2009 International Conference on Computer Engineering and Applications
IPCSIT vol.2 (2011) © (2011) IACSIT Press, Singapore

Impact of Feature Selection on the Performance of Wireless Intrusion Detection Systems

Mouhcine Guennoun [1], Zine E.A Guennoun [2] and Khalil El-Khatib [1]

[1] University of Ontario Institute of Technology
2000 Simcoe Street North, Oshawa, Ontario, Canada L1H 7K4
[2] Département Math-Info, Faculté des Sciences de Rabat
4 Avenue Ibn Battouta B.P. 1014 RP, Rabat, Maroc
{mouhcine.guennoun@uoit.ca , guennoun@fsr.ac.ma, khalil.el-khatib@uoit.ca }

Abstract. In this paper, we study the impact of the optimization of the feature set of wireless intrusion detection systems on the performance and learning time of different types of classifiers based on neural networks. The optimal set of features is selected using a hybrid selection model. In this approach, the wireless frame attributes are first ranked according to a score assigned by the information gain ratio measure. K-means classifier is then used to build the optimal subset of features that maximizes the accuracy of the detectors while reducing their learning time. Experimental results with three types of neural networks architectures show clearly that the optimization of the wireless feature set has a significant impact on the efficiency and accuracy of the intrusion detection system.

Keywords: Intrusion Detection Systems, Wireless Networks, Feature Selection.

1. Introduction

Intrusion detection in wireless networks has gained considerable attention in the last few years. Wireless networks are not only susceptible to TCP/IP-based attacks native to wired networks, they are also subject to a wide array of 802.11-specific threats. Such threats range from passive eavesdropping to more devastating denial of service attacks. To detect these 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 [1]. Unfortunately, current wireless intrusion detection solutions rely on features extracted directly from the frame headers to build the learning algorithm of the classifiers.

Feature selection is the most critical step in building intrusion detection models. During this step, the set of attributes or features that deemed to be the most effective attributes are 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 since not all features are relevant to the learning algorithm, and in some cases, irrelevant and redundant features can introduce noisy data that distracts the learning algorithm and therefore severely degrade the accuracy of the detector and cause slow training and testing process. Feature selection was proven to have a significant impact on the performance of the classifiers. Experiments in [1] show that feature selection can reduce the building and testing time of a classifier by up to 50%.

The rest of the paper is organized as follows. Section 2 lists the optimal set of features selected using a hybrid selection model. In Section 3, we go over the three different neural networks architectures we used to build the intrusion detection system. Section 4 gives an overview of the datasets that we collected to train, validate and test the classifiers. We discuss the experimental results in Section 5 and finally, conclusions and plans for future work are provided in Section 6.

2. Optimal Set of Features

In [2,3], we presented 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 [4]. 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 that maximizes the accuracy of detection of wireless attacks. To train the classifier, we first collected network traffic that contains four known wireless intrusions, which are de-authentication, duration, fragmentation, and chop-chop attacks [5,6]. As shown in table 1, the selection algorithm voted 8 features as the best set of features that maximizes the accuracy of the k-means classifier.

Table 1: List of the optimal set of features

Feature	Description
IsWepValid	Indicate if WEP ICV check is successful.
DurationRange	Indicate if duration value is low(<5ms), average (between 5-20ms), or high (>20 ms).
MoreFragment	Indicate whether a frame is non final fragment or not.
ToDS	Indicate if a frame is destined to the Distribution System.
WEP	Indicate if the frame is processed by the WEP protocol.
CastingType	Indicate whether the receiving address is a unicast, multicast or a broadcast address.
Type	Indicate the type of the frame (Mgmt, Ctrl, Data).
SubType	Indicate the subtype of the frame.

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 based on neural networks.

3. Artificial Neural Networks

Artificial Neural Networks (ANN) is a computational model that mimics 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. Activation function, called also transfer function, computes, from the output of the combination function, the output of the neuron.

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 which is an intermediary layer that contains neurons with the same combination and transfer functions. The output layer provides the output of the computation to the external applications.

An interesting property of the ANN is their capacity to dynamically adjust the weights of the synapses to solve a specific problem. There are two phases in the operation of the ANN networks. 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 back propagation algorithm. In the second phase, called generalization phase, the neural network is capable of extending the learned examples to new examples not seen before. The learning phase is resource demanding. This is 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 the learning phase and accuracy of the ANN networks, we have tested these attributes on three types of architectures of ANN networks.

3.1. Perceptron

Perceptron, fig. 1, 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 tough 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.

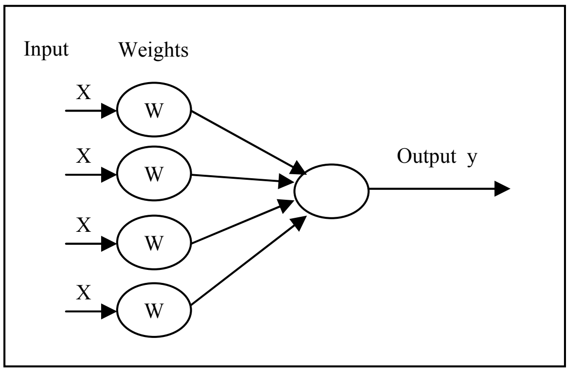

Fig. 1: Architecture of a perceptron

3.2. Multi-layer back propagation perceptrons

The multi-layer back-propagation perceptrons architecture, fig. 2, is an organization of neurons in n successive layers (n>=3). The synapses link the neurons of a layer to all neurons of the following layer. Error propagation is done in the opposite direction of the information flow. We note that we use one hidden layer composed of 8 neurons.

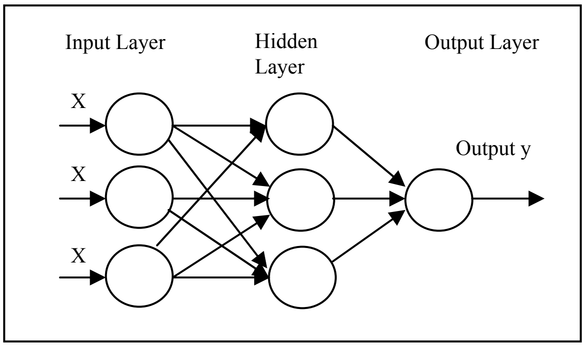

Fig. 3: Multi-Layer Back-propagation Perceptrons Architecture

3.3. Hybrid Multi-Layer Perceptrons

Hybrid Multi-Layer Perceptrons architecture, fig. 3, is the superposition of perceptron with multi-layer back-propagation perceptrons networks. This type of network is capable of identifying linear and non linear correlation between the input and output vectors [7]. We used this type of architecture with 8 neurons in the hidden layer.

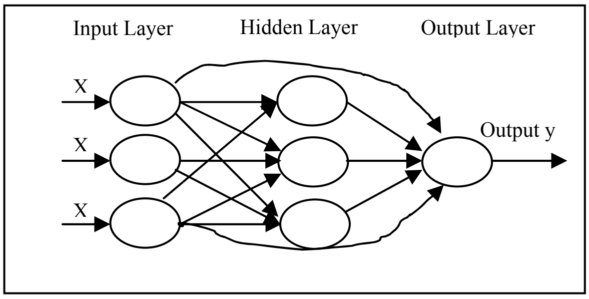

Fig. 3: Hybrid Multi-Layer Perceptrons Architecture

Transfer function of all neurons is the sigmoid function. The initial weights of the synapses is randomly chosen between the interval [-0.5,0.5].

4. Data Set

The data we used to train and test the classifiers were collected from a wireless local area network. The local network is composed of 3 wireless stations and one access point. One machine is used to generate normal traffic (HTTP, FTP). The second machine transmits simultaneously data originating from 4 types of attacks. The last station is used to collect and record both types of traffic (normal and intrusive). The attacks we used to test our system are: de-authentication, duration, fragmentation and chopchop. The source code of the attacks is available in [5].

The data collected were grouped in three sets: learning, validation and testing sets. The first set is used to reach the optimal weight of each synapse. The learning set contains the input with its desired output. By iterating on this data set, the neural network classifier dynamically adjusts the weights of the synapses to minimize the error rate between the output of the network and the desired output. Validation data set are necessary to avoid the effect of overfitting. Indeed, in some cases, the neural network classifier might produce an excellent performance on the learning data set, but still have a low performance on the testing data set. In general, the learning algorithm stops when the error between the output of the validation data set and the desired output is below a predefined threshold. Once the network is trained and validated, it should be able to predict the output of each entry of the testing data set.

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

Table 2: 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

4. Experimental results

Experimental results are obtained using NeuroSolutions software [8]. 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 (8 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. Indeed, learning time is reduced to 33% and the accuracy is increased by around 15% for the three types of neural networks architectures.

Table 3: Performance of the three types of neural networks using 8 and 38 features

	Learning Time (s)		Detection Rate (%)		False Positives (%)		False Negatives (%)	
	Optimal	Full	Optimal	Full	Optimal	Full	Optimal	Full
Perceptron	271	592	43.37	35.27	39.46	44.57	6.95	7.46
MLBP	349	967	95.99	82.87	3.02	8.93	0.38	2.37
Hybrid	356	1009	96.27	83.48	2.84	8.79	0.37	2.49

Table 3 summarizes the results of the experiments. False positives rate is the percentage of frames containing normal traffic classified as intrusive frames. False negatives rate is the percentage of frames generated from wireless attacks and classified as normal traffic.

5. Conclusion

In this paper we 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% with the reduced set of features, while the accuracy of detection is improved by 15%. In future work, we are planning to do a comparative study of the impact of the reduced feature set on the performance of classifiers based on support vector machines (SVMs), artificial neural networks (ANNs), multivariate adaptive regression splines (MARS) and linear genetic programs (LGPs).

6. References

[1] Y. Chen, Y. Li, X. Cheng, L. Guo, "Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System", Inscrypt 2006.

[2] M. Guennoun, A. Lbekkouri, K. El-Khatib, "Selecting the Best Set of Features for Efficient Intrusion Detection in Wireless Networks", 3rd International IEEE Conference on Information and Communication Technologies: From Theory to Applications, 2008.

[3] M. Guennoun, A. Lbekkouri, K. El-Khatib, “Optimizing the Feature Set of Wireless Intrusion Detection Systems”, International Journal of Computer Science and Network Security, VOL.8 No.10, October 2008.

[4] H. Liu, H. Motoda, “Feature Selection for Knowledge Discovery and Data Mining”, Boston: Kluwer Academic, 1998.

[5] J. Bellardo, S. Savage, “802.11 Denial-of-Service Attacks: Real Vulnerabilities and Practical Solutions.”, USENIX Security Symposium, pages 15-28, 2003.

[6] A. Bittau, M. Handley, J. Lackey, "The final nail in WEP's coffin", 2006 IEEE Symposium on Security and Privacy, May 2006.

[7] Z. Zhang, C. Manikopoulos, "Investigation of neural network classification of computer network attacks", International Conference on Information Technology: Research and Education, 2003. 11-13 Aug 2003 Page(s): 590 - 594

[8] NeuroSolutions Inc, http://www.neurosolutions.com/