Impact of Feature Reduction on the Efficiency of Wireless Intrusion Detection Systems

M.RAVIKANTH#1
(09491D5812 – M.Tech)
QISCET, Ongole,
PrakasamDt., A.P.,
India.
e-mail: ravi_kanth_m@yahoo.co.in

P.VENKATA RAVIKANTH#2
(09491D5816 – M.Tech)
QISCET, Ongole,
PrakasamDt., A.P.,
India.
e-mail: pvravikanth555@gmail.com

Abstract—Intrusion Detection Systems (IDSs) are a major line of defense for protecting network resources from illegal penetrations. A common approach in intrusion detection models, specifically in anomaly detection models, is to use classifiers as detectors. Selecting the best set of features is central to ensuring the performance, speed of learning, accuracy, and reliability of these detectors as well as to remove noise from the set of features used to construct the classifiers. In most current systems, the features used for training and testing the intrusion detection systems consist of basic information related to the TCP/IP header, with no considerable attention to the features associated with lower level protocol frames. The resulting detectors were efficient and accurate in detecting network attacks at the network and transport layers, but unfortunately, not capable of detecting 802.11-specific attacks such as deauthentication attacks or MAC layer DoS attacks.

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

1 INTRODUCTION

INTRUSIONS are the result of flaws in the design and implementation of computer systems, operating systems, applications, and communication protocols. Statistics [21] 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 [22]. These vulnerabilities are a result of the nature of the transmission media [26]. 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

To detect intrusions, classifiers are built to distinguish between normal and anomalous traffic.

2 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 problemof feature selection. In [7], Sung and Mukkamala used featureranking algorithms to reduce the feature space of theDARPAdata set from 41 features to the six most important features.They used three ranking algorithms based on Support VectorMachines (SVMs), Multivariate Adaptive Regression Splines(MARSs), and Linear Genetic Programs (LGPs) to assign aweight to each feature. Experimental results showed that theclassifier’s accuracy degraded by less than 1 percent whenthe classifier was fed with the reduced set of features.Sequential backward search was used in [8], [9] to identifythe important set of features: starting with the set of allfeatures, one feature was removed at a time until theaccuracy of the classifier was below a certain threshold.Different types of classifiers were used with this approachincluding Genetic Algorithms in [9], Neural Networks in [8],[10], and Support Vector Machines in [8].

3. 802.11-SPECIFIC INTRUSIONS

Several vulnerabilities exist at the link layer level of the802.11 protocol [24], [25]. In [11], many 802.11-specificattacks were analyzed and demonstrated to present a realthreat 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 simpleattack that exploits the vulnerability of the virtual carriersensing protocol CSMA/CA and it was proven in [11] todeny access to the network.

Most of the attacks we used in this work are available fordownload from [12]. The attacks we used to conduct the
experiments are:

3.1 Deauthentication Attack

The attacker fakes a deauthentication frame as if it hadoriginated from the base station (Access Point). Uponreception, the station disconnects and tries to reconnect tothe base station again. This process is repeated indefinitelyto keep the station disconnected from the base station. Theattacker can also set the receiving address to the broadcastaddress to target all stations associated with the victim basestation. However, we noticed that some wireless networkcards ignore this type of deauthentication frame. Moredetails of this attack can be found in [11].

3.2 ChopChop Attack

The attacker intercepts an encrypted frame and uses theAccess Point to guess the clear text. The attack is performed asfollows: Theintercepted encrypted frame is chopped from the last byte. Then, the attacker builds a new frame 1 byte smallerthan the original frame. In order to set the right value for the 32 bit longCRC32 checksum named ICV, the attacker makes aguess on the last clear byte. To validate the guess he/shemade, the attacker will send the new frame to the base stationusing a multicast receive address. If the frame is not valid (i.e.,the guess is wrong), then the frame is silently discarded by theaccess point. The frame with the right guess will be relayedback to the network. The hacker can then validate the guesshe/she made. The operation is repeated until all bytes of theclear frame are discovered. More details of this attack can befound in [16].

3.3 Fragmentation Attack

The attacker sends a frame as a successive set of fragments.The access point will assemble them into a new frame andsend it back to the wireless network. Since the attacker knowsthe clear text of the frame, he can recover the key stream usedto encrypt the frame. This process is repeated until he/shegets a 1,500 byte long key stream. The attacker can use the keystream to encrypt new frames or decrypt a frame that usesthe same three byte initialization vector IV. The process can berepeated until the attacker builds a rainbow key stream tableof all possible IVs. Such a table requires 23 GB of memory.More details of this attack can be found in [16].

3.4 Duration Attack

The attacker exploit a vulnerability in the virtual carrier-sensemechanism and sends a frame with theNAVfield set to a highvalue (32 ms). This will prevent any station from using theshared medium before the NAV timer reaches zero. Beforeexpiration of the timer, the attacker sends another frame. Byrepeating this process, the attacker can deny access to thewireless network. More details can be found in [11].

4 HYBRID APPROACH

Extensive work has been done to detect intrusions in wiredand wireless networks. However, most of the intrusiondetection systems examine only the network layer andhigher abstraction layers for extracting and selectingfeatures, and ignore the MAC layer header. These IDSscannot detect attacks that are specific to the MAC layer.

Some previous work tried to build IDS that functioned atthe Data link layer. For example, in [13], [14], [15], theauthors simply used the MAC layer header attributes asinput features to build the learning algorithm for detectingintrusions. No feature selection algorithm was used toextract the most relevant set of features.

In this paper, we will present a complete framework toselect the best set of MAC layer features that efficientlycharacterize normal traffic and distinguish it from abnormaltraffic containing intrusions specific to wireless networks.Our framework uses a hybrid approach for feature selectionthat combines the filter and wrapper models. In thisapproach, we rank the features using an independentmeasure: the information gain ratio. The k-means classifier'spredictive accuracy is used to reach an optimal set offeatures which maximize the detection accuracy of thewireless attacks.

To train the classifier, we first collectnetwork traffic containing four known wireless intrusions,namely, the deauthentication, 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.

chopchop 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 featuresfrom the ranked set of features F into S sequentially. Aftereach iteration, the "goodness" of the resulting set offeatures S is measured by the accuracy of the k-meansclassifier. The selection process stops when the gainedclassifier's accuracy is below a certain selected thresholdvalue or in some cases when the accuracy drops, whichmeans that the accuracy of the current subset is below theaccuracy of the previous subset.

5 INITIAL LIST OF FEATURES

The initial list of features is extracted from the MAC layerframe header. According to the 802.11 standard [17], thefields 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.

6 INFORMATION GAIN RATIO MEASURE

We used the Information Gain Ratio (IGR) as a measure todetermine the relevance of each feature. Note that we chosethe IGR measure and not the Information Gain because thelatter is biased toward the features with a large number ofdistinct values [5].

IGR is defined in [18] as

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

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, Data).
SubType	Indicate the subtype of the frame.
ToDS	Indicate if a frame is destined to the Distribution System.
FromDS	Indicate if a frame is originated from Distribution 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 the WEP protocol.
Order	Indicate if the "strict ordering" delivery is employed.
Duration	The number of microseconds the medium is expected to be busy
RA	The MAC address of the receiving station.
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

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

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 (>20 ms) .
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	< 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.

7 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 canconclude 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.

8 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.

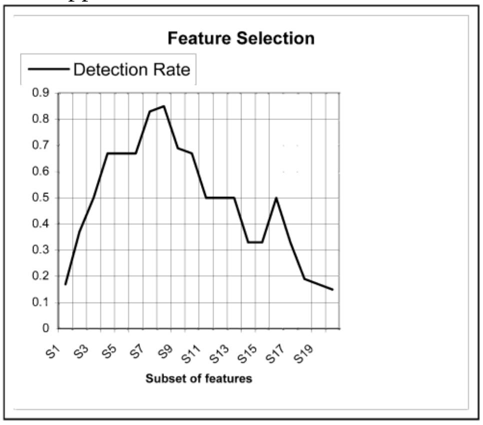

Fig. 2. Detection rate versus subset of features.

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.

TABLE 4
Distribution of Collected Data

	Learning	Valida-tion	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

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].

9 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

The data collected were grouped in three sets (Table 4): 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.

Fig. 3. Learning time (in seconds) for the three types of neural networks using 8 and 38 features.

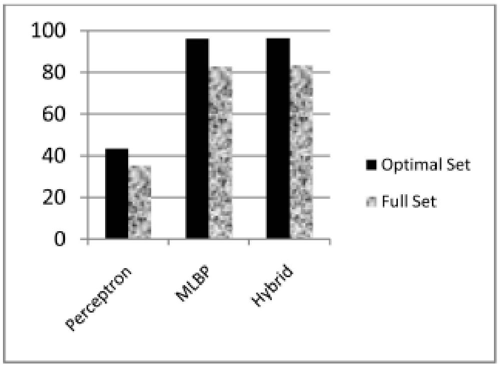

Fig. 4. Detection Rate percentage of the three types of neural networks using 8 and 38 features.

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

10 EXPERIMENTAL RESULTS

Experimental results were obtained using NeuroSolutions 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

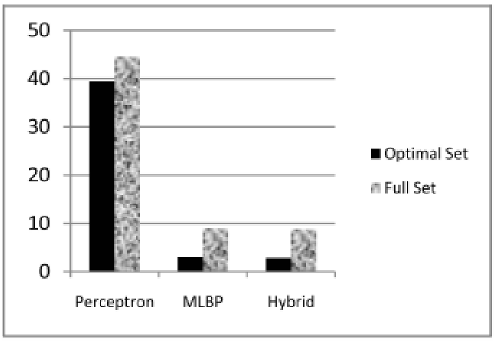

Fig. 5. False Positives Rate (%) for the three types of neural networks using 8 and 38 features.

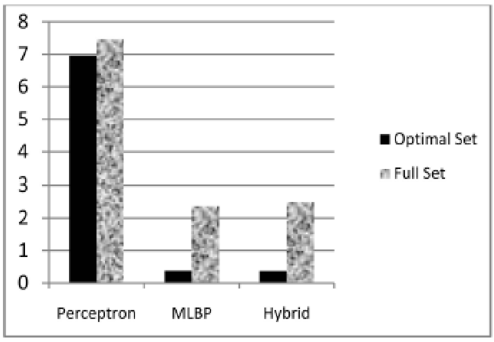

Fig. 6. False Negatives Rate (%) for the three types of neural networks using 8 and 38 features.

intrusive frames. Likewise, the false negatives rate is thepercentage 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.

11 CONCLUSION and FUTURE WORK

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. 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 ANNs, in comparison with other computational models such as the ones based on SVMs, MARSs, and LGPs.

ACKNOWLEDGMENTS

This work is partially supported by NSERC, the Natural Science and Engineering Research Council of Canada.

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.
[10] A. Hofmann, T. Horeis, and B. Sick, “Feature Selection for Intrusion Detection: An Evolutionary Wrapper Approach,” Proc. IEEE Int’l Joint Conf. Neural Networks, July 2004.
[11] J. Bellardo and S. Savage, “802.11 Denial-of-Service Attacks: Real Vulnerabilities and Practical Solutions,” Proc. USENIX Security Symp., pp. 15-28, 2003.
[12] http://www.aircrack-ng.org/, 2010.
[13] 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.
[14] T.M. Khoshgoftaar, S.V. Nath, S. Zhong, and N. Seliya, “Intrusion Detection inWireless Networks Using Clustering Techniques with Expert Analysis,” Proc. Fourth Int’l Conf. Machine Learning and Applications, Dec. 2005.
[15] S. Zhong, T.M. Khoshgoftaar, and S.V. Nath, “A Clustering Approach to Wireless Network Intrusion Detection,” Proc. 17th IEEE Int’l Conf. Tools with Artificial Intelligence (ICTAI ’05), Nov. 2005.
[16] A. Bittau, M. Handley, and J. Lackey, “The Final Nail in WEP’s Coffin,” Proc. IEEE Symp. Security and Privacy, May 2006.
[17] IEEE 802.11-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher Speed Physical Layer Extension in the 2.4 GHz Band.
[18] J.R. Quinlan, “Induction of Decision Trees,” Machine Learning, vol. 1, pp. 81-106, 1986.
[19] Z. Zhang and C. Manikopoulos, “Investigation of Neural Network Classification of Computer Network Attacks,” Proc. Int’l Conf. Information Technology: Research and Education, pp. 590-594, Aug. 2003.
[20] NeuroSolutions, Inc., http://www.neurosolutions.com/, 2010.
[21] CERT, http://www.cert.org/stats/, 2010.