisco, CA, pp 61–64. https://doi.org/10.1109/MSR.2013.6624006
4. Huang N, Xu M, Zheng N, Qiao T, Choo KR (2019) Deep android malware classification with API-based feature graph. In 2019 18th IEEE international conference on trust, security and privacy in computing and communications/13th IEEE international conference on big data science and engineering (TrustCom/BigDataSE). Rotorua, New Zealand, pp 296–303 https://doi.org/10.1109/TrustCom/BigDataSE.2019.00047
5. Jung J, et al (2018) Android malware detection based on useful API calls and machine learning. In: 2018 IEEE first international conference on artificial intelligence and knowledge engineering (AIKE). Laguna Hills, CA, pp 175–178. https://doi.org/10.1109/AIKE.2018.00041
6. Sahoo AK, Sahoo KS, Tiwary M (2014) Signature based malware detection for unstructured data in Hadoop In 2014 international conference on advances in electronics computers and communications. Bangalore, pp 1–6. https://doi.org/10.1109/ICAECC.2014.7002394
7. Sathyanarayan VS, Kohli P, Bruhadeshwar B (2008) Signature generation and detection of malware families. In: Mu Y, Susilo W, Seberry J (eds) Information security and privacy. ACISP 2008. Lecture notes in computer science, vol 5107. Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-70500-0_25
8. Shijo PV, Salim A (2015) Integrated static and dynamic analysis for malware detection. Procedia Comp Sci 46:804–811. ISSN 1877–0509 https://doi.org/10.1016/j.procs.2015.02.149
9. Utku A, DoGru IA, AkcayolMAPermission based android malware detection with multilayer perceptron. In: 2018 26th signal processing and communications applications conference (SIU). Izmir, pp 1–4. https://doi.org/10.1109/SIU.2018.8404302
10. Yerima S, Alzaylaee M, Sezer S (2019) Machine learning-based dynamic analysis of Android apps with improved code coverage. EURASIP J Info Sec 2019:4. https://doi.org/10.1186/s13635-019-0087-1
11. Ahmad M, Costamagna V, Crispo B, Bergadano F F, Zhauniarovich Y (2020) StaDART: addressing the problem of dynamic code updates in the security analysis of android applications. J Syste Softw 159: 110386. ISSN 0164–1212 https://doi.org/10.1016/j.jss.2019.07.088
12. Daniel A, Michael S, Malte H, Hugo G, Konrad R (2014) Drebin: efficient and explainable detection of android malware in your pocket. In: 21th annual network and distributed system security symposium (NDSS)
13. Zhou Y, Wang Z, Zhou W, Jiang X (2012) Hey, you, get off of my market: detecting malicious apps in official and alternative Android markets. In: Proceedings of the 19th annual network & distributed system security symposium
14. Zhou Y, Jiang X (2012) Dissecting android malware: characterization and evolution security and privacy (SP). In: IEEE symposium on security and privacy
15. MARVIN Efficient and comprehensive mobile app classification through static and dynamic analysis
16. Virus Total https://www.virustotal.com/gui/graph-overview
17. Peiravian N, Zhu X (2013) Machine learning for android malware detection using permission and API calls. In: 2013 IEEE 25th international conference on tools with artificial intelligence. Herndon, VA, pp 300–305. https://doi.org/10.1109/ICTAI.2013.53
18. Venugopal D, Hu G (2008) Efficient signature based malware detection on mobile devices. Mob Inf Syst 4(1):33–49. https://doi.org/10.1155/2008/712353
19. Zhang H, Luo S, Zhang Y, Pan L (2019) An efficient android malware detection system based on method-level behavioral semantic analysis. IEEE Access 7:69246–69256. https://doi.org/10.1109/ACCESS.2019.2919796
20. Talha KA, Alper DI, Aydin C (2015) APK auditor: permission-based Android malware detection system. Digit Investig 13:1–14
21. Li X, Liu J, Huo Y, Zhang R, Yao Y (2016) An Android malware detection method based on Android Manifest file. In: International conference on cloud computing and intelligence systems (CCIS). pp 239–243

22. Somarriba O, Zurutuza U, Uribe Etxebarria R, Delosières L, Nadjm-Tehrani S (2016) Detection and visualization of android malware behavior. J Electr Comp Eng 2016:8034967. https://doi.org/10.1155/2016/8034967
23. O'Kane P, Sezer S, McLaughlin K (2014) N-gram density based malware detection. In: 2014 world symposium on computer applications & research (WSCAR). Sousse, pp 1–6. https://doi.org/10.1109/WSCAR.2014.6916806
24. Ding Y, Zhang X, Hu J et al (2020) Android malware detection method based on bytecode image. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-020-02196-4
25. van der Veen V (2013) Dynamic analysis of android malware. https://doi.org/10.13140/2.1.2373.4080
26. Android-Apktool A tool for reverse engineering Android apk files. https://code.google.com/p/android-apktool/
27. Harris D, Harris S (2012) Digital design and computer architecture, 2nd edn. Morgan Kaufmann, San Francisco, CA, p 129. ISBN 978–0–12–394424–5
28. Gharib M, Bondavalli A (2019) On the evaluation measures for machine learning algorithms for safety-critical systems. In: 2019 15th european dependable computing conference (EDCC). Naples, Italy, 141–144 https://doi.org/10.1109/EDCC.2019.00035
29. Alahy QE, Chowdhury MNUR, Soliman H, Chaity MS, Haque A (2020) Android malware detection in large dataset: smart approach. In: Arai K, Kapoor S, Bhatia R (eds) Advances in information and communication. FICC 2020. Advances in intelligent systems and computing, vol 1129. Springer, Cham. https://doi.org/10.1007/978-3-030-39445-5_58
30. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intell Syst Appl 13(4):18–28
31. Breiman L (2001) Random Forests. Mach Learn 45:5. https://doi.org/10.1023/A:1010933404324
32. Liao Y, Vemuri VR (2002) Use of K-Nearest Neighbor classifier for intrusion detection. Comput Secur 21(5):439–448
33. Li M, Yuan B (2005) 2D-LDA: a statistical linear discriminant analysis for image matrix. Pattern Recogn Lett 26(5):527–532
34. Haifley T (2002) Linear logistic regression: an introduction. In: IEEE international integrated reliability workshop final report
35. Navada A, Ansari AN, Patil S, Sonkamble BA (2011) Overview of use of decision tree algorithms in machine learning. In: IEEE control and system graduate research colloquium

Other Related Topics

Exploring the Limitations Involved in Students' Academic Use of Facebook

Abdulsalam K. Alhazmi, Fatima Al-Hammadi, Ezzadeen Kaed, and Athar Imtiaz

Abstract Amongst students in higher education, Facebook is the most popular and preferred of all Social Networking Sites (SNSs). Facebook is free, interactive, easy to use, and well-designed. It provides a number of possibilities in terms of supporting communication among students. However, this study has shown that despite the potentially effective communication features of Facebook, and the widespread usage among students, the majority of students and lecturers only use Facebook for social purposes. Additionally, those who use Facebook for academic purposes do so infrequently, or for very small amounts of time, compared to their use of Facebook for social purposes. As a result, further investigation into the limited academic use of Facebook was conducted using case study research with several qualitative methods including semi-structured interviews, focus group discussions, and Facebook group discussions. The results of the content analysis revealed that the main reasons for the limited educational use of Facebook are related to the following five main factors: technology, distraction, content, lack of teacher support, and security and privacy.

Keywords Academic use of Facebook · Facebook in higher education · Social networking sites (SNS) · E-Learning · Social learning · Management system (LMS)

1 Introduction

SNSs are applicable for a wide range of educational endeavours due to various characteristics, for instance their ease of use, cost-effectiveness, ease of communication and interaction between members, reliance on the concept of social learning, and many more factors [1, 2]. Facebook is reported to be the leading social networking

A. K. Alhazmi (✉) · F. Al-Hammadi
University of Science and Technology, Aden, Yemen

F. Al-Hammadi · E. Kaed
Interactive Language Centre, Kuala Lumpur, Malaysia

A. Imtiaz
Massey University, Palmerston North, New Zealand

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_13

site in terms of its user base and reputation [3]. It therefore makes sense to propose that Facebook can be used as a higher education learning tool [4–6].

The potential advantages of SNS with regards to supporting education can be summarised into the usability of SNS features and tools, access to a variety of resources, communication, and interaction with peers and teachers, content creation and sharing, obtaining instant feedback on learning problems, and supporting informal and constructive learning [7, 8]. While alternative e-learning tools (i.e., LMS) are available, SNSs seem to be much better suited to student-oriented higher education, largely due to their dynamic qualities. Moreover, SNSs encourage students to join common interest communities, help each other with their academic studies, build bonds with their classmates, and promote supplementary interactions between themselves and their instructors [9]. As in [10] the main areas in which SNSs can aid learning and teaching: connectivity and social support; collaborative information discovery and sharing; content creation; knowledge and information aggregation; and content modification.

Since the number of students using Facebook is continually increasing, it is necessary to investigate how to best integrate the use of Facebook into education [11–14]. The hope is that students will eventually use Facebook just as much for studying as they do for social networking [15]. Higher education establishments must become knowledgeable regarding the best ways in which to implement SNSs into educational activities whilst making the most of the opportunities provided by sites such as Facebook, especially since Facebook offers vast potential in terms of facilitating the teaching and learning process, and the majority of students already extensively use it.

Despite the high rate of usage among higher education students, many studies have stated that social interaction and communication remains the most common reason for use, while the use of SNSs for academic purposes is still very much in the infancy stage [1, 13, 15–17].

In general, researchers who have investigated the educational use of Facebook have found that its academic use remains limited despite the popularity of the site among many students [18]. Therefore, in order to reach our research objectives, a qualitative case study is used to gain a better understanding of why the educational use of Facebook remains so limited. As a result, this study takes an inductive approach to investigating the problem, and will identify the current challenges of utilising this technology and its features to promote effective academic use and to support pedagogically-sound activities.

2 Review

2.1 *Facebook Use in Higher Education*

SNSs vary greatly, and therefore they can be used for a large number of different activities and purposes. However, many research studies have shown that Facebook is the most popular among university students [19–22]. According to a study conducted by ECAR, 90% of students reported that they used Facebook on a daily basis [23]. In another survey administered to 6,498 Malaysian university students, the results revealed that 80.8% of students held an SNS account, and the large majority used Facebook daily [24].

Up to the present time, there has been an underwhelming degree of research into student engagement with social networking, especially in relation to Facebook. [25] was critical of recent studies that did take place, noting that they were restricted by both their assessment of Facebook usage and their criteria for quantifying engagement. Previous studies into the relationship between student engagement and social network services have focused on academic activities. More recent research into the subject has concentrated on social activities. It has been established that SNSs are beneficial in encouraging student engagement, and thus increasing knowledge. However, it is apparent that this topic requires further research.

2.2 *Facebook's Educational Potential*

While the primary goal of SNSs is generally known to be for social networking activities, the results of some studies have indicated that using SNS features and tools in relevant educational ways could support student engagement and learning. Bowers-Campbell [26] demonstrates the ways in which education professionals could use Facebook to aid their teaching practices by broadcasting messages of support to their learners. This study suggested that because Facebook can serve to improve interactivity between classmates and tutors, its use in education may help to increase the prevalence of self-regulated learning and self-efficacy.

Another study examined the relationship between Facebook use and students' academic performance [25]; finding that students' GPA increased when studying for longer periods and when Facebook was used (relevantly) more often. However, students' GPA decreased in line with students increasing their social endeavours. As such, it was suggested that students' GPA and the hours spent studying have a negative correlation to the hours spent on Facebook for social purposes. Furthermore, another research study conducted by [27], showed that Facebook has been used by university students to enhance their English skills. Therefore, the literature suggests that students' learning is influenced by the use of SNSs, for better or worse [22, 28].

Essentially, if students engage in meaningful educational activities via Facebook, their academic performance may improve. On the other hand, students' academic performance may decline if time spent on SNSs is not education-oriented.

3 Research Methodology

3.1 Research Design

This research uses a single case study with the application of mixed methods [29]. The main reason for selecting the case study method is that the study of social networking is highly dependent upon the kind of contextual information which can be assessed through the case study [16, 30]. In mixed methods research "is more in line with methodology combination, which essentially requires multiple worldviews (i.e., the combination of qualitative and quantitative research methods)" [31].

The first method applied in this case study was an exploratory survey. Research shows that surveys can be used effectively in the initial phase of a study to explore the relationships and patterns in research where no assumptions or models are assumed [32]. Following the survey, qualitative methods were used: semi-structured interviews, which are extremely popular in qualitative studies [33] focus group discussions which allow researchers to gather information and gain an understanding of a certain phenomenon via the spoken opinions of a group of participants [34].

The reason that a qualitative research design was adopted is that the educational use of SNSs is a relatively new phenomenon. Typically, exploratory research is expanded via additional exploratory research or by conclusive studies [35]. Qualitative data collection approaches tend to be applied to exploratory research in order to gain greater insight into a topic and to produce new theories in the field of social science and IS studies [31, 36, 37].

3.2 Sampling

Data was collected using multiple methods. To begin with, 105 students were given an exploratory questionnaire that was designed to investigate some of the features of the educational use of SNSs. The respondents were classified as follows: male and female, undergraduate and postgraduate, and local and international students. The majority of students (97.2%) reported that they held Facebook accounts, and very few students (2.8%) stated that they did not use Facebook at all. The sample consisted of 62 (59.6%) local students and 42 (40.4%) international students. The distribution of students across genders was 56 (54.4%) male students and 47 (45.6%) female students. The distribution of students across the level of study was 61 (58.7%) undergraduate students and 43 (41.3%) postgraduate students.

Following this, participants in the semi-structured interviews, focus group discussions, and online discussions were chosen through qualitative sampling methods. In order to ensure that the findings would be well-balanced and that the insight gained would be representative of a wide range of students, the sample chosen for participation in this study were a combination of undergraduate and postgraduate students of both genders. A total of 11 participants were initially contacted via Facebook, in order to request their participation in this study. The response rate was 81%, with 9 students responding to the invitation message and agreeing to become voluntary participants in this research. Following this, face-to-face interviews were held with each of the 9 participants. A total of three groups with 17 respondents were involved in the focus group discussions. The three focus groups were comprised of 13 male students and 4 female students, of whom 5 were postgraduate students and 12 were undergraduate students.

3.3 Data Analysis

Due to the descriptive nature of the survey, the data was analysed using the descriptive analysis. In the subsequent phases of conducting the individual interviews and focus group discussions, transcription of the data occurred along with data coding, in order to organise the information by category to aid interpretation. The responses given by the participants varied from a few words to multiple paragraphs, and these were used for analysis [33], [38]. The data was assigned codes using the auto-coding function in NVivo 10, and the key categories were determined using selective coding. NVivo's selective coding function is much like open coding. Here, the interview responses go through rigorous human checking in order to select the most relevant themes [39].

It is essential that qualitative data is checked for validity and reliability in order to guarantee that the study's findings are meaningful and come from trusted sources [40, 41, 42]. Furthermore, this will help to ensure that the research can be carried out again in the future [29].

In qualitative studies, validity represents the level of accuracy achieved by an instrument in terms of addressing the research problem or phenomenon that is being studied [43]. Scholars have explained that researchers must assess the validity of their research, since this ensures that the themes that have been proposed from the data are accurate [29, 44]. Therefore, that coders are assessed in terms of their level of accordance with one another regarding the themes identified in the data [43–45].

4 Results: Descriptive Statistics of Current Facebook Use

4.1 Purposes of Use

In order to measure the key motivations of the students for using Facebook, four categories were given: to keep in touch with friends, to share what is happening in one's life, to communicate with classmates about assignments, and to share news and other issues. As Table 1 illustrates, the results show that the majority of students (84%) use Facebook to keep in touch with friends, followed by sharing what is happening in one's life (47%), and then the communication of news (44%). Other purposes of use were also identified, such as marketing, and playing games (14%). A comparatively low percentage of use (36%) was recorded for academic purposes, which was categorised as communicating with classmates on assignments in this study. This number of students using Facebook for academic purposes (36%) reflects a higher percentage than that previously reported in other studies. For example, in a study by [19], only 10% of students reported that they used Facebook for the discussion of academic work with other students. In another study, just 19 (4.9%) out of 390 students stated that they were able to acquire knowledge using Facebook [46].

Although the frequency of using Facebook for academic reasons is higher than the frequency presented in previous literature, academic use is still the least popular reason for using Facebook in comparison with other purposes, such as maintaining relationships or sharing news.

4.2 Time Spent Engaging in Non-Academic Activities on Facebook

Students were asked to estimate the time they spent each day using Facebook for both academic and non-academic activities, based on four categorical responses: less than 1 h, 1–2 h, 2–3 h, and over 3 h. As Table 2 illustrates, around one third (33.7%) of the participants stated that they spend 1–2 h per day on Facebook. In fact, when we add up the number of students who spend over 1 h per day using Facebook,

Table 1. Purposes of Facebook use

Purpose	Frequency	Percent
Keeping in touch with friends	92	84%
Sharing what is happening in one's life	51	47%
Communicating with classmates about assignments	39	36%
News	47	44%
Other (marketing, entertainment, etc.)	14	14%

Table 2. Time spent on Facebook for non-academic purposes

Experience	Frequency	Percent%
Less than 1 h	22	21.1
1–2 h	35	33.7
2–3 h	32	30.8
More than 3 h	15	14.4
Total	104	100.0

we find that this group represents the largest majority of all students (78.9%). This indicates that the majority of students spend a high amount of time on Facebook each day. The study result of [47], revealed that students spend a mean average of 1 h and 41 min per day on Facebook. In another study by [46], the results revealed that the majority of students used Facebook only for social purposes, and that their daily Facebook time amounted to around 1 h. In this sense, the more time students spend on Facebook, the more likely it is that there will be a negative impact on their academic performance and learning outcomes.

4.3 Time Spent Engaging in Academic Activities on Facebook

Table 3 shows that the majority of students spend less time engaging in academic activities on Facebook than they spend engaging in non-academic activities on this platform. Almost all of the 39 students in this study reported that they engage in under an hour of educational activity on Facebook each day; with just 3 students stating otherwise. Students' academic use of Facebook is therefore low compared with their extensive use of Facebook for social activities, sharing and reading news, and for other purposes. Out of all of the student participants in this study, only three (7.7%) of the students reported that they usually spend over one hour per day using Facebook for educational purposes, regarding the second time category.

The overall results indicate that the use of Facebook for academic purposes is still limited in terms of both time and the activities engaged in. However, the following section of this paper presents the results of an inductive investigation of the reasons behind the limited academic use of Facebook among students, despite the opportunity existing for better academic communication and interaction through Facebook.

Table 3. Time spent on Facebook for academic activities

Time spent	Frequency	Percent%
Less than 1 h	36	92.3
1–2 h	3	7.7
Total	39	100

5 Results: The Reasons for Facebook's Limited Academic Use

The results of the qualitative data collection method assisted us in studying the reasons behind the limited use of Facebook, and to go beyond the statistical results reported in the previous part of this study and in the literature. As a result of data analysis and validation, the reasons for the low level of academic use was identified based on the data of the interviews and focus group discussions. In response to the question "Why is the use of Facebook for academic purposes limited?", the results of this study reveal that the most significant factors involved in the limited academic use of Facebook are related to technology, distraction, management, privacy, content, and lack of teacher support.

5.1 Technology

During the data collection phase of the study, technological limitations such as content management and assessment, were provided as an explanation for why so few students spend a sufficient amount of time using SNSs for educational activities. Facebook offers a variety of features for communication, interaction, and information-sharing. However, if Facebook is to be used for academic purposes, users expect a number of additional supporting features, such as content management and assessment capabilities. It was reported that it is difficult for students' to keep track of specific content or activities on Facebook. In addition, it is difficult for students to retrieve specific shared content and activities. Therefore, it has been revealed that if SNS is to be used effectively for academic activities, additional features that will offer better support to academic purposes are still required.

5.2 Distraction

Distraction was another challenge which was reported by the respondents in this study. When the students were asked to provide reasons for spending less time engaging in academic activities through Facebook compared to engaging in social activities, the issue of distraction was mentioned.

The dynamic notification system is considered to be one of the most attractive features of Facebook in terms of socialising and keeping up-to-date with the activities of others. In relation to education, however, this makes it difficult for students to remain focused. The alert messages and dynamic notification settings can be used effectively for academic activities and can provide an effective course-related environment. However, the high number of students and lecturers using Facebook for non-academic purposes creates a disturbance. Respondents suggested that Facebook may require some features or settings to assist students to focus; otherwise, they may be interrupted by numerous social activities and, consequently, their attention will be diverted to such activities.

5.3 Content

According to the data collected in this study, another reason for the low academic use of Facebook is content. In response to the question, "Do you spend much time on Facebook for social or academic purposes?", one respondent stated that they only use the platform for social purposes as there are an insufficient quantity of academic materials on Facebook. Another respondent explained that students spend less time on Facebook for academic purposes because for much of the time, there are a very limited number of Facebook page which offers relevant or useful content which is assistive to studying.

Furthermore, it appears that the low level of academic Facebook use can be attributed to the type of content, accuracy of the content, and the content relevancy. Although Facebook's features facilitate content-sharing, other challenges have been identified in terms of the content's relevancy and quality.

5.4 Lack of Teacher Support

According to the respondents in this study, the high level of Facebook use for non-academic activities among students and lecturers is one of the reasons for the low level of use for academic purposes. The overall results of the individual interviews and focus group discussions revealed that the level of lecturers' participation in Facebook for academic purposes was relatively limited. Students reported that their lecturers were not very active academically on Facebook. For example, during the focus group discussion, one respondent stated that they did not believe that lecturers access Facebook for the purposes of academic work. Another participant supported this point saying that they were aware of only one lecturer who implements Facebook for academic purposes.

The responses of the participants involved in this research project go a long way in helping us to gain insight into the reasons behind the lack of academic Facebook activity among students.

5.5 Security and Privacy

The fourth reason for the low academic use of Facebook relates to privacy concerns. These play a key role in the current low level of academic Facebook use among university students. The results of this research clearly emphasise that a number of students find Facebook's privacy settings to be confusing, ambiguous, or concerning. It is noted that some of these concerns stem from students' own inexperience with Facebook settings, while some of the concerns stem from Facebook's own limitations in terms of the privacy settings which are available to users. Some of Facebook's features are more suited to academic activities than others. However, students need to be educated on these features and shown how to use the settings which are available to them with regards to privacy. In such an open social system, privacy is still a major concern in environments where more manageable privacy settings are required.

6 Conclusion and Recommendations

Driven by the results of the exploratory survey of the investigation into the current state academic Facebook use, further investigations into the reasons for the low level of academic Facebook use were conducted using in-depth interviews, focus group discussions, and online discussions. The results revealed five main reasons behind the low level of Facebook use for academic purposes.

The first reason for the limited level of academic Facebook use relates to the technological features and tools provided by Facebook. It is required more manageable features to facilitate content organisation, content classification, content filtering, and content retrieval seems to be essential for the effective use of Facebook in an educational context. The second reason relates to the disturbance which might occur as a result of the current design nature of social networking technologies, as well as the current dominant non-academic use of Facebook. It is suggested that student learning might be disturbed by Facebook's integration of social and academic activities. The third reason for the limited academic use of Facebook relates to content. The lack of relevant and resourceful content on Facebook influences its use for academic purposes. Providing interactive, resourceful, and relevant content on Facebook and other SNSs will encourage student engagement and academic use of SNSs. The fourth reason for the low level of academic Facebook use relates the lack of teacher support. The results of this research reveal that the potential application of Facebook within the educational environment is still undervalued by lecturers, as they themselves are active on Facebook socially rather than academically. The fifth reason relates to the security and privacy concern of students in this open social environment, where more manageable privacy settings are required.

The qualitative approach used in this study was shown to be useful, and it contributed a great deal in terms of understanding the phenomenon and the current challenges which are associated with the use of Facebook for academic purposes.

Therefore, this research contributes to a better understanding of SNSs and their potential applications in education. Understanding the reasons behind the current limited educational use of Facebook provides direction for educators and higher education institutions towards more successful implementation. Finally, because this study used one case study, future researchers may benefit from conducting multiple case studies and comparing the results in order to overcome this limitation.

References

1. Abdulsalam AAR, Alhazmi K (2013) Facebook in higher education: Students' use and perceptions. AISS Adv Inf Sci Serv Sci 5(15):32–41, 2013
2. Alhazmi AK, Rahman AA (2014) A framework for student engagement in social networking sites
3. Steinfield C, Ellison NB, Lampe C (2008) Social capital, self-esteem, and use of online social network sites: a longitudinal analysis. J Appl Dev Psychol 29(6):434–445. https://doi.org/10.1016/j.appdev.2008.07.002
4. Teclehaimanot B, Hickman T (2011) What students find appropriate. Tech Trends 55(3):19–30
5. Heiberger G, Harper R (2008) Have you facebooked Astin lately? Using technology to increase student involvement. New Dir Stud Serv 2008(124):19–35. https://doi.org/10.1002/ss.293
6. Menzies R, Petrie K, Zarb M (2017) A case study of Facebook use: outlining a multi-layer strategy for higher education. Educ Inf Technol 22(1):39–53. https://doi.org/10.1007/s10639-015-9436-y
7. Alhazmi AK, Rahman AA, Zafar H (2015) Conceptual model for the academic use of social networking sites from student engagement perspective. In: IC3e 2014–2014 IEEE conference on e-learning, e-Management and e-Services. pp 1–6. https://doi.org/10.1109/IC3e.2014.7081232
8. Abdulsalam Kaed Alhazmi AAR (2013) Social networking sites in higher education: potential advantages for student learning. Int J Res Educ Methodol 4(2):493–499
9. Griffith S, Liyanage L (2008) An introduction to the potential of social networking sites in education. Sites J 20Th Century Contemp French Stud 18–21
10. McLoughlin C, Lee MJW (2007) Social software and participatory learning: pedagogical choices with technology affordances in the web 2.0 era. In: ASCILITE 2007–Australas. Soc. Comput. Learn. Tert. Educ. pp 664–675
11. Reid I (2015) An education in Facebook. Inf Commun Soc 18(12):1478–1480. https://doi.org/10.1080/1369118x.2014.984744
12. Ellefsen L (2015) An Investigation into perceptions of facebook-use in higher education. Int J High Educ 5(1):2. https://doi.org/10.5430/ijhe.v5n1p160
13. Niu L (2019) Using Facebook for academic purposes: current literature and directions for future research. J Educ Comput Res 56(8):1384–1406. https://doi.org/10.1177/0735633117745161
14. Abdulsalam Kaed Alhazmi AAR (2013) Facebook in higher education: social and academic purposes. Int J Comput Technol 12(3):3300–3305
15. Toker S, Baturay MH (2019) What foresees college students' tendency to use facebook for diverse educational purposes? Int J Educ Technol High Educ 16(1) https://doi.org/10.1186/s41239-019-0139-0
16. Hew KF (2011) Students' and teachers' use of Facebook. Comput Hum Behav 27(2):662–676. https://doi.org/10.1016/j.chb.2010.11.020
17. Bamansoor S, Kayode B, Alhazmi AK, Ahmad Saany SI (2018) The adoption of social learning systems in higher education: extended TAM. https://doi.org/10.1109/ICSCEE.2018.8538371
18. Arteaga Sánchez R, Cortijo V, Javed U (2014) Students' perceptions of Facebook for academic purposes. Comput Educ 70:138–149. https://doi.org/10.1016/j.compedu.2013.08.012

19. Madge C, Meek J, Wellens J, Hooley T (2009) Facebook, social integration and informal learning at university: 'It is more for socialising and talking to friends about work than for actually doing work.' Learn Media Technol 34(2):141–155. https://doi.org/10.1080/17439880902923606
20. Selwyn N (2009) Faceworking: exploring students' education-related use of Facebook. Learn Media Technol 34(2):157–174. https://doi.org/10.1080/17439880902923622
21. Manca S, Ranieri M (2016) Facebook and the others. Potentials and obstacles of social media for teaching in higher education. Comput Educ 95:216–230. https://doi.org/10.1016/j.compedu.2016.01.012
22. Manca S (2020) Snapping, pinning, liking or texting: investigating social media in higher education beyond Facebook. Internet High Educ 44:100707. https://doi.org/10.1016/j.iheduc.2019.100707
23. Smith SD, Borreson Caruso J (2010) The ECAR study of undergraduate students and information technology. Educ Cent Appl Res 1–13
24. Embi MA, Gabarre S, Gabarre C, Hamat A, Din R (2014) Evaluating the level of diffusion of social networking sites among Malaysian university students. Asian Soc Sci 10(3):99–111. https://doi.org/10.5539/ass.v10n3p99
25. Junco R (2012) Too much face and not enough books: the relationship between multiple indices of Facebook use and academic performance. Comput Hum Behav 28(1):187–198. https://doi.org/10.1016/j.chb.2011.08.026
26. Bowers-campbell J (2008) Joy Bowers-Campbell. Reading 74–87
27. Kabilan MK, Ahmad N, Abidin MJZ (2010) Facebook: an online environment for learning of English in institutions of higher education? Internet High Educ 13(4):179–187. https://doi.org/10.1016/j.iheduc.2010.07.003
28. Peruta A, Shields AB (2017) Social media in higher education: understanding how colleges and universities use Facebook. J Mark High Educ 27(1):131–143. https://doi.org/10.1080/08841241.2016.1212451
29. Yin RK (2009) Case study research: design and methods. SAGE Publications
30. Boyd DM, Ellison NB (2007) Social network sites: definition, history, and scholarship. J Comput Commun 13(1):210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x
31. Venkatesh V, Brown SA (2013) Guidelines for conducting mixed methods research in informations systems. *MIS Q* X(X):1–34
32. Cohen L, Manion L, Morrison K (2007) Research methods in education, 7th edn.
33. Campbell JL, Quincy C, Osserman J, Pedersen OK (2013) Coding in-depth semistructured interviews: problems of unitization and intercoder reliability and agreement. Sociol Methods Res 42(3):294–320. https://doi.org/10.1177/0049124113500475
34. Benson V, Morgan S, Tennakoon H (2013) Social networking in higher education: a knowledge convergence platform. Commun Comput Inf Sci 278:416–425. https://doi.org/10.1007/978-3-642-35879-1_50
35. Wisler AK (2009) 'Of, by, and for are not merely prepositions': teaching and learning conflict resolution for a democratic, global citizenry. Intercult Educ 20(2):127–133. https://doi.org/10.1080/14675980902922143
36. Benbasat I, Goldstein DK, Mead M (1987) The case research strategy in studies of information systems. MIS Q Manag Inf Syst 11(3):369–386. https://doi.org/10.2307/248684
37. Walsham G (2006) Doing interpretive research. Eur J Inf Syst 15(3):320–330. https://doi.org/10.1057/palgrave.ejis.3000589
38. Berg BL, Lune H (2012) Qualitative research methods for the social sciences. Pearson, Boston
39. Glaser BG, Strauss AL (2009) The discovery of grounded theory: strategies for qualitative research. Aldine Trans
40. Graneheim UH, Lundman B (2004) Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 24(2):105–112. https://doi.org/10.1016/j.nedt.2003.10.001
41. Krippendorff K (2004) Reliability in content analysis: some common misconceptions and recommendations. Hum Commun Res 30(3):411–433. https://doi.org/10.1093/hcr/30.3.411

42. Perry C (1998) Processes of a case study methodology for postgraduate research in marketing. Eur J Mark 32(9/10):785–802. https://doi.org/10.1108/03090569810232237
43. Neuendorf KA (2017) The content analysis guidebook. Thousand Oaks, California https://doi.org/10.4135/9781071802878
44. Ryan GW, Bernard HR (2003) Techniques to identify themes. Field Methods 15(1):85–109. https://doi.org/10.1177/1525822X02239569
45. Ellis L (1993) Research methods in the social sciences. Brown & Benchmark, Madison, WIS
46. Wise L, Skues J, Williams B (2011) Facebook in higher education promotes social but not academic engagement. In: ASCILITE 2011–Australas. Soc. Comput. Learn. Tert. Educ. pp 1332–1342
47. Roblyer MD, McDaniel M, Webb M, Herman J, Witty JV (2010) Findings on facebook in higher education: a comparison of college faculty and student uses and perceptions of social networking sites. Internet High Educ 13(3):134–140. https://doi.org/10.1016/j.iheduc.2010.03.002

Future Network Technology

Real Network Traffic Data with PCAP in a Software-Defined Networking Test Framework for Quality of Service Mechanisms

Justin Bryce Torres, Josiah Eleazar Regencia, and William Emmanuel S. Yu

Abstract Software-Defined Networking (SDN) is an advancement in the field of computer networking. Previous studies have built SDN frameworks to test the performance of Quality of Service (QoS) mechanisms more easily than traditional networks. A limitation of previous work in this field is the use of synthetic traffic. This research aims to extend the previous study through the use of real-world network traffic data, packet captures, in achieving more accurate and realistic performance tests. This was done through the use of PCAP file replay and minor framework modifications. The results of the PCAP file tests show that the performance comparison of leaf-enforced and core-enforced algorithms are consistent as those in previous studies. The results show that in Basic CBQ 53.33% of 30 trials showed leaf advantage, in Source CBQ 70% of 30 trials showed leaf advantage, in Destination CBQ 100% of 30 trials showed leaf advantage, and in Source-Destination CBQ 70% of 30 trials showed leaf advantage.

Keywords Software-defined networking · Quality of service · Class-based queueing

1 Introduction

Software-Defined Networking (SDN) is a new technology in computer networking that can help improve network management compared to traditional networks. Instead of having the control of each network device limited to the hardware, SDN networks have a centralized control plane where network protocols can be easily implemented throughout the entire network [1]. There are many uses for this convenience given by SDN, such as firewalls [2], intrusion detection systems [4], load balancers [3], or in the case of this research, implementation of Quality of Service mechanisms to ensure better network management.

J. B. Torres (✉) · J. E. Regencia · W. E. S. Yu
Ateneo de Manila University, Quezon City, Philippines
e-mail: bryce.torres@obf.ateneo.edu

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_14

Regencia and Yu introduced a testing framework for QoS in SDN using synthetic traffic [5]. This framework tests Class-Based Queueing (CBQ) algorithms in distributed network topologies spanning through multiple layers. However, synthetic traffic data is limited in its variety of protocol and IP addresses, and does not replicate real-world applications. Other similar studies such as Akella and Xiong used iperf to also create synthetic data [8]. To more accurately measure the performance of CBQ, more realistic network traffic data must be tested in the framework. The goal of this research is to perform simulations similar to real-world applications on the test framework by Regencia and Yu by adding real network traffic test data. This is done by modifying the framework to replay PCAP files as test data.

The CBQ algorithms used for test simulations in this research were based on those used in Regencia and Yu, but modified to work with network traffic data that had IP addresses not found in the test network. The framework was made so that any PCAP can be used. This is to ensure not only the realism of the results but also the versatility of the framework in choosing network traffic test data when testing QoS mechanisms.

2 Review of Related Literature

2.1 QoS Test SDN Framework

This research is an extension of the work in Regencia and Yu. Regencia and Yu built a SDN framework to test QoS mechanisms [5]. The framework is designed to be modular in the QoS algorithms that can be implemented. The framework also provides a method to test the location of the QoS implementation, whether enforced at the core switch of a network or enforced at the edge of the network at the leaf switches. The framework allows for an easily modifiable network topology that can scale depending on the necessary size for the simulation [6]. Regencia and Yu also further tested the framework with larger traffic flows [7]. Their work extends from Chato and Yu, which explored protocol and source IP address CBQ algorithms [9]. Chato and Yu also tested the performance of the QoS algorithms in terms of latency [10].

2.2 PCAP File Replay

Similar studies on QoS implementation such as Akella and Xiong also use synthetic data [8]. A method to include real network traffic in the test framework is to use PCAP file replay, similar to those done in intrusion detection systems like that in Suba et al. [4]. Tcpreplay PCAP file replay provides realism and replayability so tests can be performed multiple times. PCAP also allow for PCAP modification, which provides control in the amount of packets to be tested, the raw bandwidth of the test, the variety of source and destination IP addresses, and almost any other relevant parameter.

2.3 Class-Based Queuing Algorithms

Lastly, the parameters to be simulated in this framework are QoS mechanisms, specifically CBQ algorithms. CBQ performs QoS by dividing bandwidth into queues and grouping data packets into classes based on certain parameters, which are assigned to queues [11]. These algorithms were originally explored in Chato and Yu [9]. Their research focused on Basic CBQ, which assigns packets to queues based on their protocol, and Source CBQ, which assigns packets to queues based on their source IP address. Destination CBQ, which groups packets based on their destination IP address, and Source-Destination CBQ, which groups packets based on a combination of source and destination IP addresses, were introduced by Regencia and Yu along with the test framework [5].

3 Framework

3.1 Framework Architecture

This study uses the framework in Regencia and Yu, but with PCAP files as the data payload. The topology was set to two clients at one layer only. Lastly, since the packets will be routed using their captured IP address parameters, the packets must move through the network without actually being sent to their correct IP addresses. Because of this, several flows were installed in the architecture to serve as default flows. If a packet passes through this network, if the IP address does not match any address on the network, it will flow from the server to the two clients, being split at the core switch based on IP ranges, the lower range being sent to the switch 1 branch and the upper range being sent to the switch 2 branch. Because of the limitations of testing data packets in this network, this study only runs data packets in one direction, from server to client. The previous studies performed QoS on packets in the direction of clients to servers and had data packets travelling in both directions as requests and replies. This difference should be noted.

3.2 Theoretical Framework

This addition of PCAP replay to the framework entailed modification to the QoS algorithms tested in Regencia and Yu. For each case, a new algorithm was created that was similar to the logic of the original but could be used on the packet files. Basic CBQ followed the same logic, but with TLSv1 protocol replacing VLC due to the lack of VLC packets in the PCAP file tested. Also, an additional "Miscellaneous" case was assigned when a packet belonged to neither HTTP nor TLSv1. The other CBQ cases were implemented using IP address ranges. The QoS bandwidth limit and the number of queues and assigned bandwidths to each follows the same in the previous studies of Regencia and Yu.

3.3 Conceptual Framework

The experiments were performed on an Ubuntu 20.04 running on a Lenovo G400, with Mininet 2.2.2, Ryu SDN Controller 4.3.4, IFSTAT for data collection, and tcpreplay 4.3.3 for PCAP replay. This study deals with the modification of the test framework to be able to use PCAP for realistic network traffic test cases. The tests use a PCAP file, bigFlows.pcap, that was captured for the purposes of simulating generic network traffic with a mix of different protocols, taken from the Tcpreplay sample captures webpage [12]. The network topology was set to use only two clients and at one layer of leaf-switches. The PCAP file was replayed from one server (see Fig. 2).

4 Methodology

4.1 Test Simulation Methdology

As an extension of the work by Regencia and Yu, the test simulation methodology of this research is the same, aside from the use of PCAP files instead of synthetic traffic. The simulations were performed for 30 trials at one minute for each CBQ case. The throughput of each test was recorded using IFSTAT bandwidth statistics.

4.2 Framework Modification

The goal is to show that PCAP replay functions within this framework and is comparable to the results of previous studies. BigFlows.pcap is replayed at real-time using tcpreplay at server automatically in the test script. Unlike synthetic testing tools, tcpreplay does not have measuring capabilities. Instead, IFSTAT was used to collect throughput data. IFSTAT is set to collect the total bandwidth reading at the ends of the leaf-switches before the clients, which are the last interface of the topology. Because the data only goes from server-to-client, the data is unidirectional. The CBQ

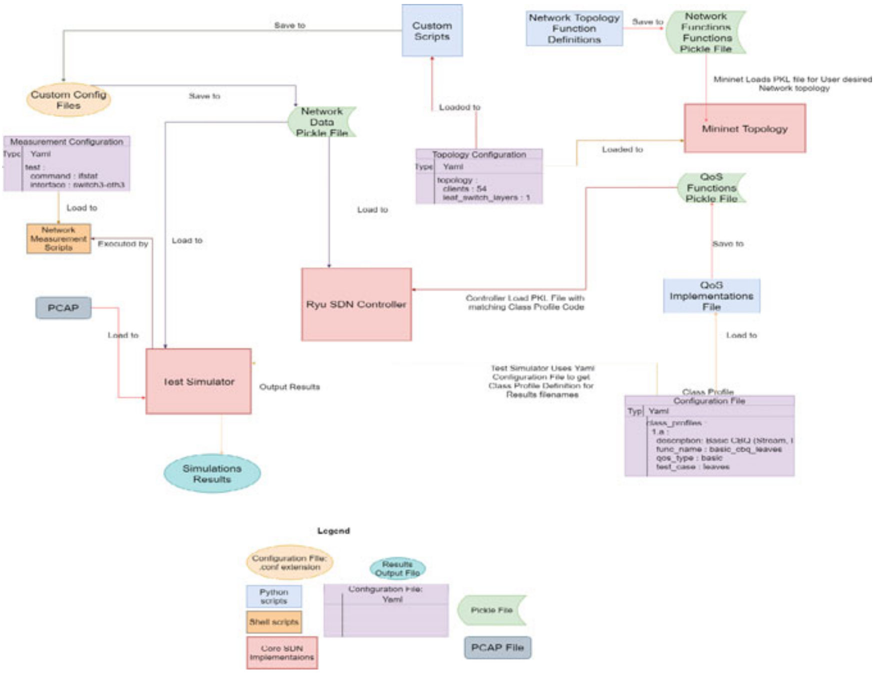


Fig. 1 Test framework architecture from Regencia and Yu with minor PCAP file addition [5]

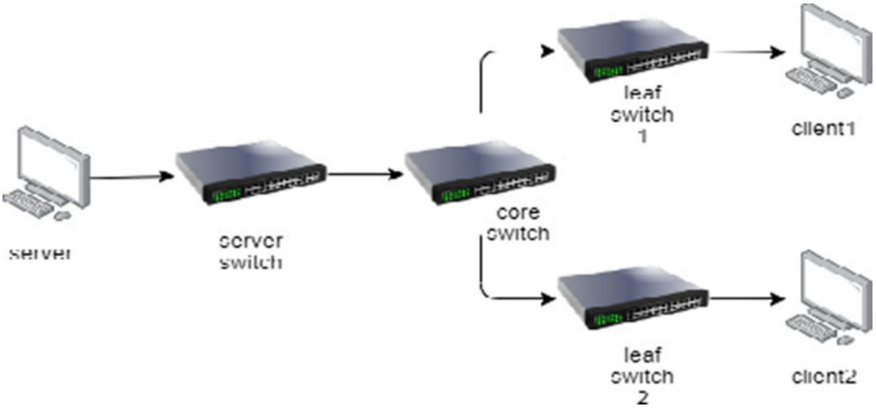


Fig. 2 Network topology used

Table 1 Class profiles

Class profile	CBQ scheduling class	Switch QoS
Basic CBQ at leaves	Traffic protocol (HTTP & TLSv1)	Client leaf switches
Basic CBQ at core	Traffic protocol (HTTP & TLSv1)	Core switch
Src CBQ at leaves	Source IP address grouping	Client leaf switches
Src CBQ at core	Source IP address grouping	Core switch
Dst CBQ at leaves	Destination IP address grouping	Client leaf switches
Dst CBQ at core	Destination IP address grouping	Core switch
Src-Dst CBQ at leaves	Src/Dst IP address grouping	Client leaf switches
Src-Dst CBQ at core	Src/Dst IP address grouping	Core switch

Table 2 Quality of service configuration in GBps

QoS parameters			Queues		
		Q (linux-htb)	q0	q1	q2
QoS interface	Min. rate		0.33	0.33	0.33
	Max. rate	1	0.33	0.33	0.33
	Priority		0	1	2

algorithms and the QoS settings are already set during the set-up of the framework prior to the execution of the test. The QoS is set at the egress interface of the relevant CBQ algorithm, at the core switch for core-enforced cases or each leaf switch for leaf-enforced cases.

5 Results and Discussion

The raw throughput results can be found in Table 3. Similar to the results of Regencia and Yu, the mean throughput, standard deviation, and the minimum and maximum throughput were recorded for each CBQ case [5]. There is an additional parameter of amplitude, the difference between maximum and minimum throughput, which was considered in order to add more metrics. It is notable that the raw throughput is significantly lower than that found in Regencia and Yu, but this is because the network traffic used in this study was of a lower bandwidth, while the synthetic traffic used in Regencia and Yu was configured to flood the network significantly [5]. However, this should not affect the findings of this study because the focus of the comparison between these results and that in Regencia and Yu is the behavior of the

Table 3 Raw ifstat results in KB/s

CBQ case	Mean	Std. dev.	Minimum	Maximum	Amplitude
Basic at leaf	1194.379	499.942	311.469	2535.634	2224.165
Basic at core	1194.021	496.596	310.623	2521.377	2210.753
Src at leaf	1208.709	505.330	301.803	2510.080	2208.276
Src at core	1208.581	503.341	309.488	2497.440	2187.952
Dst at leaf	1209.572	514.298	287.395	2595.138	2307.743
Dst at core	1208.652	504.604	307.577	2499.065	2191.488
Src-Dst at leaf	1208.772	506.780	307.259	2503.516	2196.257
Src-Dst at core	1208.646	505.722	301.082	2524.188	2223.106
Average	**1205.167**	**504.577**	**304.587**	**2523.305**	**2218.717**

Table 4 Ifstat results as percent difference of total average

CBQ case	Mean (%)	Std. dev. (%)	Minimum (%)	Maximum (%)	Amplitude (%)
Basic at leaf	–0.895	–0.918	2.259	0.488	0.245
Basic at core	–0.924	–1.58	1.981	–0.076	–0.358
Src at leaf	0.293	0.149	–0.913	–0.524	–0.470
Src at core	0.283	–0.244	1.609	–1.025	–1.386
Dst at leaf	0.365	1.926	–5.644	2.846	4.012
Dst at core	0.289	0.005	0.981	–0.960	–1.227
Src-Dst at leaf	0.299	0.436	0.877	–0.784	–1.012
Src-Dst at core	0.288	0.227	–1.150	0.035	0.197

data in order to ensure the validity of the PCAP replay data and the consistency of QoS performance.

The relative performances of each CBQ case can be found in Table 4. The best performing throughput is the Destination CBQ at the Leaf, similar to the results found from the simulations from Regencia and Yu [5]. The case with the strongest bandwidth-limiting effect is the Basic CBQ at Core. This is similar to the findings of Chato and Yu [9]. There is a trend where Leaf cases perform generally better than Core cases and the standard deviation and amplitude are consistent which shows accurate results.

According to the student’s t-test results in Table 5, all p-values are significantly higher than $\alpha = 0.05$, which means the leaf-enforced cases perform at least the same if not better than the core-enforced cases, which is similar to the findings of Regencia and Yu, aside from the Source-Destination at the core performing better than Source-Destination at the leaf [5]. When based on the number of trials where the raw throughput of leaf-enforced algorithms performed better than their core-enforced counterparts, the results show that in Basic CBQ 53.33% of 30 trials showed leaf advantage, in Source CBQ 70% of 30 trials showed core advantage, in Destination

Table 5 Core versus leaf results comparison with Regencia and Yu [5]

Algorithm	P-value	Leaf versus core	Regencia and Yu
Basic CBQ	0.9977893014	Comparable	Comparable
Src CBQ	0.9992210504	Comparable	Comparable
Dst CBQ	0.9944403091	Comparable	Comparable
Src-Dst CBQ	0.9992325734	Comparable	Comparable

CBQ 100% of 30 trials showed leaf advantage, and in Source-Destination CBQ 70% of 30 trials showed leaf advantage. When each algorithm's leaf and core-enforcement were compared using student's t-test, the results show that the difference in their performance is not statistically significant. As shown in Table 5, the leaf-enforcement performance is comparably similar to the core-enforcement, which is similar to the findings of Regencia and Yu, at least at a one layer topology.

6 Conclusion

This research extends the work done in Regencia and Yu by adding real network traffic data into the framework, allowing for more accurate performance tests for the QoS mechanisms. The results of the Ifstat bandwidth tests are consistent with the results found in Regencia and Yu, which confirms the validity of the PCAP data and the consistent performance of the QoS algorithms. Leaf-enforced Destination CBQ is the best performing case in terms of mean throughput. P-values show that the performance of Leaf-enforced algorithms are comparable, meaning as good as their core-enforced counterparts. The results show that generally leaf-enforced CBQ cases perform better than core-enforced CBQ cases with relatively consistent standard deviation and amplitude. Further tests can be done by using PCAP files of different traffic types and IP address ranges to observe their behavior.

References

1. Nadeau TD, Gray K (2013) SDN: software defined networks, 1st edn. O'Reilly Media, Sebastopol, CA
2. Pena JGV, Yu WES (2014) Development of a distributed firewall using software defined networking technology. In: 4th IEEE international conference on information science and technology. IEEE, pp 449–452
3. Atienza PV, Yu WES (2018) A framework for performance analysis of various load balancing techniques in a software-defined networking environment. In: International conference on information science and applications. Springer, Singapore, pp 67–74
4. Suba AM, Bautista KV, Ledesma JCT, Yu WES (2018) Developing a testing framework for intrusion detection algorithms using software defined networking. In: International conference

on information science and applications, pp 303–313

5. Regencia JET, Yu WES (2021) Introducing a test framework for quality of service mechanisms in the context of software-defined networking. In: Accepted in the 6th international congress on information and communication technology ICICT
6. Regencia JET, Yu WES (2020) Exploring the effects of network topology layers on quality of service mechanisms in the context of software-defined networking. In: Accepted in ICISA
7. Regencia JET, Yu WES (2021) Latency and throughput advantage of leaf-enforced quality of service in software-defined networking for large traffic flows. In: Accepted in SAI computing conference 2021
8. Akella AV, Xiong K (2014) Quality of service (QoS)-guaranteed network resource allocation via software defined networking (SDN). In: 2014 IEEE 12th international conference on dependable, autonomic and secure computing. IEEE, pp 7–13
9. Chato O, Yu WES (2016) An exploration of various quality of service mechanisms in an openflow and software defined networking environment. In: The international conference on systems and informatics, pp 768–776
10. Chato O, Yu WES (2016) An exploration of various quality of service mechanisms in an openflow and software defined networking environment in terms of latency performance. In: International conference on information science and security, pp 1–7
11. Park KI (2005) QoS in packet networks, 1st edn. Springer Science+Business Media Inc., Boston
12. TCPreplay Sample Captures, https://tcpreplay.appneta.com/wiki/captures.html. Accessed 30 Nov 2020

Intelligent Vehicular Networking and Applications

Application of Micro-frontends to Legal Search Engine Web Development

Nattaporn Noppadol and Yachai Limpiyakorn

Abstract Software systems are traditionally separated into front and rear architecture. The backend is responsible for data processing on server side, while the frontend accounts for interaction between clients and a system. Among several today backend architectural approaches, Microservices is an alternative most suitable for scalable systems. Meanwhile, the client-side applications are also growing with size and complexity. The concept of Micro-frontends has recently emerged as logical evolution of architecture for frontend side of web applications. Similar to Microservices, both concepts benefit concurrent development in addition to greater performance that result from partitioning large applications into smaller parts. This paper presents an application of Micro-frontends to Thai legal search engine web development. Since search engines are typically large-scale software projects, single page applications tend to bloat up, not well-scaled, and costly maintain. A design scheme of the client-side of legal search engine system based on Micro-frontends combined with the backend technology, Microservices, is described in practice process.

Keywords Micro-frontends · Microservices · Architecture · Web development

1 Introduction

Over the years, Microservice architecture has been introduced to solve the complexity of system maintenance and to increase the scalability of the backend and the system infrastructure. Considering the frontend side, the term Micro-frontends was first mentioned in 2016 by Thought Works Technology Radar [1]. It has been expanded into a concept and principle by Michael Geers as referred in [2, 3]. Afterwards, in 2019, Zafer [4] publicized an article on the understanding and concept of web application development with Micro-frontends architecture by expanding Geer's concept.

N. Noppadol (✉) · Y. Limpiyakorn
Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand
e-mail: 6270085221@student.chula.ac.th

Y. Limpiyakorn
e-mail: Yachai.L@chula.ac.th

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_15

He also addressed the problems incurred with this development process as well as included the guidelines for solving the problems. In the same year, Yang et al. [5] presented an approach of applying a popular Micro-frontends framework, Mooa, for developing a content management system (CMS) essential for enterprise information construction. The research of Pavlenko et al. [6] in 2020 demonstrated the advantages and disadvantages of web application development using Micro-frontends architecture. In this work, the concept of Micro-frontends has been applied for reengineering the web development of Thai legal search engine formerly implemented as single page application.

2 Background

2.1 Microservices

Due to the limitation of scalability, many organizations nowadays have evolved their technology stack from the traditional Monolithic architecture to Microservices. With Monolithic design, a change made to a small part of the application also requires the entire system to be rebuilt and deployed. This makes it harder to maintain a good modular structure over time. In contrary, the architectural design with Microservices is based on the scaling model called *scale cube* [7]. Each service can be individually scaled. As a result, the application is scaled more efficiently. The concept of Microservices enables implementation of complex software applications as a suite of independently deployable, small, modular services in which each service runs a unique process and communicates through a well-defined, lightweight mechanism, often HTTP Resource API [8]. The architecture is designed to distribute without a center and with rapid continuous delivery. Moreover, each service is independent of the others. It can be written in any languages, not necessary to be the same language. With respect to the database, there is no need to share the same database as one service owns some piece of data and can create Build, Test, and Release function. All devices have to be prepared for implementing the Release function without having any effect with existing functions.

2.2 Micro-frontends

"Micro-frontends" have been introduced from the ThoughtWorks Technology Radar at the end of 2016 [1]. The term denotes a design concept of the frontend architecture in modern web application development influenced by today backend technology so called Microservices, which promotes scalability of the development of backends and infrastructures. It extends the idea to the advent of Microservices to be used on the frontend world with developers looking at everything as a component and working out for each component to communicate independently. The developers do not have to worry about the code of other parts of the software. They just focus on the components that have been assigned to develop only. Each developer team can implement with different languages making it possible to use a variety of library programs together. The core ideas of Micro-frontends described by Geers [2, 3] include *Be Technology Agnostic; Isolate Team Code; Establish Team Prefixes, Favor Native Browser Features over Custom APIs,* and *Build a Resilient Site.*

3 Research Methodology

Lawphin is Thai legal search engine web application that serves querying the information of Judgments from various courts. In the early days, *Lawphin* has been implemented with Microservices architecture. However, the architecture of frontends is Monolithic with Single Page Application (SPA) technology using Angular Framework version 1, as shown in Fig. 1. It is very difficult to add features or even update the Angular Framework version. This paper adopts the concept of Micro-frontends to modify the *Lawphin* web application to be more modular by separating the client-side components based on the existing Microservice teams as shown in Fig. 2. Observing that each team can work concurrently and independently from each other. The developer teams can choose their own frameworks. Additionally, each team can develop and test from backend to frontend, from start to end (end-to-end).

Figure 3 illustrates the excerpt of the interface flow of all windows using Windows Navigation Diagram (WND) which visualizes the UI structure design of *Lawphin* web application. A box represents a UI component. The transition is modelled as either a single header or double header arrow. For example, pressing the Judgment & Opinions button, it will display a search page for the user to enter their keyword search and press search. The system will then display a search result page where the information provided by each development team will be described in the following subsections.

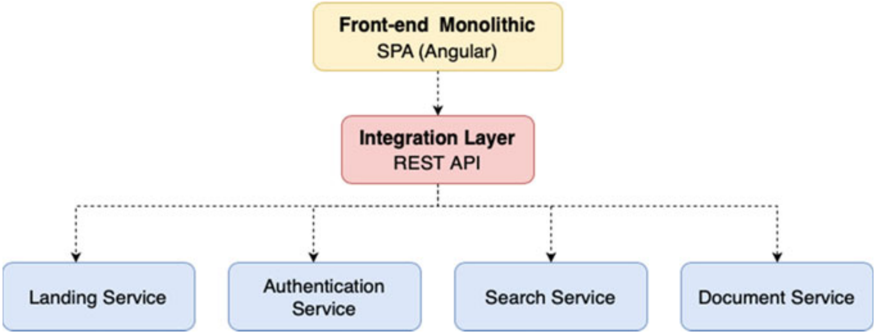


Fig. 1 Original structural design of *Lawphin* website

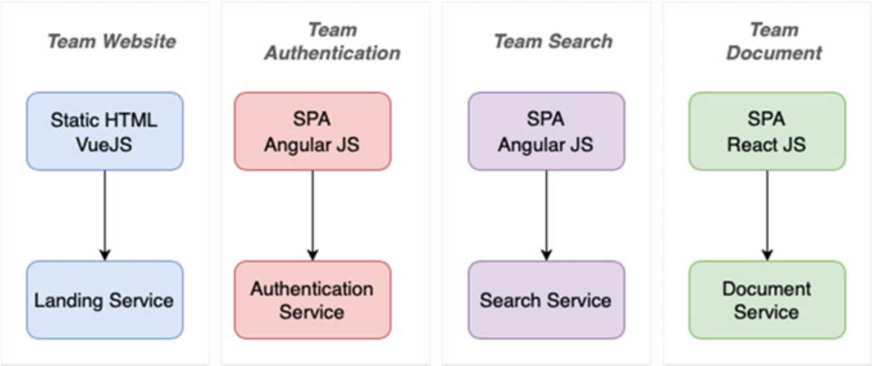


Fig. 2 Structural design of *Lawphin* website with Micro-frontends architecture

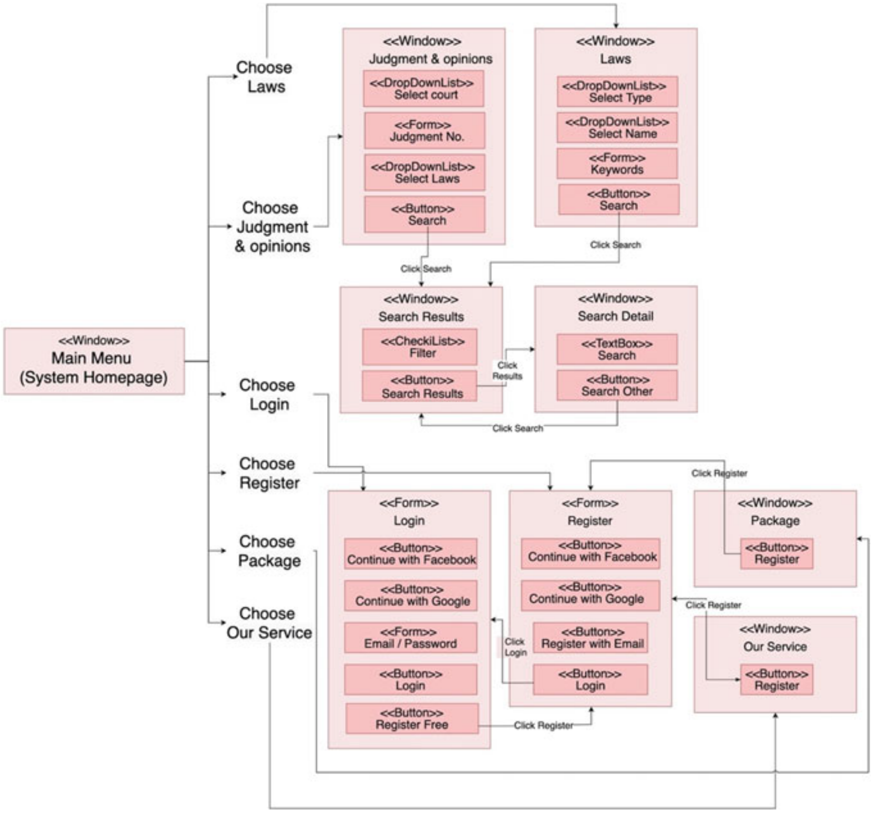


Fig. 3 Visualization of UI structure design of *Lawphin*

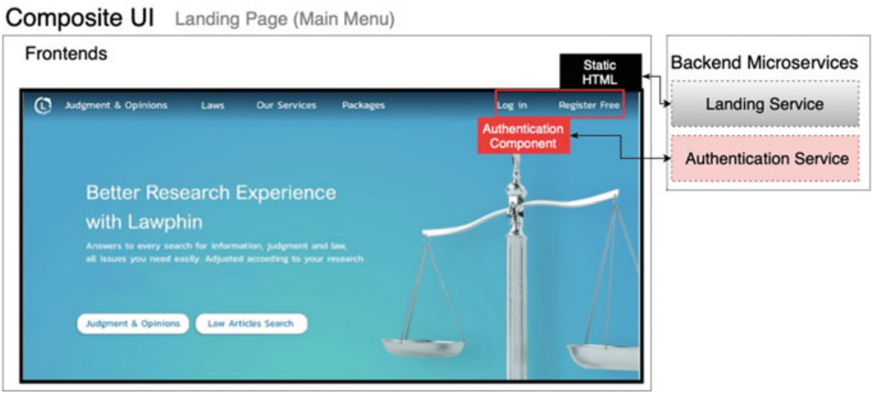


Fig. 4 Display screen developed by Website Team

3.1 *Website Team*

Referring to Fig. 3, this team is responsible for developing the home page. They opt to use static HTML for front-end development as shown in Fig. 2 in order to display the website contents that they get from the Landing Service (Fig. 4) which provides a server-side content rendering and the first web page called landing page. This is because most of the information is unchanged and needs to improve the SEO by requirements. The authentication component is located at the top right of the home page (Fig. 4). It has to import from the authentication team.

3.2 *Authentication Team*

This team has a responsibility to develop the authentication component and authentication service being invoked on every page. They have chosen to use Angular Framework (Fig. 2) for creating the custom component with the integration service. They will need to manage authentication for external login service (Social network login) (Fig. 5), and the local login service. This component will be activated on every page of this application when the user clicks login from the navigation bar (Fig. 3). It will show the login form component created by this team.

3.3 *Search Team*

The team is in charge of developing the search form component, court service to get the court list used by the search form. For the relevant law service to get the relevant law list used by the search form, they use Single Page Application to display results

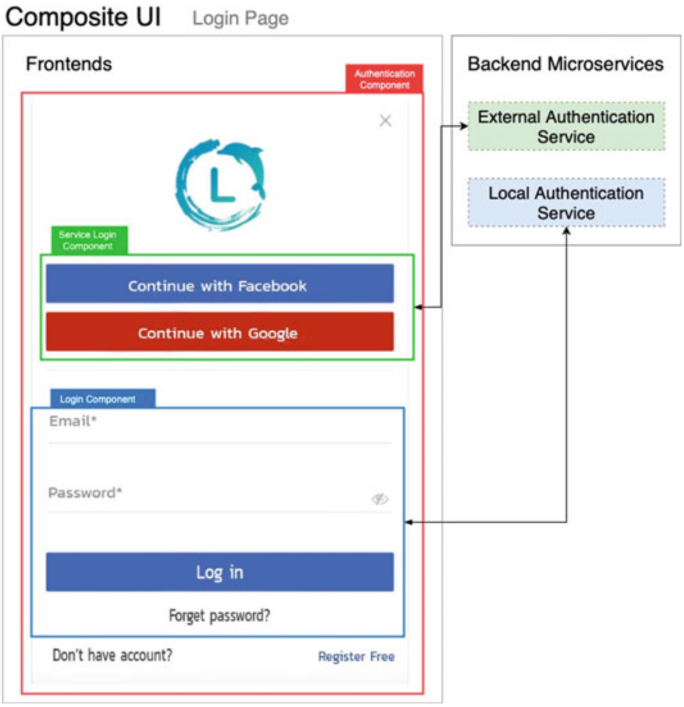


Fig. 5 Login screen for user identification

by using Angular Framework. After the component has been developed, it will be attached to the main layout component shown in Fig. 6. When the user clicks search, the search form component will be replaced by the search result component as shown in Fig. 7.

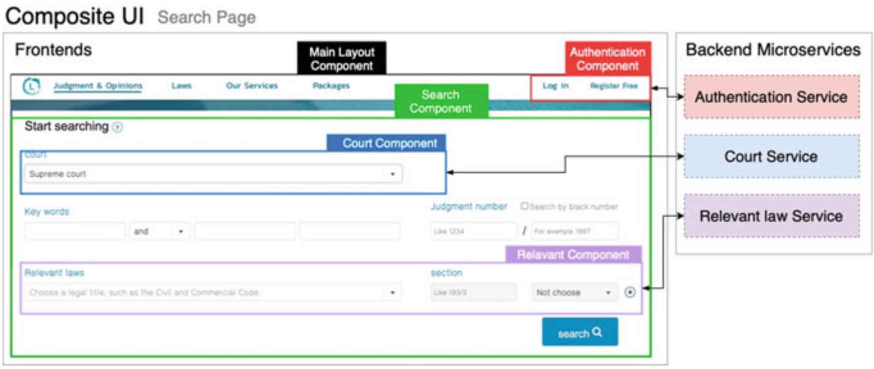


Fig. 6 Capture screen of search page

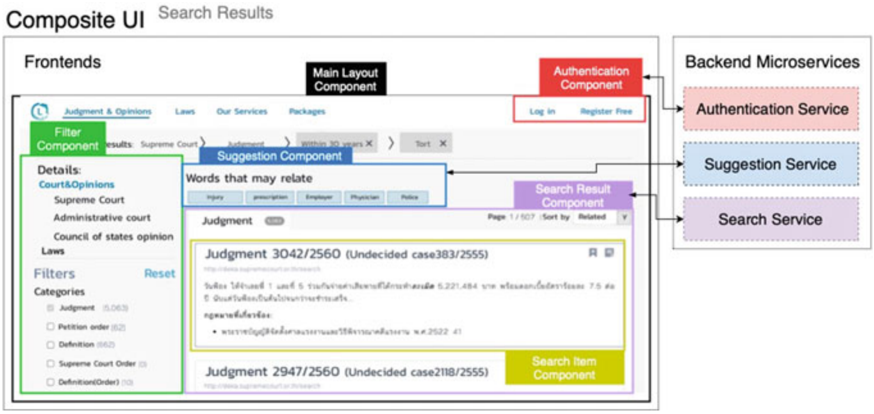


Fig. 7 Capture screen of search results

3.4 Search Result Team

The team is responsible for passing the search result from the service to show in the search result component. It will have an iteration of the resulting item component for reusable. The search result team also opts to use Angular Framework (Fig. 2) for developing the component. They need to manage showing the suggestion from the search suggestion service and filtering from the filter component in the left sidebar shown in Fig. 7. When the user selects a result from the item list, the search result component will be replaced with the detail document component shown in Fig. 8.

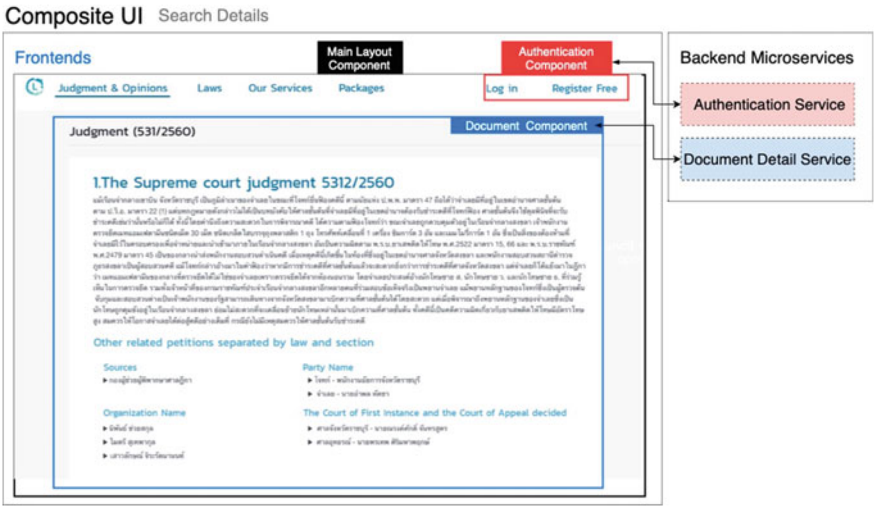


Fig. 8 Capture screen of search details

3.5 Document Team

The team is responsible for showing the whole document from the document service as shown in Fig. 8. A single page application is used to render it by using ReactJS (Fig. 2) for development. This feature is a new feature that will be added later by teams. Therefore, they have the freedom to choose the framework that suits for the team as well as the performance and characteristics of this component.

4 Conclusion and Future Work

This paper presents a UI structure design of Thai Legal Search Engine website called *Lawphin.* The reengineering of *Lawphin* takes place on the frontend part by adopting the concept of Micro-frontends architecture, expecting for better orchestration with the current backend technology, Microservices. The migration to Micro-frontends would benefit increased productivity since each part of the website is quite independent and can be implemented in parallel by several development teams. Each team can pick their own framework and programming language. The developers in each team can also develop, test and deploy in a parallel format for each team. In terms of scalability and upgrade, there is no need to upgrade the frontend system all at once, but simply upgrade the modules that are needed first. Moreover, since developers can run tests on each module in parallel, it is not necessary to wait for every module to pass the test. It is possible to release and perform a separate software versioning in each module. However, due to separate front-end development, it can cause problems with CSS and JavaScript overlap. Implementation with various technologies also incurs expense and make it more complicated to integrate all subsystems.

References

1. ThoughtWorks (2016) Technology Radar: Micro frontends. https://www.thoughtworks.com/radar/techniques/micro-frontends. Accessed 16 Jan 2021
2. Geers M (2020) Micro-frontends in action. Manning Publications
3. Geers M (2017) Micro-frontends-extending the microservice idea to frontend development. https://micro-frontends.org/. Accessed 16 Jan 2021
4. Öner Z (2019) Understanding Micro-frontends. https://hackernoon.com/understanding-micro-frontends-b1c11585a297. Accessed 16 Jan 2021
5. Yang C, Liu C, Su Z (2019) Research and application of micro frontends. In: IOP conference series: materials science and engineering, vol 490, no 6. IOP Publishing
6. Pavlenko A, Askarbekuly N, Megha S, Mazzara M (2020) Micro-frontends: application of microservices to web front-ends. J Internet Ser Inf Secur 10:49–66. https://doi.org/10.22667/JISIS.2020.05.31.049

7. Abbott M, Fisher M (2015) The art of scalability: scalable web architecture, processes, and organizations for the modern enterprise. Addison-Wesley Professional
8. Malavalli D, Sathappan S (2015) Scalable microservice based architecture for enabling DMTF profiles. In: 2015 11th international conference on network and service management (CNSM). IEEE

Blockchain and Cryptocurrency

A Trust Model for Context-Aware E-Health Services

Brenda Ayuku, George Okeyo, Agnes Mindila, and Wekesa Chemwa

Abstract Technological advancements such as ubiquitous communication, pervasive computing and ambient intelligence, have made context-aware services a reality. The health sector has adopted the use of context-aware e-health services to improve the quality of life, provide real-time patient monitoring, and in general better healthcare. However, the capability of accessing patients' context and health information via context-aware e-health services has given rise to issues of privacy and trust. The main objective of this research is to propose a Trust Model, which ensures that the privacy of the patient's health and context information is maintained and secure from unauthorised access. This means that the patient cannot only trust the service and context-providers, but also has control of their context and health information. The proposed trust model monitors the patient's health status and context information using sensors. The model is implemented using a proof-of-concept context-aware e-health service that uses blockchain technology to ensure data integrity and accountability. If successfully implemented, the security of the user's context and health information will be guaranteed, and thus lead to increased uptake of context-aware e-health services.

Keywords Context-aware services · E-health · Privacy · Trust · Context

1 Introduction

Health services include all services dealing with the diagnosis and treatment of disease, or the promotion, maintenance and restoration of health. They include personal and non-personal health services. These services are the most visible functions of any health system, both to users and to the public [1]. Technological advances in ubiquitous communication, pervasive computing and ambient intelligence present a unique opportunity for context-aware e-health services [2]. According to Dey [3] context is any information that can be used to characterise the situation of an entity.

B. Ayuku (✉) · G. Okeyo · A. Mindila · W. Chemwa
Department of Computing, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

H. Kim and K. J. Kim (eds.), *IT Convergence and Security*, Lecture Notes
in Electrical Engineering 782, https://doi.org/10.1007/978-981-16-4118-3_16

An entity being a person, place, or object that is considered relevant to the interaction between a user and an application, including location, time, activities, and the preferences of each entity. Context-awareness means one can use context information. A health system is context-aware if it can extract, interpret, use context information and adapt its functionality to the current context of use [4]. Context-aware e-health services provide a platform for patients to play an active role in their care. This is possible through communication between smart objects and patient's devices, which include smart phones, smart watches and tablets. These platforms focus on providing preventive care, proactive services and healthy lifestyle [5]. They measure physical properties of objects and the environment. The physical properties include body temperature, pulse rate, blood sugar level, blood pressure, body position, among other things. This functionality is achieved with the use of bio-medical and context sensors. Examples of context-aware e-health systems include LARIISA and ERMHAN. LARIISA is a context-aware intelligent framework for a governance model supporting decision making in public health care system [6], while ERMHAN is a multi-channel context-aware service platform designed to support care networks in cooperating and sharing information with the common goal of improving the patient's quality of life. The implementation and use of context-aware services has led the users to question whether they can trust these services. Trust can be understood as the subjectively perceived probability by a data subject that a system will perform an action before the data subject can monitor it [7]. Privacy and trust are interrelated concepts, that is, "data disclosure means loss of privacy, but an increased level of trustworthiness reduces the need for privacy" [8]. Privacy issues include the authenticity and reliability of the information receiver, the usage of the patient's information, the sensitivity of the patient's health information, and the reliability of the context environment. People are sensitive about revealing their location and activities, but context-aware systems often transmit context information without requiring a specific user action, to increase their usability. For example, the CenceMe system transmits periodically the current user's location to a group of friends [9]. It would not make sense to ask the user for permission before each transmission. Moreover, the users are comfortable with automatic transmissions because they have predefined with whom they are willing to share their location. In context-aware e-health systems, this may not be the case. Users who easily share their location among their friends will probably reject sharing their health status and records to the whole community, because they do not trust the recipients of that information [10]. Context-aware e-health services need to ensure that the patient's context and health information is secure from unauthorised access for the patients to trust the platforms. This article's proposed trust model will ensure patient's information is secure, thus making the patients trust context-aware e-health services. The contribution of this model is to ensure that only authorised personnel access the patient's context data by computing the trustworthiness of medical personnel before granting them access to the patient's data and ensuring data integrity is maintained by using blockchain to track all the changes done.

2 Motivation/Justification

The realisation of context-aware e-health services is faced with the challenge of enforcing patient's trust in the system. This problem emerges due to context-aware e-health security issues, which include patient confidentiality, secure information processing, secure transmission, secure storage, authentication and data integrity among others. If a patient's context and health information is accessed by unauthorised personnel, the information can be deleted, distorted or used for criminal activities like identity theft, blackmail, tracking the users to mention but a few. Due to these reasons, it is imperative for patients to trust the systems they are interacting with, and they should know the trustworthiness of the entities involved, be aware of who is accessing their information, and know how their information is being used. Furthermore, they should have the power of deciding the context information they want to provide and have the capability of verifying the integrity of the context information. The proposed trust model will ensure that patients' context and health information is secure from unauthorised access. The patients will have full control of their electronic medical records and context data hence increasing their trust in context-aware e-health services system. Most importantly, the health sector will also be able to provide privacy enhanced context-aware e-health services.

3 Related Works and Technologies

This section highlights related models and technologies that have been used in this research, which have helped to build a foundation for the formulation of the proposed Trust Model for Context-Aware e-health Services.

3.1 REK Trust Evaluation Model

Truong et al. [11] propose a trust evaluation model called REK. It is comprised of the triad of trust indicators (TIs): Reputation, Experience and Knowledge. The REK Model covers multi-dimensional aspects of trust by incorporating heterogeneous information from direct observation (as Knowledge TI), personal experiences (as Experience TI) to global opinions (as Reputation TI). Knowledge TI is the direct trust mentioned that renders trustor's perspective on trustee's trustworthiness in a respective environment. Experience and Reputation TIs are social features attained by accumulating previous interactions among entities over time. Reputation is considered when evaluating trust because of the propagation characteristic of trust. Truong et al. [11]'s proposed model does not adapt to the health service use case, which requires figuring out a set of trustworthiness attributes for Knowledge TI in detail as well as appropriate mathematical parameters for Experience and Reputation TI. It also lacks the capability of autonomously adapting with changes of the knowledge base, resulting in an autonomous trust computation framework and with real-time data streaming which our proposed trust model for context-aware e-health services cater for filling this gap.

3.2 Trust-Based Context-Aware Recommender Systems

Otebolaku and Lee [12], proposed a Trust Evaluation Model conceptual framework for exploiting the Internet of Things context-awareness for predicting the user's preferences for personalised services. Their Trust Evaluation Model aggregates trust-related information and feedback obtained from users whenever they consume services in specific contexts to derive an evaluation for the trustworthiness of context data. Although Otebolaku and Lee [12] model shows promise in terms of its potentials to improve the accuracy of recommendations and their relevance to the user's contextual preferences, they have only evaluated recommendations based on an extended collaborative filtering approach. However, they have not used trust in the current work to enforce security, but it has been used to improve the quality of recommendations. Consequently, trust has not been utilised as a means of ensuring that data or information from malicious entities is not allowed in the context-aware personalised service recommendation process.

3.3 A Context-Based Trust Management Model

Razavi et al. [13] proposed a Trust Management Model for a pervasive computing environment, where they assign the trust value of zero to the new entity. In this model, the interactions with the new entity can happen when other entities have negative trust values (untrustworthy entities). Recommendations help a service requester to compute an indirect trust in the case that there are not adequate records in the interaction history for direct trust computation. In the Context-Based Trust Management Model, dishonest recommenders are identified, and all recommendations provided by them are excluded from indirect trust computation. To identify a dishonest recommender, the service requester uses all recommendations, which are received from a specific recommender, and computes the mean value of the recommended trust values. In the case that the mean value is so low or so high (not in an adequate range), the service requester judges the recommender to be dishonest. The method of assigning weights to the interactions over time causes each past interaction to be effective in trust computing according to the assigned weight. Therefore, the weighting mechanism can protect the entity against the dynamic behaviour of malicious recommenders.

3.4 Context-Aware Trust Model

M'Hamed et al. [14] proposed a Context-Aware Trust Model based on user behaviour by taking into consideration both the user profiles and context attributes. They introduced protection against the malicious threats affecting the trust evaluation process. They also improved the accuracy of trust metrics based on the right human behaviour in situations that require trust.

3.4.1 Trust Degree Model

The Pervasive Trust Model of Almenárez et al. [15], applies the concept of trust degrees in the definition of access control policies. Once trust is formed, they obtain an initial trust value; this value is their belief space. However, the trust changes according to the entity's behaviour by providing feedback about entity's performance during the interaction, which is the evidence space. Almenárez et al. [15] called this process evolution, because it is constantly changing. Both trust formation and trust evolution allow the creation of trust relationships between entities, and these relationships are supported by the trustworthiness of the system and the communication. However, they neither explicitly support trust quantification for identities, nor do they target context-aware service e-health platforms.

3.4.2 Trust Modelling Computation

Yang et al. [16] trust modelling computation. The following equation is the formalised formula of updating function:

Updating function, Yang et al. [16]

$$\begin{aligned} fe_n(p,q) &= \textbf{update}\,(fe_{n-1}\,(p,q), g_n(p,q)), \\ g_n(p,q) &= \textbf{update}\,(g_{n-1}\,(p,q),\ \textbf{evidence}\,(p,q)) \end{aligned} \tag{1}$$

Here, they assume that node p (called trustor) needs to calculate the trust degree of node q (trustee). And $fe(p, q)$ denotes the nth trust decision computing in node p while $g_n(p, q)$ denotes the nth evidence modelling. Appropriate initialized data will be set for the very first iterative computing. Trust evidences in their model come from Evidence Collector including explicit feedback evidence and implicit feedback evidence. The explicit feedback evidence is from customer nodes active evaluation, while implicit feedback evidence should be excavated from transactions. They formalize trust evidence vector in their model as Evidence (p, q) to represent evidences that node p has upon node q. That is modelled by

Trust Evidence Vector, Yang et al. [16]

$$\textbf{Evidence}\,(p, q) = \langle\langle\varphi\rangle, \langle\lambda, \nu, \rho, \textbf{sim}\rangle\rangle. \tag{2}$$

In their model Yang et al. [16], explicit evidence is feedback by user ratings or satisfactions. In practice, the quantitative or qualitative feedback can be by scores (5 points or 100 points) or by satisfaction (satisfied, mostly satisfied, not satisfied, etc.).

3.5 Blockchain

Blockchain is a decentralised, distributed and public digital ledger that is used to record transactions across many computers so that the record cannot be altered retroactively without the alteration of all subsequent blocks and the consensus of the network [17].

3.5.1 Blockchain-Based Access Control Framework

Ouaddah et al. [18] implemented an access control Blockchain Model. In their framework, users comprised of an information owner or the requester and their information had an unlimited amount of cryptographic identities, called addresses. Addresses were public and shared in the network. They were used to grant and ask for an access token. Every access token is encrypted with a public key extracted from the address of the requester requesting access to which the token is designated. This ensures that only the requester with the token will be able to decrypt the information using his or her private key. All the transactions have an identifier, input and output. They uniquely identify transactions using their cryptographic hash, which is a digital fingerprint generated by hashing the transaction.

4 Method

This section provides an overview of our proposed Trust Model for Context-Aware e-health services. The model is an improvement of the existing trust models as it includes a blockchain solution, which keeps an immutable record of the patient's history, and shows who modified the data and at what time. It notifies the patient whenever their information is accessed, modified and provides feedback. Figure 1 illustrates the proposed model. As illustrated in foregoing parts of this article, none of the existing trust models in context-aware services incorporates the use of blockchain technology.

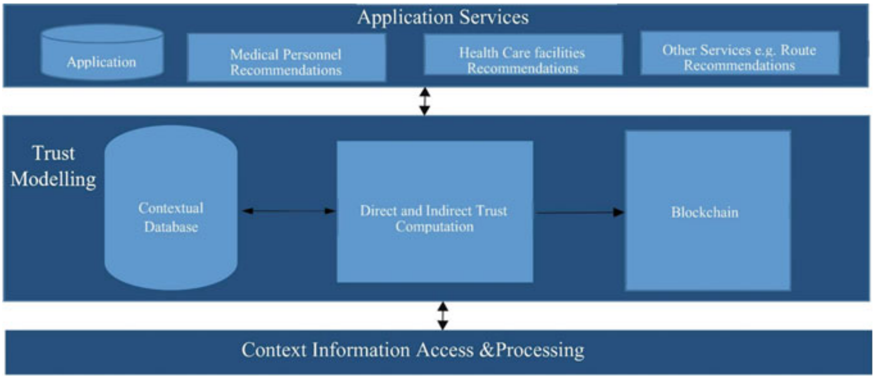

Application Services
Application
Medical Personnel Recommendations
Health Care facilities Recommendations
Other Services e.g. Route Recommendations
Trust Modelling
Contextual Database
Direct and Indirect Trust Computation
Blockchain
Context Information Access &Processing

Fig. 1 Trust model for context-aware e-health services

4.1 The Trust Model

The context information access and processing layer as shown in Fig. 1 collects the user's contextual data and manipulates it into useful information. The trust model is responsible for computing the trustworthiness of our context-aware e-health service and ensuring patients can trust the platform. It comprises of the contextual data, which includes the patient's current health status, medical history, current location and their activities. The data also comprises information of the available medical personnel, their context data, medical facilities, and the services offered. The third layer consists of the application services recommended to the user, like the healthcare facilities near them, the best route to access the facility, and the medical personnel to grant access to their data among other services.

4.1.1 Trust Modelling

The data collected is stored in the contextual database where it is categorised into the health data of the patients, including personal data, which is about the person using the service, the trust value of the medical personnel, and the location of the patient and the facility. Medical personnel without a trust value is assigned a zero. The model adopts Razavi et al.'s [13] Trust Management Model. The proposed model for this study also incorporates [16] trust modelling computation to calculate the direct and indirect trust. Direct trust is obtained from the patient's personal experience, while indirect trust is obtained from recommendations from other patients. To compute the trustworthiness of medical personnel and the context-aware e-health service the proposed model adopts the trust values stored in the context database and treats each entity as a node. When a patient node wants to obtain the trust value of medical personnel and the context-aware e-health service, it first checks its recorded list of medical personnel nodes. The direct and indirect trust metrics are used to evaluate the trustworthiness of medical personnel and context-aware e-health service nodes based on the recorded list. Here, the direct trustworthiness of the context-aware e-health system from patient evaluation, which is from the direct interaction with the medical personnel and the context-aware e-health service, is defined. Let $\boldsymbol{dt_n}$ $(\boldsymbol{p}, \boldsymbol{m}) \in [\mathbf{0}, \mathbf{1}]$ denote the direct trust of the patient (p) to medical personnel and the context aware e-health service (m). That is

Direct Trust

$$\boldsymbol{dt_n}(\boldsymbol{p}, \boldsymbol{m}) = \boldsymbol{\varphi_n}(\boldsymbol{p}, \boldsymbol{m}) \tag{3}$$

As illustrated in Eq. 3 when *m* provides exemplary service to *p* the higher the degree of trust *m* gets from *p*. On the other hand, the proposed model computes indirect trust based on the recommendations of other patients *c*. The patient node p calculates the degree of trust of the medical personnel and context-aware e-health service *mp* through the recommendation of *c*. The trust model obtains d$t(p, c)$ from the recommendation of the other patients *c*. When computing indirect trust, it considers the similarity of the patient's assessments like how [17] compute trust.

This model uses similarity when computing indirect trust. Let $\boldsymbol{it_n}$ $(\boldsymbol{p}, \boldsymbol{m}) \in$ [18] denote the indirect trust of node *p* and *mp*, which is computed by the recommended trust of set *c* to node *m* and a combination of the similarity between c and p. That is

Indirect Trust

$$(\boldsymbol{p}, \boldsymbol{m}) = \left\{ \sqrt{\frac{\Sigma_{c\varepsilon setc}\, \boldsymbol{sim}_n(\boldsymbol{p}, \boldsymbol{c}).\boldsymbol{dt}_n(\boldsymbol{c}, \boldsymbol{m})}{|\boldsymbol{setc}|}} \right., \boldsymbol{if}\ |\boldsymbol{setc}| > \mathbf{0} \tag{4}$$

4.1.2 Blockchain and Smart Contracts

After trust computation, if the medical personnel are trusted, they are allowed access to the patient's context data. To ensure increased use of the platform the use of blockchain is implemented to ensure the records are immutable and all changes done are tracked. A public blockchain is used because the proposed solution is open to the public and a private blockchain limits access to predefined users. This model borrows from existing systems like bitcoin, where a transaction is the money exchanged between parties in the blockchain network. Once a patient's medical data is recorded, it cannot be altered due to the immutable nature of the blockchain. The changes are appended as new blocks creating a chain; each block contains the previous blocks hash, its own hash, data and a timestamp as illustrated in Fig. 2.

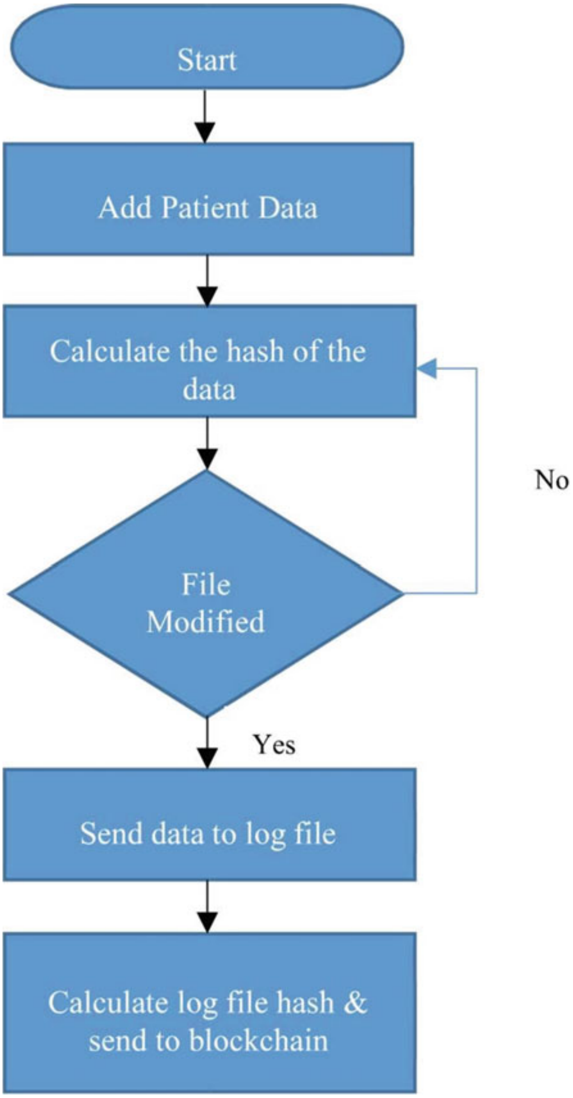

Start
Add Patient Data
Calculate the hash of the data
File Modified
No
Yes
Send data to log file
Calculate log file hash & send to blockchain

Fig. 2 Patient data hashing flow chart diagram

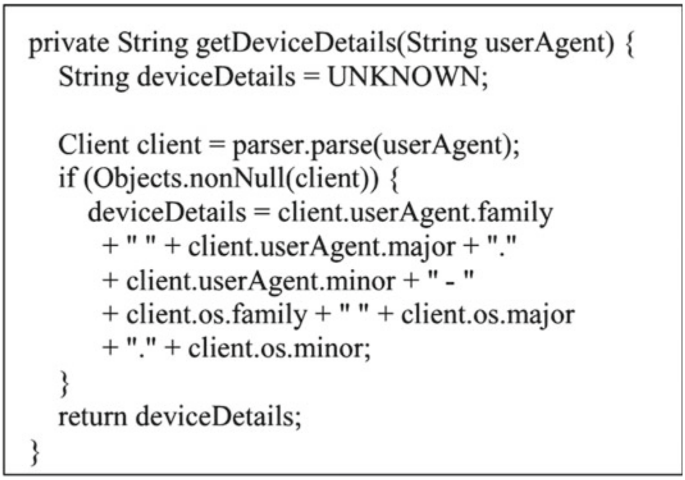
5 Experimental System Implementation

A Java-based application has been implemented and it enables the doctors to monitor the patient's health status and access their medical data. When a patient checks into the hospital, the medical personnel is granted access to their records based on his/her access rights and current location. Upon logging in verification is made regarding the device that the patient used to login including their location. To get the medical personnel's location the proposed model obtains the originating IP Address by using:

- X-Forwarded-For—the de facto standard header for identifying the originating IP address of a client connecting to a web server through an HTTP proxy or load balancer.

Once the IP address is retrieved, it is converted to a real-world location through geolocation. The User-Agent which carries the information about the device used to login to the application is used. The information it carries is used to identify the application type, software vendor and the operating system of the device used. The proposed model uses Code 1 to extract the device details.

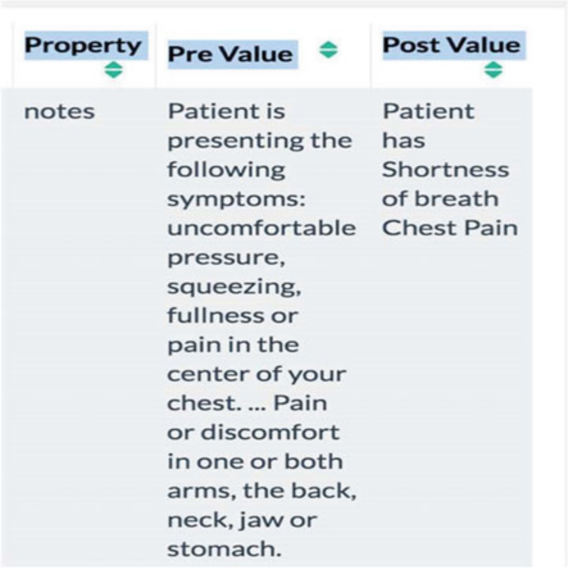
In this model, communication between the medical personnel and the patients is what is considered transactions, and this is borrowed from Ouaddah et al. [18] Access Control Block Chain Model. The transaction is composed of the patient's vital signals, their medical data and context information. An example of a transaction is as follows: Jane a patient at XYZ hospital has an underlying heart condition, her doctor John advises her to use the Context Aware e-health service provided by the hospital to monitor her health status. Two weeks later, her doctor and the hospital receive an alert that she has had a cardiac arrest. The hospital dispatches an ambulance to her location to assist her. The paramedics use the GPS tracker to get to her current location. On receipt of the alert, the doctor/hospital grants the paramedic team dispatched to the patient's location access rights to the patient's data. Once they access the data, a new record will not be appended on the blockchain, as the record was not modified. Once the patient is presented to the hospital and the doctor diagnoses her, any changes done on her medical records will be appended on the blockchain as illustrated in Fig. 4.



```
private String getDeviceDetails(String userAgent) {
   String deviceDetails = UNKNOWN;

   Client client = parser.parse(userAgent);
   if (Objects.nonNull(client)) {
      deviceDetails = client.userAgent.family
        + " " + client.userAgent.major + "."
        + client.userAgent.minor + " - "
        + client.os.family + " " + client.os.major
        + "." + client.os.minor;
   }
   return deviceDetails;
}
```

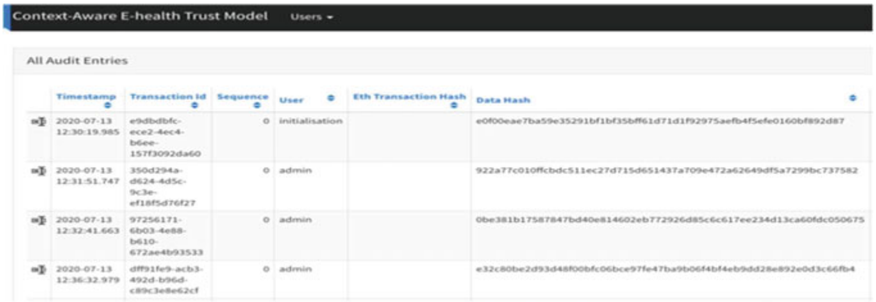
Code 1 Extract user IP



Property	Pre Value	Post Value
notes	Patient is presenting the following symptoms: uncomfortable pressure, squeezing, fullness or pain in the center of your chest. ... Pain or discomfort in one or both arms, the back, neck, jaw or stomach.	Patient has Shortness of breath Chest Pain

Fig. 4 Modified patient record

This model uses a digital file ledger that tracks all activities done on the patient's electronic records. As much as this method does not prevent modification, there is guarantee of the integrity of the data, as it will detect any changes perfumed on the records. A smart contact is used to track all the changes done on the data as illustrated on Figs. 4 and 5.



All Audit Entries

	Timestamp	Transaction Id	Sequence	User	Eth Transaction Hash	Data Hash
	2020-07-13 12:30:19.985	e9dbdbfc-ece2-4ec4-b6ee-157f3092da60	0	initialisation		e0f00eae7ba59e35291bf1bf35bff61d71d1f92975aefb4f5efe0160bf892d87
	2020-07-13 12:31:51.747	350d294a-d624-4d5c-9c3e-ef18f5d76f27	0	admin		922a77c010ffcbdc511ec27d715d651437a709e472a62649df5a7299bc737582
	2020-07-13 12:32:41.663	97256171-6b03-4e88-b610-672ae4b93533	0	admin		0be381b17587847bd40e814602eb772926d85c6c617ee234d13ca60fdc050675
	2020-07-13 12:36:32.979	dff91fe9-acb3-492d-b96d-c89c3e8e62cf	0	admin		e32c80be2d93d48f00bfc06bce97fe47ba9b06f4bf4eb9dd28e892e0d3c66fb4

Fig. 5 Patient data hashing

6 Conclusion and Future Works

In this research, the proposed Trust Model has been developed for context-aware e-health services that facilitates maintaining data privacy and promoting trust for e-health services. The model has adopted ideas from previous trust models like REK Trust Evaluation Model, Trust-Based Context-Aware Recommender Systems and Trust modelling computation. To achieve context, the proposed model used the who and where, whereby the who is the medical personnel trying to access the patient's data, and the where is their location. The model adopted the concept of direct and indirect trust to compute the trustworthiness of the medical personnel by computing the medical personnel's trust degree. Here, direct trust is achieved from the patients experience with the medical personnel while indirect trust is derived from recommendations of other patients. The patients rate the medical personnel on a scale of one to five, with five being the highest level of trust for the medical personnel. To ensure that the model is trusted, it stores the patients' records on a blockchain solution. The users in this case the patients and medical personnel register by providing their personal data. The patient's data from context aware applications is uploaded automatically and a random ounce used to encrypt the files, and a secret is stores in blockchain. The patient provides access to the medical personnel who is then able to see and fetch the patient's address and data. The major contribution of the proposed Trust Model in this research is the incorporation of the use of blockchain within the model to ensure that the data integrity is secure and maintained.

References

1. World Health Organization (2007) Delivering quality health services: a global imperative for universal health coverage
2. Bhatti A, Masud M (2014) Context aware intelligent wallet for healthcare
3. Dey AK (2001) Understanding and using context. Pers Ubiquitous Comput 5(1):4–7
4. Byan HE, Cheverst K (2004) Utilizing context history to provide dynamic adaptations. Appl Artifi Intell 18(6):533–548
5. Seppälä A (2014) Context-aware and trust-based personal wellness information framework for pervasive health
6. Oliveira M, Hairon C, Andrade O, Moura R, Gensel J, Fernandes S, Claude Sicotte, J-L D (2015) A context-aware framework for healthcare governance decision-making systems: a model based on the Brazilian digital TV
7. Gambetta D (1988) Trust: making and breaking cooperative relations. Blackwell, pp 213–237
8. Ruotsalainen PS, Blobel BG, Seppälä AV, Sorvari HO, Nykänen PA (2012) A conceptual framework and principles for trusted pervasive health
9. Miluzzo E, Lane ND, Fodor K, Peterson R, Lu H, Musolesi M, Eisenman SB, Zheng X, Campbell AT (2008) Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In: Proceedings of the 6th ACM conference on embedded network sensor systems (SenSys'08), ACM, New York, pp 337–350
10. Alves P, Ferreira PJSG (2014) Distributed context-aware systems
11. Truong NB, Lee H, Askwith B, Lee GM (2017) Toward a trust evaluation mechanism in the social internet of things

12. Otebolaku A, Lee GM (2018) A framework for exploiting internet of things for context-aware trust-based personalized services. Mob Inf Syst, 1–24. https://doi.org/10.1155/2018/6138418
13. Razavi N, Rahmani AM, Mohsenzadeh M (2009) A context-based trust management model for pervasive computing systems. Int J Comput Sci Inf Secur 6(1)
14. M'Hamed A, Zerkouk M, Husseini AE, Messabih B, Hassan BE (2013) Towards a context-aware modeling of trust and access control based on the patient behavior and capabilities. IN: ICOST 2013, June 2013, vol 7910. Singapore, pp 69–76. https://doi.org/10.1007/978-3-642-39470-69. <hal-00840462>
15. Almenárez-Mendoza F, Marín-López, A, Campo C, García C (2004) PTM: a pervasive trust management model for dynamic open environments. In: 1st workshop on pervasive security, privacy and trust in conjunction with mobiquitous
16. Yang S, Tan Z,Wang X, Wang X (2016) A novel iterative and dynamic trust computing model for large scaled P2P networks
17. Armstrong S (2016) Move over Bitcoin, the blockchain is only just getting started. Wired. Archived from the original on 8 Nov 2016. Retrieved 9 Nov 2016
18. Ouaddah A, Elkalam AA, Ouahman AA (2017) Fair access: a new blockchain-based access control framework for the internet of things