Role of Feature Reduction in Intrusion Detection Systems for Wireless Attacks

Jashuva.Chundi[*1], V.V.Gopala Rao[*2]

[1#]11A91D2503 – M.Tech Student, 2[#]Associate Proffesor

[1#, 2#]Aditya Engineering College, Jntuk, A.P., India.

Abstract*— Although of the widespread use of the WLANs, it is still vulnerable for the availability security issues. This research presents a proposal Wireless Network Intrusion Detection System (WNIDS) which is use misuse and anomaly techniques in intrusion detection. The proposal depend on Data mining is a DM-based WNIDS since mining provide iterative process so if results are not satisfied with optimal solution, the mining steps will continue to be carried out until mining results are corresponding intention results. For training and testing of WNIDS in our experiment, we used collected dataset called it W data set, the collection done on an organized WLAN 802.11 consist of 5 machines. The collection of data involved frames from all types (normal and the four known intrusions and unknown intrusion).The collected connections contain features those appear directly in the header of 802.11 frames and we added one more feature (casting) since it is critical in distinguish among intrusions. These connections are labelled as either normal or attack type, many of these features are irrelative in classification process. Here we propose Support Vector Machine SVM classifier as feature extraction to reduce no. of features to avoid time consuming in training and real-time detecting. SVM introduce 8 features as subset of correlated intrinsic features present the basic point in classification. The sets of features that have been resulted from SVM and the all features set will be the feeding of WNIDS. The results obtained from WNIDS showing that accuracy rate of ANN and ID3 classifiers are both higher with SVM (8) features than set of all features. And absolutely, ANN accuracy is higher than ID3 with both sets of features.

***Keywords*— Feature selection, intrusion detection systems, K-means, information gain ratio, wireless networks, neural networks.**

I. INTRODUCTION

WLANs suffer from a lot of vulnerabilities, some of these vulnerabilities inherited from the usual wired networks and some are new due to the broadcast connection medium. These vulnerabilities include confidentiality, integrity and availability vulnerabilities. Through the WLAN evolution, many security improvements have been added to the IEEE 802.11 standards such as: Wired Equivalent Privacy (WEP), Wi-Fi Protected Access (WPA) and IEEE 802.11i (WPA2). These techniques can only protect data frames to satisfy the confidentiality and the integrity security issues. The management and control frames still unprotected [1].

Several vulnerabilities exist at the link layer level of the802.11 protocol. Many 802.11-specificattacks were analysed and demonstrated to present a real threat to network availability. Most of the attacks on WLAN are; ***Deauthentication attack*,** where the attacker fakes a deauthentication frame as if it had originated from the base station (Access Point). Up on reception, the station disconnects and tries to reconnect to the base station again. This process is repeated indefinitely to keep the station disconnected from the base station. The attacker can also set the receiving address to the broadcast address to target all stations associated with the victim base station. However, we noticed that some wireless network cards ignore this type of deauthentication frame. ***Chop Chop attack*,** where the attacker intercepts an encrypted frame and uses the Access Point to guess the clear text.

II. FEATURE SELECTIONS

Feature selection is the most critical step in building intrusion detection models [1], [2], [3]. During this step, the set of attributes or features deemed to be the most effective attributes is extracted in order to construct suitable Detection algorithms (detectors). A key problem that many researchers face is how to choose the optimal set of features, s not all features are relevant to the learning algorithm, and in some cases, irrelevant and redundant features can introduce noisy data that distract the learning algorithm, everely degrading the accuracy of the detector and causing slow training and testing processes. Feature selection was raven to have a significant impact on the performance of he classifiers. The wrapper model uses the predictive accuracy of classifier as a means to evaluate the “goodness” of a feature set, while the filter model uses a measure such as information, consistency, or distance measures to compute the relevance of a set of features.

Different techniques have been used to tackle the problem of feature selection. In [7], Sung and Mukkamala used feature ranking algorithms to reduce the feature space of the DARPA data set from 41 features to the six most important features. They used three ranking algorithms based on Support Vector Machines (SVMs), Multivariate Adaptive Regression Splines(MARSs), and Linear Genetic Programs (LGPs) to assign a weight to each feature. Experimental results showed that the classifier’s accuracy degraded by less than 1 percent when the classifier was fed with the reduced set of features. Sequential backward search was used in [8], [9] to identify the important set of features: starting with the set of all features, one feature was removed at a time until the accuracy of the classifier was below a certain threshold. Different types of classifiers were used with this approach including Genetic Algorithms in [9], Neural Networks in [8],[10], and Support Vector Machines in [8].

III. 802.11-SPECIFIC INTRUSIONS

Several vulnerabilities exist at the link layer level of the802.11 protocol. In, many 802.11-specificattacks were analyzed and demonstrated to present a real threat to network availability. A deauthentication attack isan example of an easy to mount attack on all types of 802.11networks. Likewise, a duration attack is another simple attack that exploits the vulnerability of the virtual carrier sensing protocol CSMA/CA and it was proven in [11] to deny access to the network.

Most of the attacks we used in this work are available for download from. The attacks we used to conduct the experiments are:

3.1 Deauthentication Attack

This attack sends disassocate packets to one or more clients which are currently associated with a particular access point. Disassociating clients can be done for a number of reasons:

- Recovering a hidden ESSID. This is an ESSID which is not being broadcast. Another term for this is “cloaked”.
- Capturing WPA/WPA2 handshakes by forcing clients to re authenticate
- Generate ARP requests (Windows clients sometimes flush their ARP cache when disconnected)

Of course, this attack is totally useless if there are no associated wireless client or on fake authentications.

3.2 ChopChop Attack

This attack, when successful, can decrypt a WEP data packet without knowing the key. It can even work against dynamic WEP. *This attack does not recover the WEP key itself, but merely reveals the plaintext*. However, some access points are not vulnerable to this attack. Some may seem vulnerable at first but actually drop data packets shorter that 60 bytes. If the access point drops packets shorter than 42 bytes, aireplay tries to guess the rest of the missing data, as far as the headers are predictable. If an IP packet is captured, it additionally checks if the checksum of the header is correct after guessing the missing parts of it. This attack requires at least one WEP data packet.

3.3 IP Fragmentation Attack

IP fragmentation is the process of breaking up a single Internet Protocol (IP) datagram into multiple packets of smaller size. Every network link has a characteristic size of messages that may be transmitted, called the maximum transmission unit (MTU).

Part of the TCP/IP suite is the Internet Protocol (IP) which resides at the Internet Layer of this model. IP is responsible for the transmission of packets between network end points. IP includes some features which provide basic measures of fault-tolerance (time to live, checksum), traffic prioritization (type of service) and support for the fragmentation of larger packets into multiple smaller packets (ID field, fragment offset). The support for fragmentation of larger packets provides a protocol allowing routers to fragment a packet into smaller packets when the original packet is too large for the supporting datalink frames. IP fragmentation exploits (attacks) use the fragmentation protocol within IP as an attack vector.

3.4 Duration Attack

The attacker exploits vulnerability in the virtual carrier-sense mechanism and sends a frame with the NAV field set to a high value (32 ms). This will prevent any station from using the shared medium before the NAV timer reaches zero. Before expiration of the timer, the attacker sends another frame. By repeating this process, the attacker can deny access to the wireless network. More details can be found in [11].

IV. HYBRID APPROACH

Extensive work has been done to detect intrusions in wired and wireless networks. However, most of the intrusion detection systems examine only the network layer and higher abstraction layers for extracting and selecting features, and ignore the MAC layer header. These IDS scan not detect attacks that are specific to the MAC layer.

Some previous work tried to build IDS that functioned at the Data link layer. For example, in, the authors simply used the MAC layer header attributes as input features to build the learning algorithm for detecting intrusions. No feature selection algorithm was used to extract the most relevant set of features.

In this paper, we will present a complete framework to select the best set of MAC layer features that efficiently characterize normal traffic and distinguish it from abnormal traffic containing intrusions specific to wireless networks. Our framework uses a hybrid approach for feature selection that combines the filter and wrapper models. In this approach, we rank the features using an independent measure: the information gain ratio. The k-means classifier’s predictive accuracy is used to reach an optimal set of features which maximize the detection accuracy of the wireless attacks.

To train the classifier, we first collect network traffic containing four known wireless intrusions, namely, the de authentication, duration, fragmentation, and

```
Input:
   F – Full set of features
   IGR: Information Gain Ratio Measure
   C: K-means classifier
   T: Gained Accuracy Threshold

For each feature f compute IGR(f)
    Rank features in F according to IGR(f)

//Optimal Set Selection Algorithm
Initialize: S={}, ac=0
Repeat
        (1)  ap=ac
        (2)  f=getNext(F)
        (3)  S=S U {f}
        (4)  F=F- {f}
        (5)  ac= accuracy(C,S)
 Until (ac-ap)<T Or  ac<ap
```

Fig. 1. Best feature set selection algorithm.

Chop chop attack. The reader is referred to [11], [12], [16] fora detailed description of each attack. The selection algorithm (Fig. 1) starts with an empty setS of the best features, and then, proceeds to add features from the ranked set of features F into S sequentially. After each iteration, the “goodness” of the resulting set of features S is measured by the accuracy of the k-means classifier. The selection process stops when the gained classifier’s accuracy is below a certain selected threshold value or in some cases when the accuracy drops, which means that the accuracy of the current subset is below the accuracy of the previous subset.

V. INITIAL LIST OF FEATURES

The initial list of features is extracted from the MAC layer frame header. According to the 802.11 standard [17], the fields of the MAC header are as given in Table 1.These raw features in Table 1 are extracted directly from the header of the frame. Note that we consider each byte ofa MAC address, FCS, and Duration as a separate feature.

We preprocess each frame to extract extra features thatare listed in Table 2. The total number of features that are used in our experiments is 38 features.

VI. INFORMATION GAIN RATIO MEASURE

We used the Information Gain Ratio (IGR) as a measure to determine the relevance of each feature. Note that we chose the IGR measure and not the Information Gain because the latter is biased toward the features with a large number of distinct values [5].

IGR is defined in as

IGR(Ex,f) = Gain(Ex,F)/SplitInfo(Ex,f)

where Ex is the set of vectors that contain the header information and the corresponding class:

TABLE 1

List of Features Extracted from 802.11 Frames

Feature	Description
Version	Two bits indicate which version of the 802.11 MAC is contained in the rest of the frame
Type	Indicate the type of the frame (Mgmt, Ctrl, and Data).
SubType	Indicate the subtype of the frame
ToDS	Indicate if a frame is destined to the Distributed System.
FromDS	Indicate if a frame is originated from Distributed System.
More Fragment	Indicate whether a frame is non final fragment or not.
Retry	Indicate if the frame is a retransmitted Frame
Power Mgmt	Indicate whether the station is active or in Power Saving Mode
More Data	Indicate whether an access point has buffered frames for a dozing station
WEP	Indicate if the frame is processed by WEP protocol
Order	Indicate if the “strict ordering” delivery is employed.
Duration	The number of microsecond the medium is expected to be busy.
RA	The MAC address of the receiving
TA	The MAC address of the transmitting station.
MA .	Depending on the values of ToDS and FromDS fields, this address can be the MAC address of the Sending, Destination or Base Station

FCS A	Frame Check Sequence, which contains a 32 bit Cyclic Redundancy Code.

Gain(Ex,f) = Entropy(Ex)

For all -|Ex,v|/|Ex|*Entropy(Ex,v)

Ex,v = {x is Ex/value(x,f) = v

The entropy function is the Shannon’s entropy defined as

Entropy(Ex) = -For all Pi log2 (Pi)

Where Pi is the probability of a class i.

SplinInfo(Ex,f) is defined as

SplinInfo(Ex,f) = - For all |Ex,v|/|Ex| log2 |Ex,v|/|Ex|

TABLE 2

List of Features After Processing 802.11 Frames

Feature	**Description**
IsWepValid	Indicate if WEP ICV check is successful.
DurationRange	Indicate if duration value is low (<5ms), average (between 5-20ms), or high (>20ms).
Casting Type	Indicate whether the receiving address is a unicast, multicast or a broadcast address.

TABLE 3

Top 10 Features

Rank	Feature	IGR
1	Is Wep Valid	1.02
2	Duration Range	1.01
3	More Frag	0.98
4	To DS	0.89
5	WEP	0.85
6	Casting Type	0.82
7	Type	0.73
8	SubType	0.65
9	Retry	0.46
10	From DS	0.41
11-38	Remaining Features	<0.23

Using the data set of frames collected from our testing network, we could rank the features according to the score assigned by the IGR measure. The top 10 ranked features are shown in Table 3.

VII. THE BEST SUBSET OF FEATURES

The k-means classifier is used to compute the detection rate for each set of features. Initially, the set of features S contains only the top ranked feature. After each iteration, a new feature is added to the list S based on the rank which it is assigned by the IGR measure. Fig. 2 shows the accuracy of each subset of features. Note that Si is the i first features in the ranked list of features.

We can see that there is subset Sm of features that maximizes the accuracy of the K-means classifier. We can conclude that the first eight features (IsWepValid, DurationRange, More_Flag, To_DS, WEP, Casting_Type, Type, and SubType) are the best features to detect the intrusions we tested in our experiments.

In the rest of the paper, we report the results of our experiments related to the impact of the optimized set of features listed above on the accuracy and learning time of three different architectures of classifiers analyzed through neural networks.

VIII. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANNs) are computational models which mimic the properties of biological neurons. A neuron, which is the base of an ANN, is described by a state, synapses, a combination function, and a transfer function. The state of the neuron, which is a Boolean or real value, is the output of the neuron. Each neuron is connected to other neurons via synapses. Synapses are associated with weights that are used by the combination function to achieve a pre computation, generally a weighted sum, of the inputs. The Activation function, also known as the transfer function, computes the output of the neuron from the output of the combination function.

An artificial neural network is composed of a set of neurons grouped in layers that are connected by synapses.

There are three types of layers: input, hidden, and output layers. The input layer is composed of input neurons that receive their values from external devices such as data files or input signals. The hidden layer is an intermediary layer containing neurons with the same combination and transfer functions. Finally, the output layer provides the output of the computation to the external applications.

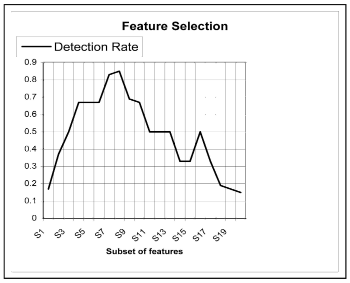

Fig. 2. Detection rate versus subset of features.

An interesting property of ANNs is their capacity to dynamically adjust the weights of the synapses to solve a specific problem. There are two phases in the operation of Artificial Neuron Networks. The first phase is the learning phase in which the network receives the input values with their corresponding outputs called the desired outputs. In this phase, weights of the synapses are dynamically adjusted according to a learning algorithm. The difference between the output of the neural network and the desired output gives a measure on the performance of the network

In order to study the impact of the optimized set of features on both the learning phase and accuracy of the ANN networks, we have tested these attributes on three types of ANN architectures.

8.1 Perceptron

Perceptron is the simplest form of a neural network. It's used for classification of linearly separable problems. It consists of a single neuron with adjustable weights of the synapses. Even though the intrusion detection problem is not linearly separable, we use the perceptron architecture as reference to measure the performance of the other two types of classifiers.

8.2 Multilayer Back propagation Perceptions

The multilayer back propagation perceptions architecture is an organization of neurons in n successive layers (n > ¼ 3). The synapses link the neurons of a layer to all neurons of the following layer. Note that we use one hidden layer composed of eight neurons.

8.3 Hybrid Multilayer Perceptrons

The Hybrid Multilayer Perceptrons architecture is the superposition of perceptron with multilayer ackpropagation perceptrons networks. This type of network is capable of identifying linear and nonlinear correlation between the input and output vectors [19]. We used this type of architecture with eight neurons in the hidden layer.

Transfer function of all neurons is the sigmoid function. The initial weights of the synapses are randomly chosen between the interval [_0:5, 0:5].

IX. DATA SET

The data we used to train and test the classifiers were collected from a wireless local area network. The local network was composed of three wireless stations and one access point. One machine was used to generate normal traffic (HTTP, FTP). The second machine simultaneously transmitted data originating from four types of attacks. The last station was used to collect and record both types of traffic (normal and intrusive

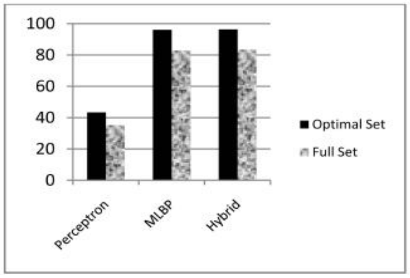

Fig. 3. Detection Rate percentage of the three types of neural networks using 8 and 38 features.

The following table shows the distribution of the data collected for each attack and the number of frames in each data set.

X. EXPERIMENTAL RESULTS

Experimental results were obtained using Neuro Solutions software [20]. The three types of classifiers were trained using the complete set of features (38 features), which are the full set of MAC header attributes, and the reduced set of features (eight features). We evaluated the performance of the classifiers based on the learning time and accuracy of the resulting classifiers. Experimental results clearly demonstrate that the performance of the classifiers trained with the reduced set of features is higher than the performance of the classifiers trained with the full set of features

As shown by the previous graph, the learning time is reduced by an average of 66 percent for the three types of classifiers. The performance of the three classifiers is improved by an average of 15 percent when they are tested using the reduced set of features. Fig. 5 and Fig. 6 show the experimental results of false positives and false negatives. The false positives rate is the percentage of frames containing normal traffic classified as

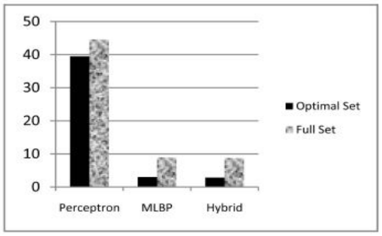

Fig. 4. False Positives Rate (%) for the three types of neural networks using 8 and 38 features.

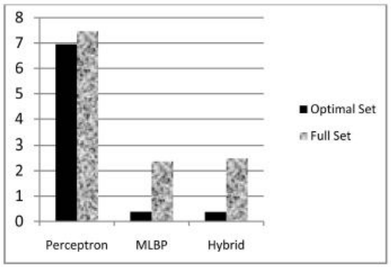

Fig. 5. False Negatives Rate (%) for the three types of neural networks using 8 and 38 features.

Intrusive frames. Likewise, the false negatives rate is the percentage of frames generated from wireless attacks which are classified as normal traffic. The false positives rate is reduced by an average of 28 percent when the reduced set of features is used. If the perceptron classifier is excluded, the combined false positives rate of the MLBP and Hybrid classifiers is reduced by 67 percent. As shown in Fig. 6, the combined false negatives rate of the MLBP and Hybrid classifiers is reduced by 84 percent.

CONCLUSION and FUTURE WORK

In this paper, Feature selection is an important task of network intrusion detection. Using SVM as a feature selection approach, intrusions are detected with less error rate and high accuracy. Usage of ANN for intrusion detection introduces high accuracy than with ID3 as in tables (2 and 3). Where notice the higher rates of detection and very less rates of false alarms especially with SVM set of features with both classifier ID3 and ANN. The added feature (casting) which is not found directly in frame header was important variable in classification, so we think if there is an additional process to the frame to extract more feature may affect in increasing the performance.

REFERENCES

1. A. Boukerche, R.B. Machado, K.R.L. Juca´ , J.B.M. Sobral, and M.S.M.A. Notare, "An Agent Based and Biological Inspired Real- Time Intrusion Detection and Security Model for Computer Network Operations," Computer Comm., vol. 30, no. 13, pp. 2649- 2660, Sept. 2007.

2. A. Boukerche, K.R.L. Juc, J.B. Sobral, and M.S.M.A. Notare, "An Artificial Immune Based Intrusion Detection Model for Computer and Telecommunication Systems," Parallel Computing, vol. 30, nos. 5/6, pp. 629-646, 2004.

3. A. Boukerche and M.S.M.A. Notare, "Behavior-Based Intrusion Detection in Mobile Phone Systems," J. Parallel and Distributed Computing, vol. 62, no. 9, pp. 1476-1490, 2002.

4. Y. Chen, Y. Li, X. Cheng, and L. Guo, "Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System," Proc. Conf. Information Security and Cryptology (Inscrypt), 2006.

5. H. Liu and H. Motoda, Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic, 1998.

6. http://kdd.ics.uci.edu/databases/kddcup99/task.html, 2010.

7. A.H. Sung and S. Mukkamala, "The Feature Selection and Intrusion Detection Problems," Proc. Ninth Asian Computing Science Conf., 2004.

8. A.H. Sung and S. Mukkamala, "Identifying Important Features for Intrusion Detection Using Support Vector Machines and Neural Networks," Proc. Symp. Applications and the Internet (SAINT '03), Jan. 2003.

9. G. Stein, B. Chen, A.S. Wu, and K.A. Hua, "Decision Tree Classifier for Network Intrusion Detection with GA-Based Feature Selection," Proc. 43rd ACM Southeast Regional Conf.—Volume 2, Mar. 2005.