Intelligent Security Schema for SMS Spam Message Based on Machine Learning Algorithms

https://doi.org/10.3991/ijim.v15i16.24197

Ali Alshahrani(✉)
Arab Open University, Kingdom Saudi Arabia
a.shahrani@arabou.edu.sa

Abstract—SMS spam messages represent one of the most serious threats to current traditional networks. These messages have been particularly prevalent overseas and are harmful to various types of devices. The current filtering scheme employed in conventional systems is unable to expose a large number of messages. To resolve this issue, a new intelligent security system is proposed to reduce the number of spam messages. It can detect novel spam messages that have a direct and negative impact on networks. The proposed system is heavily based on machine learning to explore various types of messages. The primary achievement of our study is the increase in the accuracy ratio as well as the reduction in the number of false alarms. According to the experimental results, it is clear that our system can realize outstanding results, detecting a massive number of massages.

Keywords—security, protection, internet, SMS spam, intrusion detection, attacks

1 Introduction

Security systems are considered one of the most important issues in the scientific research area [1]. A lot of modern applications have been suffered from luck protection techniques that expose various attacks. Short Message Service (SMS) or mobile text messages represent a communication service component of phone, web, or mobile communication systems that use standardized communication protocols for the exchange of short text messages between fixed-line or mobile phone devices [2][3]. Mobile text messages are used for communication between cell phone users when voice communication is undesirable or impossible. However, some of the text messages that are forwarded to the user's device are bothersome and unwanted, and these are called SMS spam.

The user stores personal and confidential information on their smartphone, such as contact lists, numbers, passwords, and credit card information. Thus, using SMS spam, hackers can attack users' devices and exploit this information. Privacy invasion and access to sensitive or unauthorized information are the main problems arising from spam messages. The privacy of the user is violated by individuals commonly known as spammers, who use various unethical activities to access user data stored on smartphones without the knowledge of the user [4].

Spam messages are unwanted but are often unavoidable. SMS spam can be undesired emails delivered as text messages across mobile devices [5]. These messages are utilized by some businesses to promote and advertise their materials in order to increase their audience. Besides promoting services or products, SMS spam can threaten users' privacy and can lead to identity theft and fraud through the use of attacks via spam text messages [6]. Spam messages originate from all regions worldwide; however, China represents a major source of these messages over other countries [7].

Recently, the popularity of SMS has increased due to the development of different communities of mobile users, which present various techniques and tools to spam mobile phones in order to maximize the desired result. Problems associated with SMS spam have inspired researchers to present different techniques for the effective detection and prevention of spam SMS. The general phases of SMS spam filtering are outlined in the following figure. Chandrasena. Premawardhena, N. (2012). Introducing Computer-Aided Language Learning to Sri Lankan Schools: Challenges and Perspectives. 15th International Conference on Interactive Collaborative Learning and 41st International Conference on Engineering Pedagogy (ICL & IGIP), Villach, Austria.

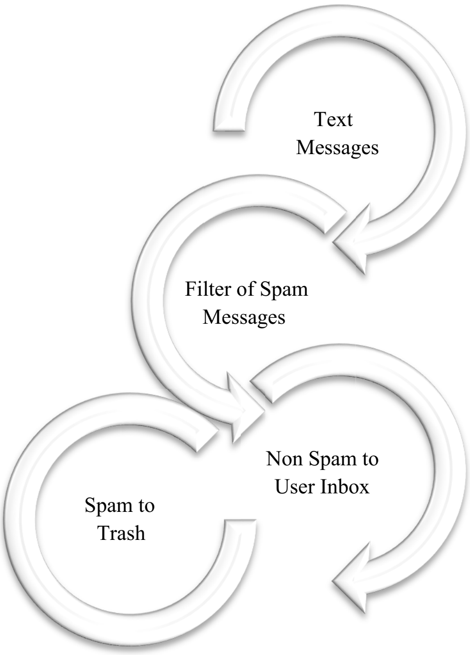

Fig. 1. SMS spam filtering

In the detection of spam messages, the availability of SMS datasets used in the training/testing techniques is still limited by the small size. Moreover, the number of features utilized for spam message detection in the text is low because of the short length of text messages. Different machine learning algorithms and techniques have been utilized to sort through spam messages and filter them. The objective of this paper is to present a system to resolve the problems associated with spam message detection. The proposed system is heavily based on the random forest algorithm and decision tree as classifiers of detection.

The rest of this paper is organized as follows: Section 2 presents related works. The methodology of the proposed approach is described in Section 3. The experimental results are given in Section 4. Finally, Section 5 presents conclusions and future directions.

2 Related works

In classifying/detecting SMS spam, many types of research have been presented and a variety of related issues have been discussed. This section presents and summarizes the earlier research related to this field.

In [4], the authors present a review to compare the performance of machine learning algorithms. The review proves that the use of support vector machine (SVM) and Naive Bayes leads to efficient performance. Another review for SMS spam classification/filtering is presented in [8]. This review focuses on expanding the number of features used for SMS classification and considers how the number of selected features affects the rate of accuracy. The authors aimed to contribute to determining spam's impact level or risk.

Machine learning has been widely utilized in the SMS spam classification field and many works have been presented. In [9], a survey is presented to prove the performance of SVM utilization to identify and filter spam messages. In [10], various methods are used to analyze SMS spam and a new pre-processing technique is utilized to obtain an actual dataset of SMS spam. Different algorithm techniques have used this dataset to develop a more suitable algorithm to achieve both recall and accuracy. The results prove that the Random Forest algorithm is capable of classifying a new dataset for ham and spam. To filter SMS spam, the Random Forest algorithm and Term Frequency–Inverse Document Frequency (TF–IDF) is used in [7]. The experimental results prove that the Random Forest algorithm achieves effective performance, with an accuracy of 97.50%.

In [11], different machine learning models are used, such as LightGBM, XGBoost, and Bernoulli Naive Bayes, to achieve greater speed and efficiency in the classification of SMS spam with low latency. The results prove that Bernoulli Naive Bayes, followed by LightGBM with the TF–IDF matrix, generated the highest accuracy of 96.5% in 0.157 seconds and 95.4% in 1.708 seconds, respectively. A Recurrent Neural Network with SVM is used in [12] to detect bot spam emails based on employing a spam dataset. The results prove that the presented solution can achieve better performance for the detection of spam emails with 98.7%. Machine learning techniques are used in [13] to develop a system for spam filtering and identification of legitimate emails. The results prove that the classification system presents 92% and 94% correct spam identification, with less identification of false positives, at 1.0%.

A framework for spam detection and risk estimation is presented in [14] based on data stream clustering and classification. The authors used Multinomial Naive Bayes and identified K-nearest neighbor algorithms for classification. In addition, the K-means algorithm is used in the clustering phase for SMS spam detection. For system evaluation, some metrics are used for performance assessment in classification/clustering methods. The WEKA text technique is used as a means of spam message classification/filtering in [15]. Different algorithms are used for SMS dataset classification and some metrics, such as accuracy, error rate, and time, are computed to select the optimal one.

In [16], analyzing of Bayesian filtering techniques is made to know to what extent these techniques are used for email spam blocking. The authors proposed two SMS spam test sets with some specific words and significant size based on using Machine Learning algorithms. The results prove that Bayesian filtering techniques can be efficiently transferred from email to SMS spam. Filtering of SMS spam and a review of modern researches in SMS spam filtering presented in [17]. This paper also studies data collection, analyses a large corpus of SMS spam, and provides results. In [18], the Naive Bayes algorithm was used to propose a spam classification model for mobile devices. This system aims to correctly filter incoming SMS received by users. Efficient and dependable results were obtained by the proposed model.

A new public, large, real, and non-encoded SMS spam collection dataset was used in [3] with a comprehensive analysis. The performance presented by several machine learning techniques is compared. A novel machine learning system for SMS spam messages detection is proposed in [19]. This paper utilized feature extraction and decision making as a method dependent on the proposed system. The results prove that high detection rates in terms of classification accuracy and F-measure can be achieved by the proposed system compared with other proposed researches.

Through analyzing and searching the earlier mentioned system, our paper is distinguished from others by presenting a detection system based on utilizing the Random Forest Algorithm and decision tree as a classifier to solve the problems associated with spam message detection.

3 SMS spam security

Many threats attack SMS security involve message disclosure, Man-in-the-middle attack, SMS viruses, and SMS spamming. In SMS spamming, SMS used as a valid marketing channel, and many people had annoyed while receiving SMS spam. It easy for virtually everyone to send out mass SMS messages because of the availability of bulk SMS broadcasting [20]. To solve this problem spam SMS detection is dependent.

The most prevalent malware used email to spread and logins by passwords to a system to steal confidential data. The Top 10 malicious programs spread by email presented as follow [21]:

- Trojan-Spy.HTML.Fraud.gen
- Email worms.Win23.Bagle.gt
- Email worms. Win23. Mydoom.m
- Email worms. Win23. Mydoom.I
- Trojan-Banker.HTML.Agent.p

- Trojan-Spy. Win23.Zbot.Ibda
- Worms. Win23. Mabezat.b
- Trojan-PSW.Win32.Tepfer.hjva
- Email worms. Win23. NetSky.q
- Trojan.Win32.Bublik.aknd

One type of threat that could lead to a hazardous situation and exploit vulnerabilities is spam when the probability of potential risk to happen is high [22]. For SMS spam management, three phases which are spam detection, classification, and severity determination level are dependent [23]. Risk management processes are necessary for SMS spam detection which are: identification of risk, assessing risk, responding to risk, and risk monitoring. To manage spam, there are three main processes which are: spam classification, spam clustering, and determination level of spam's severity [24].

4 Methodology

In this paper, the main objective of the proposed work is to classify SMS spam messages either as normal or ham spam. The proposed system starts from the collection process of a dataset that generated from real-world to classification decision whereas normal or abnormal behaviors. The proposed work includes the process presented in Figure 2 below.

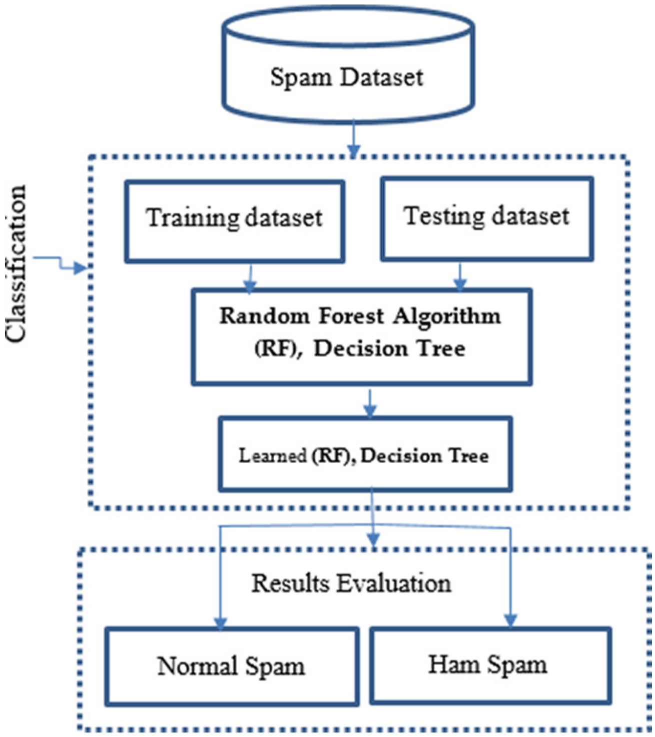

Fig. 2. Processes of SMS spam classification

In this paper, two tools of machine learning are utilized to identify normal from ham spam emails. However, these techniques are play important role in providing a secure environment among users that participate on various networks.

4.1 Dataset Source

The SMS spam dataset used in this research is obtained from the Kaggle, a machine learning repository [15]. It contains two labels, v1 and v2, and 5572 instances. The (v2) label represents the input messages, which are either ham spam or normal spam. The predicted label (v) has two classes, which are 0 = ham spam and 1 = normal spam. In the data, 4900 are ham spam instances and 672 are normal spam instances. The dataset is presented in Table 1 [24].

Table 1. SMS spam dataset description

Instance number	Input message (v2)	Predicted label (v1)
1	Cine there got a more wat…	Ham
2	Go until Jurong point, crazy… Available only in bugis *n* great world la e buffet…	Ham
3	Free entry in 2 a wkly comp to win FA cup final tkts 21st May 2005. Text FA to 87121 to receive entry question (std txt rate) T&C's apply 08452810075over18's.	Spam
. . .	. . Rofl. It is true to its name	. . .
5572		Ham

Table 1 explains the main contents of emails that were sent/received between users.

4.2 SMS message spam classification

Random forest algorithm (RF) and decision tree are machine learning algorithms used for a large number of datasets with various feature types, such as numerical, binary, and categorical [25]. In this system, RF and decision trees are utilized to efficiently classify normal and ham SMS spam. The system combines various sets of decision trees to eliminate the overfitting difficulties in the training phase. In the RF algorithm, each tree operates with randomly chosen attributes and it has the capability of providing prediction results that differ from others [26]. As a result, different levels of performance can be achieved by each tree, and the total average of their performance is generated and calculated. The processes of RF are presented in Figure 3.

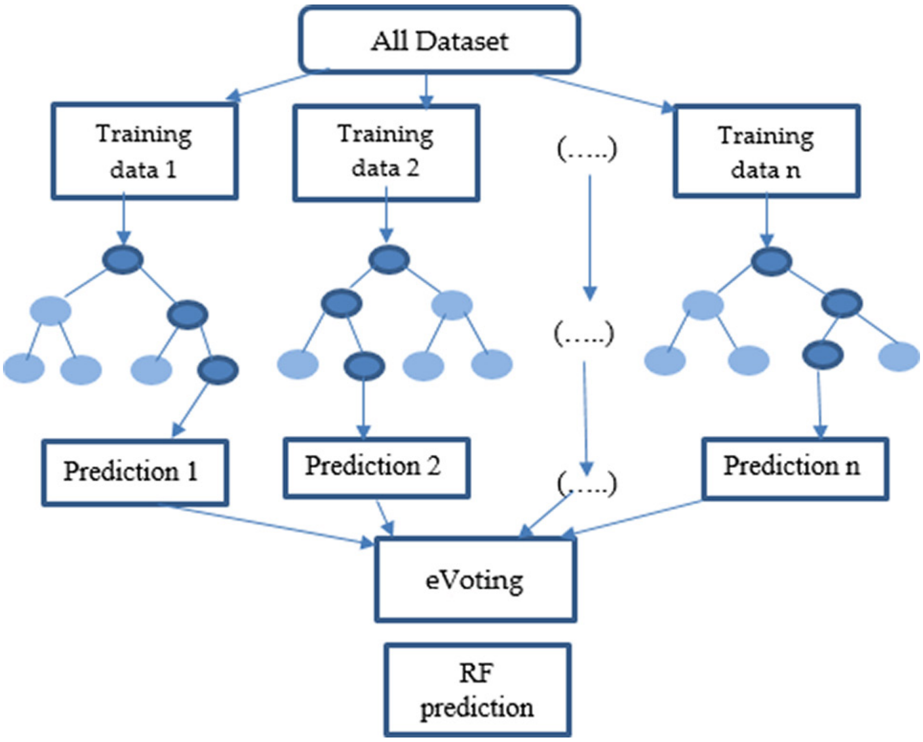

Fig. 3. The Processes of Random Forest (RF) Algorithm

4.3 Performance measurement

In this paper, some metrics are utilized for the measurement of the efficiency of the proposed system. The metrics, which are accuracy, confusing matrix, and recall, can be calculated as follows [27]:

$$Accuracy = \frac{\text{Number of correctly classified patterns}}{\text{Total number of patterns}} \tag{1}$$

To measure and evaluate the proposed system's performance, a confusing matrix is calculated, which involves four categories: true positive (TP), false positive (FP), true negative (TN) and false negative (FN). The calculation can be made by the following Equations [27]:

$$\mathrm{TP} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{2}$$

$$TN = \frac{\mathrm{TP}}{\mathrm{TN} + \mathrm{FP}} \tag{3}$$

$$FN = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}} \tag{4}$$

$$FP = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{5}$$

- True positive (TP) is the number of correctly predicted normal spam messages.
- False-positive (FP) is the number of wrongly predicted normal spam messages.
- True negative (TN) is the number of correctly predicted ham spam messages.
- False-negative (FN) is the number of wrongly predicted ham spam messages.

The recall is the last metric used in this paper, which was calculated as follows:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

All these metrics are employed in this paper to test/evaluate the performance of the proposed security system.

5 Experimental results and discussion

In order to compare the performance of the algorithms used in this experiment, this paper provides performance evaluation measures such as accuracy, precision-recall, f1 score support, and time. Table 2 shows the performance evaluation of the random forest algorithm.

Table 2. Random forest algorithm performance

Random Forest Classifier	
Accuracy	98.6%
Precision recall	86%
f1 score support	94%
Time	51s

Meanwhile, Table 3 shows the efficiency of the security system that is based on the decision tree.

Table 3. Decision tree classifier performance

Decision Tree Performance	
Accuracy	94.8%
Precision recall	79%
f1 score support	83%
Time	42s

As presented in Tables 2 and 3, the random forest and decision tree machine learning algorithms achieved the highest accuracy in the classification of SMS spam. Moreover, we achieved a 98.2% accuracy rate with the random forest machine-learning algorithm. Table 3 shows a comparison of our proposed model with an earlier model using the decision tree and random forest machine learning algorithms.

Table 4. Algorithm comparison

Reference	Algorithm	Accuracy
[6]	Decision Tree	96.57%
[6]	Random Forest	97.50%
Proposed approach	Decision Tree	94.8%
Proposed approach	Random Forest	98.6%

According to Table 4, we can see that our proposal that is based on RF is more accurate than others.

6 Conclusion and future directions

Nowadays, SMS spam detection is a major challenge due to the increase in the use of text messaging. In this paper, a technique for SMS spam detection is proposed based on utilizing the random Forest and decision tree algorithms. The dataset used in this work consists of 4900 ham spam instances and 672 normal spam instances. The experimental results show that the classification system using the random forest algorithm presents the best results, with a 98.2% accuracy rate. In future work, other machine learning methods will be employed to achieve an improved accuracy rate in the SMS spam classification field.

7 Acknowledgments

The author would like to thank Arab Open University, Saudi Arabia for supporting this study.

8 References

[1] H. Naman, N. Hussien, M. Al-dabag, and H. Alrikabi, "Encryption System for Hiding Information Based on Internet of Things," International Journal of Interactive Mobile Technologies, vol. 15, no. 2, 2021. https://doi.org/10.3991/ijim.v15i02.19869

[2] José. M. G. Hidalgo, T. A. Almeida, and A. Yamakami, "On the Validity of a New SMS Spam Collection," 2012 11th International Conference on Machine Learning and Applications, 2012, pp. 240–245. https://doi.org/10.1109/ICMLA.2012.211

[3] T. Almeida, J. M. G. Hidalgo, and T. P. Silva, "Towards sms spam filtering: Results under a new dataset," International Journal of Information Security Science, vol. 2, no. 1, pp. 1–18, 2013.

[4] S. Alqahtani and D. Alghazzawi, "A survey of Emerging Techniques in Detecting SMS Spam," Trans. Mach. Learn. Artif. Intell., vol. 7, no. 5, pp. 23–35, 2019. https://doi.org/10.14738/tmlai.75.7116

[5] Tiago A. Almeida and A. Y. Almeida, J. M. Gómez, "Contributions to the Study of SMS Spam Filtering: New Collection and Results," p. 7, 2011. https://doi.org/10.1145/2034691.2034742

[6] M. Ghulam and M. Y. Mujtaba, "SMS Spam Detection Using Simple Message Content Features," J. Basic Appl. Sci. Res, vol. 4, 2014.

[7] Amir N. N. Sjarif, N. F. Mohd Azmi, S. Chuprat, H. M. Sarkan, Y. Yahya, and S. M. Sam, "SMS spam message detection using term frequency-inverse document frequency and random forest algorithm," Procedia Comput. Sci., vol. 161, pp. 509–515, 2019. https://doi.org/10.1016/j.procs.2019.11.150

[8] K. Zainal and M. Z. Jali, "A review of feature extraction optimization in SMS spam messages classification," Commun. Comput. Inf. Sci., vol. 652, pp. 158–170, 2016. https://doi.org/10.1007/978-981-10-2777-2_14

[9] Z. S. Torabi, "Efficient Support Vector Machines for Spam Detection: A Survey," vol. 13, no. 1, pp. 11–28, 2015.

[10] S. S. Ali and J. Maqsood, "Net library for SMS spam detection using machine learning: A cross platform solution," Proc. 2018 15th Int. Bhurban Conf. Appl. Sci. Technol. IBCAST, vol. 2018-January, pp. 470–476, 2018. https://doi.org/10.1109/IBCAST.2018.8312266

[11] A. Ora, "Spam Detection in Short Message Service Using Natural Language Processing and Machine Learning Techniques," 2020.

[12] M. Alauthman, "Botnet spam e-mail detection using deep recurrent neural network," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 5, pp. 1979–1986, 2020. https://doi.org/10.30534/ijeter/2020/83852020

[13] S. J. Delany, "ECUE: A Spam Filter that Uses Machine Learning to Track Concept Drift," pp. 1–5, 1826.

[14] K. S. Adewole, N. B. Anuar, and A. Kamsin, "Ensemble based streaming framework for spam detection and risk assessment in microblogging social networks," 2016.

[15] D. R. Kawade, "SMS Spam Classification using WEKA," vol. 5, no. ICICC, pp. 43–47, 2015.

[16] G. Sethi and V. Bhootna, "SMS Spam Filtering Application Using Android," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 3, pp. 4624–4626, 2014.

[17] S. J. Delany, M. Buckley, and D. Greene, "SMS spam filtering: Methods and data," Expert Syst. Appl., vol. 39, no. 10, pp. 9899–9908, 2012. https://doi.org/10.1016/j.eswa.2012.02.053

[18] S. M. S. Spam, F. For, and M. Mobile, "Sms Spam Filtering for Modern Mobile Devices," vol. 13, no. 1, pp. 177–185, 2017.

[19] A. B. Saeid, M. T. Kheirabadi, "An Effective Model for SMS Spam Detection Using Content-based Features and Averaged Neural Network," Int. J. Eng. Trans. B Appl., vol. 33, no. 2, pp. 221–228, 2020. https://doi.org/10.5829/ije.2020.33.02b.06

[20] N. Zalpuri and M. Arora, "An Efficient Model for S.M.S Security and SPAM Detection: A Review," Int. J. Comput. Sci. Eng., vol. 3, no. 12, pp. 1–6, 2015.

[21] D. Gudkova, "Kaspersky Security Bulletin-Spam Evolution 2013," pp. 1–22, 2013.

[22] K. Zainal and M. Z. Jali, "A Perception Model of Spam Risk Assessment Inspired by Danger Theory of Artificial Immune Systems," Procedia Comput. Sci., vol. 59, no. Iccsci, pp. 152–161, 2015. https://doi.org/10.1016/j.procs.2015.07.530

[23] M. Z. Sulaiman and Jali, "Integrated Mobile Spam Model Using Artificial Immune System Algorithms," Knowl. Manag. Int. Conf., pp. 405–409, 2014.

[24] "Datasets." [Online]. Available: https://www.kaggle.com/datasets. [Accessed: 15-Mar-2021].

[25] M. Al-dabag, H. S. ALRikabi, and R. Al-Nima, "Anticipating Atrial Fibrillation Signal Using Efficient Algorithm," International Journal of Online and Biomedical Engineering (iJOE), vol. 17, no. 2, 2021. https://doi.org/10.3991/ijoe.v17i02.19183

[26] N. Choudhary and A. K. Jain, "Towards filtering of SMS spam messages using machine learning based technique," Commun. Comput. Inf. Sci., vol. 712, no. July, pp. 18–30, 2017. https://doi.org/10.1007/978-981-10-5780-9_2

[27] A. S. Hussein, R. S. Khairy, S. M. M. Najeeb, and H. T. ALRikabi, "Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression," International Journal of Interactive Mobile Technologies, vol. 15, no. 5, 2021. https://doi.org/10.3991/ijim.v15i05.17173

9 Author

Alshahrani is an associate professor in computer science studies faculty, Arab Open University - Saudi Arabia. He received his B.Sc. degree in Information Technology and Computing, in 2008. His M.Sc. and Ph.D. from University of Essex, UK in Computer science, in 2015. His research interests include network security, image processing, e-learning and mobile systems.

Article submitted 2021-05-15. Resubmitted 2021-06-18. Final acceptance 2021-06-20. Final version published as submitted by the authors.