Computers, Materials & Continua
DOI: 10.32604/cmc.2023.031047
Article

Tech Science Press

Resource Exhaustion Attack Detection Scheme for WLAN Using Artificial Neural Network

Abdallah Elhigazi Abdallah[1], Mosab Hamdan[2], Shukor Abd Razak[3], Fuad A. Ghalib[3], Muzaffar Hamzah[2,*], Suleman Khan[4], Siddiq Ahmed Babikir Ali[5], Mutaz H. H. Khairi[1] and Sayeed Salih[6]

[1]Faculty of Engineering, Future University, Khartoum 10553, Sudan
[2]Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, 88400, Malaysia
[3]Information Assurance and Security Research, School of Computing, Universiti Teknologi Malaysia, Johor, 81310, Malaysia
[4]School of Psychology and Computer Science, University of Central Lancashire, Preston PR1 2HE, UK
[5]Deaptement of Management Information Systems, College of Administrative Science, Applied Science University, Al Eker, 623, Bahrain
[6]Deaptement of Information Technology, College of Computer and Information Sciences, King Saud University, Riyadh, 11461, Saudi Arabia
*Corresponding Author: Muzaffar Hamzah. Email: muzaffar@ums.edu.my

Received: 08 April 2022; Accepted: 15 September 2022

Abstract: IEEE 802.11 Wi-Fi networks are prone to many denial of service (DoS) attacks due to vulnerabilities at the media access control (MAC) layer of the 802.11 protocol. Due to the data transmission nature of the wireless local area network (WLAN) through radio waves, its communication is exposed to the possibility of being attacked by illegitimate users. Moreover, the security design of the wireless structure is vulnerable to versatile attacks. For example, the attacker can imitate genuine features, rendering classification-based methods inaccurate in differentiating between real and false messages. Although many security standards have been proposed over the last decades to overcome many wireless network attacks, effectively detecting such attacks is crucial in today's real-world applications. This paper presents a novel resource exhaustion attack detection scheme (READS) to detect resource exhaustion attacks effectively. The proposed scheme can differentiate between the genuine and fake management frames in the early stages of the attack such that access points can effectively mitigate the consequences of the attack. The scheme is built through learning from clustered samples using artificial neural networks to identify the genuine and rogue resource exhaustion management frames effectively and efficiently in the WLAN. The proposed scheme consists of four modules which make it capable to alleviates the attack impact more effectively than the related work. The experimental results show the effectiveness of the proposed technique by gaining an 89.11% improvement compared to the existing works in terms of detection.

Keywords: 802.11; media access control (MAC); wireless local area network (WLAN); artificial neural network; denial-of-service (DoS)

1 Introduction

802.11-based wireless local area networks are characterized by usability and cost-effectiveness, leading to widespread deployment worldwide. However, due to the broadcast nature of wireless access, such networks are prone to many malicious attacks. Several security extensions to 802.11 have been proposed to counter these attacks to alleviate vulnerabilities related to unauthorized access and confidentiality breaches. The high demand for access to wireless networks makes it necessary to count for the issue of availability is another crucial security requirement [1].

Denial-of-Service (DoS) attacks to compromise the availability of part or entire wireless network resources/services. Such attacks prevent legitimate users from accessing the network [2,3]. It is worth noting that DoS attacks differ from selfish behaviour motivated by possible beneficial outcomes [4,5]. Due to the broadcast nature of wireless networks, DoS attacks can be easily carried out, particularly in the wireless domain. Besides, several 802.11-specific DoS vulnerabilities have been experimentally demonstrated in the literature in recent years [6,7]. IEEE 802.11 Wi-Fi networks are prone to many DoS attacks due to vulnerabilities at the media access control (MAC) layer of the 802.11 protocol [8,9]. DoS attacks pose a significant threat to the availability and reliability of the operation of wireless networks in general, specifically the critical information infrastructure. The world economic forum (WEF) in [10,11] has confirmed that DoS attacks lead to severe interruption of the necessary information infrastructure. In particular, the UK national infrastructure security coordination centre (NISCC) warned in 2005 [12] that denial of service attacks could impact critical national infrastructure.

DoS attacks target network availability to deny legitimate users access to network resources. It is unlike some users' selfish attitude that is motivated by ulterior motives. Given its broadcast nature, DoS attacks are easier to carry out in wireless domains. In recent years, many standard security vulnerabilities in 802.11 have been piloted in the literature [1,13]. Availability attacks are a form of DoS attack which attempts to disable access to the network using several types of MAC frames, such as de-authentication and de-association frames [14,15]. Wireless networks and technologies are generally more vulnerable to DoS attacks than wired alternatives [16].

This study aims to propose a resource exhaustion attack detection and mitigation scheme. The proposed scheme has been designed to differentiate between the genuine and fake management frames in the early stages of the attack. Thus, artificial neural networks are used to build the scheme via learning from clustered samples. Adopting artificial neural networks aims to identify the genuine and rogue resource exhaustion management frames effectively and efficiently in the wireless local area network (WLAN). The proposed scheme is evaluated using a MATLAB environment. The simulation results obtained from the experiments show the effectiveness of the proposed scheme.

The remainder of the paper is structured as follows. Section 2 provides related works. Section 3 discusses the problem formulation. Section 4 introduced the attacker type. Section 5 offers a proposed solution. Section 6 discusses the experimental results. Section 7 presents the discussion and concluding remarks.

2 Related Works

In wireless network communication, DoS attacks generally aim to deny legitimate network users access to WLAN services. Over many years several approaches have been proposed to address the problem of DoS in WLAN. For example, a study by Sounni et al. [17] presents a DoS detection scheme based on statistical process control (SPC) using the fraction non-conforming control chart; This control chart is used to monitor the packet drop ratio (PDR) variation; it collects data in a normal case, calculates graph parameters, and plots them in the same graph. The proposed method allows the detection of distributed denial of service (DDoS) attacks by monitoring the PDR metric graphs in real-time on one side; on the other side, there is no change in the IEEE 802.11 and OpenFlow standard intruder database (IDB). This is to prevent attackers from bringing down WLAN connections with DoS attacks. Although IDB can be very effective in avoiding DoS attacks, it can spoof and store an intruder's MAC address without his knowledge [18]. However, IDB cannot detect new intruders that have not registered in the database. Since it uses the MAC address to identify an intruder who usually spoofs legitimate MAC addresses to launch attacks. In [19] a comparative analysis was conducted on various proposed tools and techniques for detecting and preventing DDoS attacks.

A study in [20] presents a technique that can be easily deployed on open as well as encrypted networks. The proposed model is implemented by using the ESP8266 Wi-Fi module. It can detect these attacks in wireless LANs (WLANs) by extracting the victim's basic service set ID (BSSID). Elsabagh et al. [21] introduced a novel technique named Radmin, which works on compiled binaries for early detection of application-level resource exhaustion and starvation attacks. The idea is to limit the usage of the resources for particular programs to the learned automata and detects resource usage anomalies at their early stages. The results show that the Radmin can accurately work on both the user and kernel spaces. Ratnayake et al. [22] proposed in their study a scheme for probe request attack detection using a neural network (NN)classifier to classify a Station (STA's) real-world WLAN traffic frames. Four features, signal strength, sequence number, frame sub-type and delta time, were used to train the supervised feed forward NN classifier to differentiate a legitimate frame from a malicious one. Experimental findings show that in addition to the NN-based model detecting probe request attacks with high accuracy, it notably detects them in their early stage.

Additionally, the current approach of differentiating genuine frames from corrupted ones using real real-world traffic data for NN is manual and labor-intensive. Ding [23] proposed an approach based on a central manager (CM) to avoid DoS attacks in WLAN. CM, which uses a timer and three tables, is installed on the backend server that manages the access point (AP) to detect a DoS attack. In addition to DoS attack detection, CM can dynamically control the client and AP. The author demonstrated the performance of the CM through a Network Simulator, NS-2, while generating network traffic using case-based reasoning (CBR) applications. Throughput and delay time was the evaluation metrics used in assessing the performance of CM. The two-part attacks simulated on the network are the login (extensible authentication protocol (EAP) start and 802.11 association) and logout parts (EAP failure, EAPOL logoff, MAC disassociation). The throughput of the Network when CM was used significantly improved compared to without CM. A similar finding, which includes improved efficiency and performance, was reported on the delay time of the WLAN when CM was used [24].

3 Problem Motivation

It is challenging to detect and mitigate resource exhaustion attacks in wireless networks at their early stages. The attacker can cascade the entity of legitimate nodes or create fake entities due to the unsecured nature of the management frame in the 802.11 protocol used in wireless networks.

Unprotected management frames have motivated many adversaries to perform many types of DoS attacks. An attacker station (AST) can easily spoof a genuine station's (GST) media access control (MAC) address to deceive the access point (AP) [25,26]; thereby exhausting the resources by increasing contention on the wireless for shared media, memory and processing capacity. Adversarial stations (AST) can simply flood the access point by generating many fake frames such as a probe, authentication, and association requests. Once the AP responds to these fake frames, the resources are exhausted, and the denial of service is achieved.

Many researchers have suggested different solutions to differentiate between genuine and rogue stations. This include statistical-based, rule-based, and machine learning-based approaches. Among those approaches, machine learning-based approach has been reported as a promising approach that can be used to detect such attacks more effectively [8]. According to the MAC address, most of the works try to classify the attackers based on the profiling concept. However, when the attacker spoofs the MAC address of legitimate stations, the classification based on such profiling becomes useless and ineffective. Moreover, the attacker can manipulate its own MAC address, thus rendering attack detection unusable as it happens too late after the attacker finishes. Hence, it is better to detect fake messages instead of fake entities.

Despite the difficulty in differentiating between genuine and fake frames [8], several methods can be utilized to detect such fake frames. Many existing works (e.g., [8,25,26]) try to detect and mitigate the attack after the AP is already infected and the attack has succeeded. Most of these schemes were constructed based on profiling user activities during the attack. However, profiling user activities needs proper identification. Thus, such approaches are not able to detect the attacker's identity and mitigate fake messages without human physical intervention. Nevertheless, the attacker can simply modify and falsify the message identity.

4 Attacker Type

The attacker may actively or passively monitor the network to learn vital network information. Using such information, the attacker modifies its mac address and sends false management frames. The access point responds to the message by replying or assigning resources to the new connection. There are three possible scenarios of attacks in wireless local area networks: (i) probe request flooding attack, (ii) authentication request attack, and (iii) association request attack. In a probe request flooding attack, the attacker periodically sends probe requests to which the access point responds by sending back a probe response frame. In this case, the attacker aims to exhaust the network resources such as CPU, memory, and network bandwidth by generating a massive number of probe requests and response messages, thus increasing the network members' contention window. Such an attack decreases the throughput and increases the delay in response to the AP [27,28].

Probe request flooding attack is simple and easy to detect. However, the attacker can make the attack more sophisticated by changing its MAC address before sending the probe request frame. Changing the MAC address makes the access point monitor and profile existing network members. An authentication request attack is more severe than a probe request attack and more difficult to detect. Attackers may prefer to complicate the attack in order to avoid the detection solutions that rely on the behavioral analysis of the network members.

In such an attack, the attacker sends an authentication request attack after it receives the probe response attack. This increases the processing time and consumes the memory of the access point by overloading the authentication buffer with the fake MAC addresses. On the other hand, the association request attack is the most severe DoS attack. Detecting such an attack is challenging due to the similarity between the attacker's behavior and the normal behavior of benign stations. In this attack, the attacker sends many association messages with different MAC addresses. Upon receiving such messages, the access point assigns bandwidth, memory, and processing quota to each association request. Thus, exhausting the available resources while the access point fails to respond to genuine stations (GST).

5 Proposed Scheme

As mentioned in the previous section, the proposed data-centric resource exhaustion attack detection (READS) scheme consists of four modules, as illustrated in Fig. 1. The raw feature collection module is responsible for collecting the raw data related to receiving messages. As its name implies, the features derivation module is responsible for deriving informative features from the raw data collected in the previous module. The feature extraction module is responsible for transforming the derived features into a new form suitable for processing by the artificial neural network (ANN) classification module, which can be described as:

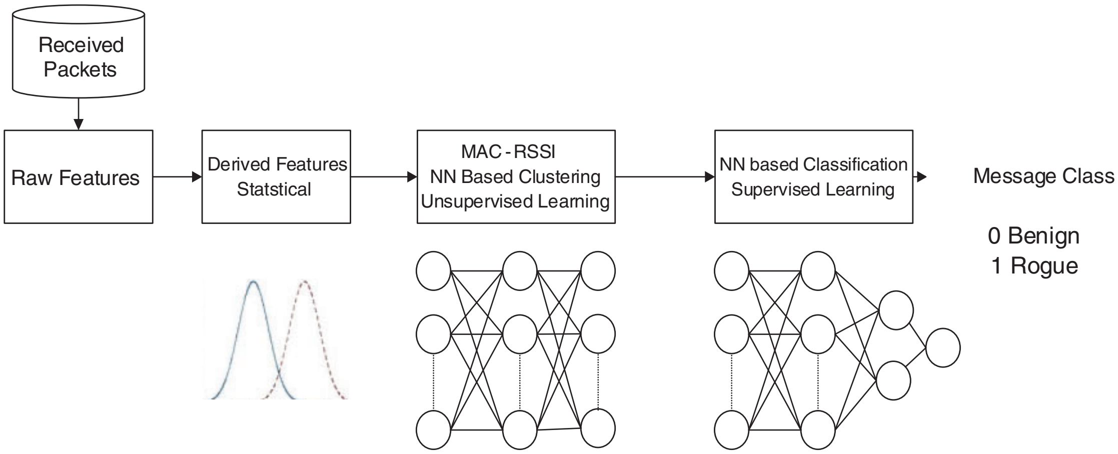
Received Packets
Raw Features
Derived Features Statstical
MAC - RSSI NN Based Clustering Unsupervised Learning
NN based Classification Supervised Learning
Message Class
0 Benign
1 Rogue

Figure 1: Description of the READ scheme

5.1 Raw Data Collocation Module

The raw features associated with each received message are collected. The five main features that are collected in this module are as follows:

(i) The source MAC address is extracted from the frame header and used as the first part of the message source identification to form the cluster.

(ii) The received signal strength indicator (RSSI) is the second part of the message identification. It is used as a complement to the MAC address because it is difficult to spoof the RSSI of a legitimate client, especially if a directional antenna is used. Hence, it becomes easy to distinguish the messages coming from legitimate stations from those of rogue ones, even if the attacker is able to spoof the MAC address.

(iii) The message sequence number is the third part of the message identification. Each station $c_{i,j}$: $i == \{1, 2, \ldots, n\}$ *and* $j = \{1, 2, \ldots, m\}$ generates a random start message sequence by which the source of the messages can be identified. This feature helps to consolidate the message protection against spoofing attacks as it is difficult for an attacker to manipulate the three tuples identification <MAC, RSSI, Sequence-Number>.

(iv) The message type feature is used to distinguish between the types of received messages which could be a probe request, authentication request, and association request.

(v) Message Creation Time and Arrival time are used to compute the message delivery time to determine the message generation rate, message arrival rate, and message delay rate.

5.2 Derived Features Module

In this module, more representative behavioural features are derived. The features have been derived according to the information recorded in the MAC address. The MAC address is used as a source of these features as it is suitable for identifying fake management frames. That is, even if the attacker has been able to change the MAC address of the frame, the message class will differ from the class of the victim's station. Thus, it becomes easy to detect and mitigate such attack. Table 1 shows the derived features.

Table 1: Summary description of the features

No#	Feature name	Description
DF1	RSSI_running _average	The running average of the RSSI signal.
DF2	RSSI_running _variance	The running statistical variance of the RSSI signal.
DF3	Sequence_step variance	The running variance of the difference between subsequent sequence numbers.
DF4	Prereq rate	The rate of the probe request frames per time.
DF5	Autreq rate	The rate of the authentication frames per time.
DF6	Asoreq rate	The rate of the association frames per time.
DF7	Data frame rate	The packet rate of the data.

5.3 The Derived Features Module

In this module, the proposed feature extraction module transforms the derived features into new feature sets representing fake and genuine frames. The idea is to group the messages received from different senders into clusters according to their derived features. Each cluster has a centroid, in which the message deviates from the cluster's centroid is deemed fake. Unsupervised learning using an artificial neural network clustering method is used to learn the input features and extract the complex relations that exist between the input features. A variety of neural network architectures have been used for clustering. The most widely used one is the self-organizing maps (SOM) [29], in which each neuron represents a cluster center in the same distribution. Thus, the total number of neurons should be equal to the total number of clusters [30]. The neurons are arranged in a two-dimensional array known as maps. One of the important features of SOM is topology order. SOM is organized into a rectangular grid topology [31]. The SOM applies competitive learning whereby the neurons compete to respond to the input data. The weights are used to represent the neurons of the centroid as a newly generated feature [29–32].

The clustering problem can be formulated as follows: Given a set of input management frames, $x_k : k = \{1, 2, \ldots, p\}$, where p is the number of the given samples, find the cluster centers, where $n; m$ is the number of required clusters which are arranged in $(n \times m)$ matrix. The input feature, x_k, is called the original problem space or $x - space$, while $c_{i,j}$ is called the $map - space$. The clustering algorithm groups the input samples into ($n \times m$) clusters such that the samples with the SOM are trained in two iterations. In the first iteration, the nearest cluster, $c_{i,j}$, is found for each training sample, x_k, according to the following equation:

$$||c_{i,j} - x_k|| < ||c_{p,q} - x_k\ ||\forall p, q,\ p \neq i \ and \ q \neq j \tag{1}$$

In the next iteration, also called the neighbouring update phase, the centre, $c_{i,j}$, and its neighbouring centres, $c_{p,q}$, are updated according to the following equation:

$$c_{i,j} = c_{i,j} + \alpha\,(t)\left(c_{i,j} - x_k\right),\ c_{p,q} = c_{p,q} + \alpha\,(t)\left(c_{p,q} - x_k\right) \tag{2}$$

where α is the learning rate that decays as the training progresses, Fig. 2 illustrates how the updating phase makes the map learn to be aligned to the original space to represent the input space.

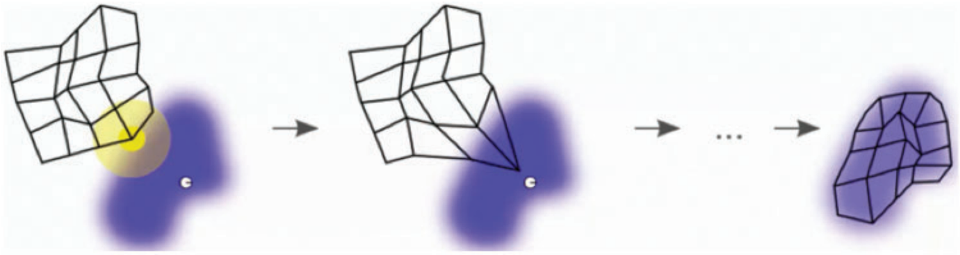

Figure 2: Illustration of the mapping of neighboring nodes update

After the training phase completes, the representative feature vectors for each input, x_k, can be generated as illustrated in the pseudo-code in Algorithm 1.

Algorithm 1: Features Extraction Algorithm

```
Input: Input features $x$
Output: $f_k \ : k = \{1, 2, .., n \times m\}$ (the extracted features)
SOM Construction Phase
Create $n \times m$ clusters according to the number of the unique MAC addresses in the dataset. The cluster $c_{i,j}$ is randomly initialized using $x_k$.
for each $x_k$ in the data − set
        Find the nearest cluster
                $||c_{i,j} - \ x_k|| < ||c_{p,q} - \ x_k\ ||\forall p, q,\ p \neq i \ and \ q \neq j$
        Update the cluster centroid
                $c_{i,j} = c_{i,j} + \alpha\,(t)\left(c_{i,j} - \ x_k\right)\forall i, j\ \in SOM\,(n \times m)$
Stop when there $\alpha\,(t)$ approaches zero
Features Extraction Phase
for each $x_k$ in the data − set
        for each i in $\{1, 2, \ldots, n\}$
                for each j in $\{1, 2, \ldots, m\}$
                        $f_k.append\left(||c_{i,j} - \ x_k||\right)$
```

5.4 ANN Classification Module

This module aims to construct a data-driven classifier that can differentiate between fake and genuine management frames in WLAN. The classifier is constructed using a supervised artificial neural network architecture, i.e., the feed-forward back propagation with a scaled conjugate gradient. The ANN classification module consists of two phases: training and testing phase. The classifier was trained using labelled dataset samples from real and simulated network traffic data.

The classifier architecture is a multilayer perceptron with four layers; one input layer, two hidden layers, and one output layer. Only two hidden layers were selected for the training to avoid overtraining (overfitting) caused by using a higher number of hidden layers. This, in turn, adversely affects the model's generalisation ability and consequently decreases the detection accuracy on unseen data. Thus, the network was trained using the scaled conjugate gradient and backpropagation methods. The conjugate gradient method, which is a learning algorithm for feed-forward neural network, was selected because it adaptively assigns the learning rate in each iteration, as opposed to the gradient descent, which needs to adjust the learning rate manually. The backpropagation was used to calculate the derivative concerning the weights and biases, while the scaled conjugate gradient method [33] was used to update the weight and bias values.

The success of the learning algorithm depends on the value of specific parameters, such as the learning rate. In the Conjugate gradient method, a search is made along with the conjugate gradient direction to determine the learning rate. To reduce the computational complexity resulting from the linear search of the conjugate gradient method, the scaled conjugate gradient algorithm was selected, which avoids such a linear search when adjusting the learning rate. This algorithm is more efficient than the linear search employed by the standard conjugate gradient method. With the scaled conjugate gradient, the line search was avoided using a Levenberg-Marquardt approach [34].

6 Experimental Setup

This section describes the experimental setup for evaluating the proposed PRFADS scheme. Further, the section introduces the datasets and procedures employed for data collection, pre-processing, and attack simulation.

6.1 Traffic Capturing and Normal Traffic Generation

Due to the lack of a ground truth dataset for evaluating the DoS attacks in a Wireless Network, the common practice in this domain is to generate the datasets using real experiments. Real traffic data has been captured from a TP-LINK DG834GT wireless AP with MAC address F81A67DF22B2. The AP is configured with open access wireless LAN, so any wireless station can associate with this access point. Five wireless stations were used to communicate with the access points. One station was used to capture the wireless traffic. The capturing machine is MacBook Pro with OS X EI Caption v10.11.6 operating system. The wireless adapter in the machine was put into the monitoring mode to capture all frame types, including control and management frames. The monitoring mode can sniff the MAC layer traffic on the AP working channel. Unlike the promiscuous mode, the monitoring mode allows the adapter to sniff the traffic in the air without having to associate it with the access point. Wireshark v2.6.5 was used for capturing the network traffic.

The captured data contains as many noises as access points that persist in the neighbourhood of the current access point. The noise was removed to ensure that the data is free of attacks. Since no attack has been launched in the real experiment, the data represents the normal behavior and can be used as a ground truth. However, the data collected from the real scenarios are not enough to train a model that can generalize well. Therefore, computer-generated data have been synthesized to simulate WLAN behavior during normal operation. Statistical techniques have been used to model the captured traffic variables' station and access point behavior. Thus, using the client and access point model, 100 workstations were simulated in the MATLAB environment. The simulation has been carried out based on the ground truth data collected from real scenarios. Fig. 3 shows the conceptual structure of the dataset collection scenario.

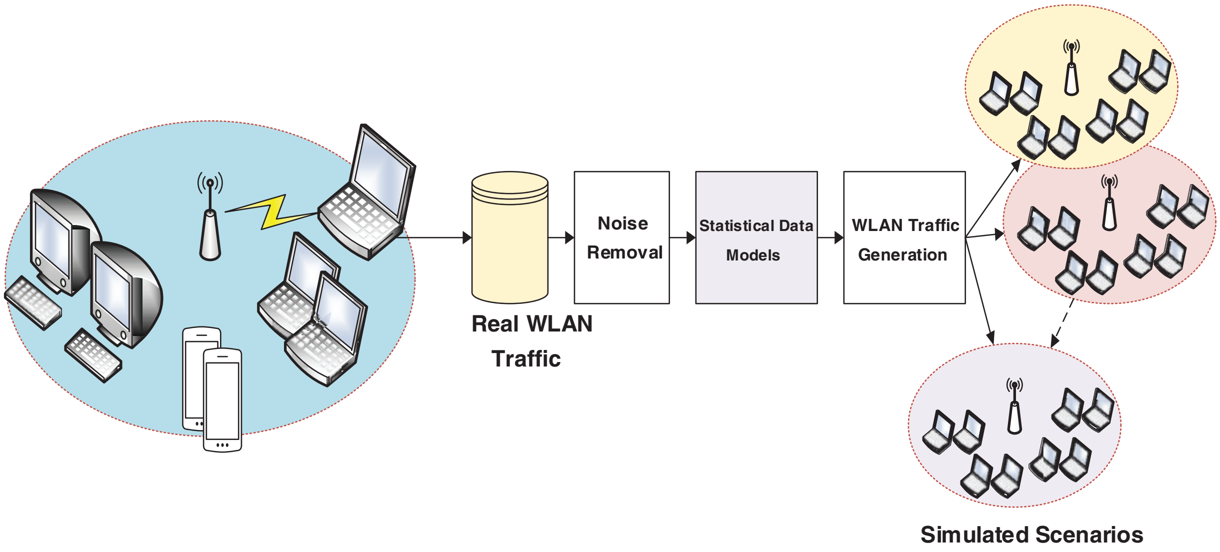

Figure 3: Capturing and simulating WLAN scenarios

6.2 DoS Attack Simulation and Traffic Generation

Datasets that contain DoS attacks and reflect the real ability of the attackers in the WLAN are lacking. Hence, most of the existing WLAN experiments depend on limited attack scenarios that are already implemented in the existing attack tools. However, the attacker can create many sophisticated attack scenarios far from the current implementations. These attacks can easily bypass the security measures put in place by the existing solutions. In this study, more sophisticated attacks have been simulated and their traffic data recorded. Four types of DoS attacks that have been simulated are as follows:

(a) The attacker aims to exhaust the processing resources of the access point by forcing the access point to spend a substantial amount of time processing and reply to the attacker's messages. This attack can be detected through probe request rate/authentication request rate/association request rate. The AP can be protected from this attack through delayed response to the MAC+RSSI Entity for a while.

(b) The attacker aims to exhaust the AP's processing and/or memory resources. Thus, the attacker sends many probe requests from a single MAC and different RSSI Vectors. It changes the RSSI to avoid RSSI blocking-based methods. The message is considered genuine if the received signal strength belongs to the corresponding claimed model. The proper action against such an attack is to drop such messages.

(c) The attacker sends many probe requests from different MAC and single RSSI vectors. If the received signal strength belongs to the corresponding claimed model, the message is genuine; otherwise, the message is dropped.

(d) The attacker sends many probe requests from different MAC and RSSI vectors. If the received signal strength belongs to the corresponding claimed model, the message is genuine; otherwise, the message is dropped.

The AP disassociates the clients that are not active for a pre-defined time threshold determined based on the traffic. The access point disassociates the clients that have similar RSSI values.

6.3 Performance Evaluation

To evaluate the performance of the proposed scheme, the average cross-entropy error (ACE) was used as it gives a more accurate indication of the performance of the training and testing phases than other measurements like to mean squared error (MSE).

It also shows whether the trained classifier has suffered from the problem of overfitting or underfitting based on the behavior of the variance and biases. Although ACE can generally give intuition about the performance of the classifier in both training and testing sets, it cannot give a clear picture of classification accuracy. Therefore, four performance indicators based on classification errors have been used to evaluate the performance. This includes classification accuracy, false-positive rate (FPR), false-negative rate (FNR), and F-score.

6.4 Results Analysis and Discussion

Fig. 4 shows the performance in terms of Cross-Entropy error. The X-axis represents the number of epochs, while Y-axis corresponds to cross-entropy. As shown in the figure, the cross-entropy of training, validation and testing decreases as the number of epochs increases. The best validation performance was 0.26 at epoch 87, while the best training performance was 0.25 at epoch 93. Moreover, the best test performance was 0.26 at epoch 93.

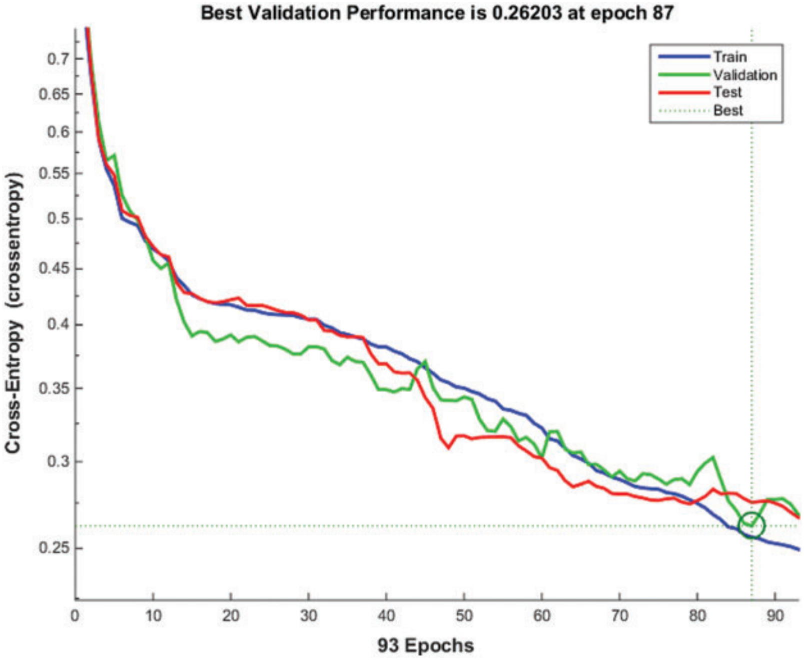

Figure 4: Performance in terms of cross-entropy error

Fig. 5 illustrates the performance in terms of the receiver operating characteristic (ROC) curve. The area under the ROC curve measures how well a parameter can distinguish between two classes of frames (forge/genuine). According to [35], the closer the curve to the upper left corner, the higher the test's overall accuracy. The X-axis represents the false positive rate (FPR), while Y-axis represents the true positive rate (TPR). The ROC was determined for four variants of the datasets, namely, training 5(a), validating 5(b), testing 5(c), and the whole datasets all together 5(d). Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a specific decision threshold. As the results show for all dataset variants, the area under the curve is almost close to the upper left corner of the plots. This indicates the high performance achieved by the proposed scheme.

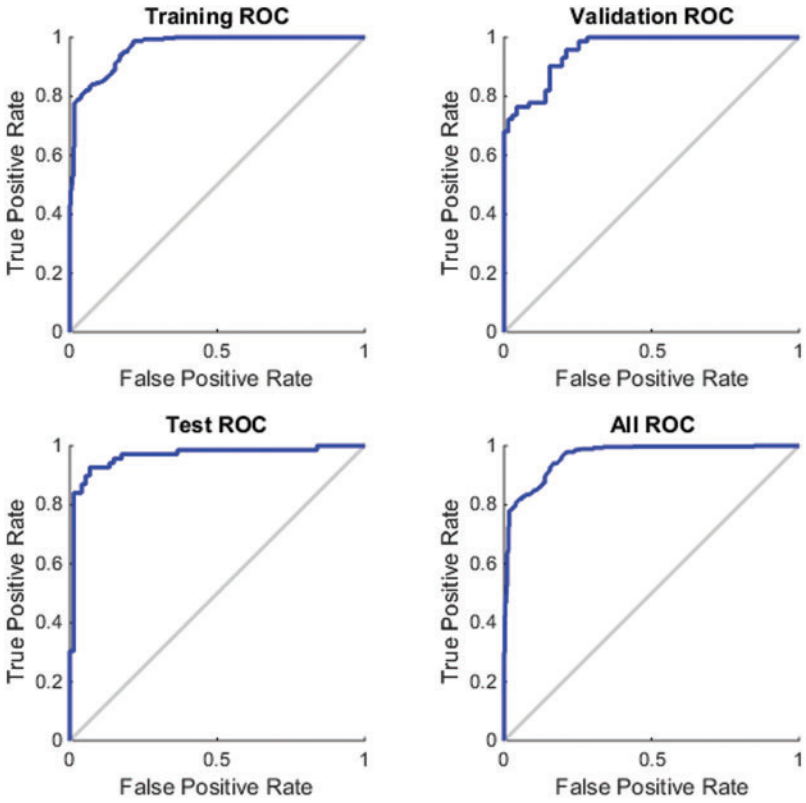

Figure 5: Performance in terms of the receiver operating characteristic (ROC) curve

6.5 *Comparison and Result Analysis*

To show the improvement that the proposed READS scheme has achieved, a comparison with a related scheme, PRFADS, was carried out in terms of the overall accuracy, false-positive rate, false-negative rate, and F-measure. As mentioned earlier, attack scenarios have been simulated to evaluate the proposed scheme's true potential compared to the existing ones. The results of the four-evaluation metrics are listed in Table 2. Figs. 6–9 compare the performance of the proposed and the related PRFADS scheme.

Table 2: Average effectiveness results of the four attack scenarios

Scheme	Attack scenario	Accuracy	FPR	FNR	F SCORE
READS (Proposed technique)	Attack scenario 1	99.26%	1.48%	0.04%	99.29%
	Attack scenario 2	92.46%	6.67%	8.55%	91.81%
	Attack scenario 3	80.59%	7.55%	35.13%	73.27%
	Attack scenario 4	84.11%	9.02%	23.13%	82.27%
Average		**89.11%**	**6.18%**	**16.71%**	**86.66%**
PRFADS	Attack scenario 1	99.51%	1.04%	0.00%	99.53%
	Attack scenario 2	79.43%	3.11%	40.93%	72.61%
	Attack scenario 3	69.86%	8.73%	53.49%	55.14%
	Attack scenario 4	64.08%	6.52%	65.61%	44.76%
Average		**78.22%**	**4.85%**	**40.01%**	**68.01%**

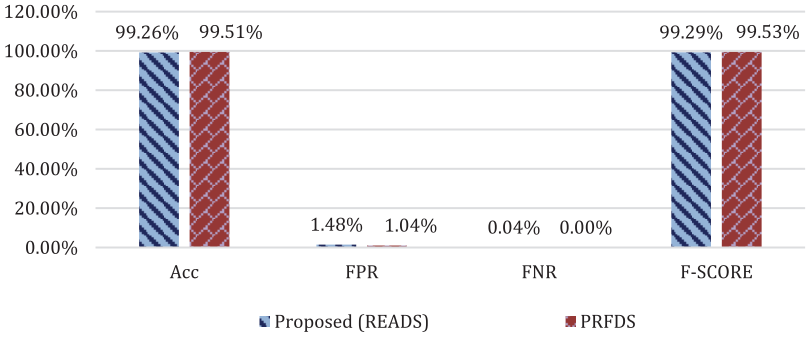

Figure 6: Results of proposed READS scheme (Attack scenario 1)

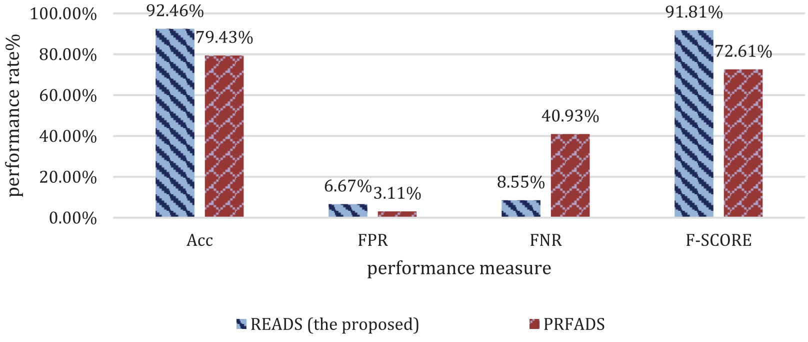

Figure 7: Comparison between READS and PRFADS (Attack scenario 2)

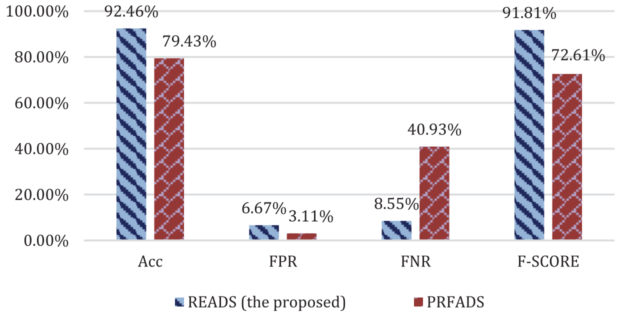

Figure 8: Comparison between READS and PRFADS (Attack scenario 3)

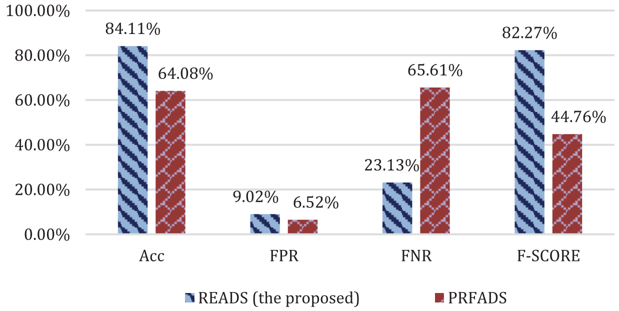

Figure 9: Comparison between READS and PRFADS (Attack scenario 4)

As illustrated in Table 2, the averaged accuracy of 89.11% of the proposed READS scheme outperforms the 78.22% accuracy achieved by the PRFADS scheme. This is because PRFADS considers only simple attack scenarios, in which the representative features of the basic attacks cannot be generalized to different, more sophisticated attack scenarios. In contrast, the proposed scheme has been trained on more expansive data, including advanced attack behavior. In addition, the features represented by the distance between the instance and the cluster centroid allow the classifier to learn the unseen pattern of the attack behavior. In terms of the false positive rate, the proposed scheme achieved 6.18%, slightly higher than the PRFADS scheme. Although there is a 1.33% increment in the false positive rate, the detection accuracy improves by 11.2%. Moreover, the proposed scheme achieved a 16.71% false-negative rate compared to the PRFADS scheme which achieved 40.01%.

The high false-negative rate of the PRFADS scheme is due to the absence of the representation of the advanced attacks during learning the classifier. As such, the PRFADS scheme leads to a misclassification rate of 40.01% of the false message attacks i.e., PRFADS cannot maintain a balanced trade-off between the false positive rate and false-negative rate. This explains the low false-positive rate of the PRFADS scheme compared to the proposed READS scheme. The overall performance of both schemes has been evaluated in terms of F-score, which show that the proposed READS scheme outperforms the PRFADS scheme. The proposed model has an average of 27.3% overall performance improvement over the PRFADS scheme. Figs. 6–9 show the performance comparison between the proposed and the related PRFADS schemes concerning the four attack scenarios.

Fig. 6 compares the performance of proposed READS and PRFADS schemes in detecting basic flooding attacks. As shown in Fig. 6, both the proposed READs and the related PRFADS schemes have similar performance with respect to basic attack scenarios. This is because the false management frames produced by attackers have distinct features compared to the normal management frames coming from benign stations.

Fig. 7 shows a comparison between the proposed READS scheme's performance and the PRFADS scheme's performance in detecting flooding attacks raised by attackers who exploit the vulnerability of the management frames and manipulate the radio signal strength of the messages. Because the attackers start manipulating the RSSI thus detecting, the detection resource which relies on the RSSI as the main distinguishing feature of the attacker is ruined. The attacker creates an illusion to puzzle the RSSI-based solutions, the detection scheme than cannot differentiate between a moving station and the attacker. Thus, such a solution can misclassify many attacks due to the high similarity between the attacker and the mobile stations. Accordingly, most schemes that rely on the RSSI assume that the stations are stationary.

The results illustrated in Fig. 8 clearly show that the PRFADS scheme has failed to misclassify 40.93% of this type's false management frames, leading to poor detection performance. In contrast, compared with the PRFADS scheme, the proposed READS scheme has better accuracy performance and has misclassified only 8.55% of the false management frames.

Fig. 9 demonstrates a comparison between the proposed READS scheme's performance and the PRFADS scheme's performance in detecting flooding attacks raised by masquerading attackers who manipulate their signal strength (Attack Scenario 4). In this scenario, the attackers send too many frames from the spoofed MAC address. As mentioned earlier, the attackers who manipulate the MAC address, which considers the identity of the message, are challenging to be detected due to the illusion between the false and correct management frame. The results illustrated in Fig. 9 show that the PRFADS scheme has misclassified 65.61% of the false management frames. In contrast, the proposed READS scheme has misclassified 23.13% of the false management frames. The reason behind the high misclassification of both schemes is the illusion that the attackers can create when they can copy the behavior of the legitimate station. However, the proposed scheme outperforms the PRFADS scheme in overall detection performance. The F-score achieved by the proposed READS scheme is 82.27% which is higher than that was achieved by the related PRFADS scheme. The proposed READS scheme has improved the overall detection accuracy by 83%.

The results of the four-evaluation metrics using four different attack scenarios listed in Table 2 and illustrated in Figs. 6–9 suggest that the proposed scheme is more effective than the compared scheme in detecting and mitigating the attack. Therefore, these results indicate that the proposed scheme can effectively detect and mitigate false management frames early because it can represent the distinct features of both legitimate and fake messages more accurately.

7 Conclusions

Low detection accuracy in the early stages of the existing resource exhaustion attack detection and mitigation schemes was attributed to the poor feature representation of the attacks during the training phase. This has led to creating an illusion when it comes to distinguishing between false and legitimate management frames. This paper proposes a solution to improve detection accuracy in the early stages of the attack. The proposed scheme consists of four modules: The first module is the features collection module, in which the raw features are collected. In the second module, new features are derived using statistical techniques. The third module focuses on feature representation, through which an unsupervised neural network-based clustering algorithm was trained to categorize the messages according to the behavioral activities of the network participants such that each entity is allocated to a respective cluster. In the fourth module, a supervised machine learning algorithm was used to train a model to differentiate between the distinct features of the clustering output. The aim is to improve the detection performance to mitigate the attacks and prevent their consequences. The results show that the proposed READ scheme outperforms the related scheme. The proposed scheme achieved an average of 27% improvement in terms of the overall detection performance. Although the proposed scheme alleviates the attack impact more effectively than the related work, the scheme is still vulnerable to other attacks related to the management frames. The attackers can exploit the vulnerability of the management frames and send fake de-authentication and disassociation frames to interrupt the communication of the legitimate stations and carry out the DoS attack. Therefore, improving the proposed scheme to mitigate such attacks is essential.

Acknowledgement: The manuscript APC is supported by Universiti Malaysia Sabah, Jalan UMS, 88400, Kota 599 Kinabalu, Malaysia, under the name of grant (UMS No. DFK2005) "Smart Vertical farming Technology for Temperate vegetable cultivation in Sabah: practising smart automation system using IR and AI technology in agriculture 4.0".

Funding Statement: The manuscript APC is supported by the grant name (UMS No. DFK2005) "Smart Vertical farming Technology for Temperate vegetable cultivation in Sabah: practising smart automation system using IR and AI technology in agriculture 4.0".

Conflicts of Interest: The authors declare they have no conflicts of interest to report regarding the present study.

References

[1] K. Bicakci and B. Tavli, "Denial-of-service attacks and countermeasures in IEEE 802.11 wireless networks," *Computer Standards & Interfaces*, vol. 31, no. 5, pp. 931–941, 2009.

[2] A. Mohammed and G. George, "Vulnerabilities and strategies of cybersecurity in smart grid-evaluation and review," in *2022 3rd Int. Conf. on Smart Grid and Renewable Energy (SGRE)*, Doha, Qatar, pp. 1–6, 2022.

[3] W. Li, W. Meng and L. F. Kwok, "Surveying trust-based collaborative intrusion detection: State-of-the-art, challenges and future directions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 280–305, 2021.

[4] E. Chatzoglou, G. Kambourakis and C. Kolias, "How is your Wi-Fi connection today? DoS attacks on WPA3-SAE," *Journal of Information Security and Applications*, vol. 64, pp. 103058, 2022.

[5] R. Nazir, A. A. Laghari, K. Kumar, S. David and M. Ali, "Survey on wireless network security," *Archives of Computational Methods in Engineering*, vol. 29, pp. 1591–1610, 2022.

[6] F. N. Nwebonyi, R. Martins and M. E. Correia, "Reputation based approach for improved fairness and robustness in P2P protocols," *Peer-to-Peer Networking and Applications*, vol. 12, no. 4, pp. 951–968, 2019.

[7] S. M. Cheng and P. Y. Chen, "Ecology-based DoS attack in cognitive radio networks," in *2016 IEEE Security and Privacy Workshops (SPW)*, San Jose, CA, USA, pp. 104–110, 2016.

[8] M. Agarwal, D. Pasumarthi, S. Biswas and S. Nandi, "Machine learning approach for detection of flooding DoS attacks in 802.11 networks and attacker localization," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 6, pp. 1035–1051, 2016.

[9] K. Lounis, S. H. H. Ding and M. Zulkernine, "Cut It: Deauthentication attacks on protected management frames in WPA2 and WPA3," in *Int. Symp. on Foundations and Practice of Security*, vol. 13291, Cham, Switzerland, Springer, pp. 235–252, 2022.

[10] F. H. Ferreira and M. Walton, "World development report 2006: Equity and development," in *World Bank Publications*, 28st ed., vol. 28, Washington D.C, USA: A copublication of The World Bank and Oxford University Press, pp. 1–340, 2005.

[11] S. J. Kim and G. Jeong, "A study of distributed denial of service attack on government infrastructure," *International Journal of Internet, Broadcasting and Communication*, vol. 8, no. 2, pp. 55–65, 2016.

[12] R. Bruce, S. Dynes, H. Brechbuhl, B. Brown, E. Goetz *et al.,* "International policy framework for protecting critical information infrastructure: A discussion paper outlining key policy issues," *The Hague: TNO*, vol. 1, pp. 73, 2005.

[13] C. Kolias, G. Kambourakis, A. Stavrou and S. Gritzalis, "Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 184–208, 2016.

[14] M. S. Hajar, M. O. Al kadri and H. K. Kalutarage, "A survey on wireless body area networks: Architecture, security challenges and research opportunities," *Computers & Security*, vol. 104, pp. 102211, 2021.

[15] E. Chatzoglou, G. Kambourakis, C. Kolias and C. Smiliotopoulos, "Pick quality over quantity: Expert feature selection and data preprocessing for 802.11 intrusion detection systems," *IEEE Access*, vol. 10, pp. 64761–64784, 2022.

[16] A. N. Kadhim and S. B. Sadkhan, "Security threats in wireless network communication-status, challenges, and future trends," in *2021 Int. Conf. on Advanced Computer Applications (ACA)*, Maysan, Iraq, pp. 176–181, 2021.

[17] H. Sounni, E. Najib and L. Fatima, "Distributed denial of service attacks detection using statistical process control in centralized wireless networks," *Journal of Engineering Science and Technology*, vol. 17, no. 2, pp. 1436–1446, 2022.

[18] A. Persia, M. Durairaj and S. Sivagowry, "Study of thwarting DoS attacks by detecting MAC spoof in WLAN infrastructure networks," in *Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE Int. Conf. on. IEEE*, Ramanathapuram, India, pp. 264–268, 2012.

[19] N. Singh, A. Dumka and R. Sharma, "Comparative analysis of various techniques of DDoS attacks for detection & prevention and their impact in MANET," in *Performance Management of Integrated Systems and Its Applications in Software Engineering*, 1st ed., vol. 1, Singapore: Springer Nature, pp. 151–162, 2020.

[20] I. Joseph, P. B. Honnavalli and B. R. Charanraj, "Detection of DoS attacks on Wi-Fi networks using IoT sensors," in *Sustainable Advanced Computing*, vol. 840, Singapore: Springer, pp. 549–558, 2022.

[21] M. Elsabagh, D. Barbará, D. Fleck and A. Stavrou, "On early detection of application-level resource exhaustion and starvation," *Journal of Systems and Software*, vol. 137, pp. 430–447, 2018.

[22] D. N. Ratnayake, H. B. Kazemian and S. A. Yusuf, "Identification of probe request attacks in WLANs using neural networks," *Neural Computing and Applications*, vol. 25, no. 1, pp. 1–14, 2014.

[23] P. Ding, "Central manager: A solution to avoid denial of service attacks for wireless LANs," *International Journal of Network Security*, vol. 4, no. 1, pp. 35–44, 2007.

[24] J. Kaur and P. Sondhi, "Study of thwarting DoS attacks by detecting MAC spoof in WLAN infrastructure networks," *Journal of Positive School Psychology*, vol. 6, no. 3, pp. 5937–5942, 2022.

[25] M. A. C. Aung and K. P. Thant, "Detection and mitigation of wireless link layer attacks," in *2017 IEEE 15th Int. Conf. on Software Engineering Research, Management and Applications (SERA)*, London, UK, pp. 173–178, 2017.

[26] A. E. Abdallah, S. A. Razak and F. A. Ghalib, "Deauthentication and disassociation detection and mitigation scheme using artificial neural network," in *Emerging Trends in Intelligent Computing and Informatics*, 1st ed., vol. 1073, Cham, Switzerland: Springer Nature Switzerland AG, pp. 857–866, 2019.

[27] D. N. Ratnayake, H. B. Kazemian, S. A. Yusuf and A. B. Abdullah, "An intelligent approach to detect probe request attacks in IEEE 802.11 networks," in *Engineering Applications of Neural Networks*, 1st ed., vol. 363, Heidelberg, Germany: Springer, pp. 372–381, 2011.

[28] A. Elhigazi, S. Abd Razak, M. Hamdan, B. Mohammed, I. Abaker *et al.,* "Authentication flooding DOS attack detection and prevention in 802.11," in *2020 IEEE Student Conf. on Research and Development (SCOReD)*, Batu Pahat, Malaysia, pp. 325–329, 2020.

[29] M. Al Mehdhara and N. Ruan, "MSOM: Efficient mechanism for defense against DDoS attacks in VANET," *Wireless Communications and Mobile Computing*, vol. 2021, no. 4, pp. 3771–3778, 2021.

[30] M. H. Ghaseminezhad and A. Karami, “A novel self-organizing map (SOM) neural network for discrete groups of data clustering,” *Applied Soft Computing*, vol. 11, no. 4, pp. 3771–3778, 2011.

[31] W. Bhaya and M. E. Manaa, “Review clustering mechanisms of distributed denial of service attacks,” *Journal of Computer Science*, vol. 10, no. 10, pp. 2037–2046, 2014.

[32] W. Bhaya and M. E. Manaa, “DDoS attack detection approach using an efficient cluster analysis in large data scale,” in *2017 Annual Conf. on New Trends in Information & Communications Technology Applications (NTICT)*, Baghdad, Iraq, pp. 168–173, 2017.

[33] S. Abdulkarim and A. Garko, “Effectiveness of firefly algorithm based neural network in time series forecasting,” *Bayero Journal of Pure and Applied Sciences*, vol. 9, no. 1, pp. 6–10, 2016.

[34] J. J. Moré, “The levenberg-marquardt algorithm: Implementation and theory,” in *Numerical Analysis*, 1st ed., vol. 1, Heidelberg, Germany: Springer, pp. 105–116, 1978.

[35] M. H. Zweig and G. Campbell, “Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine,” *Clinical Chemistry*, vol. 39, no. 4, pp. 561–577, 1993.