A comparative study of Cascaded Forward Back Propagation and Hybrid SOFM-CFBP Neural Networks based Intrusion Detection Systems

Afrah Nazir

Abstract- Intrusion detection technology is an effective approach to deal with the problems of network security. The key idea is to aim at taking advantage of classification abilities of supervised learning based neural networks and clustering abilities of unsupervised learning based neural network. The neural network algorithms are popular for their ability to 'learn' the patterns in a given environment and thus can be trained to detect intrusions by recognizing patterns of an intrusion. In this work we perform a comparative study of Cascaded Forward Back Propagation neural network based intrusion detection system and a Hybrid neural network based intrusion detection system, where cascade connections of two different types of neural networks, namely Self Organizing Feature Map and Cascaded Forward Back Propagation , are used for intrusion detection .In this study we work on the well structured KDD CUP 99 dataset.

Index Terms- Cascaded Forward Back Propagation (CFBP), Self Organizing Feature Map (SOFM), Intrusion Detection Systems (IDSs).

1. INTRODUCTION

In a computer network there are a lot of data exchanges between computers within a local network and between a computer and another network (e.g. the Internet). Being connected to a large network like the Internet plunges the computers into a world where the risk of getting in touch with harmful network traffic activity is relatively high. Several security precautions can be taken, like deploying antivirus, firewall, and access control etc. in order to prevent such activities from intruding upon your computer or network. They all concentrate on different aspects of how to protect and secure a computer/network. Damage caused by these intrusions is the unauthorized modifications of system files, user files or exploiting design flaw of specific protocols etc.

According to sources of data there are two types of intrusion detection systems namely (i) Network Intrusion Detection System (NIDS) and (ii) Host Intrusion Detection System (HIDS). NIDS capture all the network traffic and analyses the contents of individual packets for malicious traffic. HIDS run on individual hosts or devices on the network and monitors the inbound and outbound packets only from the individual host.

First advantages of using neural networks for intrusion detection is that the intrusion detection systems operate by making results in the sense of predictions based on known as well as unknown patterns. With the use of neural network models it is possible to comply with this process, since these models offer the option to train a custom network and use it as some sort of a strainer for new incoming network connection and thereby detect abnormal behaviors.

The second advantage lies in the fact that when working with intrusion detections one will realize that the dimension of the data of a network connection is high. There are many different protocols on different layers of the internet with different services, destinations and sources, etc. The property of dimensionality reduction and data visualization in neural networks can be very useful to reduce the many dimensions of a network connection to 2-dimension. This will help to visually discover connections which do not fall into the same category or group (clusters) with the trained and trusted ones and thus will be classified differently.

This work compares and evaluates the performance of single neural network based CFBP intrusion detection system and Hybrid neural network approaches to intrusion detection, based on classification rate, false positive rate, and false negative rate for each of the four classes of attacks present in KDD CUP 99 Dataset.

This paper is organized as follows. In next section, we discuss related work in the area of intrusion detection using Neural Networks. In the next two sections, we discuss the

- *Afrah Nazir pursued masters degree program in Computer engineering in A.M.U, INDIA, PH-09897705922. E-mail: afrahnazir@gmail.com*

details of different neural networks used for intrusion detection in this study, namely, Cascaded Forward Back Propagation Neural Network and Self Organizing Feature Map Neural Network. In section 5, we discuss the Hybrid SOFM-CFBP neural network based IDSs. In section 6, we discuss the training and test datasets used in this study. In section 7, we discuss the various performance evaluation criteria. In section 8, we show our system's results and finally conclude in last section.

2. RELATED WORKS

ANN for intrusion detection provides the potential to identify and classify network activity based on limited, incomplete, and nonlinear data sources.

In [10], SOM was used to map the network connections onto 2-dimensional surfaces, which were displayed to the network administrator. The intrusions were easily detected in this view. However, the approach needs a visual interpretation by the network administrator.

The artificial neural networks have also been proposed in the detection of the computer viruses. A self-organizing map was selected in [11] for intrusion detection. In that work, the self-organizing map was designed to learn the characteristics of normal activities. The variations from normal activities provided an indication of a virus.

In [1] a hybrid model of the SOM (Self Organizing Map) and the MLP (Multi Layer Perceptron) was proposed. In that work, the self-organizing map was combined with the feed-forward neural network for detecting the intrusions in their home network. In [2] the same proposed approach has been implemented and tested using DARPA 1999 data set but finds trouble in detecting all types of attacks simultaneously.

In this work, we develop and perform a comparative study of Cascaded Forward Back Propagation neural network and Hybrid SOFM-CFBP neural networks based IDSs.

3. KOHONEN'S SELF ORGANIZING FEATURE MAPS (SOFM)

The Self-Organizing Feature Map is a competitive network where the goal is to transform an input data set of arbitrary dimension to a one- or two-dimensional topological map. The model was first described by the professor Teuvo Kohonen and is thus sometimes referred to as a Kohonen Map. The SOFM aims to discover underlying structure, e.g. feature map, of the input data set by building a topology preserving map which describes neighborhood relations of the points in the data set.

The Self Organizing Feature Maps (SOFM) neural networks is based on unsupervised learning i.e. the ability to learn from unlabeled data and create new classes automatically. Training of SOFM Neural Network occurs in several steps and over much iteration.

1. For each node on the SOFM neural network, weights are initialized.
2. An input vector is chosen at random from the set of training data.
3. Examine every node to calculate whose weights are most like the input vector.
4. The winning node is commonly known as the Best Matching Unit (BMU).
5. The radius of the neighborhood of the BMU is now calculated.
6. Any nodes found within this radius are deemed to be inside the BMU's neighborhood.
7. This radius keeps on decreasing until the size is just 1 node.
8. This winner neuron has minimum Euclidean distance from input vector.
9. Each neighboring node's weights are adjusted to make them more like the input vector.

$$W (t+1) = W (t) + L (t) (V (t) - W (t)) \qquad (1)$$

The new adjusted weight for the node is equal to the old weight (*W*), plus a fraction of the difference (*L*) between the old weight and the input vector (*V*).

4.CASCADEDFORWARD BACK PROPAGATION NEURAL NETWORK (CFBP)

A cascade correlation net consists of input units, hidden units, and output units. Input units are connected directly to output units with adjustable weighted connections. Connections from inputs to a hidden unit are trained when the hidden unit is added to the net and are then frozen. Connections from the hidden units to the output units are adjustable consequently.

Cascade correlation network starts with a minimal topology, consisting only of the required input and output units. This net is trained until no further improvement is obtained. The error for each output until is then computed. Next, one hidden unit is added to the net in a two-step process. During the first step, a candidate unit is connected to each of the input units, but is not connected to the output units. The weights on the connections from the input units to the candidate unit are adjusted to maximize the correlation between the candidate's output and the residual error at the output units.

The residual error is the difference between the target and the computed output, multiplied by the derivative of the output unit's activation function, i.e., the quantity that would be propagated back from the output units in the back propagation algorithm. When this training is completed, the weights are frozen and the candidate unit becomes a hidden unit in the net.

The second step in which the new unit is added to the net now begins. A second hidden unit is added using the same process. The process of adding a new unit, training its weights from the input units and the previously added hidden units, and then freezing the weights, followed by training all connections to the output units, is continued until the error reaches an acceptable level or the maximum number of epochs (or hidden units) is reached.

5. A HYBRID NEURAL NETWORK APPROACH TO INTRUSION DETECTION

In this study we developed a hybrid SOFM-CFBP neural networks based IDS. This model uses two different network structures SOFM and CFBP. Both are Feed Forward Neural Network Structure but uses different learning algorithm and meant for different tasks. The various steps involved in intrusion detection through neural network are:

1. The intrusion detection process starts with extracting and selecting desired features from input vector representing connection records on which training occurs.
2. Next step is preprocessing the input vector. This preprocessing involves transformation i.e. converting textual attributes to numeric attributes and normalization i.e. scaling the data to an acceptable range.
3. SOFM Neural Network is trained until the desired performance criteria is met. Corresponding to each input vector a winner neuron is found in SOFM network and its weight information is stored.
4. Weight information from SOFM is fed into the CFBP neural network for further training and then the class of the each trained input vector is stored.
5. Next step is to test the neural network through test vector.
6. The test vector is simulated on the trained network to classify it as a normal or an attack connection.
7. Evaluate the performance of the neural network approach to intrusion detection.

6. DATASET FOR TRAINING AND TESTING

The dataset used in this study for intrusion detection is the KDD Cup 99 Dataset [9]. The KDD 99 intrusion detection datasets are based on the 1998 DARPA [7] initiative, which provides designers of intrusion detection systems (IDS) with a benchmark on which to evaluate different methodologies. The “10% KDD” dataset is used for the training of different intrusion detection systems. It includes 22 types of attacks connections. The “Corrected KDD” dataset is used for testing purpose. The “Corrected KDD” dataset provides a data with different statistical distributions as compared to the data present in either “10% KDD”

6.1 CATEGORIES OF ATTACKS IN KDD 99 DATASET

Attacks in KDD 99 Dataset fall in the following four major categories:

1. **Denial of Service (DoS) attacks: -** DoS is a class of attacks where an attacker makes some computing or memory resource too busy or too full to handle legitimate requests, thus denying legitimate users access to a machine Example Smurf, Teardrop, Neptune, pod are the common DoS attacks.
2. **Probe: -** Probe is a class of attacks where an attacker scans a network to gather information or find known vulnerabilities. Example Portsweep and Satan are the common Probing attacks.
3. **User to root attacks (U2R): -** User to root exploits are a class of attacks where an attacker starts out with an access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system. Buffer_overflow, rootkit, Load modules, perl, are the common User to Root attacks.
4. **Remote to user attacks (R2L): -** A remote to user is a class of attacks where an attacker sends packets to a machine over a network, then exploits machine's vulnerability to illegally gain local access as a user. Multihop, Spy, is the common Remote to User attacks.

7. PERFORMANCE EVALUATION CRITERIA

The various performance evaluation criteria for the different neural network approaches to intrusion detection are discussed here.

7.1. CLASSIFICATION RATE (CR)

It denotes true-positives rate or true-negatives rate.
True-positives = (Total Number of Normal Instances detected & classified by the system) / (Total Number of Normal Instances present in the Test Dataset)

True-negatives = (Total Number of Attack Instances detected & classified by the system) / (Total Number of Attack Instances present in the Test Dataset)

7.2. FALSE POSITIVE RATE AND FALSE NEGATIVE RATE

A False-positive (FPR) occurs when the system classifies an action as anomalous (a possible intrusion) when it is a legitimate action. Although this type of error may not be completely eliminated, a good system should minimize its occurrence.

A False-negative (FNR) occurs when an actual intrusive action has occurred but system allows it to pass as non-intrusive behavior. This implies malicious data is not detected and alerted. It is a more serious error.

8. RESULTS AND DISCUSSION

Both the CFBP and Hybrid SOFM-CFBP neural networks based IDSs are developed with the help of MATLAB 7.0 Neural Network Toolbox. Table 1 shows the training parameters for CFBP neural network used in two IDSs.

Table 1 Parameters for "41 41 40 1" neural network Architecture

Parameters for training	**Value of parameter**
Number of Nodes in Input Layer	41
Number of Nodes in Output Layer	1
Number of Nodes in First Hidden Layer	41
Number of nodes in second Hidden Layer	40
Training Functions	trainrp
Learning Rate and Number of Epochs	0.2 & 1000

8.1 CLASSIFICATION RATE (%) FOR CFBP AND HYBRID SOFM-CFBP NEURAL NETWORK BASED IDSs

Table 2 shows CR percentages for two different IDSs. With Hybrid SOFM-CFBP highest classification rate (%) of 100 % is achieved for DoS attack.

Table 2: Classification Rate (%) for CFBP and Hybrid SOFM-CFBP IDSs

Class	**CFBP**	**Hybrid SOFM-CFBP**
Normal	93.86	97.21
DoS	94.73	97.91
Probe	94.05	96.22
U2R	90.08	95.31
R2L	92.36	96.73

The highest classification rate percentage is for classes Normal (99.95 %) and DoS (100 %) by using Hybrid SOFM-CFBP neural networks based intrusion detection systems, while the lowest classification rate percentage is for the class U2R (92.89 %) with CFBP neural network based intrusion detection systems.

Figure 2 shows that the classification rate percentage (Table 3) of hybrid SOFM-CFBP is better than the single neural network based IDS for all the different classes of data present in the KDD 99 dataset.

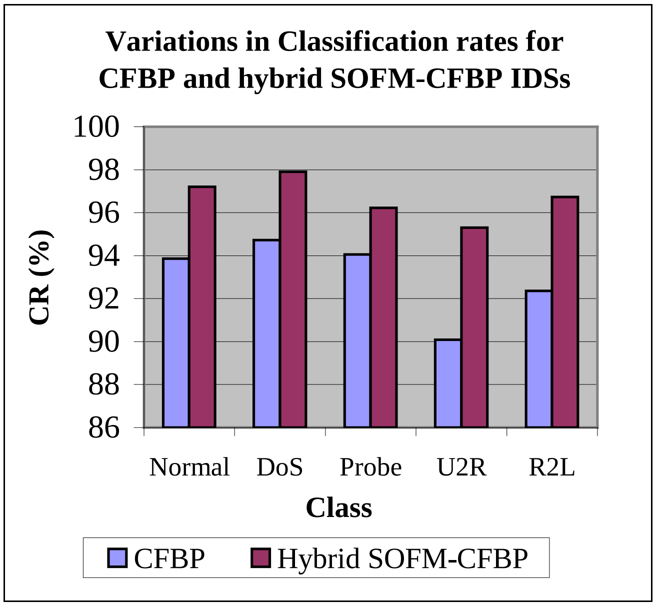

Figure 2: Classification Rate (%) for CFBP and Hybrid SOFM-CFBP neural networks based IDSs

8.2 FPR (%) AND FNR (%) FOR CFBP AND HYBRID SOFM-CFBP NEURAL NETWORK BASED IDSs

Table 3 and Table 4 shows the FPR and FNR performances for the 3 hybrid Neural Networks. Lowest FPR i.e. 6.7 % (Table 3) is achieved for Probe attack using Hybrid SOFM-CFBP network. And lowest FNR i.e. 9.2 % (Table 4) is achieved for DoS attack using Hybrid SOFM-CFBP network

Table 3: FPR (%) for CFBP and Hybrid SOFM-CFBP IDSs

Class	**CFBP**	**Hybrid SOFM-CFBP**
Normal	33	13.4
DoS	24	11.2
Probe	21	9.8
U2R	49	14.6
R2L	37	12.7

It is clear from the data recorded in the Table 3 that the performance of CFBP neural network based IDS for attack type U2R is 15 % (FPR), is highest of all the other records; this means that the number of misclassification for U2R is more with CFBP neural network than with hybrid IDS..

It is clear from the false positive rate performance results given in Table 3 that the performance of Hybrid SOFM-CFBP neural network is better than the single neural network based CFBP IDS, because the number of misclassifications in terms of classification of a normal data to an intrusive class is lesser i.e. FPR is small

Table 4: FNR (%) for CFBP and Hybrid SOFM-CFBP IDSs

Class	CFBP	Hybrid SOFM-CFBP
Normal	36	12.1
DoS	25	12
Probe	22	13.2
U2R	51	13.3
R2L	11	10.4

It is clear from the false negative performance results given in Table 4 that the performance of Hybrid SOFM-CFBP neural network is better than the other IDS using CFBP network, because the number of misclassifications is lesser. Like, for example, the number of times Dos attack is bypassed as the non intrusive action by CFBP neural network IDS is 10%, but with Hybrid SOFM-CFBP neural network this percentage is least i.e. 9.2 %.

Figure 3 and 4 shows the FPR (%) and FNR (%) plots for the two different neural networks. The false positive rate (figure 3) is lowest for hybrid SOFM-CFBP neural network and the lower false negative rates (figure 4) is again achieved through hybrid SOFM-CFBP intrusion detection system. For all the classes the FPR (%) for CFBP neural network based IDS are much greater than with Hybrid IDS. This means that the number of misclassifications is less with hybrid IDS.

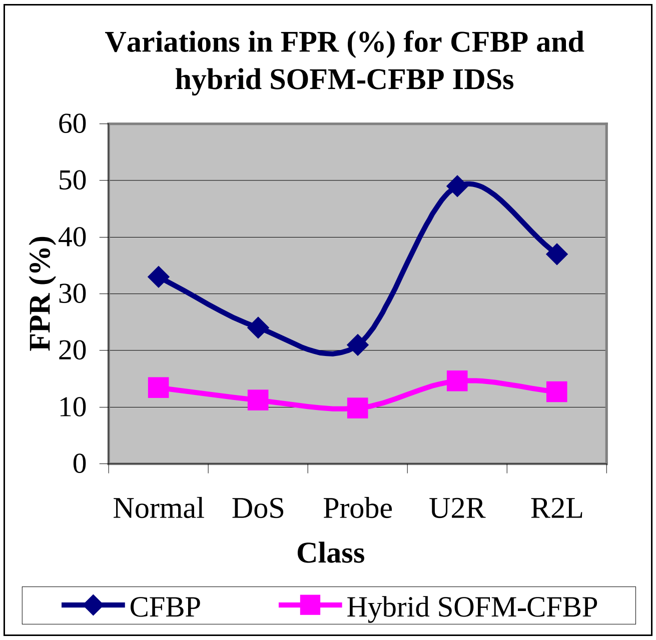

Figure 3: False Positive Rate (%) plot for CFBP and Hybrid SOFM-CFBP neural networks based IDSs

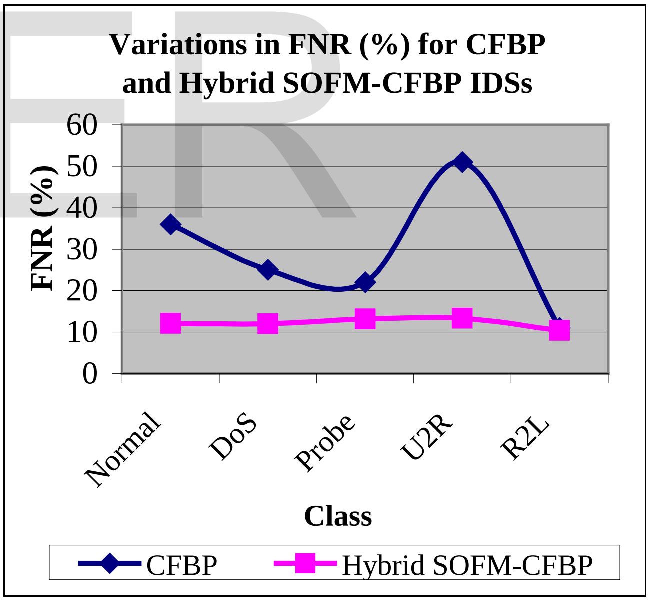

Figure 4: False Negative Rate (%) for CFBP and Hybrid SOFM-CFBP neural networks based IDSs

From these results we conclude that the performance of the Hybrid SOFM-CFBP network is better over single CFBP neural network based IDS, in terms of lowest false positive rate and lowest false negative rate. The lower the FPR and FNR the better the network, this means that the numbers of misclassifications are less.

9. CONCLUSIONS

This paper has presented a hybrid neural networks based intrusion detection system using Self Organizing Feature Map and Cascaded Forward Back Propagation neural networks.

The data required for the development of neural network intrusion detection systems have been obtained from KDD Cup' 99 data. Totally 4 category of attacks which include 22 number of intrusion from the computer network were considered in the developed models. The SOFM network is used to visualize and study the characteristics of each input features and the weights information from SOFM is fed into CFBP network for classifying the attacks.

For all kinds of attacks considered the Hybrid SOFM-CFBP intrusion detection system shows very good classification rate, smaller false positive and false negative rates, as compared to the results reported by the simple CFBP neural network based IDS.

10. REFERENCES

[1] James Cannady and Jim Mahaffey, “*The application of Artificial Intelligence to Misuse Detection*”, in proceedings of the first Recent Advances in Intrusion Detection (RAID Conference) ,1998.

[2] C.Jirapummin, N.Wattanapongsakorn and P.Kanthamanon, “*Hybrid Neural Networks for Intrusion Detection System*”, in the Proceedings of International Technical Conference on Circuits / Systems, Computers and Communications, 2002.

[3] Alan Bivens, Chandrika Palagiri, Rasheda Smith, Boleslaw Szymanski, Mark Embrechts, "Network Based Intrusion Detection using Neural Networks", in the proceedings of ANNIE Intelligent Engineering Systems through Artificial Neural Networks, 2002.

[4] Fox, Kevin L., Henning, Rhonda R., and Reed, Jonathan H., “*A Neural Network Approach Towards Intrusion Detection*”, in the Proceedings of the 13th National Computer Security Conference, 1990.

[5] P.Ganesh Kumar and D. Devaraj, “*Network Intrusion Detection using Hybrid Neural Networks*”, in the proceedings of IEEE International Conference on ICSCN, 2007.

[6] P.GaneshKumar, D.Devaraj, V.Vasudevan, “*Artificial Neural Network for Misuse Detection in Computer Network*”, in the proceedings of the International Conference on Resource Utilization and Intelligent Systems (INCRUIS-2006), 2006.

[7]DARPA Dataset documentation: http://www.ll.mit.edu/IST/ideval/data/data_index.html .

[8] The MathWorks-MATLAB and Simulink for Technical Computing, MATLAB online help http://www.Mathworks.com/product/matlab/tryit.html .

[9] KDD-cup' 99 dataset, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[10] L. Girardin,”*An eye on network intruder-administrator shootouts*”, in Proceedings of the Workshop on Intrusion Detection and Network Monitoring (ID ’99), pages 19 .28, Berkeley, CA, USA, 1999.USENIX Association.

[11] K.Fox, R.Henning, J.Reed, and R.Simonian, “*A neural network approach towards intrusion detection*”, in Proceedings of the 13th National Computer Security Conference, *1990*.