Performance of Intrusion Detection System using GRNN

[1]**Amit Kumar Choudhary and [2]Akhilesh Swarup**

[1]National Institute of Technology, Kurukshetra, Haryana, India
[2]Galgotias College Engineering and Technology, Greater Noida, on leave from NIT, Kurukshetra, India.

Summary

The steady growth in research on intrusion detection systems has created a demand for tools and methods to test their effectiveness. Intrusion Detection System (IDS), is based on the belief that an intruder's behaviour will be noticeably different from that of a legitimate user and would exploit security vulnerabilities. This paper proposes a novel intrusion detection approach by applying Generalized Regression Neural Network (GRNN) for feature selection and detection. The MIT's KDD Cup 99 dataset is used to evaluate the present method. The results clearly demonstrate that the method can be an effective way for intrusion feature selection and detection and promises a good scope for further research.

Key words:

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

1. Introduction

Intrusion is defined as "a set of actions that attempts to compromise the confidentiality, integrity or availability of a resource". Intrusion detection is the problem of identifying unauthorized use, misuse, and abuse of computer systems by both system insiders and external penetrators. Intrusion detection systems (IDS) are an important component of defensive measures protecting computer systems and networks from abuse. 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.

Interest in the research and development of IDSs has been growing over the last several years, with the publication of John Anderson's *Computer Security Threat Monitoring and Surveillance* followed by D.Denning's seminal paper, "An Intrusion Detection Model," published in 1980 and 1987, 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. In order to evaluate the performance of intrusion detection, the 1998 DARPA Intrusion Detection Evaluation project from MIT Lincoln Labs were reduced and processed by domain experts to yield KDD Cup 99 dataset [1].

Proposals have been made to extend current research to a wide area network (WAN), but no significant products have resulted. The problem lies in the fact that the intruder is an intelligent and flexible agent while the rule-based IDSs obey fixed rules. This problem can be tackled by the application of soft computing techniques in IDSs. A number of soft computing based approaches have been proposed for detecting network intrusions [2]-[3]. 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) .

In this paper, the Generalized Regression Neural Network based intrusion feature selection and detection algorithm is proposed. The rest of paper is organized as follows. In section 2, Intrusion Detection System is briefly described, followed by types of networking attacks. In section 3, the development of ANN with Intrusion Detection is being focused. Section 4 presents the intrusion detection data and introduction of GRNN followed by the methodology and experimental analysis. At last, section 5 concludes the paper with future scope.

2. Intrusion Detection System

Different but complementary technologies have been developed and deployed to protect organizations' computer systems against network attacks, for example anti-virus software, firewall, message encryption, secured network protocols, password protection, and so on. Despite different protection mechanisms, it is nearly impossible to have a completely secured system. Therefore, intrusion detection is becoming an increasingly important technology that monitors network traffic and identifies network intrusions such as anomalous network behaviours, unauthorized network access, and malicious attacks to computer systems.

An Intrusion Detection System is a computer program that attempts to perform ID by either misuse or anomaly detection, or a combination of techniques. Anomaly

Manuscript received December 5, 2009
Manuscript revised December 20, 2009

detection is based on the premise that an intruder's behaviour will differ noticeably from that of a typical user. In misuse detection, the IDS's goal is to recognize "specific, precisely representable 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). Once an intrusion has been detected, IDS issues alerts notifying administrators of this fact. The next step is undertaken either by the administrators or the IDS itself. Intrusion detection may sometimes produce false alarms, for example as a result of malfunctioning network interface or sending attack description or signatures via email.

There are four major categories of networking attacks. Every attack on a network can be placed into one of these groupings.
Denial of Service (DoS): A DoS attack is a type of attack in which the hacker makes a memory resources too busy to serve legitimate networking requests and hence denying users access to a machine. e.g. Neptune, etc.
Remote to User attacks (R2L): A remote to user attack is an attack in which a user sends packets to a machine over the internet, and the user does not have access to in order to expose the machines vulnerabilities and exploit privileges which a local user would have on the computer. e.g. guest, etc.
User to Root Attacks (U2R): These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges. e.g. phf, etc.
Probing: Probing is an attack in which the hacker scans a machine or a networking device in order to determine weaknesses or vulnerabilities that may later be exploited so as to compromise the system. This technique is commonly used in data mining, e.g. satan, etc.

3. Artificial Neural Network and Intrusion Detection

Neural networks are algorithmic techniques used to first learn the relationship between the two sets of information, and then "generalize" to obtain new input-output pairs in a reasonable way. Neural networks are a uniquely powerful tool in multiple class classification, especially when used in applications where formal analysis would be very difficult or even impossible, such as pattern recognition, hand-written character recognition, nonlinear system identification and control. Provided the neural network has been given sufficient time to train, the property of generalization ensures that the network will be able to classify patterns that have never been seen before. In intrusion detection, neural networks have mainly been used learn the behavior of actors (e.g. users, daemons) in the system.

Fox, Henning, Reed, and Simmonian [4] 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 [5] 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 [6] 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 [7] 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 [8] described an off-line anomaly detection system, which utilized a back-propagation MLP neural network. Another study by Mukkamala [9], 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.

4. Experimental Results

4.1 Intrusion Detection Data

To evaluate the performance of proposed real-time IDS system, we use Knowledge Discovery in Database (KDD) Cup 99data supplied by the Defence Advanced Research Projects Agency (DARPA) and the Massachusetts Institute of Technology's Lincoln Labs in 1998. For each TCP/IP connection, 41 various quantitative and qualitative features were extracted. 22 of these features describe the connection itself and 19 of them describe the properties of connections to the same host in last two seconds. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type.
Two different attack types were included for this study: SYN Flood (Neptune) and Satan. These two attack types

were selected from two different attack categories (denial of service and probing) to check for the ability of the intrusion detection system to identify attacks from different categories. 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 non-occurrence. 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.

4.2 Introduction of GRNN

The Generalized Regression Neural Network (GRNN) paradigm has been proposed [10] as an alternative to the popular back-propagation training algorithm for feed forward neural networks. It is closely related to the probabilistic neural network. Regression can be thought of as the least-mean-squares estimation of the value of a variable based on available data. The GRNN 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 if $f(x,Y)$ is known, the expected value of Y given x (the regression of Y on x) can be estimated as

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

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

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

Where x_i and y_i are the ith 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

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

where d_i is given by the distance function of the input space.

Now let us consider each element of the vector K, namely k_i, to be estimated by an individual GRNN. If the weighted average approach is used to construct the output of GRNN, then each k_i can be written as

$$\hat{k}_i = \frac{\sum_{j=1}^{m}[k_j \exp(d_j)]}{\sum_{j=1}^{m}\exp(d_j)} \qquad (4)$$

Where d_j, the distance function and here can be written as

$$d_j = \left[-\left(\frac{s-s_j}{\sigma}\right)^2\right] \qquad (5)$$

In the above expression s is the new input and s_j is the stored input, σ is the spread factor. In (4) k_j is the stored output corresponding to s_j and $\hat{k}_i$ implies the estimated value of true k_i. s_j is defined as [1 0 0] for normal conditions, [0 1 0] for Neptune attack and [0 0 1] for the Satan attack. Here $\hat{k}_i$ is the output of the GRNN and a good estimation of k_i depends on the selection of spread factor σ.

4.3 Methodology and Analysis

A GRNN is used for function approximation in the present study. It has a radial basis layer and a special linear layer. The best two layer neural network used in this study was {41 41 41}. The performance for different epochs with a good amount of validation was obtained for the given set of dataset. It has been observed that the validation is obtained earliest at 4th epochs out of various simulations by changing the values of spread factor. The best result was attained in a training session that was stopped on 4th epoch as shown in fig. 1. The result was 100% correct classification on training and 100% on the testing set, giving a more accurate performance compared to the result reported earlier [3]. This is a preliminary result with a static data.

In a previous study [9], a result of more than 99% correct classification on this dataset using the neural network structure {41-40-40-1}was reported. In another similar study with different dataset [6], the success rate was comparable to the results of the present study (89-99%) and again a two class problem was implemented.

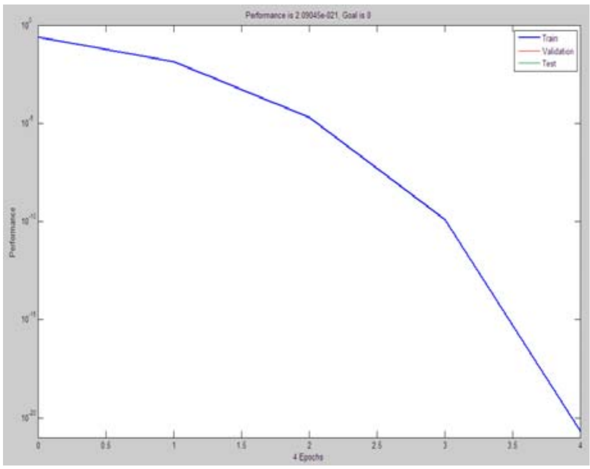

Fig. 1. The training process error. Here, validation and training dataset coincide each other (darker part).

5. Conclusion

This paper presents an intrusion detection system based on neural networks. The neural network model was used to solve a three-class problem, that is, normal, attack patterns, and the type of the attack. When given data is presented to the model, the results obtained revealed a great deal of accuracy approximately 100%. Since this has been done offline with static data, the results are encouraging. Efforts are being made to improve it with online simulation, with a fast validation. As a possible future development to the present study, more attack scenarios can be included in the dataset. Practical IDSs should include several attack types. In order to avoid unreasonable complexity in the neural network, an initial classification of the connection records to normal and general categories of attacks may be the first step. The records in each category of intrusions can then be further classified to the attack types. The intrusion detection is expected to become a practical and effective solution for protecting information systems.

References

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

[2] S. M. Bridges and R. B. Vaughn, "Fuzzy Data Mining And Genetic Algorithms Applied To Intrusion Detection," *Proceedings of 12th Annual Canadian Information Technology Security Symposium*, 2000, pp. 109-122.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] J. Ryan, M. Lin and R. Miikkulainen, "Intrusion Detection with Neural Networks," *AI Approaches to Fraud Detection and Risk Management: Papers from the 1997 AAAI Workshop*, Providence, RI, pp. 72-79.

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

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

Amit Kumar Choudhary received his M.Tech degree in Control System in Electrical Engineering Department from National Institute of Technology, Kurukshetra. He is continuing as a lecturer at NIT, Kurukshetra with the Electrical Engineering Department. His field of interest are Control Systems, Network Security.

Dr. A. Swarup obtained his Ph.D. degree from IIT Delhi. Presently he is Dean at Galgotias College of Engineering and Technology, Greater Noida. He is on leave from National Institute of Technology Kurukshetra where he is serving in various capacities since 1981. Few Ph.D. thesis have been completed under his supervision and about 40 research publications to his credit. His areas of interest are Robust Control, Intelligent Control and Information Security. He is Senior Member of IEEE and member of its Control System Society.