DESIGNING INRUSION DETECTION SYSTEM USING BACK PROPOGATION NETWORK AND EVALUATING PARAMETERS

DEVENDRA SINGH[1], SARTHAK GUPTA[2],PRATIK KUMAR GUPTA[3] and MANISH SHRIVASTAVA[4]
[1]ASSISTANT PROFESSOR,RESEARCH SCHOLAR,[2]STUDENT, [3] STUDENT,
[4] ASSISTANT PROFESSOR
[1]DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF STUDIES GURU GHASIDAS VISHWAVIDYALAYA,BILASPUR,C.G.

ABSTRACT: Intruders have become the major pain point as far as network security is concerned. Security is a vital element of any network because enormous intimate data is residing on our network and the intruders try to intervene between the host and the network attempting to steal or modify this data. During the past decades several soft computing methods are formulated for strengthening the intrusion detection system.

In this study, we are concerned about detecting the presence of intrusion and for the same we have used BPN. BPN along with KDD 99[1] data sets of DARPA are analyzed for classification between attacks. The result was obtained with an accuracy of 90.2%.

Keywords: ***Back Propagation Network(BPN), Intrusion Detection System(IDS).***

INTRODUCTION

The world is growing with a faster pace. Internet these days have helped people extend their limits and resources by providing enormous facilities to flourish. Several important business deals are done over the network within shorter durations hence saves a lot of user's time. But every coin has two faces. With this advancements and lot of information circulating over the network, the possibilities of theft has also increased. The illegal access to any network with the intention to steal or break into the security of the network is termed as **INTRUSION.** Intruders try to gain access over the network trying to hinder the user's work pace and access the crucial information stored in the network.

So, the major point of concern is to detect these intrusions and try to protect the network to prevent the confidentiality of the same. IDS[3], the most prevailing technique serves the purpose for us. The working of IDS can be related to a security alarm. As our bank lockers are installed with a security alarm and only the authorized person can access these lockers, and if any third person tries to do the same, the alarm indicates the action. Similar is the case with IDS. Whenever an intruder tries to intervene into the network illegally, it informs us about this activity.

An intrusion detection system works paralelly with firewall. Firewall prevents the system from any malicious attacks from the internet, and intrusion detection system detects if someone tries to break through the security of firewall.

Hence an intrusion detection system is an application software that monitors the entire traffic over the network to analyze the network for intrusions (if any) prevailing within or outside the network, trying to pirate or modify the confidential information in the network.

The two different types of Intrusion Detection System[8] are:

1) **Host Based Intrusion Detection System (HIDS) [8]**: The host based intrusion detection works on a single host or a device in a network and it inspects the inbound and the outbound packets of data and will alert the administrator upon any malicious activity or modification of data. It compares the newly generated output with the previous data and if a change in the pattern is observed, it will quickly alert the administrator about the same.

2) **Network Intrusion Detection System (NIDS) [8]** : Network intrusion detection systems are installed at some certain points in a network to analyze the passing traffic to and from the devices that are connected in a network. A detailed analysis of the entire traffic

is done and is matched with the library of the known attacks. And if match is found then the administrator is alarmed about the presence of intrusion in the network.

TYPES OF ATTACKS

1. **Denial of Service (DoS)** :- A DoS[3][5] attack is a type of attack in which the hacker makes a computing or memory resources too busy or too full to serve legitimate networking requests and hence denying users access to a machine e.g. apache 2 , smurf , Neptune , teardrop , back, mail bomb , UDP storm etc. are all DoS attacks.

2.**Remote to User Attacks (R2L)**: A remote to user attack[3][5] is an attack in which a user sends packets to a machine over the internet, which she/he 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. named, ncftp, phf, netcat etc.

3. **User to Root Attacks (U2R)[3][5]**: Under these attacks the hacker log in the system as a normal user and tries to exploit the valuable information as an authorized user of the system. e.g. yaga, sechole etc.

4. **Probing**: Probing[3][5] is a technique in which the hacker checks for the loop-holes in the system or the entire network of the organization and tries to exploit the valuable information which is often very crucial.

ATTACK	TYPES	GROUPS	CLASS
NORMAL	normal	A	1
R2L	dictionary,ftp_write,guess_password,imap,named,sendmail,spyxlock,xsnoop,snmpgetattack,httptunnel,worm,snmpguess, multi-hop,phf, warezclient,wrazemaster.	D	2
DOS	smurf,teardrop,pod,back,land, apache2,udpstrom,mailbomb, processtable, neptune	B	3
PROBE	ipsweep, portsweep,nmap,satan,saint, mscan	C	4
U2R	perl,ps,xterm,loadmodule,ejct,buffer_overflow, sqlattack	E	5

TYPES AND CLASSIFICATION OF THE DIFFERENT ATTACKS [2]

REVIEW

We have concluded and conducted our study after analyzing the previous results of

[1] H.GUNES KAYASICK Et., Al.,2005in their research paper with the title "*Selecting Features for Intrusion Detection : A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets*" mentioned their study on the KDD 99 datasets and worked upon the feature optimization techniques in an attempt to reduce the 41 features of the dataset and were successful in their venture with 91% results and false positive rates less than 1%.

[2] **MOHAMMAD SAZZADUL HOQUE Et., Al.,2012,**in their paper with the title "*An Implementation of Intrusion Detection System using Genetic Algorithm"* conducted their study on intrusion detection system and tried to optimize the features using genetic algorithm. They have used 494,021 records out of which they found that 97,280 are normal connection records, while the test set contains 311,029 records among which 60,593 are normal connection records.

[3] R. GRAHAM, 2000"FAQ: Network Intrusion Detection Systems", gave a detailed description regarding the Intrusion Detection Systems and the various types of Intrusion Detection System.

[4] MEHEDI MORADI Et., Al.,2014"A Neural Network Based system for Intrusion Detection and classification of attacks", conducted an experiment to design an optimal neural network with minimal number of hidden layers and were able to achieve 91% accuracy with two hidden layers and 87% with one hidden layer. They used Early stopping validation method for the same.

[5] RAJEEV SINGH,2014 "*Introduction to Intrusion Detection System*" He examined and applied intrusion detection system to various system architectures and evaluated key mechanisms in the architecture, such as anomaly detection and misuse detection for both wired and wireless networks.

[6] M. MADHVI,2012" *An Approach For Intrusion Detection System in Cloud Computing",* she through her works, proposed a multi-threaded cloud IDS model which can be administered by a third party monitoring service for a better optimized efficiency and transparency for the cloud user.

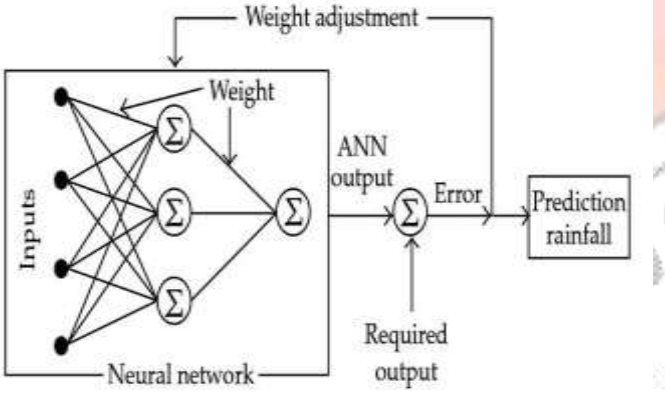
[7] NING-QING SUN Et., Al.,2009 "*Intrusion Detection Based on Back-propagation Neural Network and Feature Selection Mechanism",* They emphasized upon developing a Intrusion Detection System using Back Propagation. Firstly they conducted feature optimization operation upon the available data set and then applying feeding the BPN with the optimized data set to yield accurate and precise results.

KDD'99 DATASET [1]:

With the advancement in the technology, the risk to the security of data has also increased and has been a critical issue over times. Several researches have been conducted in the recent times using machine learning for formulating a solution out of this. For the training purpose we have used KDD 99 intrusion detection datasets, which are based on DARPA 98 dataset, and it provides labelled data for the researchers working in the field of intrusion detection and is the only labelled dataset publicly available. The KDD 99 Dataset is based on the 1998 DARPA initiative, which helps the designers of the intrusion detection systems with a relevant set of data as a benchmark for analysis. In our research we will be using dataset with 84 fields, with 41 features in each field.

BACK PROPAGATION NETWORK

Back-propagation[4] is a method used in artificial neural networks to calculate the error contribution of each neuron after a batch of data (in image recognition, multiple images) is processed. This is used by an enveloping optimization algorithm to adjust the weight of each neuron, completing the learning process for that case.[2] The aim of our research is to utilize back propagation network to prepare a predictive model for the detection of intrusions in the network. The system has been repeatedly trained with the KDD 99 dataset and is extensively used to generate an optimized result.



[6]Fig 1. Back Propagation Network

METHODOLOGIES

Initially the network is trained using KDD 99 dataset. a small portion of the data is initially used for the training purpose. After the successful preprocessing, training and testing of the data it then classifies the internal connections into 23 categories.

Working:

We have used back propagation algorithm for the training purpose. The working is discussed below:

(i) Design the network and initialize the parameters.

(ii) Feed the network with the appropriate data from the dataset upon which the experiment is to be conducted.

(iii) Specify the number of hidden layers in the network.

(vi) The training is continued until the desired result with the minimum possible deviations is obtained.

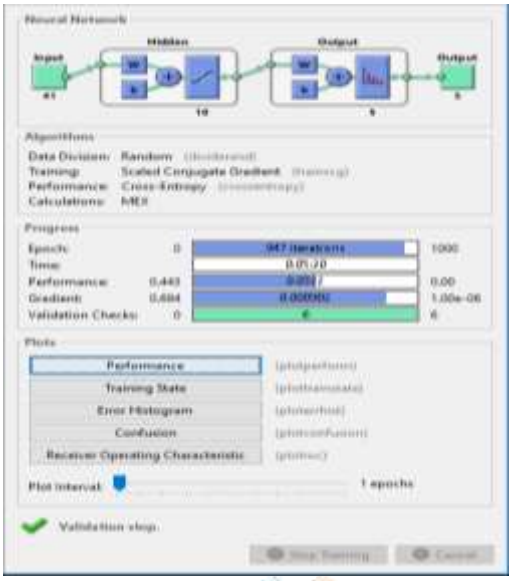


Fig 2. TRAINING SCREENSHOT

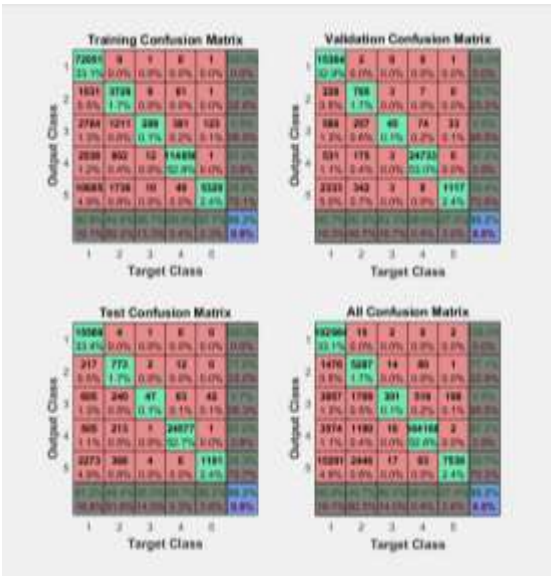


Fig 3. CONFUSION MATRIX

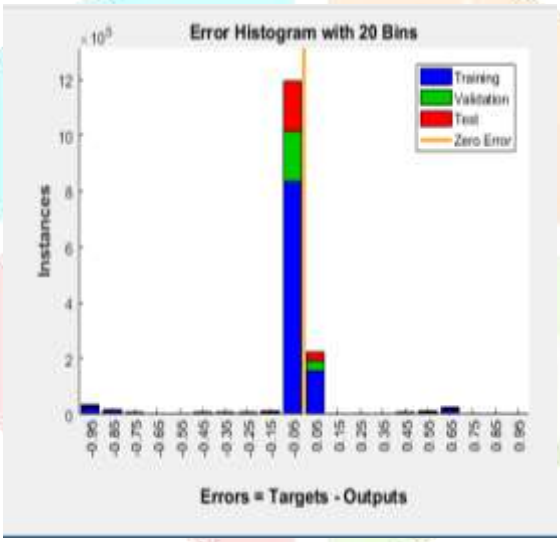


Fig 4. ERROR HISTOGRAM

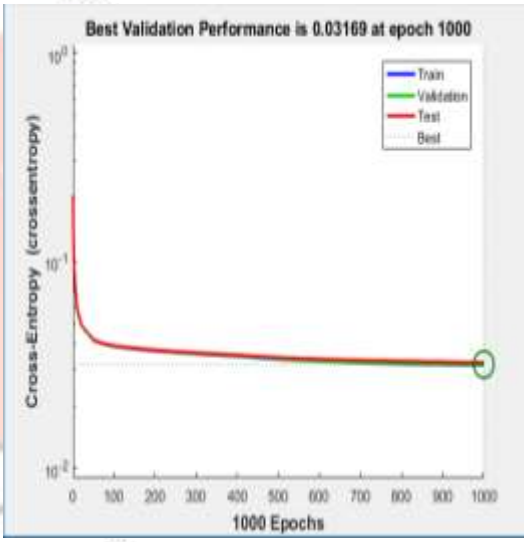


Fig 5. PERFORMANCE PLOT

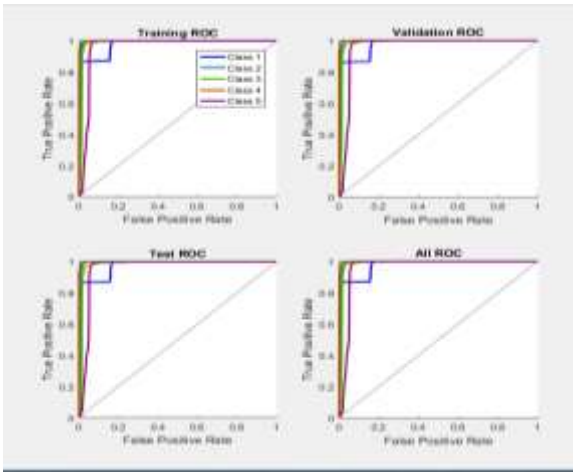


Fig 6. ROC

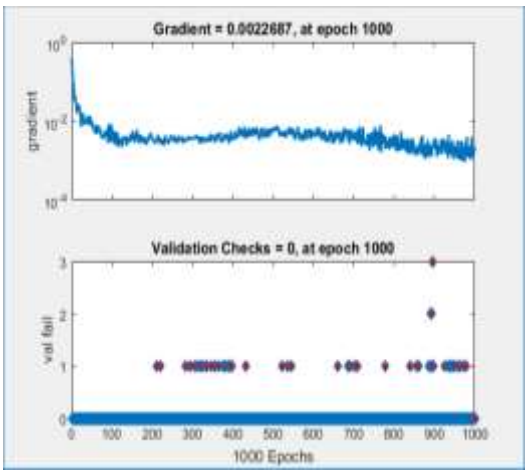


Fig 7. TRAINING STATE

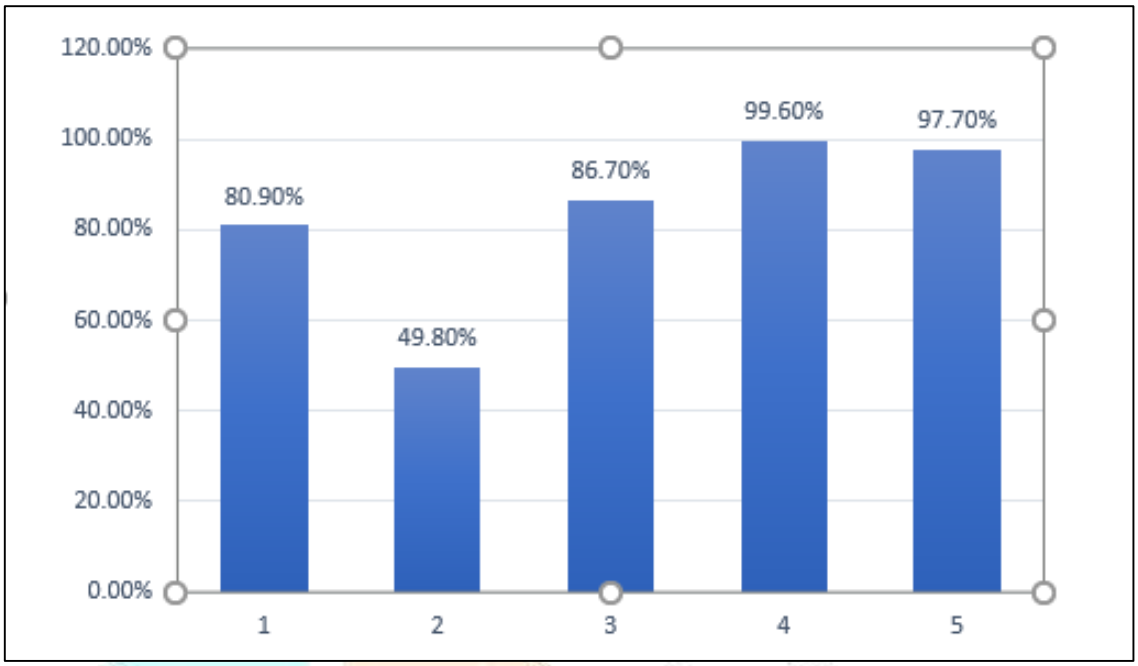


Fig 8. BAR GRAPH

CALCULATIONS

Things to notice :

- Attacks correctly detected as Attacks are termed as True Positive (TP), and its value for each is obtained through all the diagonal values of the confusion matrix (shown in green boxes).
- Attacks misclassified as Normal is termed as False Negative (FN) and the total number of FN's for a class is the sum of values in the corresponding row (excluding the TP)
- Normal correctly classified as Normal is classified as True Negative (TN), and the total number of TN's for a class will be the sum of all columns and rows (excluding same class's row and column).
- Normal misclassified as Attack is termed as False Positive (FP), and the total number of FP's for a class is the sum of values in the corresponding column of that class(excluding the TP).
-

Now we are interested in finding the Recall value and Specificity value through our obtained confusion matrix.[7]

1. Recall value, commonly called as "**Sensitivity**" of a particular class gives us the true positive rate of that particular class, i.e.,

$$\text{Recall} = \text{Sensitivity} = TP/(TP + FN)$$

For example, in the above given Confusion Matrix, the Recall Value for class 1 can be calculated as

$$\text{Recall 1} = \text{Sensitivity 1} = TP_1/(TP_1 + E_{12} + E_{13} + E_{14} + E_{15})$$
$$= 72051/(72051+9+1+0+1)$$
$$= 0.9998$$

2. Specificity of any given class corresponds to the true negative rate of the corresponding class, i.e.,

$$\text{Specificity} = TN / (TN+FP)$$

For example, in the above given confusion matrix, the specificity for class 1 can be calculated as

$$\text{Specificity 1} = TN_1/(TN_1 + E_{21} + E_{31} + E_{41} + E_{51})$$
$$= 128510/(128510+1031+2764+2538+10685)$$
$$= 0.8830$$

3. The third parameter which can be evaluated through the confusion matrix is the precision, which is calculated as

$$\text{Precision} = TP / (TP+FP)$$

For example, the precision for class 1 can be calculated as

$$\text{Precision 1} = TP_1 / (TP_1 + E_{21} + E_{31} + E_{41} + E_{51})$$
$$= 72051/(72051+1031+2764+2538+10685)$$
$$= 0.8089$$

Note: Where Eij corresponds to the value in the confusion matrix in i^{th} row and j^{th} column.

REFERENCES

[1] H.GUNES KAYASICK, A. NUR ZINCIR-HEYWOOD and MALCOM I. HEYWOOD," Selecting Features for Intrusion Detection : A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets": Proceedings of the third annual conference on privacy, security and trust, 10/2005
[2]M.E. Elhamahmy, Hesham N. Elmahdy and Imane A. Saroit, " A New approach for Evaluating Intrusion Detection system": CiiT International Journal Of Artificial Intelligent Systems and Machine Learning, Vol 2, No. 11, November 2010.
[3] Rajeev Singh, " Introduction to Intrusion Detection System": International Journal of Electrical and Electronics Research (IJEER), Vol.2 Issue 1,pp:(1-6), Month: January-March 2014.
[4] NING-QING SUN and YANG LI, FGIT 2009"Intrusion Detection Based on Back-propagation Neural Network and Feature Selection Mechanism".
[5] M.Madhavi," An approach For Intrusion Detection System in Cloud computing": International Journal of Computer Science and Information Technologies, Vol.3 (5), 2012, 5219-5222.
[6]https://www.google.co.in/search?q=back+propagation+network&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjv3cWl4vXYAhXkJsAKHYaLDFIQ_AUICygC&biw=767&bih=780#imgrc=Acgm985E12M_GM:
[7]https://www.youtube.com/watch?v=FAr2GmWNbT0.
[8] R. Graham, "FAQ: Network Intrusion Detection Systems", March 21, 2000.