IJCSC International Journal of Computer Science & Communication Vol. 1, No. 1, January-June 2010, pp. 165-169

Anomaly Intrusion Detection System using Hamming Network Approach

[1]Muna M. Taher Jawhar & [2]Monica Mehrotra

Department of Computer Science, Jamia Millia Islamia, New Delhi, India
[1]muna.taher@gmail.com, [2]drmehrotra2000@gmail.com

ABSTRACT

Intrusion detection is an interesting approach that could be used to improve the security of network system. IDS detects suspected patterns of network traffic on the remaining open parts through monitoring user activities. The major problems of existing models is recognition of new attacks, low accuracy, detection time and system adaptability. In this paper, evolving anomaly intrusion detection system is constructed using hamming and MAXNET Neural Network for recognize attack class in the network traffic. The result is encouraging, the detection rate is 95% which is relatively high. We describe another approach based on Multilayer Perceptrons (MLP) network and compare the results of the two approaches to evaluate the system. The experimental results demonstrate that the designed models are promising in terms of accuracy and computational time of real word intrusion detection. Training and testing data obtains from the Defense Advanced Research Projects Agency(DARPA) intrusion detection evaluation datasets.
Keywords: Intrusion Detection system, Hamming Network, Neural Network, KDD dataset.

1. INTRODUCTION

Incessant distribution of application of information technologies to all spheres of human activity constantly puts new requirements to a level of security of information system. The number of attacks and criminals concerning computer network is increases[1]. So the network security has become a very important issue. The intrusion detection has become research focus of the network security. The intrusion detection technology uses the trace information which are left by the intruder such as the failure records of attempt to log to find the illegal intrusion from the outsider or insider effectively. The intrusion detection system is the computer system which can realize the intrusion detection technology[2].

Intrusion detection systems(IDS) can be classified as network based and host-based according to the information source of the detection. Network-based IDS monitors the network traffic and looks for network-based attacks, while host-based IDS is installed on host and monitors the host audit trail. Intrusion detection systems can be roughly classified as anomaly detection and misuse detection. Anomaly detection is based on the normal behavior of a subject (e.g., a user or a system), any action that significantly deviates from the normal behavior is considered intrusive. Misuse detection is based on the characteristics of known attacks or system vulnerabilities, which are also called signatures. Any action that matches the signature is considered intrusive. Misuse-base detection detects attacks based on signatures (known attacks signatures), at which the traffic pattern compared with these signatures, if a match is found, then it is reported as an attack, otherwise it is not. So misuse detection cannot detect novel attacks. On the other hand, anomaly-based detection depends on monitoring system activity and classifying it as either normal or anomalous. The classification is based on heuristics or rules, rather than patterns or signatures, and will detect any type of misuse that falls out of normal system behavior. Thus, it is able to detect not only known intrusion but also unknown intrusion. In addition, this approach can detect the intrusion that is achieved by the abuse of legitimate users or masqueraders without breaking security policy[3][4].

However, most available commercial IDS's use only misused detection and the major problem of existing models is recognition of new attacks, low accuracy and detection time. In this paper, we propose a new method for anomaly detection by using hamming and MAXNET network, the results from this model compared with the MLP Neural Network which depends on the deferent type of learning algorithms. Training and testing data were obtained from the KDD intrusion detection evaluation datasets.

2. PREVIOUS WORK

Neural Networks(NNs) approach is one of the most interesting in this area. An increasing amount of research in the last few years has investigated the application of Neural Networks to intrusion detection. If properly designed and implemented, Neural Networks have the potential to address many of the problems encountered

by rule-based approaches. Neural Networks were specifically proposed to learn the typical characteristics of system's users and identify statistically significant variations from their established behavior. In order to apply this approach to Intrusion Detection, we would have to introduce data representing attacks and non-attacks to the Neural Network to adjust automatically coefficients of this Network during the training phase. In other words, it will be necessary to collect data representing normal and abnormal behavior and train the Neural Network on those data. After training is accomplished, a certain number of performance tests with real network traffic and attacks should be conducted[5]. And instead of processing program instruction sequentially, Neural Network based models on simultaneously explorer several hypotheses making the use of several computational interconnected elements (neurons), this parallel processing may imply time savings in malicious traffic analysis[6].

In particular several Neural Networks based approaches were employed for Intrusion Detection. Tie and Li[7] used the BP network with GAs for enhance of BP, they used some type of attack with some feature of KDD data. The detection rate was for satan 90.97, guess-password 85.60 and peral 90.79. Jimmy and Heidar[8] used feedforward Neural Networks with Back Propagation training algorithm, they used some feature from TCP Dump and the classification result is 25/25. Dima, Roman and Leon[9] used Radial Based Function (RBF) Neural Network for classification and the accuracy result is 93.2. Iftikhar, Sami and Sajjad[10] used Resilient Back propagation for detect each type of attack along the accurse detection rate was 95.93. Mukkamala, Andrew, and Ajith[11] used Back Propagation Neural Network with many type of learning algorithm the performance of the network is 95.0.

3. KDD INTRUSION DETECTION EVALUATION DATASETS

KDD 99 data set are used as the input vectors for training and validation of the tested neural network. It was created based on the DARPA(Defense Advanced Research Project Agency) intrusion detection evaluation program[12]. MIT Lincoln Lab that participates in this program has set up simulation of typical LAN network in order to acquire raw TCP dump data[13]. They simulated LAN operated as an normal environment, which was infected by various types of attacks. The raw data set was processed into connection records. For each connection, 41 various features were extracted. Each connection was labeled as normal or under specific type of attack. There is 39 attacker types that could be classified into four main categories which summarized in Table 1.

The four main categories of attacks :

- *DOS (Denial of Service):* An attacker tries to prevent legitimate users from using a service e.g. TCP SYN Flood, Smurf (229853 record).
- *Probe:* An attacker tries to find information about the target host. For example: scanning victims in order to get knowledge about available services, using Operating System (4166 record).
- *U2R (User to Root):* An attacker has local account on victim's host and tries to gain the root privileges (230 record).
- *R2L (Remote to Local):* An attacker does not have local account on the victim host and try to obtain it (16187 record).

Table 1
Attacks Type in KDD Dataset

Dos	*U2R*	*Probe*	*U2L*
Back Land Neptune Pod Smurf Teardrop	Buffer-overflow Load module Perl Rootkit	Ipsweep Nmap Portsweep Satan	ftp-write guess-passwd imap multihop phf spy warezclient warezmaster

4. THE SYSTEM MODEL ARCHITECTURE

The Hamming Network in cooperation with a MAXNET is used to identify the class that contain a given input pattern. The pattern is identified through a stored prototype pattern set (examplar dataset). The input data pattern is associated to the most similar examplar, in other words, the class that has the smallest hamming distance. The hamming distance[6] is the number of bits within an input pattern that does not match the corresponding bits in an examplar. The architecture and main components of our system is shown in the following Figure 1.

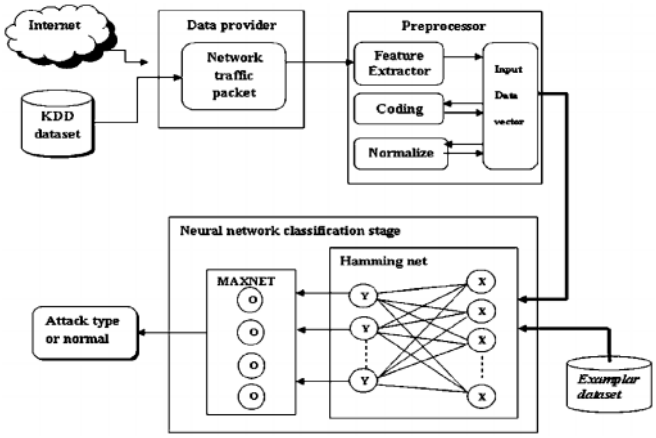

Fig 1: The Model of Hamming and MAXNET System

The components of the model of hamming and MAXNET system are:

a. *Data Provider:* For the experimental environment, it was difficult to obtain real life dataset directly from the internet. Therefore we are using the data from KDD cup99 data set. But in the future work and after test the system we will be using real data from the internet by using one of the sniffer tools. Data provider collects data from KDD dataset file and send text data to the preprocessor component.

b. *Preprocessor:* The Preprocessor component gets traffic data from data provider, extracts appropriate features, convert features into numeric form and then convert to binary bipolar form in order to feed the neural net sensors in Neural Network based analyzer component. And, sent them to the network classification stage. Therefore the following operation are applied to the original dataset.

- *Feature Extractor:* During inspection of the data it turned out that the values of six features (num-shell, land, urgent, num-outbound-cmds, is-host-login, and su-attempted) were constantly zero over all data records see[15] for descriptions. Clearly these features could not have any effect on classification and only made it more complicated and time consuming. They were excluded from the data vector. Hence the data vector was a 35 dimensional vector.
- *Encoding:* The KDD data set are used in the work including symbolic attributes, such as "*protocol type*", with values "TCP","UDP", and "ICMP", feature "*services*" with values "private", "HTTP","ECR_I", ect …, and other feature "*flag*", with values "SF","REJ","RSTR",… At the beginning this features convert to numerical as TCP=1,UDP=2, ICMP=3 and so on.
- *Normalization:* In the normalization, each numerical value in the data set is normalized in form 1 and -1. All the features which have integer value or continuous convert to binary bipolar.

c. *Hamming and MAXNET Network Classification:* In this stage there are two components first is the Hamming net while the second is MAXNET. The input vector to the hamming net is composed of 84 bits (vector X), exampler data set have m pattern taken from KDD dataset in random way. The output is a Y vector having the length m, Y can be calculated as:

$$Y_m = (1/2)X^m + n/2 \quad (1)$$

Where $n = 84$, the number of bits in X. The strongest response of a neuron is indicative of the minimum hamming distance between the input vector and the category which neuron represents.

The second layer of the classifier component is MAXNET and it operates as a recurrent recall network. The output Y vector is input to MAXNET which will strengthen the largest value and will eliminate the others. In other words, only one neuron in MAXNET will be the winner corresponding to the examplar index that matches the input. The output of MAXNET is 5 nodes which represents the normal, DOS, R2L, Probe, and R2L classes. The equation using in this net is:

$$Y^{k+1} = F_{net}(W_m Y^k) \quad (2)$$

0 if output <0

$$F_{net} = \begin{cases} 0 & \text{if output} < 0 \\ \text{Output} & \text{if output} >= 0 \end{cases} \quad (3)$$

5. THE EXPERIMENTAL RESULTS

To assess the effectiveness of proposed intrusion detection approaches, the series of experiments were performed. We have used 6186 training samples for learning of neural network. Proposed intrusion detection approaches are implemented to detect 5 classes of attacks from the dataset including Dos, U2R, Probe, U2L and normal. The distribution of an attack and normal records are 80%-20%. To evaluate our system, there are two major indicators of performance: the detection rate for each attack class, false positive and false negative rate. The detection rate (true attack alarms) is defined as the number of intrusion instances detected by the system divided by the total number of intrusion instances present in the test set. The false positive rate (false attack alarms) represents the total number of normal instance that were classified as intrusion divided by the total number of normal instances. The overall classification rate of the system is 95.0% and false negative is 4.94% for all type of the attack class and the time cover the operation of detection is 7.1414ms. Table 2 represents the results of these experiments.

Table 2
Error Detection for each Attack Class

Attack class	DOS	U2R	Probe	U2l	Normal
Detection error	0.096	0.428	0.145	0.138	0.9

The error detection rate for each type of the attack class is shown above. And table 3 describe true negative(TN) and true positive(TP) for all type of attack class.

Table 3
The TN and TP of each Attack Class

Attack class	DOS	U2R	Probe	U2l
TN	0.912	0.571	0.854	0.861
TP	0.088	0.429	0.146	0.139

To perform the modeling work, compare the result of hamming model with other architecture of neural network like MLP which is the most famous network used by most of the researchers. The structure of MLP which are used is explained as follows: input layer is 84 nodes, two hidden layers with 10 nodes and output layer is 5 nodes. We used the same data input which we have used in hamming net. We used different learning algorithms. The result of MLP network is shown in Table 4.

Table 4
The Result of Running the MLP Network for different Learning Algorithms

Training algorithm	*No. of Epochs*	*Time of detection*	*Accuracy (%)*
Resilient back propagation	63	6.0759	95.6
Gradient descent with momentum	1000	28.9332	71.7
Levenberg-Marquardt	12	6.1841	99.1
One step secant	37	7.7795	94.5
Scaled conjugate gradient	25	5.3339	93.4
BFGS quasi-newton	7	7.0108	84.6

We can see the performance of each approach of the network in the following figures. By choosing from MLP approach the best two learning algorithm which is Levenberg-Marquardt and Resilient algorithms.

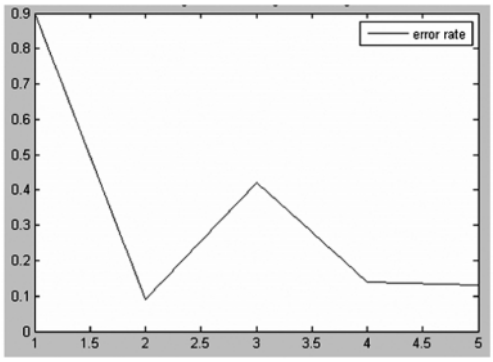

Fig 2 : The Performance of Hamming and MAXNET Approach

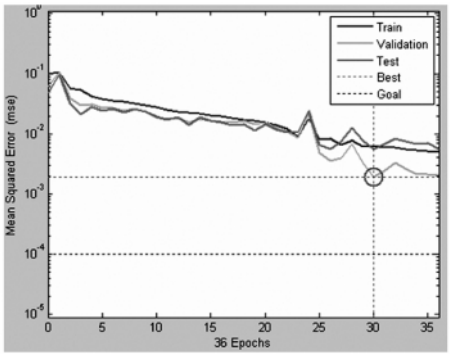

Figure 3: The Performance of MLP with Resilient Function

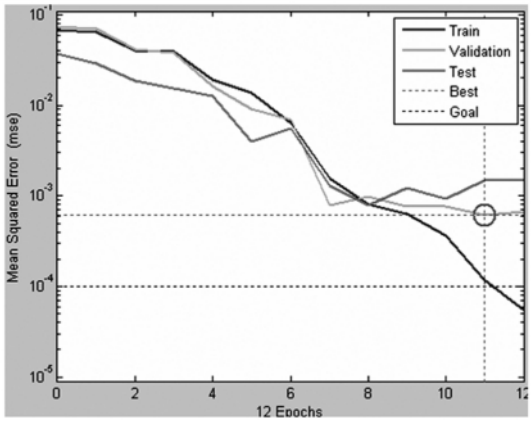

Figure 4 : The Performance MLP Approach with Levenberg-Marquardt Function

The system are implemented under windows XP operating system by using Matlab R2008a as a programming language.

6. CONCLUSION

Network Intrusion Detection System is a hot field of the network security research, and it is a new kind of defense technology of the network security. Usage of neural network for intrusion detection was present in many publication. Unfortunately, in description of simulation process very often is lack of recognition of new attacks, low accuracy detection rate. In this paper, we propose a new method by using Hamming and MAXNET for anomaly Intrusion Detection System and comparison with MLP network working with the same assumed parameters and testing with the usage of KDD dataset. The results are encouraging. The detection rate of the model is 95.0% and false negative is 4.94% which is relatively high when compared with conventional IDS and other design with neural network.

7. ACKNOWLEDGMENT

This work has been financially supported by the Indian Council of Cultural Relations (I.C.C.R.), India. It has been also partially subsidized by the University of Mousl, Ministry of Higher Education and scientific Research, Iraq.

REFERENCE

[1] Vladimir Golovko, Pavel Kachurka, and Leanid Vaitsekhovich, 2007, "Neural Network Ensembles for Intrusion Detection", *IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Dortmund, Germany.

[2] Jingwen Tian and Meijuan Gao, 2009, " Network Intrusion Detection Method Based on High Speed and Precise Genetic Algorithm Neural Network", *2009 International Conference on Networks Security, Wireless Communications and Trusted Computing, IEEE*.

[3] Khattab M. Ali, Venus W, and Mamoun Suleiman Al Rababaa, 2009, "The Affect of Fuzzification on Neural Networks Intrusion Detection System", IEEE.

[4] Jawhar, Muna M. T. and Monica M., "Intrusion Detection System: A Design Perspective", *the Proceeding of International Conference for Data Management*, IMT, Gaziabad, India, 2009.

[5] Dima Novikov, Roman V. Yampolskiy, and Leon Reznik, 2006, "Artificial Intelligence Approaches For Intrusion Detection", IEEE.

[6] Lília de Sá Silva, Adriana C. Ferrari dos Santos, José Demisio S. da Silva, and Antonio Montes, 2004, "A Neural Network Application for Attack Detection in Computer Networks", Instituto Nacional de Pesquisas Espaciais – INPE, BRAZIL.

[7] TIE-JUN Zhou and LI Yang, 2008, "The Research of Intrusion Detection Based on Genetic Neural Network", *Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition*, Hong Kong, IEEE, 30-31 Aug.

[8] Jimmy Shum and Heidar A. Malki, 2008, "Network Intrusion Detection System Using Neural Networks", *Fourth International Conference on Natural Computation*, IEEE.

[9] Dima Novikov, Roman V. Yampolskiy and Leon Reznik, 2006, " Anomaly Detection Based Intrusion Detection", *Proceedings of the Third International Conference on Information Technology: New Generations, IEEE*.

[10] Iftikhar Ahmad, Sami Ullah Swati and Sajjad Mohsin, 2007, " Intrusions Detection Mechanism by Resilient Back Propagation (RPROP)", *European Journal of Scientific Research,* ISSN 1450-216X, **17**, No.4, pp.523-531.

[11] Srinivas Mukkamala, Andrew H. Sung, and Ajith Abraham, 2005, " Intrusion Detection using an Ensemble of Intelligent Paradigms", *Journal of Network and Computer Applications,* **28**, p167–182.

[12] KDD-cup Dataset, http://kdd.ics.uci.edu/data base/ kddcupaa/kddcup.html.

[13] PrzemysBaw KukieBka and Zbigniew Kotulski, 2008, " Analysis of Different Architectures of Neural Networks for Application in Intrusion Detection Systems", *Proceedings of the International Multiconference on Computer Science and Information Technology, IEEE*, pp. 807– 811.

[14] Morteza Amini and Rasool Jalili, 2005, "Network-Based Intrusion Detection Using Unsupervised Adaptive Resonance Theory (ART)", *Computer Engineering Department, Sharif University of Technology*, Tehran, Iran.

[15] Mehdi Moradi and Mohammad Zulkernine, 2004, "A Neural Network Based System for Intrusion Detection and Classification of Attacks", School of Computing, Queen's University, Kingston, Ontario, Canada.