A new proactive feature selection model based on the enhanced optimization algorithms to detect DRDoS attacks

Riyadh Rahef Nuiaa[1,2], Selvakumar Manickam[2], Ali Hakem Alsaeedi[3], Esraa Saleh Alomari[1]
[1]Department of Computer, College of Education for Pure Sciences, Wasit University, Al Kut, Iraq
[2]National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia, George Town, Malaysia
[3]College of Computer Science and Information Technology, Universitas of Al-Qadisiyah, Al Diwaniyah, Iraq

Article Info

Article history:

Received May 7, 2021
Revised Sep 17, 2021
Accepted Oct 11, 2021

Keywords:

Adaptive threshold
Cybersecurity
DRDoS attacks
DRDoS detection mechanism
Feature selection

ABSTRACT

Cyberattacks have grown steadily over the last few years. The distributed reflection denial of service (DRDoS) attack has been rising, a new variant of distributed denial of service (DDoS) attack. DRDoS attacks are more difficult to mitigate due to the dynamics and the attack strategy of this type of attack. The number of features influences the performance of the intrusion detection system by investigating the behavior of traffic. Therefore, the feature selection model improves the accuracy of the detection mechanism also reduces the time of detection by reducing the number of features. The proposed model aims to detect DRDoS attacks based on the feature selection model, and this model is called a proactive feature selection model proactive feature selection (PFS). This model uses a nature-inspired optimization algorithm for the feature subset selection. Three machine learning algorithms, i.e., k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM), were evaluated as the potential classifier for evaluating the selected features. We have used the CICDDoS2019 dataset for evaluation purposes. The performance of each classifier is compared to previous models. The results indicate that the suggested model works better than the current approaches providing a higher detection rate (DR), a low false-positive rate (FPR), and increased accuracy detection (DA). The PFS model shows better accuracy to detect DRDoS attacks with 89.59%.

Corresponding Author:

Selvakumar Manickam
National Advanced IPv6 Centre, Universiti Sains Malaysia
11800 USM, Penang, Malaysia
Email: selva@usm.my

1. INTRODUCTION

Users are exploring smartphone and portable technologies in today's computing age to access banking, network, online shopping, retail, gaming, and media content resources using web apps over the internet [1]. Using online apps to navigate resources to execute such functions also expanded the number of users. Cybersecurity attackers develop their methodology to bring down the network or prevent legitimate users from using the resources or the victim network's services [2]. Therefore [3] will lead to the loss of business and finance. The growth line curve for cyber threatens calls for concern and thinking to find successful solutions to reduce the risks involved in these threats and reduce the economic impacts. Distributed denial of service (DDoS) is a considerable scale cyber-attack when hackers launch their attack by utilizing more than one attack point to produce a massive volume of malicious traffic from several sources. The hacker uses different instruments or programs to create a huge flood of malicious traffic with one or more attack vectors [4]. Denial of service (DoS) and distributed DoS attacks are primarily used by hackers to disable or degrade service performance. The existence of more than one attack vector from several sources produces a challenge to security administrators in the intrusion detection mechanism [5]. These attacks can occur at the network, transport, and application level of the open systems interconnection (OSI) model. DDoS attacks continued to be the dominant challenge seen by the overwhelming majority of service providers, according to the Arbor Networks Inc. (2019) report, and the biggest attack in 2018 was 1.7 Tbps [6]. These DDoS attacks are further abused utilizing reflection and exploitation behavior at the application level employing transport level protocols. Those assaults are indicated as distributed reflection and exploitation-based DoS attacks (distributed reflection denial of service (DRDoS) and declarative dispersion-oriented software (DEDoS) [7]. DRDoS attacks are performed at the application level utilizing transmission control protocol (TCP), user datagram protocol (UDP), or a mixture of both. An attacker sends falsified demands to multiple servers with the prey spoofed source address in DRDoS attacks. In reply, replies will be sent to the prey by the servers. And these responses are frequently (many times) far greater than the requests [8]. Furthermore, DRDoS attacks appear to be more dedicated and complex with greater diversity, resulting in the need for a fast, intelligent and powerful cyber-attack identification system for security control for the frequently vulnerable network [9].

Securing data is very important, especially with the rapid increase of the users and devices connected to the Internet. All this led to the rise in the amount of data. In recent years, new types of cyberattacks have emerged called DRDoS attack. Fast development in those attacks and their methodology variety led researchers and cyber security companies to focus on detecting those attacks. Many studies were done to find a solution to address this issue. The classical detection methods that use static threshold are not suitable with the high dimension of data; therefore, we propose a new detection method based on an adaptive threshold for designing a proactive feature selection model.

This paper presents a new model called a proactive feature selection (PFS) model to detect multi classes of DRDoS attacks then classify them. The PFS model is based on swarm optimization and evolutionary algorithms (SWEVO), machine learning (ML), and the fitness function is the adaptive threshold. The primary function of the PFS model is to reduce the number of features. Therefore the adaptive threshold (fitness function) will be updated every search in the population to eliminate irrelevant or redundant features.

This paper aims to reveal the measures used to address issues related to the dataset and actions taken to enhance the detection of DRDoS attacks. In this study, the significant contributions are summarized as follows: i) the proposed feature selection algorithm focused on the metaheuristic optimization algorithm and adaptive threshold to optimize the detection mechanism performance by reducing the number of features; ii) the new PFS model intends to detect the DRDoS attacks and achieve this usefulness by diagnosing vulnerabilities in the intrusion detection system exploiting those attacks. Therefore improving detection accuracy. The results have proved that the PFS can detect several types of those attacks with high accuracy; iii) testing of the PFS model has been done, and then the results are comparing with three famous metaheuristic optimization algorithm (particle swarm optimization (PSO), bat algorithm (BA), and differential evolution (DE)). The section of results and discussion show the comparison tables.

The CICDDoS2019 dataset was used to test the current method's performance, reliability, and validity in detecting the DRDoS attack. The results indicate that the PFS model attains a high degree of accuracy of 89.59% detection rate in detecting several kinds of DRDoS attacks on protocols; and a drop in the false alarm rate as follows: i) evaluation strategy: The CICDDoS2019 dataset includes a multi-class of DRDoS attacks. Essential evaluation metrics include accuracy, precision, recall, F1-score, confusion-matrix, and the number of features. The PFS model has been used to improve and enhance the accuracy metrics and reduce the number of features compared to the original dataset's number of features. The evaluation indicated that the PFS model achieves a high true-positive rate and a low false-negative rate. Moreover, the proposed model's accuracy metrics are investigated and compared with other techniques' accuracy metrics, which is found to be significantly higher than other models and ii) paper organization: the remainder of this paper is organized as follows: in section 2, the authors review the related works. Section 3 describes the swarm optimization algorithms and evolutionary algorithms used to develop our model also the machine learning algorithms used as classifiers. Section 4 describes the proposed model for feature selection. In section 5, the authors describe the experiment finally, our conclusion in section 6.

2. RELATED WORK

Earlier research used the SWEVO to improve detection and reduce false positives. Therefore, the number of features used plays a critical role in the quality of DRDoS detection. Sharafaldin *et al.* [7] suggested a new model that can detect several types of DDoS attacks. Moreover, it can detect various kinds of DRDoS attacks. The proposed model was designed using four types of machine learning algorithms which are ID3, random forest (RF), naive Bayes, and multinomial logistic regression. The model is tested on the CICDDoS2019 dataset that contains 88 features. Table 1 shown the accuracy metrics of the model and its effectiveness against DRDoS attacks. The main limitation of this work is only using the classifier algorithm with the same number of features, therefore often contributes to the classifier's poor detection and high misclassification rates.

Sharafaldin [10] suggested a feature selection model enhance the intrusion detection system (IDS) by using four SWEVO algorithms: PSO, grey wolf optimizer (GWO), firefly optimization (FFA), and genetic algorithm (GA) name. The features extracted from the suggested model are assessed based on the support vector machine (SVM) and J48 ML classifiers and the UNSW-NB15 dataset. The model contains 13 rules, and the two essential rules are R12 and R13, which are shown high accuracy and reduce the number of features. These lead to the enhancement of the IDS performance. The main limitation of this work is that the dataset user does not contain the essential attack types like DDoS attacks. Therefore, the IDS may be inefficient with modern cyberattacks.

Vijayanand and Devaraj [11] proposed an approach based on the modified whale optimization algorithm. The improved approach's performance was evaluated using SVM, and two standard datasets are intrusion detection evaluation dataset (CICIDS2017) and Australian defense force academy Linux dataset (ADFA-LD). The selected features were the basis to identify kinds of intrusion. The informative features were select to help increase the accuracy of the IDS dependent on the SVM. By choosing the informative features with the enhanced whale optimization algorithm, the efficiency of the IDS was improved. The identification ratio for attacks was better than that of the regular whale optimization algorithm (WOA).

Ghasemi *et al.* [12] proposed a new hybrid model that builds on GA and four classification algorithms. This model is called kernel extreme learning machine (KELM) for feature selection in IDS. By using network security layer-knowledge discovery in database (NSL-KDD) standard datasets that is an enhanced version of the KDD CUP 99 dataset, the performance of the KELM model was evaluated. Through the implementation of the KELM, a new dataset is produced called GA-dataset. The KELM enhanced by GA on the GAdataset achieved high accuracy and low false alarm.

Sarvari *et al.* [13] suggested a new feature selection approach called mutation cuckoo fuzzy (MCF) to select the optimal features. For the purposed of classification, multiverse optimizer-artificial neural network (MVO-ANN) is used. The suggested search algorithm utilizes a mutation to examine the search space more accurately. The validation of the performance and relevance MFC model for IDS problems uses the NSL-KDD standard datasets.

Patil and Kshirsagar [14] suggest a system based on feature selection to detect DDoS attacks. Information Gain has applied the process of feature selection with the Ranker algorithm. The proposed method uses RF, J48, and logistic model tree (LMT) classifiers to detect DDoS attacks. With the assistance of the CICIDS2017 dataset, the suggested system has been tested. The outcome of the experiments reveals that the J48 classifier has major features with an increased detection performance relative to Random Forest and LMT. The main limitations of previous research works are due to the use of static threshold with the high dimensional dataset, and some of the studies used only machine learning algorithms when proposed a new model for feature selection. We propose a new proactive feature selection model based on an adaptive threshold to enhance the detection accuracy rate to address these shortcomings.

3. SWARM OPTIMIZATION EVOLUTIONARY ALGORITHMS (SWEVO) AND MACHINE LEARNING ALGORITHMS MACHINE LEARNING (ML)

The proposed mechanism to detect the DRDoS attacks is based on two swarm optimization algorithms and one evolutionary algorithm besides three machine learning algorithms as classifiers. Machine learning-based IDSs can reach satisfactory detection levels, and machine learning models have sufficient generalizability to detect attack variants and novel threats. The promising research area in computer science, derived from SWEVO algorithms, is motivated by the natural evolution of biological organisms [15]. Many heuristic algorithms obtained from the natural behavior of biological or physical systems were suggested as robust methods for global optimization [16]. Cybersecurity challenges have been commonly applied to machine learning techniques. ML combines statistics and artificial intelligence with learning a data model [17], [18]. Cybersecurity ML techniques effectively suggest the correct decision for analysis and even automatically perform the appropriate response [19]. Thus, we can also differentiate between supervised, semi-supervised, and unsupervised approaches [20], [21].

3.1. Particle swarm optimization (PSO)

Kennedy and Eberhart presented the PSO in 1995. It gives a unique mechanism to imitate swarm behavior in flocking birds and fish schooling to direct the particles searching for optimal global solutions

[22]. PSO is a common swarm intelligence algorithm employed in the continuous search space to solve global optimization [23], [24].

3.2. A new metaheuristic bat-inspired algorithm (BA)

Xin-She Yang presented the BA in 2010. The bat algorithm relies on the bat behavior, which is based on echolocation. The bats employ this feature to determine prey's location and distinguish several bugs even in absolute darkness [25].

3.3. Differential evolution (DE)

A simple and efficient heuristic for global optimization over continuous spaces. The DE algorithm was presented by Storn and Price in 1997 and belonging to the family of evolutionary computation algorithms that apply biologically inspired one of the search algorithms. Moreover, this algorithm based on population also utilizes three operators utilized in the new heuristic algorithm: mutation, crossover, and selection. This method is robust, simple to use, and well-suited for parallel computing because fewer control variables are required [26], [27].

3.4. K-nearest neighbor (KNN)

The KNN is a lazy learning algorithm aimed at classifying a new object based on the current classes in which the previous training points are categorized. It categorizes the latest data points based on metrics of resemblance [24], [28].

3.5. Support vector machine (SVM)

The SVM technology offers the best approach for classifying clean and invasive data forms. High-class precision in detecting data intrusions is solved by SVM technologies [29]. Initially, SVM is an application of the systemic risk minimization (SRM) concept of Vapnik, which is considered to have a low generalization error or is not necessary to overfit the training data collection [30].

3.6. Random forest (RF)

A machine learning algorithm that incorporates two principles of decision tree and ensemble learning is a RF. RF achieves high accuracy in the detection and can accommodate outliers and data noise [31]. The RF Algorithm focuses on creating many decision trees, each of which acts as a classifier. The outcome of the final decision is decided by the balloting of all decision trees [32].

4. PROPOSED PROACTIVE FEATURES SELECTION MODEL (PFS)

Our previous work [33] provided a critical review paper of the designed and implemented mechanisms to detect DRDoS attacks. PFS is inspired by the principle of adaptive system parameters dynamically with changes in the algorithm's behavior. Dynamic systems often prefer static because they have significant performance, are flexible, and are suitable for solving many problems [x1]. The probability of finding the optimal solution and enhance stagnation on local optimum in a dynamic system is high [x2]. Metaheuristic has been recognized as fast, flexible, easy to implement, and successful in optimizing different fields [x3]-[x5]. The limitations of several metaheuristics models are often suffering from stagnation during the search process. Therefore, the applied dynamic system will reduce the stagnation of the local optimum and enhance the performance of systems that use metaheuristic algorithms. The wrapper model is a compelling feature selection method [x6]. It selects features based on trial and error, then updates corresponding features after each iteration.

Furthermore, the features are selected randomly. The wrapper models select features based on ranking given by metaheuristic techniques to these features. Generally, it selected only features corresponding to rank higher than the threshold is a fixed value often set by 0.5. The proposed system sets the threshold dynamically; it has been changed adaptively with search progress. Figure 1 shown the graphical abstract of the proposed model and Figure 2 illustrates the main steps of the proposed system.

4.1. Data preparation stage

Data preprocessing: the first stage is to convert raw data into an analysis-ready format by implementing preprocessing to the CICDDoS2019 datasets. There are several procedures for data preprocessing to be performed: i) import the dataset into python IDE; ii) search for incomplete data and outliers; iii) eliminate the data noise from the preprocessing; and iv) split the data used to build the model into training and testing collection.

Data normalization: the mechanism by which the data values of each function are converted or scaled into a proportional set. According to (1), the used dataset was normalized to the range [0, 1]; normalizing the data is important in eliminating the biased features of greater values for the dataset. We used 20% of the original dataset CICDDoS2019, and it consists of two parts: 70% train size and 30% test size.

$$X_{normalized} = \frac{X - X_{min}}{Xmin_{max}} \tag{1}$$

Where min is the minimum value in feature, max is the maximum value in feature.

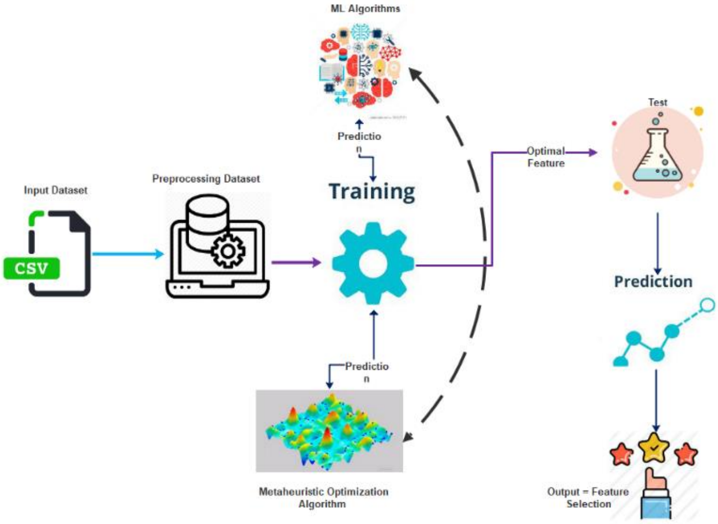


Figure 1. The process design of the proposed PFS model

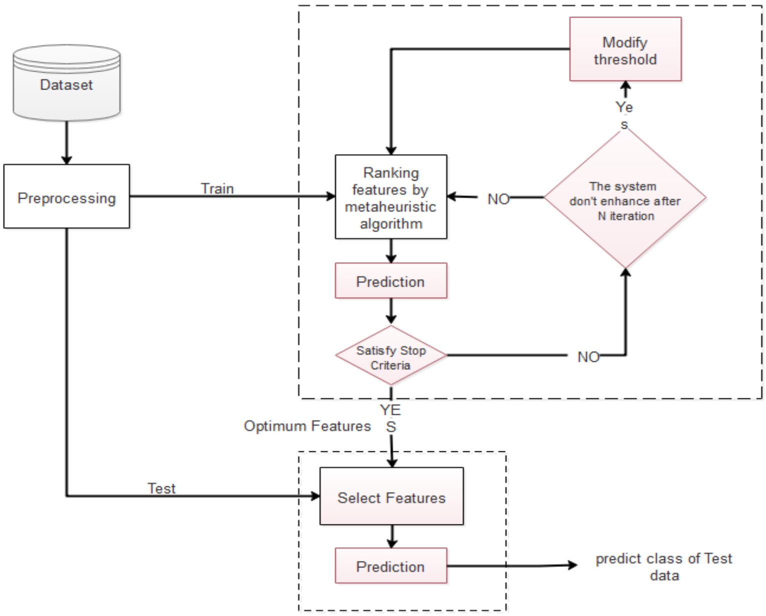


Figure 2. Flowchart of the proposed PFS model

4.2. Enhance feature selection by using the PFS model

A challenging issue is the selection of features. When the dimensionality of the feature is high, the choice of the appropriate features is crucial. To address this, SWEVO metaheuristic algorithms are most suited. Centered on the PSO, BA, and DE algorithms, three subsets were derived from the proposed model. The proposed system used dynamic behavior for setting the value of θ that selected only features that corresponding to rank higher than the threshold to select an optimal value to the θ. Equation (2) calculates the θ for wrapper models,

$$\varDelta\theta = \theta Max - \theta Min \tag{2}$$

where θMax and θMin can be determined by the user

$$\theta' = \theta Min + \varDelta\theta \times (1 - (\frac{current\ Iter}{MaxIter})^2 \times \lambda \tag{3}$$

where θMin and $\varDelta\theta$ can be computed from (2), $current\ Iter$, $MaxIter$, λ is a random variable in the interval [-1, 1].

In the initial stage of the search process, the system needs to apply as high of θ as it is possible to restrict to features with high-rank values only. The value of θ is reduced during search progress. Figure 3 shows the probability of θ during search iterations.

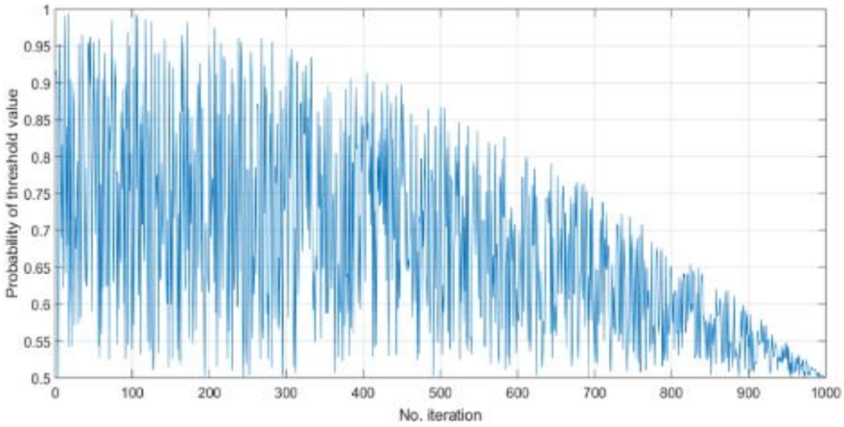


Figure 3. Probability of θ

Optimal feature selection is influenced by the appropriate θ value chosen based on the initial value of θ. Therefore, the upper and lower bound of the period range of θ is not constant and can be changed by the user based on the resulting quality of the model; therefore, the user can specify the range or period within which the θ is located. For the above reason, we used the adaptive threshold because the initial θ represents the threshold. We set an adaptive threshold with an initial value used to distinguish between normal and abnormal behaviors. The user can change this adaptive threshold value based on the result and expand the search space if the results do not satisfy and set an acceptable threshold is not straightforward. The adaptive threshold is more useful than other thresholds because its ability to adapt to the changes that may occur in network traffic during the attack and set an initial value of the threshold has become challenging.

5. EXPERIMENTS

The data collection used along with the data pre-processing protocol is presented in detail in this section. We also provide the metrics of performance used in our experiments. Furthermore, we show our model's architecture. Finally, we provide a comparative analysis of our model and that of various classifiers. All experiments were performed on a 2.90 GHz, i7, 16 GB RAM, and Windows 10 pro-64 bits operating system. PyCharm ide python and python 3.8 are used to execute our model. The total number of features in the dataset CICDDoS2019 [7] is 88. We have used 20% per cent from the dataset based on each attack's time in the original dataset. When the researchers designed the PFS model, they focus on enhancing the detection accuracy and reducing the number of features; these two factors are very important when designing a new approach for the detection mechanism. The metrics used are based on its ability to categorize network traffic into a correct category; IDS efficiency is measured.

Algorithm 1. A proactive features selection model (PFS)

```
Input: Initialization Parameters"
Output: Optimal Features"
theta←0.5//initially set the threshold (theta=0.5)
segma←0//segma is a stagnation sensitive parameter
while $iter \leq Max - Iteration$ do
     Update population according to optimiaztion method
      $Local_{best}$←find(best)
      if $Global_{best}$<$Local_{best}$ then
            $Global_{best}$←$Local_{best}$
             segma←0
       else
             segma←segma+1
       if segma≤2*populationsize then
               segma←0
                theta←select, randomaly, max-min, eqx
return Optimal Features
```

$$Accuracy = \frac{\sum TP_{For\ Each\ Attack\ Class} + TN_{Benign}}{\sum TP_{For\ Each\ Attack\ Class} + \sum FN_{For\ Each\ Attack\ Class} + TN_{Benign} + FP_{Benign}} \tag{4}$$

$$Precision = \frac{\sum TP_{For\ Each\ Attack\ Class}}{\sum(TP_{For\ Each\ Attack\ Class}, FP_{Benign})} \tag{5}$$

$$Recall = \frac{\sum TP_{For\ Each\ Attack\ Class}}{\sum(TP_{For\ Each\ Attack\ Class}, FN_{For\ Each\ Attack\ Class})} \tag{6}$$

$$F1\ scor = \frac{2 * (Recall\ *\ Precision)}{\sum(Recall + Precision)} \tag{7}$$

5.1. Performance analysis (results and discussion)

The proposed PFS model results have been compared with the model's results in the base paper [7] based on the accuracy metrics shown in Table 1. The results had proved that the PFS model is better than the model in the base paper, depending on the accuracy measures. Also, the selected dataset was tested on both the machine learning algorithms such as KNN, RF, and SVM without the PFS model, SWEVO metaheuristic algorithms such as PSO, BA, and DE without PFS. Moreover, reducing the number of features selected has been done.

Table 1. Results of the performance test for the reference [7]

Algorithms	Precision	Recall	F1 Score
Decision Tree ID3	0.78	0.65	0.69
Random Forest	0.77	0.56	0.62
Naïve Bayes	0.41	0.11	0.05
Multinomial Logistic Regression	0.25	0.02	0.02

Table 2 shows that when running the KNN only and then PSO_KNN without PFS, BA_KNN without PFS, and DE_KNN without PFS. We implement the PFS model with the three previous models. The result indicates that the PFS model shows that the accuracy achieved is better than that of other models than PFS. The number of features was reduced when running PFS, and the details are also shown that: the PFS_PSO_KNN the number of features is 19 features, the PFS_BA_KNN the number of features is 34 features, and the PFS_DE_KNN the number of features is 45 features. The PFS_PSO_KNN is better than the other two proactive models PFS_BA_KNN and PFS_DE_KNN, in terms of accuracy and number of features.

Table 3 shows that when running the Random Forests RF only and then PSO_RF without PFS, BA_RF without PFS, and DE_RF without PFS. Then we implement the PFS model with the three previous models. The results indicate that the PFS shows that the accuracy achieved is better than that of other models than PFS. The PFS_PSO_RF reduced the number of features to 49 features while the PFS_BA_RF reduced the number to 41 features, and finally, the PFS_DE_RF had reduced the number of features to 53 features. The PFS_DE_RF is better than the other two proactive models PFS_PSO_RF and PFS_BA_RF, in terms of accuracy and other accuracy metrics such as precision, recall, and F1 score.

Table 2. PFS model performance and KNN with the three SWEVO algorithms relative to the different attack types

	DRDoS_DNS	DRDoS_LDAP	DRDoS_NetBIOS	DRDoS_SSDP	DRDoS_UDP	Normal	Accuracy	Precision	Recall	F1 score	Number of features
Model						KNN					
	89.79	78.77	96.31	66.77	99.8	99.61	81.94	80.64	80.66	80.65	88
						PSO_KNN without PFS					
	89.02	73.47	97.46	85.58	99.8	99.54	85.23	84.97	84.6	84.78	45
						PFS_PSO_KNN					
	91.43	79.55	97.21	89.79	98.48	99.45	**89.59**	**90.04**	**89.64**	**89.84**	**19**
						BA_KNN without PFS					
	89.1	82.75	96.47	74.07	99.81	99.71	84.62	84.53	83.75	84.14	37
						PFS_BA_KNN					
	91.27	82.67	96.99	87.39	98.47	99.53	**89.56**	**90.28**	**89.54**	**89.91**	**34**
						DE_KNN without PFS					
	93.55	81.22	96.73	72.96	99.8	99.57	85.11	83.97	84.22	84.09	65
						PFS_DE_KNN					
	91.8	81.12	97.13	84.38	99.78	99.57	**88.27**	**88.71**	**87.85**	**88.28**	**45**

Table 3. PFS model performance and RF with the three SWEVO algorithms relative to the different attack types and benign

	DRDoS_DNS	DRDoS_LDAP	DRDoS_NetBIOS	DRDoS_SSDP	DRDoS_UDP	Normal	Accuracy	Precision	Recall	F1 score	Number of features
Model						RF					
	81.89	76.04	92.69	93.19	95.97	95.49	83.07	84.91	83.49	84.2	88
						PSO_RF without PFS					
	84.74	80.61	94.46	90.63	97.67	97.17	85.78	86.74	86.16	86.45	46
						PFS_PSO_RF					
	86.69	83.26	96.43	91.82	99.74	99.15	**87.89**	**88.61**	**88.3**	**88.45**	**49**
						BA_RF without PFS					
	86.73	80.91	96.41	93.58	99.75	99.32	87.63	89.03	88	88.51	37
						PFS_BA_RF					
	86.65	82.07	96.44	92.08	99.74	99.13	**87.69**	**88.37**	**88.08**	**88.22**	**41**
						DE_RF without PFS					
	81.03	77.8	90.72	86.13	94.08	93.54	82.13	82.61	82.52	82.57	59
						PFS_DE_RF					
	86.65	83.44	96.43	92.11	99.71	99.19	**87.98**	**88.76**	**88.39**	**88.57**	**53**

Table 4 shows that when running the SVM only and then PSO_SVM without PFS, BA_SVM without PFS, and DE_SVM without PFS. Then we implement the PFS model with the three previous models. The results indicate that the PFS shows that the accuracy achieved is better than that of other models than PFS. The PFS_PSO_SVM reduced the number of features to 48 features while the PFS_BA_SVM reduced the number to 30 features, and finally, the PFS_DE_SVM reduced the number of features to 45 features. The PFS_BA_SVM is better than the other two proactive models PFS_PSO_SVM and PFS_DE_SVM, in terms of accuracy and number of features and accuracy metrics precision, recall, and f1scor.

Figure 4 shows that the accuracy line curve, although in the early iterations, PSO_KNN without PFS seems to be performing better in terms of detection accuracy, after iteration 11, PSO_KNN with PFS yields a higher detection accuracy rate. The final accuracy rate achieved by PSO_KNN with PFS stands at 89.59% compared to the one without PFS at 85.23%. Figure 5 shows that the accuracy line curve, although in the early iterations, BA_KNN with PFS, seems to be performing better in detection accuracy from the first iteration. It yields a higher detection accuracy rate. The final accuracy rate achieved by BA_KNN with PFS stands at 89.56% compared to the one without PFS at 84.62%. Figure 6 shows that the accuracy line curve, although in the early iterations, DE_KNN with PFS, performs better in detection accuracy from the first iteration. It yields a higher detection accuracy rate. The final accuracy rate achieved by DE_KNN with PFS stands at 88.27% compared to the one without PFS at 85.11%.

Figure 7 shows that the accuracy line curve, although in the early iterations, PSO_RF without PFS seems to be performing better in detection accuracy; after iteration 5, PSO_RF with PFS yields a higher detection accuracy rate. The final accuracy rate achieved by PSO_RF with PFS stands at 87.89% compared to the one without PFS at 85.78%. Figure 8 shows that the accuracy line curve, although in the early iterations, BA_RF without PFS seems to be performing better in detection accuracy; after iteration 161, BA_RF with PFS yields a higher detection accuracy rate. The final accuracy rate achieved by BA_RF with

PFS stands at 87.69% compared to the one without PFS at 87.63%. Figure 9 shows that the accuracy line curve, although in the early iterations, DE_RF with PFS seems to be performing better in detection accuracy from the first iteration, and it yields a higher detection accuracy rate. The final accuracy rate achieved by DE_RF with PFS stands at 87.89% compared to those without PFS at 82.13%.

Table 4. PFS model performance and SVM with the three SWEVO algorithms relative to the different attack types and benign

	DRDoS_DNS	DRDoS_LDAP	DRDoS_NetBIOS	DRDoS_SSDP	DRDoS_UDP	Normal	Accuracy	Precision	Recall	F1 score	Number of features
Model						SVM					
	80.05	74.23	81.47	83.51	72.83	93.76	70.3	73.31	72.13	72.72	88
						PSO_SVM without PFS					
	83.13	77.29	85.08	86.83	76.08	96.85	73.61	76.85	75.46	76.15	31
						PFS_PSO_SVM					
	85.39	79.6	87.48	89.04	78.38	99.15	**75.92**	**79.14**	**77.78**	**78.41**	**48**
						BA_SVM without PFS					
	85.4	79.55	87.64	88.92	78.43	99.15	75.91	79.18	77.78	78.47	35
						PFS_BA_SVM					
	85.35	81.4	88.76	81.99	82.56	99.15	**76.38**	**79.45**	**77.77**	**78.6**	**30**
						DE_SVM without PFS					
	85.4	79.63	87.83	88.92	78.5	99.15	76	79.23	77.87	78.54	46
						PFS_DE_SVM					
	85.4	79.62	87.7	89.04	78.47	99.18	**76.01**	**79.23**	**77.88**	**78.55**	**45**

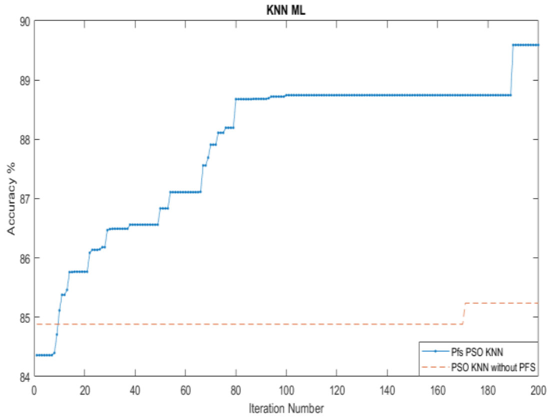


Figure 4. Resultant accuracy line the curve of the PSO and KNN

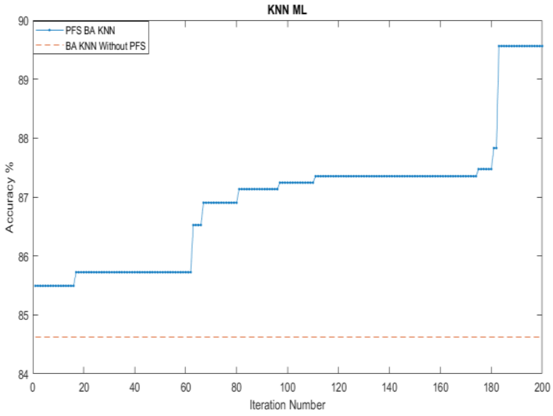


Figure 5. Resultant accuracy line the curve of the BA and KNN

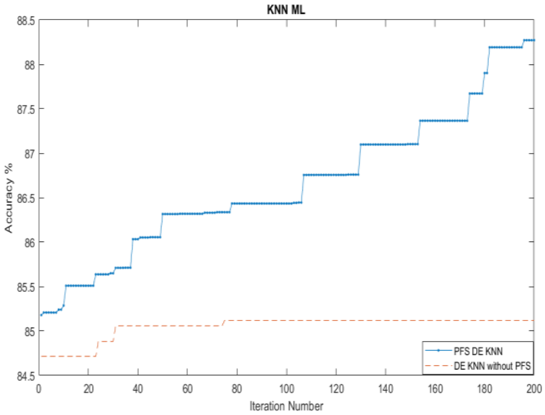


Figure 6. Resultant accuracy line the curve of the DE and KNN

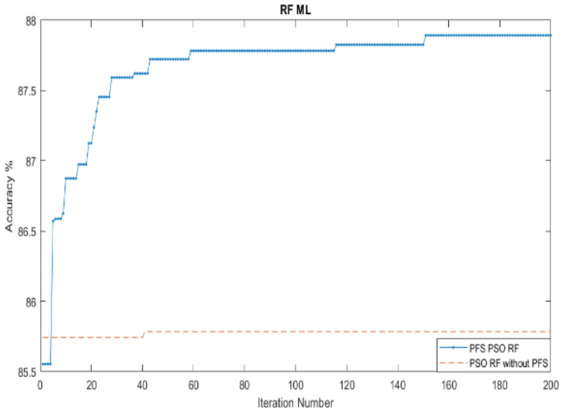
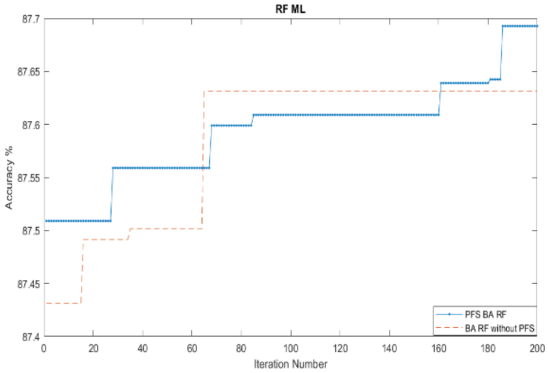


Figure 7. Resultant accuracy line the curve of the PSO and RF



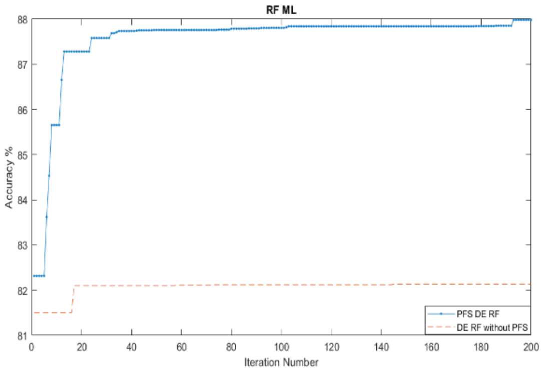
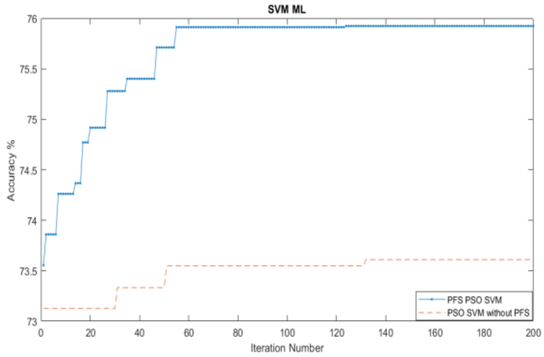


Figure 8. Resultant accuracy line the curve of the BA and RF

Figure 9. Resultant accuracy line the curve of the DE and RF

Figure 10 shows that the accuracy line curve, although in the early iterations, PSO_SVM with PFS, seems to be performing better in detection accuracy from the first iteration. It yields a higher detection accuracy rate. The final accuracy rate achieved by PSO_SVM with PFS stands at 75.92% compared to the one without PFS at 73.61%. Figure 11 shows that the accuracy line curve, although in the early iterations, BA_SVM with PFS, seems to be performing better in detection accuracy from the first iteration. It yields a higher detection accuracy rate. The final accuracy rate achieved by BA_SVM with PFS stands at 76.38% compared to the one without PFS at 75.91%. Figure 12 shows that the accuracy line curve, although in the early iterations, DE_SVM without PFS seems to be performing better in detection accuracy; after iteration 178, DE_SVM with PFS yields a higher detection accuracy rate. The final accuracy rate achieved by DE_SVM with PFS stands at 76.01% compared to those without PFS at 76%.



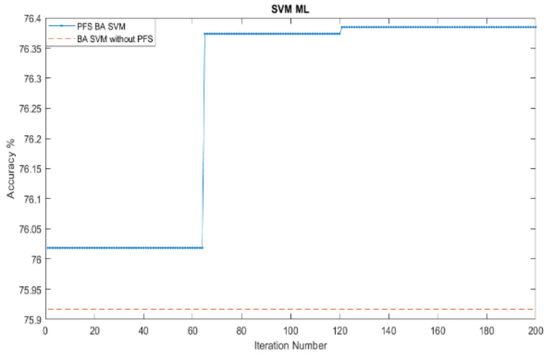


Figure 10. Resultant accuracy line the curve of the PSO and SVM

Figure 11. Resultant accuracy line the curve of the BA and SVM

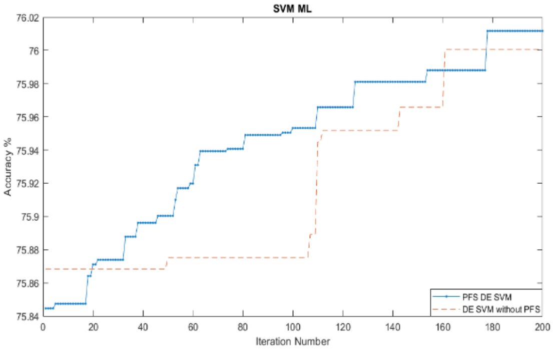


Figure 12. Resultant accuracy line the curve of the DE and SVM

6. CONCLUSION

The number of features in the dataset influences the detection mechanism performance; therefore, reducing the number of features is necessary to improve the detection accuracy rate. In the PFS model, we use an adaptive threshold to enhance accuracy of detection by distinguishing normal from abnormal in the dataset. A few models are suggested to detect the DRDoS attacks, but some failed, or the detection accuracy rate is very low; for those reasons, the authors are suggested the new model PFS that can detect the DRDoS attacks. The PFS model is based on optimization algorithms and classifiers machine learning algorithms. The tables of comparisons and figures of accuracy line curve prove that the PFS model is the best never to detect the DRDoS attacks with high true positive rate and low false-negative rate. Our future work will focus on enhancing the detection rate and minimizing the false alarm rate by using other techniques such as clustering or neural networks.

REFERENCES

[1] I. Sreeram and V. P. K. Vuppala, “HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm,” *Applied Computing and Informatics*, vol. 15, no. 1, pp. 59–66, Jan. 2019, doi: 10.1016/j.aci.2017.10.003.

[2] M. T. Manavi, “Defense mechanisms against distributed denial of service attacks: a survey,” *Computers and Electrical Engineering*, vol. 72, pp. 26–38, Nov. 2018, doi: 10.1016/j.compeleceng.2018.09.001.

[3] N. Agrawal and S. Tapaswi, “Detection of low-rate cloud DDoS attacks in frequency domain using fast Hartley transform,” *Wireless Personal Communications*, vol. 112, no. 3, pp. 1735–1762, Jan. 2020, doi: 10.1007/s11277-020-07125-4.

[4] D. Kshirsagar and S. Kumar, “A feature reduction based reflected and exploited DDoS attacks detection system,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–3, Jan. 2021, doi: 10.1007/s12652-021-02907-5.

[5] M. Prasad, S. Tripathi, and K. Dahal, “An efficient feature selection based bayesian and rough set approach for intrusion detection,” *Applied Soft Computing*, vol. 87, Feb. 2020, doi: 10.1016/j.asoc.2019.105980.

[6] NETSCOUT, “14th annual worldwide infrastructure security report, 2019,” 2019.

[7] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, “Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy,” in *2019 International Carnahan Conference on Security Technology (ICCST)*, Oct. 2019, pp. 1–8, doi: 10.1109/CCST.2019.8888419.

[8] Y. Gao, Y. Feng, J. Kawamoto, and K. Sakurai, “A machine learning based approach for detecting DRDoS attacks and its performance evaluation,” in *Proceedings - 11th Asia Joint Conference on Information Security, AsiaJCIS 2016*, Aug. 2016, pp. 80–86, doi: 10.1109/AsiaJCIS.2016.24.

[9] R. Xu, J. Cheng, F. Wang, X. Tang, and J. Xu, “A DRDoS detection and defense method based on deep forest in the big data environment,” *Symmetry*, vol. 11, no. 1, pp. 1–22, Jan. 2019, doi: 10.3390/sym11010078.

[10] O. Almomani, “A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms,” *Symmetry*, vol. 12, no. 6, pp. 1–20, Jun. 2020, doi: 10.3390/sym12061046.

[11] R. Vijayanand and D. Devaraj, “A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network,” *IEEE Access*, vol. 8, pp. 56847–56854, 2020, doi: 10.1109/ACCESS.2020.2978035.

[12] J. Ghasemi, J. Esmaily, and R. Moradinezhad, “Intrusion detection system using an optimized kernel extreme learning machine and efficient features,” *Sādhanā*, vol. 45, no. 1, Dec. 2020, doi: 10.1007/s12046-019-1230-x.

[13] S. Sarvari, N. F. Mohd Sani, Z. Mohd Hanapi, and M. T. Abdullah, “An efficient anomaly intrusion detection method with feature selection and evolutionary neural network,” *IEEE Access*, vol. 8, pp. 70651–70663, 2020, doi: 10.1109/ACCESS.2020.2986217.

[14] A. Patil and D. Kshirsagar, “Towards feature selection for detection of DDoS attack,” in *Advances in Intelligent Systems and Computing*, vol. 1025, Springer Singapore, 2020, pp. 215–223.

[15] A. Thakkar and R. Lohiya, “Role of swarm and evolutionary algorithms for intrusion detection system: a survey,” *Swarm and Evolutionary Computation*, vol. 53, Mar. 2020, doi: 10.1016/j.swevo.2019.100631.

[16] R. Y. M. Nakamura, L. A. M. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa, and X.-S. Yang, “BBA: a binary bat algorithm for feature selection,” in *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, Aug. 2012, pp. 291–297, doi: 10.1109/SIBGRAPI.2012.47.

[17] H. Liu and B. Lang, “Machine learning and deep learning methods for intrusion detection systems: a survey,” *Applied Sciences*, vol. 9, no. 20, Oct. 2019, doi: 10.3390/app9204396.

[18] J. Camacho, G. Macia-Fernandez, N. M. Fuentes-Garcia, and E. Saccenti, “Semi-supervised multivariate statistical network monitoring for learning security threats,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 8, pp. 2179–2189, Aug. 2019, doi: 10.1109/TIFS.2019.2894358.

[19] S. Hosseini and M. Azizi, “The hybrid technique for DDoS detection with supervised learning algorithms,” *Computer Networks*, vol. 158, pp. 35–45, Jul. 2019, doi: 10.1016/j.comnet.2019.04.027.

[20] R. Magán-Carrión, D. Urda, I. Díaz-Cano, and B. Dorronsoro, “Towards a reliable comparison and evaluation of network intrusion detection systems based on machine learning approaches,” *Applied Sciences*, vol. 10, no. 5, Mar. 2020, doi: 10.3390/app10051775.

[21] M. Wang, Y. Lu, and J. Qin, “A dynamic MLP-based DDoS attack detection method using feature selection and feedback,” *Computers & Security*, vol. 88, Jan. 2020, doi: 10.1016/j.cose.2019.101645.

[22] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95 - International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.

[23] K. M. Ang, W. H. Lim, N. A. M. Isa, S. S. Tiang, and C. H. Wong, “A constrained multi-swarm particle swarm optimization without velocity for constrained optimization problems,” *Expert Systems with Applications*, vol. 140, Feb. 2020, doi: 10.1016/j.eswa.2019.112882.

[24] H. Xu, K. Przystupa, C. Fang, A. Marciniak, O. Kochan, and M. Beshley, “A combination strategy of feature selection based on an integrated optimization algorithm and weighted k-nearest neighbor to improve the performance of network intrusion detection,” *Electronics*, vol. 9, no. 8, pp. 1–22, Jul. 2020, doi: 10.3390/electronics9081206.

[25] X.-S. Yang, “A new metaheuristic bat-inspired algorithm,” in *Studies in Computational Intelligence*, vol. 284, Springer Berlin Heidelberg, 2010, pp. 65–74.

[26] R. Storn and K. Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997, doi: 10.1023/A:1008202821328.

[27] M. N. Ab Wahab, S. Nefti-Meziani, and A. Atyabi, "A comprehensive review of swarm optimization algorithms," *PLOS ONE*, vol. 10, no. 5, May 2015, doi: 10.1371/journal.pone.0122827.

[28] C. Raghuraman, S. Suresh, S. Shivshankar, and R. Chapaneri, "Static and dynamic malware analysis using machine learning," in *Advances in Intelligent Systems and Computing*, vol. 1045, Springer Singapore, 2020, pp. 793–806.

[29] P. Nagar, H. K. Menaria, and M. Tiwari, "Novel approach of intrusion detection classification deeplearning using SVM," in *Advances in Intelligent Systems and Computing*, vol. 1045, Springer Singapore, 2020, pp. 365–381.

[30] A. A. Aburomman and M. Bin Ibne Reaz, "A novel SVM-KNN-PSO ensemble method for intrusion detection system," *Applied Soft Computing*, vol. 38, pp. 360–372, Jan. 2016, doi: 10.1016/j.asoc.2015.10.011.

[31] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.2841987.

[32] X. K. Li, W. Chen, Q. Zhang, and L. Wu, "Building auto-encoder intrusion detection system based on random forest feature selection," *Computers and Security*, vol. 95, Aug. 2020, doi: 10.1016/j.cose.2020.101851.

[33] R. R. Nuiaa, S. Manickam, and A. H. Alsaeedi, "Distributed reflection denial of service attack: a critical review," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6, pp. 5327–5341, 2021, doi: 10.11591/ijece.v11i6.pp5327-5341.

BIOGRAPHIES OF AUTHORS

Riyadh Rahef Nuiaa received a B.Sc. in Computer Sciences in 2004 from Baghdad College of Economics sciences University, Baghdad, Iraq. He has completed the M.Sc. in Information System/Computer Sciences in 2014 from the college Osmania University, India. He is enrolled as a Ph.D. student from December 2019 in the National Advanced IPv6 Centre/Universiti Sains Malaysia. He has worked as a lecturer at Wasit University, Iraq in cloud computing, computer theory, and operation systems. His research interests network cloud computing, cybersecurity, and data mining. He can be contacted at email: riyadhrahef@gmail.com and riyadh@uowasit.edu.iq.

Selvakumar Manickam is an associate professor and researcher at National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia. He has authored and co-authored almost 170 articles in journals, conference proceedings and book reviews. He has graduated 13 Ph.Ds. and many Masters and FYP students. He has given several key note speeches as well as dozens of invited lectures and workshops at conferences, international universities and for industry. his research interest includes cybersecurity, cloud computing, software defined network, IPv6, internet of things (IoT) and open source technology. He can be contacted at email: selva@usm.my.

Ali Hakem Alsaeedi is completed B.Sc. in Computer Sciences in 2006 from the college of sciences at University of Al-Qadisiyah, Diwaniya, Iraq. Received his M.Sc. (master) in computer sciences in the year 2016 from the college of computer sciences at the Yildiz Technical University (YTU), Istanbul, Turkey. He has worked as a lecturer at a number of the Iraqi Universities in the areas of artificial intelligent, data mining, and signal processing. He currently works as a lecturer in the University of Al-Qadisiyah. His research interests machine learning, smart optimization algorithms, and optimization of big data. Ali has several publications in the areas of the binary of metaheuristic optimization and data mining. He can be contacted at email: ali.alsaeedi@qu.edu.iq.

Esraa Alomari has done a Bachelor in Computer Science from computer collage-Al-Anbar University-IRAQ in 2003 and received her M.Sc. Degree in Computer Science from University of Technology-IRAQ in 2006. Received her Ph.D. from National Advanced Center of Excellence (Nav6) in University Sains Malaysia (USM). Currently she is an Assis. Prof in Wasit University/College of Education and Pure Sciences as a lecturer. Her research interest includes Advances internet security and monitoring, Botnet and Cyber-attacks and IoT security. She has authored and co-authored around 16 publications. She is member of the IPv6 Forum Global Education including Certified IPv6 Network Engineer (CNE6) Level 1 and 2. She can be contacted at email: ealomari@uowasit.edu.iq.