JKAU: Comp. IT. Sci., **Vol**. **9** No. 1, pp: 45 – 62 (1441 A.H. / 2020 A.D.)
Doi: 10.4197/Comp. 9-1.4

A Hybrid Intrusion Detection System for 802.11 Networks with Effective Feature Selection

Isaac S. Ikram and **Mohamed A. I. Madkour**

Faculty of Computing and Information Technology, King Abdulaziz University
Jeddah, Saudi Arabia

isaac.sayid@gmail.com

Abstract. The IEEE 802.11i protocol is the current security standard for WLANs. While it has strong security mechanisms such as Advanced Encryption Standard for encrypting and the four-way handshake protocol for authentication, it is still vulnerable to a number of serious attacks such as deauthentication and disassociation flooding. Various intrusion detection techniques are proposed by the research community to detect known and zero-day WLAN attacks. Nevertheless, further efforts are needed to improve the detection performance using a benchmark 802.11 dataset that contains both normal traffic and intrusive traffic of all known attacks. In this paper, we introduce a hybrid real-time network based WLAN intrusion detection system that employs signature and anomaly detection methods. Using signature detection can increase the true positive rate while anomaly detection can detect zero-day attacks. In addition to the signature rules, we considered both C4.5 classifier and Averaged One-Dependence Estimator (AODE) for anomaly detection. The developed system is evaluated in terms of precision and recall, providing three contributions. Firstly a novel technique is developed for effective feature selection based on filtering model and knowledge of WLAN attack footprints. Secondly, it improves classification accuracy, compared with recently published results, and dramatically reduces the classification speed by minimizing the training time and the classification attributes. Thirdly, it offers a high performance real time hybrid WLAN intrusion detection system. A prototype is implemented and tested on 1.7 GHz i5 PC with 12 GB RAM. The experimental results show that the implemented system has a fast learning time of 45 seconds and a high classification performance of 99.6% precision, 98.11% recall, and an overall accuracy of 99.82%.

Keywords: IEEE 802.11; WLAN Security; Intrusion Detection System (IDS); C4.5 Classifier; AODE Classifier; Effective Feature Selection; Information Gain Ratio (IGR).

1. Introduction

Wireless local area networks WLANs, also called Wi-Fi, are today's prevalent wireless network environment for enterprises, small offices and home (SOHO), and smart devices. The reasons may include the economic price, portability and suitability for certain types of business processes. WLAN is released as IEEE 802.11 standard since 1997 [1]. A number of security flaws have been discovered in the earlier WEP standard and the most serious one was cracking the access point encryption key. IEEE 802.11i was ratified to replace WEP with Wi-Fi protected access (WPA2) which provided stronger confidentiality, integrity, and robustness against denial of service (DoS) attacks. WPA2 is a commonly used standard in modern devices nowadays. However, WLANs are still vulnerable to a number of serious security attacks including DoS, impersonation attacks, flooding attacks …etc. [2][3][4][5][6]. Most

of these attacks are due to the unprotected frames available such as management, control and EAP frames. The recently ratified standard 802.11w in 2009 adds cryptographic protection to the management frames which prevented spoofing and helped to address some of the DoS attacks [7]. However, 802.11w requires code update not only on Access Point (AP), but also on client side. Thus, the standard is still incompatible with most of legacy devices. Also, it is still vulnerable to a number of new attacks [8].

To help mitigate the WLAN vulnerabilities, intrusion detection systems (IDS) are needed. Conceptually, a wireless IDS examines the physical and data link layers of the WLAN whereas wired IDS analyzes the Internet protocol layer and above. WLAN intrusion detection system is a software or hardware system that can monitor wireless network traffic and analyze it to identify anomalies and intrusions [9]. Researchers proposed several Wireless IDSs that can mitigate a number of WLAN attacks. However, the detection and classification accuracy of many proposed classification-based IDSs needs to be assessed and improved using datasets with known attacks such as Aegean Wi-Fi Intrusion Dataset (AWID) [2]. Moreover, the performance and speed of an IDS can be further improved by minimizing both the training time of the classification model and the selected attributes of the dataset. With the existing vulnerabilities in WLAN security, developing and implementing enhanced wireless IDS model is definitely a step in the right direction. This research aims to enhance the detection accuracy and speed of the existing network-based WLAN intrusion detection systems. Aegean datasets are used for developing and testing a **h**ybrid real-time network-based **W**LAN **i**ntrusion **d**etection **s**ystem (HWIDS). By effective selection of features from the Aegean datasets and using a hybrid classifier that adopts signature and anomaly detection approaches, the developed HWIDS system achieved better detection accuracy and speed compared with the performance of recently published solutions [10] [11] [12]. The architecture of HWIDS allows using one or more trained models (classifiers) to classify the monitored wireless MAC frames in parallel. HWIDS combines signature and anomaly detection approaches to make a decision.

The contribution of the present work is manifold. Firstly, it chooses the most effective attributes to detect attacks among one-hundred fifty-five attribute vector of the Aegean dataset by using an algorithm that is based on filtering and analyzing attack patterns. Secondly, it improves classification accuracy, compared with recently published results, and dramatically increases the classification speed by minimizing the training time and the classification attributes. Thirdly, it offers a high performance real time hybrid WLAN intrusion detection system. The rest of this paper is structured as follows. A background is given in Section 2 and the related work is reviewed in Section 3. Section 4 presents the design features of the developed system, followed by the prototype implementation and performance evaluation in Section 5. Conclusions and future work are stated in Section 6.

2. Background

2.1 Overview of Intrusion Detection Techniques

In general, intrusion detection systems can be designed in two flavors, host-based (HIDS) and network-based (NIDS) depending on the protected environment; and it is possible to combine both types if needed. The present work concentrates on NIDSs being the obvious choice for detecting attacks in wireless networks. An NIDS provides external approach of protection to a network against intrusions in a timely manner. There are several techniques in the literature about intrusion detection systems,

as shown in [12][13][14][15][16], to mention a few. Regardless of the network type whether it is wired or wireless, any intrusion detection system can be one of two types: anomaly-based IDS or signature-based IDS. Anomaly-based IDS can detect new attacks, known as zero-day attacks, due to its capability to create profiles of normal behavior to distinguish any later deviations. In contrary, signature-based IDS is only capable to detect known attacks which have predefined signatures.

2.2 Anomaly-based Detection

The work in [17] provided a comprehensive taxonomy of anomaly based intrusion detection systems techniques with the advantages and disadvantages of each technique. The four main categories proposed by the authors for anomaly-based IDSs are: Statistical based, data mining based, knowledge-based and machine learning based detection. Statistical based detection is a set of techniques that use statistical properties to create the normal behavior or profile. Markov process is an example of statistical based techniques. Data mining based detection is a good technique to find specific patterns of data or anomalies within large historical datasets. Clustering is a data mining based technique which is used to group events of similar properties. Knowledge based detection technique can be used to detect known attacks and/or anomalies. It depends on rules generated from the knowledge base to detect the attacks. This knowledge base should be updated periodically to cope with newly identified attacks. Machine learning based detection is a type of IDS that improves its detection performance over time. It also depends on previous results, and its major drawback is the high resource consumption.

The present work uses two different classification based intrusion detection techniques for anomaly detection. Firstly, C4.5 classifier developed by Quinlan [18], which is a commonly used technique that has efficient classification accuracy compared to other learning methods. The model generated from this classifier after the learning phase is represented as a decision tree and it classifies testing instances using the rules induced from the decision tree. Secondly, Average One-Dependence Estimator (AODE) which is proposed by Webb *et al*. [19]. AODE classifier weakens the attribute independence assumption of Naïve Bayes classifier by averaging over all models in which all attributes depend upon the class and one other attribute [20]. The experiments proved that AODE is computationally efficient and has lower error rates. Moreover, AODE supports incremental learning directly as shown next in Section 4.

2.3 Signature-based Detection

Signature-based detection techniques create a rule base by analyzing the attack types, and performing the detection based on the obtained rules. Snort-Wireless is a sub-project of the main Snort distribution [21]. It is developed to detect 802.11 intrusions by looking for specific frame patterns, and matches incoming wireless MAC frames to its "wifi" rule engine. The detection rules of Snort-Wireless are defined by the standard format syntax which is composed of the rule header and the rule options. Following is the rule syntax:

<alert>wifi<mac_addr><direction><mac_addr>(<option:value>)

Example of "wifi" Snort rule is: when deauthentication frames are directed to the broadcast address from any client in WLAN, Snort can detect such attacks using the following rule:

alert wifi any -> FF:FF:FF:FF:FF:FF (msg:"example rule";stype:" STYPE_DEAUTH")

Although Snort-Wireless can detect NetStumbler [22] and rogue AP attacks, it has

limited rule options and unable to detect attacks like WEB key cracking. To mitigate these limitations, the work in [23] offered WLAN intrusion prevention system WIPS that extends the rule options to include radio options, and predict the future actions of an attack based on plan recognition models.

2.4 Evaluation Methods of IDS Systems

Several performance measures are available to evaluate the performance and accuracy of IDS systems. To calculate performance measures, we need to compute four main statistical measures as shown in Table 1. Performance measures may include True Positive Rate (TPR), False Positive Rate (FPR), Learning Time, Classification Time, Detection Accuracy, Precision and F-Score. Table 2 shows the purpose and formulas of several performance measures for binary classification [24].

Table 1. Statistics for Computing IDS Performance Measures.

Statistical Measure	Definition
TP	The number of correctly classified instances as attack
TN	The number of correctly classified instances as normal
FP	Instances that were incorrectly classified as attack
FN	Instances that were incorrectly classified as normal

It is interesting to differentiate between the binary classification problems and the multi-classification problems. The formula to calculate these measures for binary classification problems are different than multi-classification problems. In the binary classifier, there are only two output classes and the class of interest is considered the positive class. For an IDS classifier, the "attack" is the positive class and the "normal" is the negative. In multi-classification IDS problems, in addition to the normal class there can be several classes to differentiate between different attack types. We can easily define positive and negative examples in binary classification problems because there are only two non-overlapping classes. While in multi-classification problems, there are more than two non-overlapping classes depending on the number of considered attack types. The interested reader can find the related formulas for multi-class classification performance measures in [25].

2.5 Aegean Wireless Intrusion Dataset

Kolias *et al*. [2] introduced new 802.11 dataset called Aegean WiFi Intrusion Dataset (AWID). It is a publicly available collection of sets of data in easily distributed format, which contain real traces of both normal and intrusive IEEE 802.11 traffic. The data is gathered for this task from a realistic and physical SOHO lab. The lab consists of several wireless supported devices with single AP, secured using WEP, and supports 54MBps transfer rate. For sniffing wireless traffic, dedicated device was running in monitor mode. Attacker device was outside the lab boundaries and used different tools to conduct the attacks. Also, they categorized and evaluated almost all known wireless 802.11 network attacks. Additionally, they have generated reduced version of the dataset for quick experiments. Finally, the authors evaluated their datasets and compared them using different machine learning techniques. However, the procedure used for feature selection was not defined clearly, and the obtained results showed low classification performance of flooding and impersonation attacks.

In general, there are two types of AWID datasets: AWID-CLS and AWID-ATK. Frames in the former are classified into four classes based on attack pattern and includes normal, flooding, impersonation and injection, while frames of the latter are classified into specific attack names. Each dataset type comes with reduced and full subset (specifically, AWID-CLS-R, AWID-CLS-F). Each one of the AWID

datasets comprises a training set and a testing set with names suffixed by -Trn and -Tst respectively. This facilitates building and testing the classification models.

Table 2. Performance Measures for Binary Classification.

Measure	Formula for binary classification	Evaluation Purpose
Accuracy	$\frac{tp+tn}{tp+tn+fp+fn}$	Classifier's overall effectiveness
Precision	$p=\frac{tp}{tp+fp}$	How accurate the detected positive instances are
Recall	$r=\frac{tp}{tp+fn}$	How complete the true positive instances are
F-score	$\frac{2pr}{p+r}$	Balance between precision and recall
AUC	$\frac{1}{2}(\frac{tp}{tp+fn}+\frac{tn}{tn+fp})$	Classifier's ability to avoid misclassification
Error Rate	$\frac{fp+fn}{tp+tn+fp+fn}$	Classifier's overall error
False Positive Rate (FPR)	$FPR=\frac{fp}{fp+tn}$	How accurate the detected negative instances are
Learning time	-	The total time taken to build a classifier model (profile), using given training set
Classification time	-	The total time taken to test a classifier model (profile) on the given testing set

2.6 WLAN Attacks

Wireless local area networks are vulnerable to various types of attacks. Table 3 lists the definition of the most serious attacks and the impact of each attack on confidentiality, privacy and availability [2]. Launching such attacks has become nowadays affordable even for non-experts for several reasons. WLAN equipment has software-changeable MAC address which will permit the equipment owner to spoof any MAC address. Also, ordinary persons often misconfigure WLAN settings, exposing the network to attacks even if they use the latest WLAN IEEE 802.11w protocol. Finally, the broadcast nature of WLAN expands its access range to exceed organization boundaries unlike wired networks which are confined to the physical network boundaries. Consequently, nearby adversaries can easily attempt to attack WLANs even if they are outside an enterprise building. Aegean AWID dataset can be considered as a reliable testbed for IDS evaluations and experiments in wireless networks because it contains traces of almost all attacks mentioned in Table 3. The AWID data set considers the following three classes of attack, with several attack names in each class.

- **Flooding attacks**: Attempt to transmit large number of frames within short time intervals to cause denial of service attack (DoS).

- **Impersonation attacks**: Affect the privacy of wireless networks by attempting to steal WEP key or introducing new AP with identical information as valid networks.

- **Injection attacks**: Help to crack the network key by transmitting small data frames.

Table 3. WLAN Attacks Definition [2].

Attack Name	Impact	Definition
FMS	Confidentiality	Repeatedly collect weak IVs condition to crack all bytes of the WEP key
KoreK Family of Attacks	Confidentiality	Cracking the WEP key similarly to FMS but based on statistical observation
PTW	Confidentiality	Cracking the WEP key like FMS but requires less IVs which makes it more efficient
ARP Injection	Confidentiality	Tools cracking the WEP key require this step repeatedly to work
Dictionary	Confidentiality	Brute force method to crack WPA/WPA2 keys at first place.
ChopChop	Confidentiality	Allows an attacker to retrieve the last x bytes of both the WEP keystream and the corresponding plaintext of a packet without knowing the Key.
Fragmentation	Confidentiality	Revealing a large part of the keystream by sending very less messages which make it more efficient than the ChopChop
Caffe Latte	Confidentiality	Cracking the WEP key by taking the ESSID of an AP that has been probed by victim client, even if both client and attacker are not in the AP range.
Hirte	Confidentiality	Cracking the WEP key using a client independently of AP similarly to Caffe Latte.

Deauthentication	Availability	After learning the MAC addresses of both client and AP, the attacker sends de-authentication messages to either client or AP on behalf to cause DoS.
Disassociation	Availability	It has the same effect as de-authentication but a disconnected client needs longer time to re-connect to AP.
Deauthenticati on Broadcast	Availability	The attacker uses broadcast address as a target of de-authentication frames. This will cause all clients to disconnect and de-authenticate
Disassociation Broadcast	Availability	The attacker uses broadcast address as a target of dis-association frames. This will cause all clients to dis-associate
Block ACK flood	Availability	This attack causes AP to drop all network packets by using the Add Block Acknowledgment (ADDBA) feature presented in 802.11n.
Authentication Request Flooding	Availability	This attack stops disconnected clients from joining the network
Fake Power Saving	Availability	The attacker tricks an AP by sending spoofed null data frames repeatedly on behalf of a specific STA, forcing AP to drop buffered frames.
CTS Flooding	Availability	Attacker keeps sending CTS frames, causing annoyance to the rest of the STAs in the form of message postponing.
RTS Flooding	Availability	Attacker keeps sending RTS frames with large transmission duration, forcing the rest of the STAs to stop transmitting.
Probe Request Flooding	Availability	AP is required to reply to Probe Requests with Probe Responses. Attacker misuse this feature by transmitting fake Probe Request packets constantly, thus AP will struggle to respond to legitimate requests.
Probe Response Flooding	Availability	Attacker monitors for probe request messages coming from valid clients and then responds with stream of fake probe responses to the STAs causing annoyance.
Beacon Flooding	Availability	Attacker transmits a constant flood of fake beacons that advertise false ESSIDs.
Honeypot	Privacy	Honeypot is an AP created by adversary with attracting ESSID to fool users into connecting to it.
Evil Twin	Privacy	Evil Twin is a Honeypot AP created by adversary with known ESSID to fool users into connecting to it instead of the valid one
Rogue Access Point:	Privacy	Unauthorized AP in the home or corporate area created by an insider to violate the policies defined.
Amok	Availability	Attacker sends a constant stream of de-authentication and disassociation frames to the network to cause DoS.

2.7 Feature Selection Methods

The presence of noise in data may affect the performance of classification algorithms in terms of detection accuracy and learning time. Noise may result due to attribute redundancy and irrelevant attributes. Feature selection plays an important role in developing intrusion detection models [26][27]. It includes an automatic process of selecting the most relevant attributes in model construction. As a result, feature selection increases the detection accuracy and shortens the computation times of learning and testing by selecting the effective feature set only [28]. A feature selection algorithm consists of an attribute evaluator (a mining algorithm) to measure the worth of a set of attributes, and a searching technique which finds different subsets from the attribute vector. There are three feature selection methods categorized based on the attribute evaluator algorithm: Filters, wrappers and embedded methods [29]. Filters select and rank attributes based on their contribution to the predict class using independent measure for evaluation regardless of the classifier. Thus, filters are not optimized to a given classifier but they are computationally efficient and immune to overfitting. Examples include the information gain ratio and correlation coefficient scores. Wrappers evaluate different combinations of attribute subsets according to a given classifier, and choose the subset of attributes with highest detection accuracy [30]. There are efficient searching techniques used to find attribute subsets such as best-first search and genetic search algorithms. Although wrappers select efficient attributes for a particular classifier, they require high amount of computation time. Embedded methods are similar to the wrapper methods except that they select optimal

variables during the learning phase unlike the aforementioned methods which work independently. Hence, it is more efficient than wrapper methods. In the present work we proposed a novel algorithm for feature selection based on hybrid scheme that merges filtering and wrapper methods.

3. Related Work

This section reviews the data mining based intrusion detection techniques for WLAN networks only and the methods used to select the optimal feature set. Guennoun *et al.* [31] utilized a hybrid approach of filter and wrapper methods (HFW) to select the optimal features for determining WLAN intrusions. The dataset involved normal frames and five types of intrusion frames. The information gain ratio IGR is used as an independent measure to rank the features, and k-means classifier is used to conclude eight optimal features from the IGR ranked features. The experimental results showed that the accuracy of k-means classifier used for intrusion detection is increased using eight optimal feature set and the learning time is reduced. However, the proposed system needs further tests using different types of attacks and classifiers. Moreover, detection accuracy and learning time was not compared before and after selecting the feature set. El-Khatib [32] reported the results of the previous work using three types of Artificial Neural Networks (ANNs). The same eight optimal features set are used. The experimental results showed that the accuracy of the three artificial neural networks classifiers is increased by an average of 15% using the optimal feature set. In addition, the learning time is reduced by an average of 66% for the three classifiers. However, further tests are necessary using more types of attacks such as ARP injection and beacon flooding. Moreover, the detection accuracy was displayed in the experimental results only without true positive or false negative rates.

Danziger and de Lima Neto [10] proposed a hybrid WLAN intrusion detection system using multi-agent system (MAS) which included danger theory for signature detection, and Naïve Bayes classifier for anomaly intrusion detection. The signature detection represented in basic agent detects the attack signs first. If an unknown intrusion takes place, the intermediary agent embedded with Naïve Bayes classifier will try to identify it. The results of experiments showed that the efficiency for Naïve Bayes classifier was 83.9%. However, the proposed system has been tested on a limited number of attacks.

A hybrid technique, called G-LDA process, integrating genetic algorithm and Latent Dirichlet Allocation to identify the anomalies in network traffic is proposed by Kasliwal *et al*. [11]. The attributes selected for intrusion detection fulfill two conditions: they have different mode values for the anomaly and normal packets, and their mean is close to their mode value. G-LDA needs more efforts to reduce its high false positive rate. Usha and Kavitha [12] proposed normalized gain based WIDS, called NMI, which ranks the features first using normalized gain measure and then selects the optimal features using particle swarm optimization (PSO). PSO is also used to cluster unlabeled attacks under classes based on the optimal features. NMI utilizes support vector machine (SVM) classifier to detect WLAN intrusions. SVM classifier can identify the attack classes by creating margins for each feature. NMI achieved higher detection accuracy with 99.25% compared to HFW and G-LDA which achieved 98.5% and 98.1%, respectively. The learning time of NMI decreases surprisingly when the training data size is increased. When the training data size is high, NMI can measure its parameters accurately to select optimal features for learning. NMI and HFW takes about 1.6h to learn the whole dataset, while G-LDA required less learning time of 1.42h only because the underlying metrics include estimating the mean

and mode values which are light processes. The classification time increases proportional to the testing data size for NMI, HFW and G-LDA. For example, NMI increases the classification time by almost 1.25h when the testing data set changes from 280.7MB to 886MB. Classification time of G-LDA is less than NMI and HFW by 41.6% and 46.7%, respectively, because of the light underlying processes. The training and classification time of the aforementioned IDS techniques make them unpractical with common PCs. Moreover, having a relatively high classification time would miss some attacks when the network traffic is high.

4. HWIDS Design

This section presents the salient design features of a hybrid WLAN intrusion detection system (HWIDS) that employs misuse and anomaly detection methods using an effective feature set. This includes the systems architecture, the developed algorithm for the effective features selection, and the hybrid classification technique.

4.1 HWIDS Architecture

HWIDS consists of typical IDS modules that model the intrusion detection system. Figure 1 depicts an adapted form of generalized IDS model showing the proposed HWIDS [9][33]. The process of intrusion detection starts by matching the defined signature patterns with captured frames first. If there is no matching signature, the anomaly detection module will take over to classify the incoming frames. This module comprises a set of parallel machine learning classifiers; currently AODE and C4.5. As explained next, in the anomaly detection module, each classifier is considered best for a unique subset of the attack types and its decision will be considered for that subset. If an attack is detected, related packet information is logged by active processing module. There are eight main modules in HWIDS as stated next.

4.1.1 Monitored Entity

This represents the data source for IDS. HWIDS can monitor wireless network activity of one or more AP devices in the range.

4.1.2 Audit Collection

HWIDS detects and captures the MAC layer information of WLAN frames such as Source Address, Type and Subtype of the frame. Additionally, physical layer information is collected such as channel frequency and Signal Strength.

4.1.3 Audit Storage and Available Datasets

Its purpose is to store captured WLAN packets permanently as a raw data for later referencing, and keeping a temporal copy of the packets that are waiting processing. There are defined policies in the configuration data module to control the volume of data before exceeding the local storage limits. For example: when to remove older audit files. In addition to real-time capturing, HWIDS evaluates offline datasets.

4.1.4 Anomaly Detection

Anomaly detection and signature detection are the core modules in HWIDS. In anomaly detection, HWIDS first preprocesses captured frames and then selects the effective feature set. Next, a set of parallel classifiers (currently two) analyze the preprocessed frames to detect deviations from normal behavior. Some classifiers are better for detecting specific types of attack. In other words, they may have higher True Positive rate and lower False Positive rate with specific attack class. Thus, we designed hybrid anomaly classifier algorithm which assigns priority number to the attack classes of each classifier, and the classifier with highest priority is the predict class.

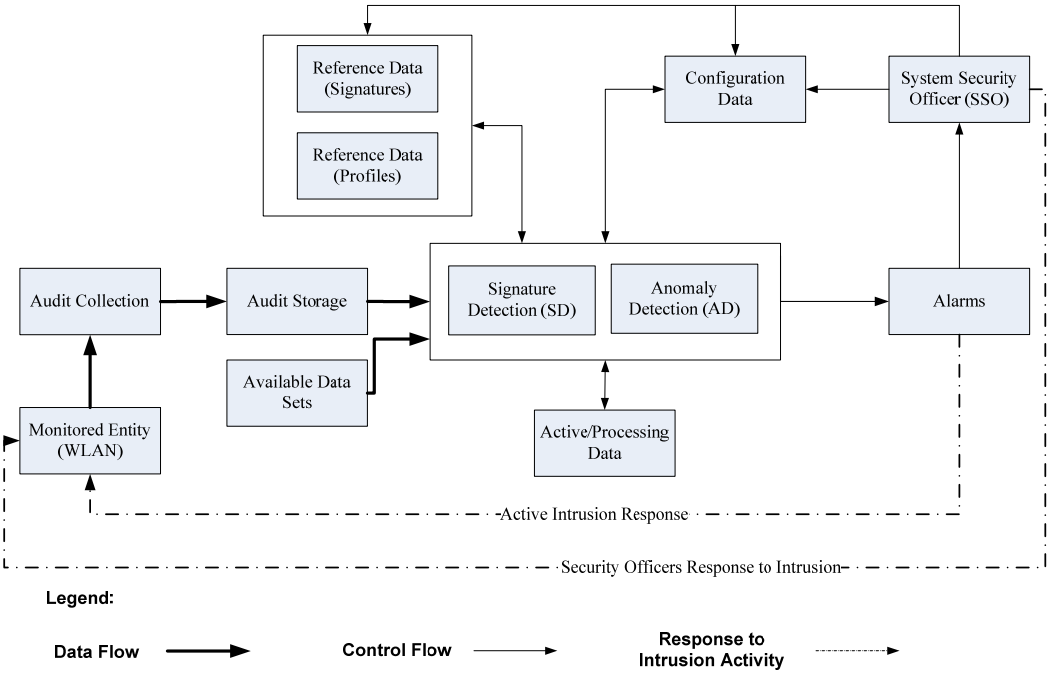

Fig. 1. The proposed HWIDS model.

4.1.5 Signature Detection

HWIDS incorporates a rule-based system to detect known attacks by matching the incoming frames to the known attack signatures. We argue that using the signature detection can increase the true positive rate because it can check statistically illegal behavior over multiple frames. A signature in HWIDS comprised of rule header and rule options, like Snort-Wireless. The rule header includes signature name and description, targeted attack class, signature priority to solve conflicts when more than one signature is matched, matching frequency which indicates how many frames must match within a window before the signature is triggered, frame type and subtype to scan. The rule options may include any MAC or radio feature. Following is the rule syntax**:**

{name, desc, className, priority, freq, frameTypeSubtype, [{;option:value ...}]}

When beacon frames with sequence field equals to zero are detected at least three times within a window of ten frames, HWIDS considers these frames as beacon flooding attack using the following rule:

{"beacon flooding", "beacon flooding frames", "flooding", 1, 3, 0x0C, {"wlan.seq":0}}

4.1.6 Reference Data

This module stores both classifier profiles and known attack signatures. The profiles can be updated later with new training data or replaced with updated profiles if the corresponding classifier doesn't support incremental learning.

4.1.7 Configuration Data

This is the control panel of HWIDS. System Security Officer (SSO) can configure the settings of other modules through this sensitive module. Example of configurations include where to collect the raw captures, when to remove old captured files, what actions

performed when an intrusion is detected, what signatures should be enabled, etc.

4.1.8 Active / Processing Data

This module is responsible for storing intermediate results of the HWIDS such as information about detected attacks and whether detected through signature approach or anomaly approach, etc. Moreover, this module helps to draw statistics about the intrusions.

4.1.9 Alarms

This module controls the outputs of HWIDS by either taking automated actions to the intrusions or informing the SSO.

4.2 Effective Feature Selection

HWIDS preprocess the captured PCAP files or AWID dataset records. Preprocessing includes conversion of 48-bit MAC addresses to integer values, conversion of EUI-64 addresses to integer values and conversion of any hexadecimal values to long numbers. The preprocessing of data is necessary because several machine-learning systems can only process numbers which makes it computationally efficient. To select the most effective features in AWID dataset, we proposed a novel algorithm for feature selection based on hybrid scheme that merges filtering and wrapper methods. The algorithm consists of three main processes. Firstly, it inspects the training data and finds the redundant attributes and attributes with constant values. Secondly, univariate filtering method is employed with information gain ratio (IGR) as a measurement of attribute contribution to the class. Thus, the algorithm produces ranked feature subset and excludes features with IGR value less than the defined threshold value ε. A threshold value of 0.05 was found empirically in our experiments to give the highest performance. We selected the filtering method because AWID datasets have large number of attributes and the filtering methods are computationally efficient as explained previously in Subsection 2.5. Moreover, we selected IGR as a measure because it is not biased toward features with distinct values [34]. However, to overcome the disadvantages of the filtering methods, our selection algorithm adds two important steps: The first step is to include features from WLAN Knowledge Subset (WKS) to the ranked feature subset, regardless of the feature IGR. To construct WKS, a subset of significant features is selected after examining the footprints of WLAN attacks [2][3][4][5][6]. Table 4 shows WKS features. The result of this step is the union of WKS features and the ranked feature subset. The second and final step is using wrapper methods to evaluate different combinations of feature subsets using AODE classifier, and choose the subset of attributes with highest detection accuracy.

The algorithm developed for feature selection is shown in Fig. 2. This algorithm begins with empty sets of effective feature S_{best} and ranked features S_{IGR}, and then continues to rank features from the training set D. For each iteration, the algorithm skips the current feature D_i if it is redundant in S_{IGR}. Also, if the feature has one distinct value or its IGR is below the defined threshold value ε, it will discard the feature. The ranking of features iterates over training set until the last feature of D is reached. The algorithm takes the union of the ranked feature set S_{IGR} and significant WLAN feature set S_{WKS} to produce initial effective subset S_{best}. Lastly, the algorithm starts the wrapper selection method by reducing features from the subset S_{best} sequentially. For each iteration, the overall accuracy of the new subset Acc_{new} is computed and compared with the accuracy of the subset before reducing the current feature Acc_{best}. If the new accuracy is dropped, then the reduced feature is added back to effective feature set S_{best_tmp}. The process continues until last feature in S_{best}. The result of applying effective feature selection algorithm on AWID

dataset was an effective feature set with eighteen attributes including the class as listed in Table 5. To the best of our knowledge, these effective features play a crucial role to identify WLAN attacks.

5. Performance Evaluation

The anomaly detection classifiers are developed using the AWID datasets which contain recent WLAN attacks. As shown in Table 6 there are three attack groups, namely flooding, impersonation, and injection. More description about the attacks was given in the background. In this research, we considered the reduced dataset with four classes because it is smaller in size, more efficient and suitable for experiments. The training version of the dataset (AWID-CLS-R-Trn) has 1,795,575 rows, among them 90.96% are normal frames and the remaining 9.04% are intrusive frames. The training set is used during the learning phase of a classifier to construct the model, while the testing set (AWID-CLS-R-Tst) is used to evaluate the classifier's accuracy.

5.1 Implementation

We have implemented HWIDS modules in Java. HWIDS integrates libraries of Waikato Environment for Knowledge Analysis (WEKA) [35] with Tshark software [36]. WEKA framework is a collection of machine learning algorithms to perform different data mining tasks. We used WEKA libraries to develop the anomaly detection module and effective feature selection algorithm. Moreover, we used pure java object-orientation to develop the rule-based signature detection module. Tshark was mainly used to capture WLAN frames. We conducted our experiments on Windows 10 64-bit operating system with 1.7 GHz i5 CPU, 12 GB RAM.

Table 4. WKS feature set.

wlan.fc.type_subtype	wlan.da
wlan_mgt.fixed.reason_code	wlan.ra
wlan_mgt.fixed.beacon	wlan.fc.ds
radiotap.dbm_antsignal	frame.len
frame.time_delta	wlan_mgt.fixed.listen_ival
wlan.seq	wlan_mgt.fixed.timestamp
wlan.fc.retry	wlan.wep.iv

Table 5. Effective feature set.

frame.len	wlan.da
radiotap.datarate	wlan.sa
radiotap.dbm_antsignal	wlan.frag
wlan.fc.type_subtype	wlan.seq
wlan.fc.ds	wlan_mgt.fixed.timestamp
wlan.fc.frag	wlan_mgt.fixed.beacon
wlan.fc.retry	wlan_mgt.fixed.reason_code
wlan.fc.pwrmgt	wlan.wep.iv
wlan.duration	class

Table 6. Attack Distribution over reduced AWID datasets.

Instance Class	Attack Name	Training Dataset (Instances)	Testing Dataset (Instances)
flooding	amok	31,180	477
	authentication_request	3,500	0
	beacon	1,799	599
	cts	0	1,759
	deauthentication	10,447	4,445
	disassociation	0	84
	power_saving	0	165
	probe_request	0	369
	probe_response	1,558	0
flooding	rts	0	199
impersonation	cafe_latte	45,889	379
	evil_twin	2,633	611
	hirte	0	19,089
injection	arp	64,609	13,644
	chop_chop	0	2,871
	fragmentation	770	167
normal	normal	1,633,190	530,785

Effective Feature Selection Algorithm

Input: $D\{F_0, F_1, ... ,F_{n-1}\}$ // training dataset with n number of features

$ś0$ // threshold by which features are discarded

M // a Machine Learning model

S_{WKS} // significant WLAN features set

Output:

```
begin
    initialize: S_best = {};          // initialize effective features set
    S_IGR = {};    // initialize the set of ranked features with IGR ≥ ś
    for i=1 to n do
        for each feature f ∈ S_IGR do    // discard redundant features
            If redundant(D_i , f) then
                continue;
            end
        If not(distinct(D_i)) OR IGR(D_i) < ś then
            continue ; // discard constant features or below threshold
        end
        S_IGR = S_IGR ∪ D_i ;
    end for
    sort(S_IGR);          // sort features in descending order based on IGR
    S_best = S_IGR ∪ S_WKS;
    Acc_best = accuracy(S_best , M) ;  //wrapper method feature selection
    S_best_tmp = S_best ;
    for i=1 to n do
        S_best_tmp = S_best_tmp - S_best[i] ;
        Acc_new = accuracy(S_best_tmp, M) ;
        If Acc_new ≥ Acc_best then
            Acc_best = Acc_new ;
        else
            S_best_tmp = S_best_tmp ∪ S_best[i] ; // dropped accuracy feature
        end
    end for
            S_best = S_best_tmp ;
    return: S_best
end
```

Fig. 2. Effective feature selection algorithm.

5.2 Experimental Results

We conducted different experiments on Aegean data sets using the proposed HWIDS. We selected binary performance measurements to evaluate HWIDS in the experiments by considering instances with all attack classes (impersonation, injection and flooding) as positive examples, and normal instances as negative examples. Figures 3 and 4 illustrate the learning time and classification time of HWIDS when tested using effective feature set and full feature set. HWIDS using AODE with effective features learns the dataset in 45 seconds, which is 10 times faster than the AODE with full feature set (90.02% reduction). Moreover, HWIDS using C4.5 with effective features learns the dataset in 160 seconds, which is 8.66 times faster than the full feature set (88.46% reduction). Additionally, HWIDS with effective features decreases the testing (classification) time. For example, C4.5 with effective features requires less than a second to test 575,643 instances. We deduce that the effective feature selection algorithm increased the computing efficiency of learning time and classification time of both classifiers.

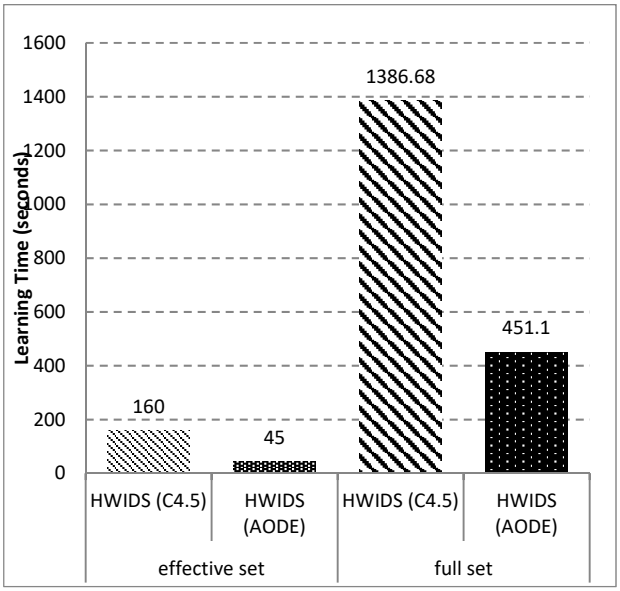

Fig. 3. Learning time versus effective and full feature sets.

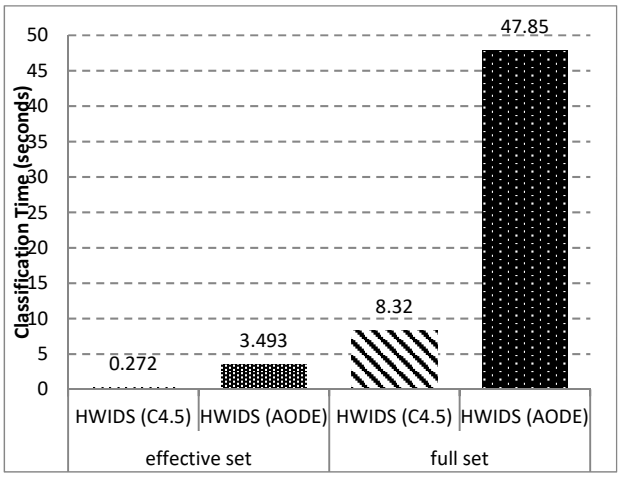

Fig. 3. Classification time versus effective and full feature sets.

5.2.1 Preliminary results

Figures 5 and 6 compare the learning time and classification time of HWIDS with the recent G-LDA [11] and NMI [12] classifiers, as they have been tested on the same AWID datasets. We notice that HWIDS with AODE learns the whole training set of 1,795,575 instances in less than a minute, which is faster than the other two classifiers. The reason is that AODE has a linear computational complexity with respect to the number of training instances and therefore can efficiently process large numbers of training instances. Moreover, the number of features affects greatly the training and classification times of AODE according to the following order of computational complexity [20]:

- $O(tn^2)$ is the training time complexity, where t is the number of training examples, and n is the number of features.
- $O(kn^2)$ is the classification time complexity, where k is the number of classes.

Figure 7 and 8 depict the recall, precision, and FPR results for each IDS system. NMI and G-LDA achieve the highest recall with 98.75% and 97.60%, respectively, at the cost of FPR which is 1% for NMI and 14% for G-LDA. In other words, NMI and G-LDA can detect most of WLAN attacks at the cost of classifying 1% and 14% of normal frames as intrusions. Conversely, C4.5 and AODE has the highest precision among others with 99.84% and 99.80%, respectively. This means that HWIDS can accurately detect intrusions with low false positives. The initial testing of HWIDS showed lower recall with 52.88% for C4.5 and 93.37% for AODE. However, this deficiency is improved as shown next. Figure 9 depicts the F-score for HWIDS (C4.5), HWIDS (AODE), G-LDA and NMI. The F-score combines both precision and recall in a single measure. NMI has a high F-score of 99% due to the high recall and precision values, followed by HWIDS (AODE) with F-score lower than NMI by 2.52%.

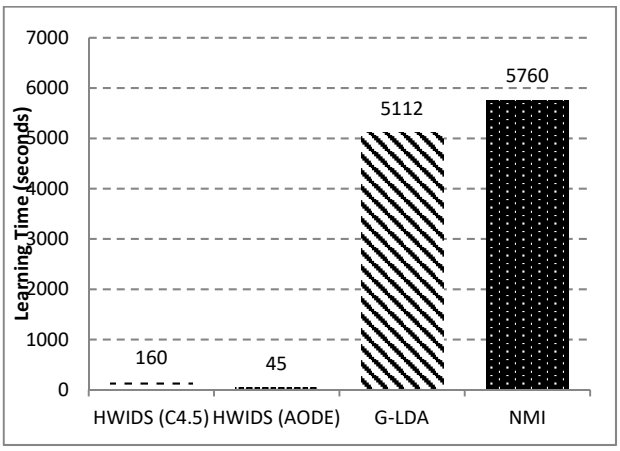

Fig. 5. Learning time versus WLAN IDSs.

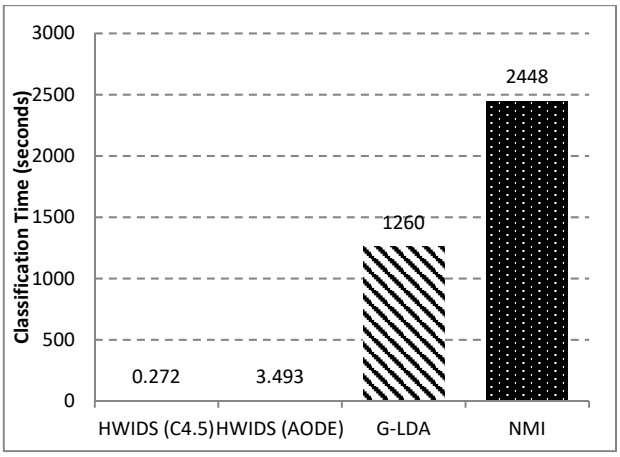

Fig. 6. Classification time versus WLAN IDSs.

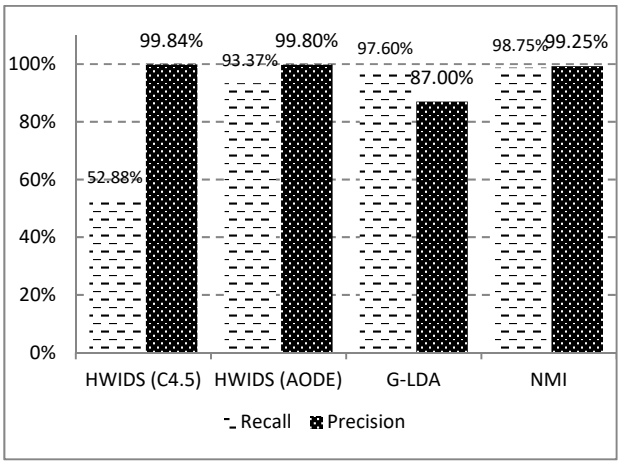

Fig. 7. Recall and precision versus WLAN IDSs.

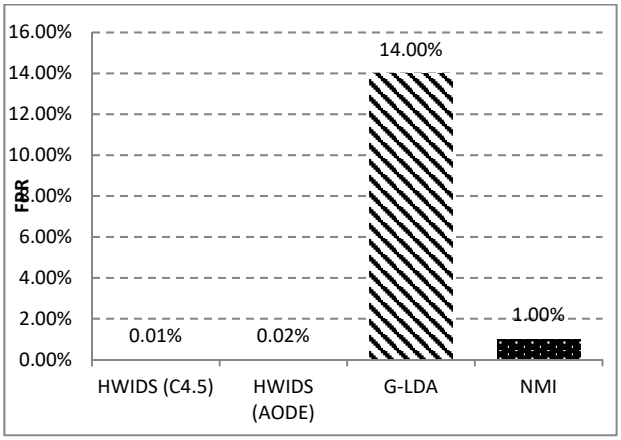

Fig. 8. FPR versus WLAN IDSs.

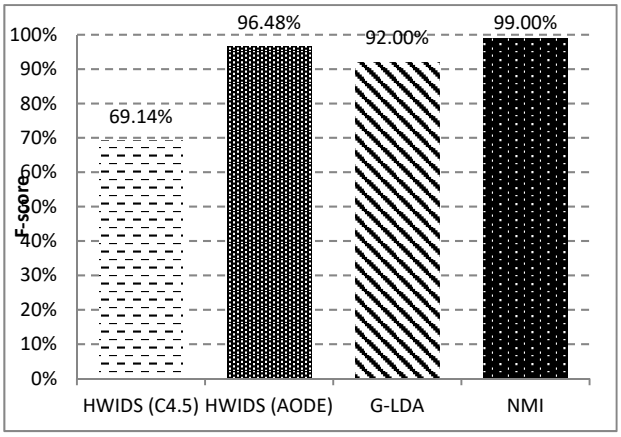

Fig. 9. F-score versus WLAN IDSs (before HWIDS improvement).

Figure 10 compares the detection accuracy of HWIDS, G-LDA and NMI. HWIDS using AODE achieves the highest detection accuracy with 99.47% due to the effective feature selection algorithm which employs wrapper methods to evaluate the best subset of features. NMI and G-LDA come next with overall accuracy of 99.25% and 98.1%, respectively. We conclude from the previous results that AODE is the appropriate classifier to be embedded in the anomaly detection module of HWIDS.

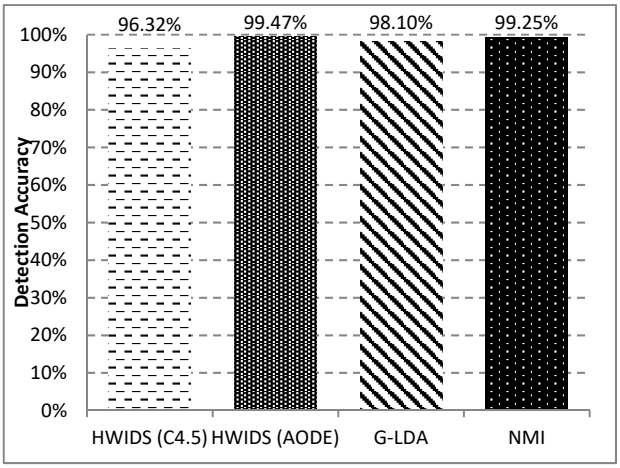

Fig. 10. Accuracy versus WLAN IDSs (before HWIDS improvement).

5.2.2 Improved results

Despite the satisfactory results of HWIDS using AODE with effective features, we need further efforts to improve its recall value. For this purpose, we analyzed the confusion matrix resulted from testing HWIDS with AODE and we found that AODE misclassified some types of flooding attacks that were absent in the training set. These attacks include CTS, Disassociation, Power_saving, Probe_request and RTS. Moreover, flooding attacks have footprints like normal frames when they are tested independently. Therefore, we studied CTS and Probe_request behavior as examples and we developed corresponding signature rules to overcome these attacks.

- The CTS signature is: when CTS control frames with duration field equals to zero are detected at least 10 times within a window of 20

frames, HWIDS considers these frames as CTS flooding attack using the following rule:

{"CTS flooding", "CTS flooding frames", "flooding", 1, 10, 0x1C, {"wlan.duration ":0}}

- The Probe_request signature is: when Probe_request management frames with sequence field equals 10, fragmentation field equals 10 and frame length field equals 75 are detected, HWIDS considers these frames as Probe_request flooding attack using the following rule:

{"Probe_request flooding", "Probe_request flooding frames", "flooding", 2, 1, 0x4, {"wlan.frag":10, {"wlan.seq":10, {"frame.len ":75}}

HWIDS (AODE) with previously defined signatures could detect almost all CTS and Probe_request flooding attacks successfully. Figure 11 compares the performance measures of HWIDS and NMI after improving HWIDS signatures. We notice that HWIDS has higher precision and detection accuracy of 0.35% and 0.57% more than NMI. Also, HWIDS has almost the same recall and F-score values as NMI with much faster learning time and classification time. It is necessary to note that the signature detection effect on classification time of HWIDS was almost negligible. The experimental results of activating the signature rules showed that the classification time is increased by 54 milliseconds only. This is because a full signature matching is done only for specific frame types excluding other frames; namely CTS control frames and probe request frames.

6. Conclusions and Future Work

In this research, we made three valuable contributions throughout the development of a network-based hybrid WLAN intrusion detection system (HWIDS) that uses both signature and anomaly detection. Firstly, we proposed a novel algorithm for feature selection based on hybrid scheme that merges filtering and wrapper methods. The algorithm resulted in smaller training and classification times because the full 155 features (of the considered benchmark dataset) are reduced to 18 features only, which in turn reduced the dataset size. Secondly, HWIDS can work in real-time and the developed prototype achieved higher precision and almost equal recall value compared with the best published recent system (NMI). However, NMI is very slow compared to HWIDS and cannot work in real-time. We adopted aggregating one-dependence estimator (AODE) for anomaly detection due to its high efficiency, high detection accuracy and its support to incremental learning. The experimental results show that the HWIDS has a fast learning time of 45 seconds and a higher detection accuracy of 99.82%. Thus, we achieved the third contribution to improve the detection accuracy and efficiency of the recent classifiers.

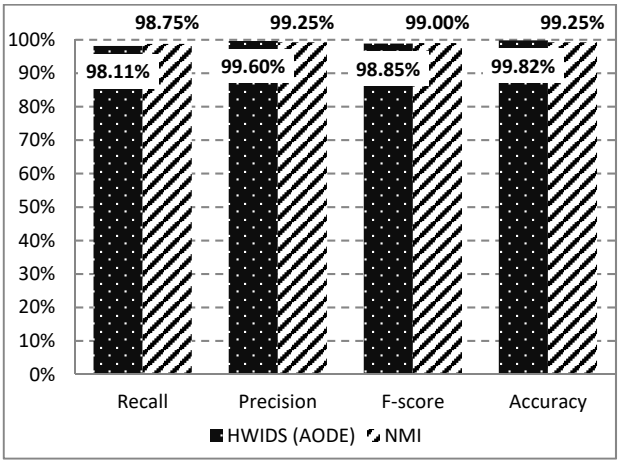

Fig. 11. HWIDS with signature improvement versus NMI.

We evaluated the developed HWIDS prototype using AWID datasets. Experimental results show that HWIDS with AODE learns the whole training set of 1,795,575 instances in less than a minute. Moreover, HWIDS with effective features takes 3.49 seconds to classify all instances in the testing set. Results also proved that HWIDS with AODE has higher precision with 99.80% compared to NMI. This means

HWIDS can accurately detect intrusions with low FPR of 0.03% only. Moreover, we improved the recall value of HWIDS to become 98.11%, which is almost the same as NMI. The low recall resulted from some types of flooding attacks that were absent in the training set and have similar footprints as normal frames. Therefore, we developed two more signature rules to overcome CTS and Probe_request flooding attacks, which consequently improved HWIDS recall, F-score and detection accuracy by 4.74%, 2.37% and 0.35%, respectively. In the future work, we plan to test HWIDS using other machine learning techniques that can detect flooding attacks without relying on signature rules. Also, patterns of attacks that exist within sequence of frames need to be studied to improve the capability of HWIDS to detect more types of attacks.

Acknowledgment

My thanks and appreciation to Prof. Mohamed Ashraf Madkour for persevering with me as my advisor throughout the time it took me to complete this research. As my advisor, Prof. Mohamed provided detailed guidance and encouragement and steered me in the right the direction whenever he thought I needed it. I also thank Dr. Kamal Jambi and Dr. Omar Batarfi, who have generously given their time and expertise to better my work. I thank them for their contribution and their useful suggestions. I would also like to thank Constantinos Kolias from University of the Aegean for providing me the necessary datasets that were used in my experiments.

References

[1] **"**IEEE. 802.11-1997 IEEE Standard," IEEE. 802.11-1997 IEEE Standard for Information Technology, Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks-Specific Requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=654749. [Accessed: 05-May-2015].

[2] **Kolias, C., Kambourakis, G., Stavrou, A.** and **Gritzalis, S.**, "Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset," *IEEE Commun. Surv. Tutor.*, vol. PP, no. 99, pp. 1–1, 2015.

[3] **Guo, F.** and **Chiueh, T.**, "Sequence Number-Based MAC Address Spoof Detection," in: *Recent Advances in Intrusion Detection*, A. Valdes and D. Zamboni, Eds. Springer Berlin Heidelberg, 2006, pp. 309–329.

[4] **John, M. C.** and **He, C.,** "Security Analysis and Improvements for IEEE 802.11 i," in: *The 12th Annual Network and Distributed System Security Symposium (NDSS'05) Stanford University, Stanford, Stanford: Electrical Engineering and Computer Science Departments, Stanford University, 2005*, pp. 90–110.

[5] **Wang, L.** and **Srinivasan, B.**, "Analysis and Improvements over DoS Attacks against IEEE 802.11i Standard," in: *Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010*, vol. 2, pp. 109–113.

[6] **Faria, D. B.** and **Cheriton, D. R.,** "Detecting Identity-based Attacks in Wireless Networks Using Signalprints," in: *Proceedings of the 5th ACM Workshop on Wireless Security, New York, NY, USA, 2006*, pp. 43–52.

[7] "*IEEE. 802.11w-2009*," IEEE. 802.11w-2009 - IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Protected Management Frames. [Online]. Available: http://standards.ieee.org/findstds/standard/802.11w-2009.html. [Accessed: 20-May-2015].

[8] **Ahmad, M. S.** and **Tadakamadla, S.** "Short Paper: Security Evaluation of IEEE 802.11W Specification," in: *Proceedings of the Fourth ACM Conference on Wireless Network Security, New York, NY, USA, 2011*, pp. 53–58.

[9] **Patcha, A.** and **Park, J. M.**, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput Netw.*, vol. 51, no. 12, pp. 3448–3470, Aug. 2007.

[10] **Danziger, M.** and **de Lima Neto, F. B.,** "A hybrid approach for IEEE 802.11 intrusion detection based on AIS, MAS and naïve Bayes," in: *10th International Conference on Hybrid Intelligent Systems (HIS), 2010,* pp. 201–204.

[11] **Kasliwal, B., Bhatia, S., Saini, S., Thaseen, I. S.** and **Kumar, C. A.**, "A hybrid anomaly detection model using G-LDA," in: *Advance Computing Conference (IACC), 2014 IEEE International, 2014*, pp. 288–293.

[12] **Usha, M.** and **Kavitha, P.**, "Anomaly based intrusion detection for 802.11 networks with optimal features using SVM classifier," *Wirel. Netw.*, pp. 1–16, May 2016.

[13] **Baharudin, N., Ali, F. H. M., Darus, M. Y.** and **Awang, N.**, "Wireless Intruder Detection System (WIDS) in Detecting De-Authentication and Disassociation Attacks in IEEE 802.11," in: *5th International Conference on IT Convergence and Security (ICITCS), 2015*, pp. 1–5.

[14] **Torres, L. M., Magana, E., Izal, M., Morato, D.** and **Santafe, G.**, "An anomaly-based intrusion detection system for IEEE 802.11 networks," in: *Wireless Days (WD), 2010 IFIP*, 2010, pp. 1–6.

[15] **Parish, D. J., Aparicio-Navarro, F. J.** and **Kyriakopoulos, K. G.,** "Manual and Automatic assigned thresholds in multi-layer data fusion intrusion detection system for 802.11 attacks," *IET Inf. Secur.*, vol. 8, no. 1, pp. 42–50, Jan. 2014.

[16] **Ma, J., Le, F., Russo, A.** and **Lobo, J.,** "Detecting distributed signature-based intrusion: The case of multi-path routing attacks," in: *IEEE Conference on Computer Communications (INFOCOM), 2015*, pp. 558–566.

[17] **Manasi, G., J.L., R.** and **R.N., Y.,** "Taxonomy of Anomaly Based Intrusion Detection System: A Review," *Int. J. Sci. Res. Publ.*, vol. 2, no. 12, Dec. 2012.

[18] **Salzberg, S. L.,** "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240.

[19] **Webb, G. I., Boughton, J.** and **Wang, Z.**, "Averaged one-dependence estimators: preliminary results," in: *Proceedings of the Australasian Data Mining Workshop 2002*, 2002.

[20] **Webb, G. I., Boughton, J. R.** and **Wang, Z.,** "Not so naive Bayes: aggregating one-dependence estimators," *Mach. Learn.*, vol. 58, no. 1, pp. 5–24, 2005.

[21] " *Snort - Network Intrusion Detection & Prevention System*." [Online]. Available: https://www.snort.org/. [Accessed: 01-Dec-2016].

[22] "*The award-winning wireless networking tool and the best source for your daily Wi-Fi, WiMAX, 3G and VoIP news. | NetStumbler* ".

[23] **Chen, G., Yao, H.** and **Wang, Z.,** "An Intelligent WLAN Intrusion Prevention System Based on Signature Detection and Plan Recognition," in: *2010 Second International Conference on Future Networks*, 2010, pp. 168–172.

[24] **Zaki, M. J.** and **Jr, W. M.,** "Data Mining and Analysis: Fundamental Concepts and Algorithms," in: *Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014*, pp. 553–557.

[25] **Sokolova, M.** and **Lapalme, G.** "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

[26] **Boukerche, A., Machado, R. B., Jucá, K. R. L., Sobral, J. B. M.** and **Notare, M. S. M. A.,** "An agent based and biological inspired real-time intrusion detection and security model for computer network operations," *Comput. Commun.*, vol. 30, no. 13, pp. 2649–2660, Sep. 2007.

[27] **Boukerche, A., Jucá, K. R. L., Sobral, J. B.** and **Notare, M. S. M. A.**, "An artificial immune based intrusion detection model for computer and telecommunication systems," *Parallel Comput.*, vol. 30, no. 5, pp. 629–646, 2004.

[28] **Chen, Y., Li, Y., Cheng, X.-Q.** and **Guo, L.,** "Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System," in: *Information Security and Cryptology*, H. Lipmaa, M. Yung, and D. Lin, Eds. Springer Berlin Heidelberg, 2006, pp. 153–167.

[29] **Guyon, I.** and **Elisseeff, A.,** "An introduction to variable and feature selection," *J. Mach. Learn. Res.,* vol. 3, no. Mar, pp. 1157–1182, 2003.

[30] **Kohavi, R.** and **John, G. H.,** "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1, pp. 273–324, Dec. 1997.

[31] **Guennoun, M., Lbekkouri, A.** and **El-Khatib, K.**, "Selecting the Best Set of Features for Efficient Intrusion Detection in 802.11 Networks," in: *3rd International Conference on Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008*, 2008, pp. 1–4.

[32] **El-Khatib, K.,** "Impact of Feature Reduction on the Efficiency of Wireless Intrusion Detection Systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 8, pp. 1143–1149, Aug. 2010.

[33] **Axelsson, S.,** "*Research in Intrusion Detection Systems: A Survey*!," 1999.

[34] **Liu, H.** and **Motoda, H.,** *Feature Selection for Knowledge Discovery and Data Mining*. Springer Science & Business Media, 2012.

[35] **Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P.** and **Witten, I. H.,** "The WEKA data mining software: an update," *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.

[36] "*Wireshark*." [Online]. Available: https://www.wireshark.org/. [Accessed: 11-Jan-2017].

نظام الكشف عن التسلل الهجين لشبكات802.11 مع التحديد الفعال للميزات

إسحاق سيد إكرام و **محمد أشرف إسماعيل مدكور**

كلية الحاسبات والمعلومات، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية

isaac.sayid@gmail.com

المستخلص. بروتوكول IEEE 802.11i هو معيار الأمان الحالي للشبكات المحلية اللاسلكية. ويقدم هذا البروتوكول آليات أمنية قوية، مثل معيار التشفير المتقدم AES للقيام بالتشفير وبروتوكولات المصافحة رباعية الاتجاه للمصادقة، إلا أنه لا يزال عرضة لعدد من الهجمات الخطيرة، مثل فيضان إلغاء المصادقة والانفصال. تم اقتراح العديد من تقنيات كشف التسلل من قبل المجتمع البحثي للكشف عن هجمات الشبكات المحلية اللاسلكية المعروفة وغير المعروفة. ومع ذلك، هناك حاجة لبذل مزيد من الجهود لتحسين أداء الكشف باستخدام مجموعة بيانات قياسية لبروتوكول 802.11 والتي تحتوي على كل من بيانات حركة انتقال البيانات العادية وحركة انتقال البيانات المتسللة لجميع الهجمات المعروفة. يبدأ البحث الحالي من خلال التعرف على جميع الهجمات الخطيرة ومواطن الضعف في شبكات IEEE 802.11i. بعد ذلك نقدم مسحا شاملا لأنظمة كشف التسلل المقترحة في الأدبيات لمعرفة مزاياها وقيودها. ويلي ذلك تصميم وتنفيذ نموذج أولي لنظام كشف تسلل هجين يعمل في الزمن الحقيقي ويستخدم أساليب الكشف عن البصمة والكشف عن السلوك المستغرب. استخدام أسلوب الكشف عن البصمة يمكن أن يحسّن أداء نظام الكشف عن التسلل الذي طورناه من خلال زيادة المعدل الإيجابي الحقيقي، في حين يمكن لأسلوب الكشف عن السلوك المستغرب إيجاد الهجمات غير المعروفة. بالإضافة إلى قواعد بصمة الهجمات، أخذنا بعين الاعتبار كلا من خوارزمية التصنيف C4.5 وخوارزمية معدل التقدير أحادي الاعتماد (AODE) للكشف عن السلوك المستغرب. تم تقييم النظام المطّور من حيث دقة اكتشاف الهجمات (Precision) وشموليتها (Recall)، مقدمين بذلك ثلاث مساهمات. أولا تم تطوير خوارزمية جديدة لاختيار خواص فعالة بالاعتماد على نموذج التصفية ومعرفة الآثار التي تتركها هجمات الشبكات اللاسلكية. وثانيا، فإن الخوارزمية المطورة تحسن دقة التصنيف والاكتشاف، مقارنة مع النتائج المنشورة مؤخراً، وتقلل بشكل كبير من وقت التصنيف عن طريق تقليل وقت التدريب وعدد الخواص. ثالثاً، فإن البحث يقدم نظام كشف تسلل هجين عالي الأداء للشبكات اللاسلكية يعمل في الزمن الحقيقي. تم تنفيذ النموذج الأولي واختباره على حاسب شخصي 1.7 جيجاهرتز i5 مع 12 جيجابايت من ذاكرة الوصول العشوائي. وأظهرت النتائج التجريبية أن النظام المنفذ لديه وقت تعلم سريع قدره 45 ثانية وأداء تصنيف عالي بدقة 99.6٪، وشمولية 98.11٪، ودقة شاملة قدرها 99.82٪.