A Survey and Analysis of
Neural Network approaches to Intrusion Detection

Hussam O. Mousa

November 12, 2002

Abstract

Intrusion Detection Systems have gone a long way since their necessity was first realized at the onset of cyber revolution. Numerous shortcomings of this technology were recognized and Neural Networks have been identified as a potential model to address those shortcomings.

Research in Neural Networks as applied to Intrusion Detection has been relatively limited, and largely focused on a single NN architecture. This basic architecture was veryy effective in addressing the problem, especially when applied to Anomaly Detection on user behavior. In addition, initial reports and experiments identified other NN topologies to be very promising to the IDS problem. Furthermore, the exploration of hybrid system, to which a NN analysis engine is a component, has yielded positive results.

This paper introduces Neural Networks as applied to Intrusion Detection Systems. A survey is provided of present research and projects performed to this end. An analysis of present efforts is presented, as well as a discussion of trends and future possibilities in this area.

1 Introduction

Since the inception of research into the detection of malicious and anomalous computer behavior, various models and systems have been proposed, and since then the community have significantly developed; perhaps at a more rapid rate than the remainder of the networking and systems community. However since then and until this day the primary plague of Intrusion Detection Systems (IDSs) remains to be the high rates of False Positives. The basic origin of this problem stems from the dynamic nature of systems and networks. It is very easy, and common, for legitimate traffic to be mistaken for an intrusion attempt, and action on such a conclusion may even yield self-inflicted denial of Service incidents. Almost all solutions addressing this problem involve either a long (and often unpractical) customization and tuning period, or alternatively the reduction in detection sensitivity at the expense of an increased false negative rate. Currently the IDS community addresses these shortcomings by assigning extra analysts to IDS systems to verify alerts, or to use additional management and correlation systems that can at best prioritize and perhaps slightly reduce the false positive rate.

Another shortcoming of IDSs is their inability to detect new and modified intrusion attempts (commonly referred to as day-0 attacks). This is primarily due to the

technologies most dominantly used in present IDSs which rely on the formulation of an attack signature and attempting to match it to real time network and system behavior. This limitation of IDSs has caused many organizations that had deployments of IDSs to suffer at the onset of famous internet worms such as the Nimda and Code Red worms. Anomaly Detection Systems (ADSs) have been proposed to remedy this limitation. ADSs function on the assumption that for every system there is a normal pattern of behavior. Deviations from this pattern can then be detected and flagged as anomalous. In the majority of current ADS implementations, the false positive rate is too prohibitively high for wide scale adoption and deployment.

Neural Networks (NNs) have been identified since the beginning as a very promising technique of addressing the Intrusion Detection problem. Limited research has been performed to this end, and the results varied from inconclusive to extremely promising. The primary premise of NN that initially made it attractive was its generalization property, which makes it suitable to detect day-0 attacks. In addition NN also posses the ability to classify patterns, and this property can be used in other aspects of IDSs such as attack classification, and alert validation.

2 Artificial Neural Networks

Artificial Neural Networks is a name given to a large family of computational models that are inspired by the human nervous system. They are generally hailed as being the hope of escaping the theoretical limit imposed on the current computational architecture (as modeled by the Turing machine) due to their reliance on analogue data representation as the primary storage unit (presently binary is the basic form of data representation in current computational architecture). Of course, performance issues are a critical concern, and potential solutions (such as dedicated hardware coprocessors) are actively being researched in the NN community.

There are several criteria to classify NN; the two that are most relevant to the IDS problem are the training mode and the feedback mechanism. It is important to note that the two systems of classification are overlapping.

- Training Mode:
    - *Supervised learning*: These types of Neural Networks rely on a previously compiled and sanitized data “training set”. The NN is told that if it sees this pattern it should report the following conclusion. The NN then tweaks its internal “weights” such that it will try to accurately classify the largest majority of training vectors. The benefit of this technique over expert systems, and traditional signature based systems, is that it has a more generalized coverage of its problem domain. Instead of looking for a specific match it looks for a pattern match. IDS and ADS systems that utilize Neural Networks are most commonly reliant on this type of neural network.

- *Unsupervised learning*: Instead of being told what they should be looking for or what to report, NN based on unsupervised training models try to find patterns within a data set and seeks to group them according to the most relevant features. This technique is very useful when dealing with large volumes of raw data with little or no knowledge of the inter relation between the various fields in a vector. Several research projects based on this model are underway and will be discussed later on.

- Feedback Mechanism
  - *Feed-Forward networks* (non-recurrent networks): In this model, the NN tries to draw its conclusion solely on the data vector it is presented with, in other words no reliance on other data vectors or previous results is utilized, and there is no system memory. This model is suitable for problems where the data vectors are independent of each other. NN based on this model are fairly simple to train and deploy and are the most commonly used models within the scope of IDS and ADS.
  - *Recurring networks* employ active feedback mechanisms. In such a model the results of an earlier "decision" may influence subsequent decisions. Such networks are heavily – and successfully- employed in optimization problems with excellent performance, however introducing them to the field of IDS is not without obstacles. The primary difficulty in using recurring networks is training them on data sets, as well as the basic production of suitable data sets. Data sets need to be carefully designed to address the interrelationship between different vectors that is sought to be modeled by the neural network. This data model is very promising for IDS applications in modeling problems of state transition, and interdependent network activities.

3 Research efforts in IDS based on Neural Networks

The most common Neural Network used in Intrusion Detection Research is the Multi Layer Perceptron[1] (MLP). IDSs based on this model were researched in [1] [2] [4] [6] [7] [8] and were differentiated mainly by the domain of information they were monitoring.

The earlier works [1] [4] [7] were focused mainly on the application of Anomaly Detection on user behavior analysis. Subsequent research focused on using the MLP as an alternative to signature based misuse detection systems [6][8].In the latter category of research, the NN was focused on either generalized detection, or alert verification. One genuine approach to the problem was to use two different Neural Networks for attack detection and attack classification [6].

[1] This network is also refered to as Back Propagation network (Backprop) and Multi-Layer Feed Forward network (MLFF)

More recent approaches to IDS based on Neural Networks were based on the unsupervised learning model, primarily relying on the Self Organizing Map (SOM). This technique was applied to user behavior analysis in [3] [5]. Presently research direction in using SOM for Intrusion Detection is geared towards application and process analysis.

Among the most promising IDS architectures based on Neural Networks is their application in a hierarchical structure in conjunction with other detection engines. Such idea was proposed in [2] [3] [6] , and an architecture was proposed in [2]. This approach has the feature of harnessing the benefits of multiple technologies; however the impact on performance and the training complexity in online application are yet to be studied.

The following are brief summaries of several prominent research efforts and projects in the application of NN to IDS/ADS.

3.1 Neural Network Intrusion Detector (NNID)

This system originally developed in 1998, and was based on the analysis of user behavior using an MLP network [1]. The basic working of the system is the collection of individual user information based primarily on the commands they run. At fixed intervals the collected data would be used to train the system. Several training methodologies were simultaneously employed to try and arrive at the optimal trained Neural Network. The system was then made to monitor user behavior and attempt to detect anomalous behavior. The reported false positive rate for the system was 7% and the false negative rate was 4%.

3.2 Application of Neural Networks to UNIX Security

This project was one of the earliest systems to utilize NN to the problem of user anomaly detection [7]. The system used a MLP network to attempt, in real-time, to both train and detect anomalies. After a brief initial training session, the system is designed to continuously modify and “adapt” to its users’ normal behavior.

The system’s monitoring domain was as follows:

- User Activity times
- User login hosts
- User foreign hosts
- Command set
- CPU usage
- The system was able to successfully detect a wide variety of anomalies in its test environment which was on a student host in a university setting. One of the primary concerns noted by the system’s author was its heavy reliance on system resources which makes it unpractical in wide scale implementations

3.3 Anomaly Detection Using Neural Networks

In this research project, the MLP network was used to examine applications at the process level in an attempt to detect anomalous behavior [4]. The basic concept of this approach is the assumption that regardless of user characteristics, and anomalous behavior at the application level will generate activities at the process level that are distinguishable and identifiable as anomalous. In this project process states and input/output combination where used as input. A Training set was generated using simulated normal data input and known attacks to the application being monitored.

The results of the system were amazing in that the False positive rate was 0% for an experimental data set different from the training set. On the other hand false negatives were 20% of all legitimate attacks. The results of this experiment demonstrate the suitability of using Neural Networks in layered detection architectures to verify alerts and minimize/eliminate false positives.

3.4 Hierarchical Anomaly Network IDS using NN classification

This system's analysis engine relies on the synergy of two technologies: Statistical Analysis and MLP [2]. In addition the system consists of a hierarchy of agents at multiple tiers with each level reporting its alerts to the higher level.

The analysis engine consists of multiple tiers as follows:

- Probe to collect network traffic and abstracts it into statistical variables
- Event Preprocessor collects data from probes and other agents and formats it for the statistical analyzer
- Statistical Model compares the data to reference models previously compiled describing the normal state of the system. A "stimulus vector" is formed of the discrepancy and forwarded to the NN

Neural Network analyzes the vector and decides if it is anomalous or normal.

3.5 Artificial Neural Networks for Misuse Detection

This project was one of the first attempts to apply Neural Network to Misuse Detection [8]. The system was designed as a network IDS which relies on basic packet information as its inputs and an MLP network for analysis. The data first goes through 3 levels of pre-processing that select certain fields of the packet, normalize and group the data fields and convert it to a NN readable format. In addition every packet was labeled with a flag that indicates weather it is an attack or not. The preprocessed data is then divided into a training set and a test set.

While this system was not intended as a complete system, it indicated the potential for the use of NN in the field of misuse detection, which until then was dominated by rule based expert systems.

3.6 Anomaly and Misuse Detection using Neural Networks

This system was one of the earlier systems that shifted focus in Anomaly detection from user behavior analysis to process analysis [6]. The concept of the system is to have individual MLP networks trained on the normal behavior of varying programs. The goal of the system is to generalize from incomplete data and classify data as anomalous or normal. This system was experimented with on a SUN platform and used the Basic Security Module (BSM) as its source of data.

The input data was extracted from the BSM module and then distance metric was devised which measures the difference between the data item in question and several “exemplar” strings. The collection of these distance vectors constituted the input to the various networks. Then for each program that is to be monitored, a network was constructed, tuned and trained.

At operational time, the system is monitored per session. During each session, several programs are run with several inputs. The processes spawned by these events are fed to the various neural networks and an anomaly grade is computed for each. Finally a post-processing Leaky Bucket Algorithm was devised which accumulated the anomaly scores for the different event, while applying a timed decay function. The use of this post-processor allows for the detection of temporally co-located anomalies, while ignoring sparsely located ones. The post processor is controlled by a threshold, and, as expected variation on the threshold can lead to different combinations of accuracy levels.

The system demonstrated promising results with a 77% detection rate at only a 2.2% false positive rate.

3.7 UNIX Host Based User Anomaly Detection using SOM

This system relies on the assumption that normal behavior is consistent and concentrated in a limited feature space [5]. Conversely, scattered behavior will denote and irregular behavior which may ultimately signal an anomalous activity. The UNIX based system, designed to monitor user activity over extended periods, relies on an SOM for analysis and detection.

Features describing the user or “object” being investigated are collected, normalized and reduced. The SOM is then trained on a collection of data that is assumed to be normal. The resulting network is then considered to be representing the valid feature space of legitimate use.

At the operational level the system is designed to analyze user activities for a period of time, e.g. a day. Data from this period is collected and passed through the trained SOM. An event within the object is flagged as anomalous if it sufficiently deviates from the trained feature space. The threshold for flagging an anomaly is an operator defined variable, and like similar systems can be set to high at the expense of false negatives, or too low at the expense of false positives.

3.8 Host Based Intrusion Detection using SOM

The basic premise of this system is to examine session data by users on a UNIX system in search of behavioral anomalies [3]. This system utilized an unsupervised learning model provided by the Self Organizing Map. The system collects the following session data for analysis:

- User group
- Connection type
- Connection source
- Connection time

The collected data is preprocessed and normalized for presentation to the SOM analysis engine. This engine consists of two levels, a 3-map tier which summarizes the first three input domains with respect to time, and the second aggregates and correlates the conclusions of the first tier. The result of the analysis engine is a grouping of sessions with respect to the variables examined. Each group can then be examined and associated with a particular user behavior – weather it be normal or anomalous.

3.9 Elman Networks for Anomaly Detection

This study into the use of Elman Networks in analysis of program behavior is one of the pioneer researches into the use of recurrent networks in Intrusion/ Anomaly Detection [9]. Elman Networks are similar to the MLP with additional context nodes which maintains a state of the system. The main added benefit of using recurrent networks was the ability to maintain state information between inputs.

The Elman network works by predicting the next sequence given a present input and the context. The actual next sequence is compared to the predicted sequence, and the difference between them represents the measure of anomaly.

For the application of this system, a leaky bucket algorithm similar to the one employed in [6] was used to reduce the rate of false positives. Experimental results were extremely promising, with the elimination of false positives at a detection rate of 77%, and the achievement of 100% detection rate with around 9% false positives.

4 Discussion and Trends

The common goal of applying Neural Networks to IDS is to overcome the primary deficiency of the present IDS systems: the inability to detect novel and modified attacks, and the increasing rates of false positives. Research has been performed to achieve both of these goals; however the vast majority of NN research into IDS has been limited to the application of the Multi Layer Perceptron network. Only recently has the unsupervised learning model of the Self Organizing Map been investigated, and only to user behavior analysis.

Future research focus in NN applications in IDS should be focused on two distinct fronts: investigation of other NN topologies, and especially recurrent networks, and the extension of the NN monitoring domains to include more system level inputs in addition to the present focus on user behavior analysis.

In addition, while the majority of present IDS analysis modules have to compromise one of the two primary accuracy measures: False Positive and False Negative rates, NNs can be made to specialize to focus on one of those primary metrics. The current trends in NN IDS research have been focusing on the goal of training the analysis module with a specific detection approach: generalization, verification, or classification. These goals can actually present the IDS community with a promising solution in the form of layered analysis engines with a specialized function at each layer.

NN based IDS should be viewed as a pre or post analysis module to a more stable and traditional IDS analysis engine (or maybe even another NN engine). NNs can be used to preprocess large volumes of input in an attempt to normalize and present them to, for example, statistical based detection engines. In a similar manner, NNs can be used as a post processing layer for the verification of alerts and reduction/elimination of False positives.

Furthermore, as the Neural Network community continues to develop more efficient implementations of NN engines, in the form of specialized integrated circuits and physical co-processors, the usage of NN based application will become more commonplace. This can be very beneficial to the IDS community since there is already a trend of moving towards IDS appliances, in addition to this being an opportunity for further improving the performance benchmarks of IDSs.

As a final note, while neural network research into IDS has been active for several years now, the majority of projects were performed by primarily Neural Network people with the secondary aid of security specialists, or by Security researchers with the aid of Neural Network tools. In order to escalate the levels of NN research into IDS, research groups based on teams containing experts from both fields should be formed, and novel NN architectures should be utilized in conjunction with the more advanced IDS techniques currently popular amidst the security community.

5 Conclusion

Neural Networks have been extensively studied in the application of Intrusion Detection Systems with varying degrees of success. While Feed Forward Back propagation Networks have been the dominant topology used, recent research into the unsupervised model of the Self Organizing Map (SOM), and recurrent Networks (Elman Network) have been successfully studied.

Future research initiatives should focus on the following:

- Focus on other NN topologies, particularly recurrent networks

- Enclosure of NN IDS into layered systems with diverse detection engines
- Specialization of the functions of the NN detection engine

In addition, more involvement of security specialists is necessary for NN IDS research in order to determine and design optimal monitoring and operational scenarios

References:

1. Lin, M., Miikkulainen, R., Ryan, J., “Intrusion Detection with Neural Networks.” http://citeseer.nj.nec.com/ryan98intrusion.html (1998).
2. Jorgenson, J., Manikopoulos, C., Li, J., Zhang, Z., “A Hierarchical Anomaly Network Intrusion Detection System Using Neural Network Classification.” http://www.itoc.usma.edu/Workshop/2001/Authors/Submitted_Abstracts/paperT2A2(19).pdf (2001).
3. Heywood, M., Lichodzijewski, P., Zincir-Heywood, N., “Host-Based Intrusion Detection Using Self-Organizing Maps.” http://www.itoc.usma.edu/Workshop/2001/Authors/Submitted_Abstracts/paperT2A2(19).pdf .
4. Charron, F., Ghosh, A., Wanken, J., “Detecting Anomalous and Unknown Intrusions Against Programs.” http://citeseer.nj.nec.com/tan95application.html (1998).
5. Hätönen, K., Höglund, A., Sorvari, A., “A computer Host-Based User Anomaly Detection System Using the Self-Organizing Map.” Proceedings of the IEEE-INNS-ENNS International Joint Conference of Neural Networks (IJCNN’00).
6. Ghosh, A., Schwartzbard, A., “A study in using Neural Networks for Anomaly and Misuse Detection.” http://www.usenix.org/events/sec99/ghosh.html (1999)
7. Tan, K., “The Application of Neural Networks To Unix Computer Security.” http://citeseer.nj.nec.com/tan95application.html (1995)
8. Cannady, J., “Artificial Neural Networks for Misuse Detection.” http://citeseer.nj.nec.com/cannady98artificial.html (1998).
9. Ghosh, A., Schatz, M., Schwartzbard, A., “Learning Program Behavior Profiles for Intrusion Detection.” http://citeseer.nj.nec.com/ghosh99learning.html (1999).
10. Axelsson, A., “Intrusion Detection Systems: A survey and Taxonomy” http://citeseer.nj.nec.com/ghosh99learning.html (2000).
11. Zurada, J. M., “Introduction to Artificial Neural Systems”, WEST publishing company (1992).