A Novel Approach for Intrusion Detection

Vikram Chopra[1], Sunil Saini [2] and Amit Kumar Choudhary [3]

[1]Department of Electrical and Instrumentation Engineering
Thapar University, Patiala
Punjab, 147004, India

[2] Department of Computer Science and Engineering
Kurukshetra Institute of Technology and Management, Kurukshetra
Haryana, 136119, India

[3] Department of Electrical Engineering
National Institute of Technology, Kurukshetra
Haryana, 136119, India

Abstract

Research in the field of computer and network science demands for tools and methodology to test their security effectively. Intrusion Detection System is used to perform the same with a fact that an intruder's behavior will be noticeably different from that of a legitimate user and would exploit security vulnerabilities. Proposed here is a novel intrusion detection approach with the application of Generalized Regression Neural Network and the MIT's KDD Cup 99 dataset. The result clearly demonstrates an efficient way for intrusion feature selection and detection and promises a good scope for further research.

Keywords: Intrusion Detection System, Networking Attacks, Intrusion Detection, Generalized Regression Neural Network.

1. Introduction

Explosive growth in Internet connectivity and accessibility has created a tremendous threat to information systems worldwide, thus increasing network security. Despite of careful policies, network security is still difficult to guarantee. Intrusion Detection (ID), act as the "second line of defense" is becoming a critical technology to help protecting computer network systems. Essentially speaking based on its functionality; intrusion detection can be considered as a classification problem to define TCP/IP connections as normal/intrusion, therefore GRNN can play an important role on it. The goal is to discover breaches of security, attempted breaches, or open vulnerabilities that could lead to potential breaches. IDS's are based on the belief that an intruder's behaviour will be noticeably different from that of a legitimate user.

Publication by John Anderson's *Computer Security Threat Monitoring and Surveillance* followed by D.Denning's seminal paper, "An Intrusion Detection Model," provided a methodological framework that inspired many researchers and laid the groundwork for commercial products. The analysis relies on sets of predefined rules that are provided by an administrator or created by the system.

A number of soft computing based approaches have been proposed for detecting network intrusions. Soft computing refers to a group of techniques that exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve robustness and low solution cost. The principle constituents of soft computing are Fuzzy Logic (FL), Artificial Neural Networks (ANNs), Probabilistic Reasoning (PR), and Genetic Algorithms (GAs) . Fox, Henning, Reed, and Simmonian [1] were the first to attempt modeling system using neural networks. Their choice of neural network was Kohonen's *self-organizing map* (SOM). In another attempt to apply neural network to anomaly detection Ghosh, Wanken, and Charron [2] proposed backpropagation network to monitor running programs. Some recent studies on the application of the Neural Network approach to the scope of Intrusion Detection are as Cannady [3] of Georgia Technical Research Institute conducted research to apply Multi-Level Perceptron (MLP) model and SOM for misuse detection. The final result succeeded in 89-91% of the cases. In yet another study by Cunningham and Lippmann [4] of the MIT Lincoln Laboratory used a MLP model. With the Neural Network approach, false alarms were reduced and the detection rate increased to roughly 80% with the DARPA database. Then Ryan, Lin, and Miikkulainen described an off-line anomaly detection system, which utilized a back-propagation MLP neural

network. Another study by Mukkamala [5], described the three and four layer neural networks and reported results of about 99.25% correct classification for their two class (normal and attack) problem. This paper is aimed to solve an off-line multi class problem using regression method in which not only the attack records are distinguished from normal ones, but also the attack type is identified. The promising results of the present study show the potential applicability of ANNs for developing high efficiency practical IDSs.

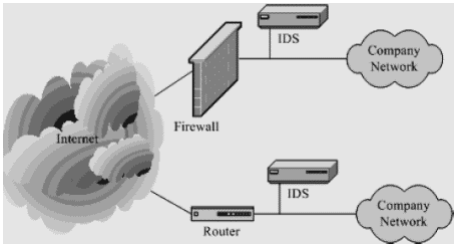

Fig 1. Intrusion Detection System

In order to evaluate the performance of applying varied GRNN based methods to intrusion detection, the 1998 DARPA Intrusion Detection Evaluation project was conducted by MIT Lincoln Labs, which set up a LAN network and logged normal and attack network traffic. These records were reduced and processed by domain experts to yield KDD Cup 99 dataset for competition [6].
In this paper, the GRNN based intrusion feature selection and detection algorithm is proposed with the motivation to start. The rest of paper is organized with section 3, a brief about Intrusion Detection System and focus on the development of ANN to GRNN. Section 4, presents the methodology and experimental analysis. At last, section 5 concludes the paper with future scope.

2. Motivation

Network security has number of enemies namely hackers, unaware disgruntled staff and snoopers, creating computer programs like viruses, Trojan horse programs, vandals and spam thus, harming the network security in one or the other ways. Thus, there has been an extensive choice of technologies, ranging from anti-virus software packages to dedicated security hardware such as firewalls. Organizations continue to deploy firewalls as their central gatekeepers and are increasingly looking to additional security technologies to counter risk. An intrusion detection system provides around-the-clock network surveillance analyzing and searching for unauthorized activity. Thus, Intrusion detection is a viable and practical approach for providing a different notion of security in today's huge and existing infrastructure of computer and network systems.

3. Brief Study

3.1 Intrusion Detection System

Act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a resource can be a possible definition of Intrusion Detection. More specifically, the goal of intrusion detection is to identify entities attempting security controls. Intrusion detection is receiving considerable attention since two decades; as a mechanism for keeping administrators informed on potential security breaches and suspicious network activity.
An IDS (Intrusion Detection Systems) is the term for a mechanism which quietly listens to network traffic in order to detect abnormal or suspicious activity, thereby reducing the risk of intrusion. The IDS's goal is to recognize "specific, precisely represent able techniques of computer system abuse." Data generated by intrusion detection systems is carefully examined (this is the main task of each IDS) for detection of possible attacks (intrusions). The step after the detection of intrusion is taken either by the administrators or the IDS itself. ID systems are being developed in response to the increasing number of attacks on major sites and networks, including those of the Pentagon, the White House, NATO, and the U.S. Defense Department. Today, some of the free IDS available are Snort, Endian Firewall, Untangle, Bro NIDS and many more.

3.2 Generalized Regression Neural Network

Neural Network has an important property to automatically grasp/produce coefficients according to inputs and outputs, provided it has been given enough time to train, ensuring that the network will classify the patterns that have never been seen before. The main importance of the neural network approach to intrusion detection is to learn the behavior of actors in the system (e.g., users, daemons) to obtain new relationship between input-output in a descent way. While studying about NN, feed forward and back propagation method are always in focus. Then Specht proposed an alternative to these by introducing the Generalized Regression Neural Network (GRNN) paradigm [8] which is very closely related to the probabilistic neural network. This is based on the estimation of a probability density function. It utilizes a probabilistic model between the independent vector random variable X with dimension D, and dependent scalar random variable Y. Assume that x and y are the

measured values for X and Y variables, respectively. If $f(x,Y)$ represents the known joint continuous probability density function, and is known, the expected value of Y given x (the regression of Y on x) can be estimated as

$$E[Y/x] = \frac{\int_{-\infty}^{\infty} Yf(x,Y)\,dY}{\int_{-\infty}^{\infty} f(x,Y)\,dY}. \quad (1)$$

Based on p sample observations that are available, i.e., on the training set given by x and y, further assuming that the underlying density is continuous and the first partial derivatives of the function evaluated at any x are small, the probability estimator $\hat{f}(x,y)$ can be written as

$$\hat{f}(x,y) = \frac{1}{(2\pi)^{D+1/2}\sigma^{D+1}}\frac{1}{p}\times\sum_{i=1}^{p}\left[\exp\left(\frac{(x-x_i)^T(x-x_i)}{2\sigma^2}\right)\exp\left(\frac{(y-y_i)^2}{2\sigma^2}\right)\right] \quad (2)$$

Where x_i and y_i are the i_{th} training set data, and x_i denotes the vector form of variable x.

A physical interpretation of the probability estimate $\hat{f}(x,y)$ is that it assigns a sample probability of width σ for sample x_i and y_i, after that, the probability estimate is the sum of those sample probabilities. Substituting (2) into (1), the desired conditional mean of Y given $x, \hat{y}$ can be calculated as

$$\hat{y}(x) = E[Y/x] = \sum_{i=1}^{n}[y_i \exp(d_i)] / \sum_{i=1}^{n}\exp(d_i) \quad (3)$$

where d_i is given by the distance function of the input space and can be written as

$$d_i = \left[-\left(\frac{s-s_i}{\sigma}\right)^2\right]. \quad (4)$$

In the above expression s is the new input and s_i is the stored input, σ is the spread factor. A good estimation of d_i depends on the selection of spread factor σ.

4. Experiment Results

Evaluation of the proposed off-line IDS system, is performed with the use Knowledge Discovery in Database (KDD) Cup 99dataset [6] supplied by the Defence Advanced Research Projects Agency (DARPA) and the Massachusetts Institute of Technology's Lincoln Labs in 1998. Various quantitative and qualitative features were extracted, for each TCP/IP connection. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type.

4.1 Training of the dataset using GRNN

The symbolic representation has been used to express each of the three conditions in such a way that, a "1" in a column indicates the occurrence of the column's corresponding string and a "0" indicates a nonoccurrence. Thus, we have three cases of classification probabilities, that is, [1 0 0] for Normal conditions, [0 1 0] for Neptune attack and [0 0 1] for the Satan attack.

Input: Protocol Types, Training and testing dataset.

Output: Desired Attack.

Initialize the features of the dataset, Including protocol and also Initialize the condition for normal, Neptune and satan. State the syntax for GRNN and declare the real-time input. Simulate the real-time neural network input and Check to obtain result for any of the three attacks.
Above explained is the algorithm on which GRNN is used for function approximation in the present study.

4.2 Methodology and Experimental Analysis

Moradi and Zulkernine [9] were the first to implement intrusion detection with a three layer MLP (two hidden layers with 35 neurons in each) referred to as: {35 35 35 3}. At this stage, early stopping validation was not applied and the training was performed for 200 times and the process took more than 25 hours. Then he used 900 datasets (300 of each type) and trained it. It took almost 5hours because of the implementation of early stopping validation method. With this, the training error decreased at the 45th epochs that can be seen in fig 2. Figure shows the mean square error of the back propagation training process versus the progress of training epochs. The correct classification rate was more than 90% showing an 11% increase from the former experiment.

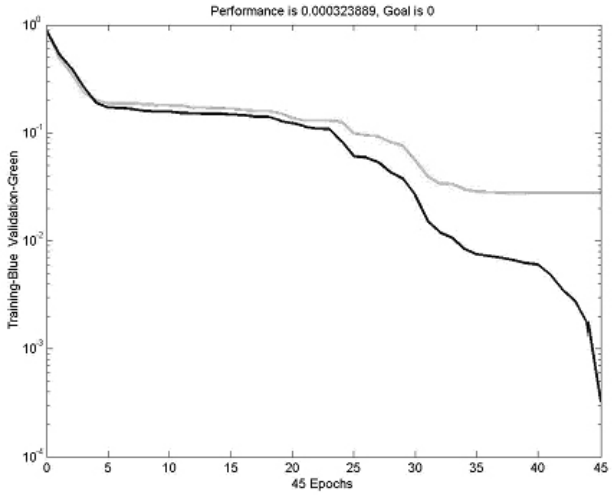

Fig. 2. The training process error when the early stopping validation method is applied. The darker curve shows the error on the training set and the brighter curve presents the error on validation set.

More than 99% of correct classification on the dataset using the neural network structure was reported before [5]. In another study with different dataset, the success rate was comparable to the results of the present study (89-99%) and has a two class problem.

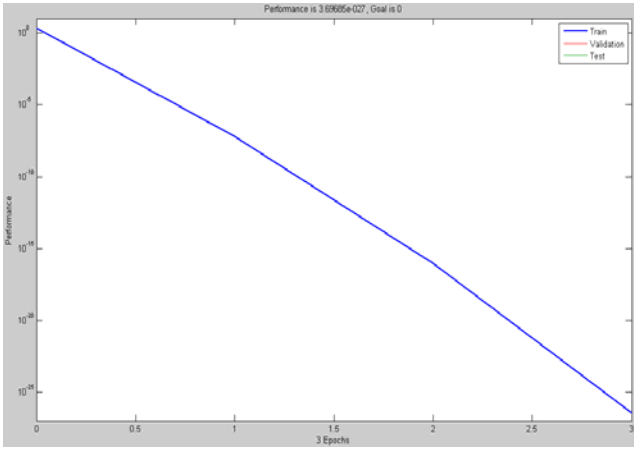

Fig 3. The Training process error. Here, validation and training dataset coincide each other (Darker part).

Use of GRNN achieved the best performance among the neural network learning algorithms in terms of accuracy (100 %) and faster convergence (3 epochs) as compared to the earlier result reported [9]. Result obtained here is with a static data i.e. offline data fig 3.

5. Conclusion

Neural network has given a lot to the computing world. And intrusion detection is an open scenario which has been, and will continue to, develop rapidly. Some of the present techniques and solutions found in current systems are outlined in this paper. Here a multi class problem was solved by using NN. The presented technology shows a significant promise as the results obtained reveals an accuracy of approximately 100%, and our future work will involve the refinement of this approach and the development of a full-scale demonstration system, with fast validation. The intrusion detection is expected to become a practical and effective solution for protecting information systems.

References

[1] K.L.Fox, R.R. Henning, J.H. Reed, and R.P.Simonian, " A neural network approach toward intrusion detection," Proceedings of 13th National Computer Security Conference, 1990, pp. 125-134, National Institute of Standards and Technology (NIST), Baltimore, MD.

[2] A.K. Ghosh, J. Wanken, and F. Charron, " Detecting anomalous and unknown intrusions against programs," In K. Keus (Ed), Proceedings of the 14th annual computer security applications conference IEEE Computer Society, Los Alamitos, CA. 1998, pp.259-267.

[3] James Cannady, "Artificial neural networks for misuse detection," Proceedings of the 1998 National Information Systems Security Conference (NISSC'98), Arlington, VA.

[4] R. Cunningham and R. Lippmann, "Improving intrusion detection performance using keyword selection and neural networks," Proceedings of the International Symposium on Recent Advances in Intrusion Detection, Purdue, IN, 1999.

[5] Srinivas Mukkamala, "Intrusion detection using neural networks and support vector machine," Proceedings of the 2002 IEEE International Honolulu, HI.

[6] "University of California at Irvine, 1999. KDD Cup." [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/task.html.

[7] A. Choudhary and A. Swarup, "Performance of Intrusion Detection System using GRNN", International Journal of Computer Science and Network Security," Vol.9, No.12, pp.219, December 2009.

[8] D. F. Specht, "A general regression neural network," IEEE Trans. on Neural Network, vol. 2, no. 6, pp. 568–576, 1991.

[9] M. Moradi, and M. Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks," IEEE International Conference on Advances in Intelligent Systems -Theory and Applications, Luxembourg-Kirchberg, Luxembourg, November 15-18, 2004.

[10] A. Choudhary and A. Swarup, "Neural Network Approach for Intrusion Detection" ACM 4th International Conference on Computer Science and Convergence Information Technology, Korea, pp. 1297, 2009.