International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.2, Issue.4,July-Aug 2012 pp-1552-1556 ISSN: 2249-6645

Feature Selection for Wireless Intrusion Detection System Using Filter and Wrapper Model

Mohanabharathi R[*1], Mr T.Kalaikumaran[#2] Dr.S.Karthi[*3]
[1,2]*Department of Computer Science Engineering*
SNS College of Technology, Coimbatore.

Abstract: Intrusion detection systems are applied to detect network intrusions identified from different sources. Anomaly and signature based schemes are used for the intrusion detection process. Signature based intrusion detection schemes uses the predefined signature collection for the detection process. The anomaly-based model detects the intrusions by learning the network transaction patterns. Feature selection schemes are used to reduce the network transaction features. Performance, time, accuracy and reliability are improved by the feature selection schemes. Current intrusion detection systems use the TCP/IP header information for the intrusion detection process. Network layer and transport layer attacks can be easily detected using TCP/IP header information's. MAC layer is not considered in the intrusion detection process. The hybrid approach is used for the feature selection process. Information gain ratio measure and K-means classifiers are used in the feature selection process. Back propagation perceptron based neural network algorithm is used for thee learning and testing process. Scalability and high learning error rate problems are identified in the neural network method.
The proposed system is designed to perform the feature reduction and intrusion detection process under wireless LAN environment. The recurrent neural network is used for the intrusion detection process. The feature reduction process is also enhanced to improve accuracy. Real Time Recurrent Learning (RTRL) algorithm is used to solve the scalability problems.

Keywords: Intrusion detection, anomaly based scheme, signature based scheme, feature selection, RTRL Algorithm.

I. INTRODUCTION

Intrusions are the result of flaws in the design and implementation of computer systems, operating systems, applications, and communication protocols. Statistics [2] show that the number of identified vulnerabilities is growing. Exploitation of these vulnerabilities is becoming easier because the knowledge and tools to launch attacks are readily available and usable. It has become easy for a novice to find attack programs on the Internet that he/she can use without knowing how they were designed by security specialists.

The emerging technology of wireless networks created a new problem. Although traditional IDSs are able to protect the application and software components of TCP/IP networks against intrusion attempts, the physical and data link layers are vulnerable to intrusions specific to these communication layers. In addition to the vulnerabilities of wired networks, wireless networks are the subject of new types of attacks which range from the passive eavesdropping to more devastating attacks such as denial of service. These vulnerabilities are a result of the nature of the transmission media [13]. Indeed, the absence of physical boundaries in the network to monitor, meaning that an attack can be perpetrated from anywhere, is a major threat that can be exploited to undermine the integrity and security of the network. It is, therefore, essential to take into account these considerations when designing and deploying an intrusion detection system.

To detect intrusions, classifiers are built to distinguish between normal and anomalous traffic. It has been proved that optimizing the feature set has a major impact on the performance, speed of learning, accuracy, and reliability of the intrusion detection system. Unfortunately, current wireless intrusion detection solutions rely on features extracted directly from the frame headers to build the learning algorithm of the classifiers.

II. Feature Selections

Feature selection is the most critical step in building intrusion detection models [1]. 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, as 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, severely degrading the accuracy of the detector and causing slow training and testing processes. Feature selection was proven to have a significant impact on the performance of the classifiers. Experiments in [4] show that feature selection can reduce the building and testing time of a classifier by 50 percent.

There are currently two models in the literature for feature selection: the filter model and the wrapper model. The wrapper model uses the predictive accuracy of a 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. These approaches suffer from many drawbacks: the first major drawback is that feeding the classifier with arbitrary features may lead to biased results, and hence, we cannot rely on the classifier's predictive accuracy as a measure to select features. A second drawback is that for a set of N features, trying all possible combinations of features (2N combinations) to find the best combination to feed the classifier is not a feasible approach. For example, the DARPA data set contains 41 features [6], and the data set would be larger if we add to it the OSI Layer 2 (MAC layer) features, resulting in thousands of billions of different feature combinations.

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 theDARPA 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 [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, Neural Networks and Support Vector Machines.

III. 802.11-SPECIFIC INTRUSIONS

Several vulnerabilities exist at the link layer level of the 802.11 protocol [8], [5]. Many 802.11-specific attacks were analyzed and demonstrated to present a real threat to network availability. A deauthentication attack is an example of an easy to mount attack on all types of 802.11 networks. Likewise, a duration attack is another simple attack that exploits the vulnerability of the virtual carrier sensing protocol CSMA/CA and it was proven to deny access to the network.

Many free tools are available on the Internet which allow novice hackers to exploit these protocol weaknesses to deny access to a network, as can be seen in [12], where a collection of tools to attack 802.11-based networks is available for download. These tools operate on WEP and WPA-protected networks. Most of the attacks we used in this work are available for download from [12]. The attacks we used to conduct the experiments are:

A. Deauthentication Attack

The attacker fakes a deauthentication frame as if it had originated from the base station (Access Point). Upon 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. More details of this attack.

B. ChopChop Attack

The attacker intercepts an encrypted frame and uses the Access Point to guess the clear text. The attack is performed as follows: Theintercepted encrypted frame is chopped from the last byte. Then, the attacker builds a new frame 1 byte smaller than the original frame. In order to set the right value for the 32 bit longCRC32 checksum named ICV, the attacker makes a guess on the last clear byte. To validate the guess he/she made, the attacker will send the new frame to the base station using a multicast receive address. If the frame is not valid (i.e., the guess is wrong), then the frame is silently discarded by the access point. The frame with the right guess will be relayed back to the network. The hacker can then validate the guess he/she made. The operation is repeated until all bytes of the clear frame are discovered. More details of this attack can be found in [10].

C. Fragmentation Attack

The attacker sends a frame as a successive set of fragments. The access point will assemble them into a new frame and send it back to the wireless network. Since the attacker knows the clear text of the frame, he can recover the key stream used to encrypt the frame. This process is repeated until he/she gets a 1,500-byte long key stream. The attacker can use the key stream to encrypt new frames or decrypt a frame that uses the same three byte initialization vector IV. The process can be repeated until the attacker builds a rainbow key stream table of all possible IVs. Such a table requires 23 GB of memory.

D. Duration Attack

The attacker exploits vulnerability in the virtual carrier-sense mechanism and sends a frame with theNAVfield 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.

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 IDSs cannot 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 [3], [11] 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 deauthentication, duration, fragmentation, and chopchop attack. The reader is referred to [12] for a detailed description of each attack. The Best feature set selection algorithm is shown below

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

The selection algorithm starts with an empty set S 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, 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 of a MAC address, FCS, and Duration as a separate feature. We preprocess each frame to extract extra features that are 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.
IGR is defined as

$$IGR(Ex, f) = \frac{Gain(Ex, f)}{SplitInfo(Ex, f)}$$

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.

$$Gain(Ex, f) = Entropy(Ex) - \sum_{v \in Values(f)} \frac{|Ex,v|}{|Ex|} * Entropy(Ex,v),$$

$$Ex,v = \{x \in Ex / value(x, f) = v$$

The entropy function is the Shannon's entropy defined as

$$Entropy(Ex) = -\sum P_i \log_2(P_i)$$

where Pi is the probability of a class i.
SplinInfo(Ex, f) is defined as

$$Split\inf o(Ex, f) = -\sum_{v \in values(f)} \frac{|Ex,v|}{|Ex|} \log_2\left(\frac{|Ex,v|}{|Ex|}\right)$$

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	**IsWepValid**	**1.02**
2	**DurationRange**	**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**	**<.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. 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. Increasing the number of features does not contribute to the improvement of the accuracy. In fact, irrelevant features distract the classifier and the accuracy drops to 17 percent with 19 features.

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 precomputation, 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.

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. The most used learning algorithm is the retro backpropagation algorithm. In the second phase, called the generalization phase, the neural network is capable of extending the learned examples to new examples not seen before. The learning phase is resource demanding, explained by the iterative nature of the operation mode of the ANN. Once the network is trained, the processing of a new input is generally fast. 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.

A. 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.

B. Multilayer Backpropagation Perceptrons

The multilayer backpropagation perceptrons 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.

TABLE 4
Distribution of Collected Data

	Learning	Validation	Test
Normal	**6000**	**4000**	**5000**
De-authentication	**900**	**600**	**800**
Duration	**900**	**600**	**800**
Fragmentation	**900**	**600**	**800**
Chopchop	**900**	**600**	**800**
Total	**9600**	**6400**	**8200**

C. Hybrid Multilayer Perceptrons

The Hybrid Multilayer Perceptrons architecture is the superposition of perceptron with multilayer backpropagation perceptrons networks. This type of network is capable of identifying linear and nonlinear correlation between the input and output vectors. 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. Intrusion Detection Using Recurrent Neural Network

The proposed system is designed to perform the feature selection and intrusion detection process. The feature selection scheme is used to filter the irrelevant fields in network transactions. The feature selection selects suitable fields for the intrusion detection process. Filter, wrapper and hybrid feature selection schemes are used in the system. The intrusion detection process is performed using the artificial neural networks. The back propagation perception algorithm is used for the intrusion detection process. The recurrent neural network is used to for the intrusion detection process. The system also performs the intrusion detection process on the supervised feature selection model transactions. Detection latency, false positive and false negative measures are used for the performance evaluation.

The system is divided into four major modules. Feature reduction module is used to select optimized features. Artificial neural network based intrusion detection is performed using filtered data sets. Supervised features based IDS module is applied on user selected features. Recurrent neural network algorithm is used for intrusion detection process.

A. Feature Reduction Process

The feature selection is applied to reduce the fields that are used in the intrusion detection process. It will improve the accuracy of the system. The process time is also reduced in the learning and testing process. The

features selection scheme uses filter, wrapper and hybrid feature selection techniques. The filter model uses the information consistency ratio or distance measure. Wrapper model uses predictive accuracy. Information gain measure is used in the hybrid feature selection model.

B. Intrusion Detection Using ANN

Artificial neural networks are used with reduced features. Back probgation algorithm is used for the learning and testing process. Perceptron based neural network model uses 8 neurons for each layer. Multilayer based back propagation perceptron algorithm is used for intrusion detection. Learning process supports limited transactions only.

C. IDS Using Supervised Features

Supervised feature selection model uses features retrieved using experts' knowledge. Listed attributes are separated from user transactions. Artificial neural network is used for the intrusion detection process. Supervised features based IDS scheme is compared with automated feature selection process.

D. Intrusion Detection Using RNN

Selected features are used in recurrent neural network based intrusion detection model. RNN supports scalability in learning process. Learning errors are reduced in RNN. Real-Time Recurrent Learning Algorithm (RTRL) is used in RNN. The RNN technique is applied to verify the accuracy level of feature selection and learning process. The system improves the accuracy for the intrusion detection process. The detection latency is also reduced by the system.

X. CONCLUSION

In this paper, we have presented a novel approach to select the best features for detecting intrusions in 802.11-based networks. Our approach is based on a hybrid approach, which combines the filter and wrapper models for selecting relevant features. We were able to reduce the number of features from 38 to 8. We have also studied the impact of feature selection on the performance of different classifiers based on neural networks. Learning time of the classifiers is reduced to 33 percent with the reduced set of features, while the accuracy of detection is improved by 15 percent. The system reduces the feature selection complexity. Detection period is reduced by the system. The system integrates all layers in the IDS. Wireless LAN attacks are controlled by the system.

References

[1] Boukerche, 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] CERT, http://www.cert.org/stats/, 2010.
[3] Y.-H. Liu, D.-X. Tian, and D. Wei, "A Wireless Intrusion Detection Method Based on Neural Network," Proc. Second IASTED Int'l Conf. Advances in Computer Science and Technology, Jan. 2006.
[4] Y. Chen, and L. Guo, "Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System," Proc. Conf. Information Security and Cryptology (Inscrypt), 2006.
[5] Boukerche, Handbook of Algorithms for Wireless Networking and Mobile Computing. CRC/Chapman and Hall, 2005.
[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] Boukerche, Algorithms and Protocols for Wireless and Mobile Ad Hoc Networks. Wiley, 2008.
[9] G. Stein, 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.
[10] Bittau and J. Lackey, "The Final Nail in WEP's Coffin," Proc. IEEE Symp. Security and Privacy, May 2006.
[11] T.M. Khoshgoftaar and N. Seliya, "Intrusion Detection inWireless Networks Using Clustering Techniques with Expert Analysis," Proc. Fourth Int'l Conf. Machine Learning and Applications, Dec. 2005.
[12] http://www.aircrack-ng.org/, 2010.
[13] Khalil El-Khatib "Impact of Feature Reduction on the Efficiency of Wireless Intrusion Detection Systems" IEEE Transactions on Parallel and Distributed Systems, Vol. 21, no. 8, August 2010.