Gradient Controlled-BP Algorithm for Effective Intrusion Detection

Priyanka
C.S.E. Department
P.I.E.T, Kurukshetra University,
Haryana, India

Shekhar Sengar
A.P., C.S.E. Department
P.I.E.T, Kurukshetra University,
Haryana, India

S.C. Gupta
H.O.D, C.S.E. Department
P.I.E.T, Kurukshetra University,
Haryana, India

ABSTRACT
High level security maintenance is very important nowadays for safe and trusted communication over the internet but due to enormous interconnectivity this task has become very complex. Threat of intrusions and misuses is always present in communication over the internet and any other network. These intrusions are occurring at higher rates than before and additionally existing security products are not able to detect these. Neural networks can help in this problem and not only can the known but unknown intrusions also be detected with certain efficiency. Due to high error rate and low detection rate BP algorithm's efficiency is not unto mark. So this research has used a Gradient based BP algorithm for detection of intrusions considering all 41 inputs from dataset. It shows how learning with Gradient-based BP algorithm and testing it in real time can improve efficiency. The desired results that are low false detection and high accuracy are achieved with this and for better results KDD99 datasets are also filtered.

Keywords
Network Security, Intrusion Detection, Neural Networks, Back Propagation, Gradient.

1. INTRODUCTION
It is unfortunate that the dangers and chances of malicious intrusions are rising date by date as the benefits of internet and supplementary webs are increasing. Like the past easy protection methods are not adequate now. Alongside firewalls and virus protection arrangements Intrusion detection has safeguarded a colossal space in web protection landscape. Intrusion Detection concept[1] that seems to be extremely convoluted is easy one truly – all the inbound and outbound data is examined and malicious outlines that facts of an arrangement or web aggressions are recognized. Intrusion Detection can be requested in two ways- Whichever by ideal of lawful deeds (anomaly-based IDSs) or by employing by now recognized signatures (misuse-based IDSs) [2].A analogy of the continuing traffic records alongside the pending data is the key instrument in misuse-based IDSs and on the supplementary hand recognizing deviation from the normal deeds is the key instrument in Anomaly-based IDSs.

Unknown Intrusions stay undetected in misuse-based IDSs that is the main drawback of this but this setback can be resolved by employing Behavior-based IDSs whereas the detection is not completed by merely contrasting the pending data packets alongside the continuing recognized signatures[3].Anomaly-based methods are extra effectual as contrasted to misuse-based methods because anomaly-based methods observes the deeds and if deviation occurs from the anticipated or normal deeds next it considers it as intrusion.

Neural webs are extremely flexible and good classifiers and that's why these are flawlessly suitable for IDS as the main work of IDS is alike that of neural webs that is classification. Different kinds of training can be utilized for this purpose. Different types of training are present. In supervised training desired output is known but in unsupervised training it is not known [4].

2. RELATED WORK
Robert Mitchell et al., 2014 [5] This paper includes Cyber-Physical Systems (CPSs) whose examples are unmanned aircraft systems, Pervasive healthcare systems, smart grids have become an inseparable part of our lives nowadays. More is the integration of these techniques to day to day life more will be rise in the importance of security of these. To classify modern CPS Intrusion Detection System (IDS) techniques in two design dimensions: detection technique and audit material is the basic approach used there. The effectiveness of IDS techniques as they apply to CPSs is main highlight of the work. Most and least studied CPS IDS techniques are also summarized in this work.

Amin Dastanpour et al., 2014 [6] This paper concludes how the detection of threats in network based systems have become very complex and important. Genetic Algorithms (GA) with Artificial Neural Networks classifier, Modified Mutual Information Feature selection (MMIlFS), Linear Correlation Feature Selection (LCFS), and Forward Feature Selection (FFS) are the methods to detect the attacks which are studied, implemented and compared in this paper. Datasets used in this process are KDD CPU datasets .Requirement of features used is different for the different methods. In case of GA-ANN 18 features are required for this purpose but for MMIFS, LCFS and FFSA 24, 21, and 31 features are required.

Alina Oprea et al., 2014 [7] sophisticated attacks including advanced persistent threats (APTs) which pose severe risks to organizations and governments by targeting confidential proprietary information and new malware strains that appearing at a higher rate than ever before are the topics of investigation. Anti-virus, firewalls, intrusion detection systems, often fail at detecting infections at an early stage which makes the problem more severe. Selected packet inspection scheme is used for mining large scale log data.

Liyuan Xiao et al., 2014 [8] This paper shows how to detect intrusion with reasonable accuracy and efficiency BN network classifiers play an important role. Two problems which limits the efficiency of these networks: heuristic methods used for training of data and time consuming process are also discussed in this paper. A Bayesian classifier by Bayesian Model Averaging (BMA) over the k-best BN classifiers,

called Bayesian Network Model Averaging (BNMA) classifier is suggested to reduce this problem.

M .Govindarajan et al., 2014 [9] Homogeneous ensemble classifiers using bagging and heterogeneous ensemble classifiers using arcing classifier and their performances are analysed in terms of accuracy are proposed worked in this research work. By the means of real and benchmark data sets of intrusion detection the benefits of given approach are shown.

Weiming Hu et al., 2014 [10] Requirement of improvement in Intrusion detection systems lacking adaptability to the frequently changing network environments and intrusion detection in the new distributed architectures is the main content of this paper. For this purpose improved online Adaboost process is proposed in this paper week classifiers used here are online Gaussian mixture models (GMMs).

Mradul Dhakar et al., 2014 [11] Hybrid model of intrusion detection has been proposed by this paper. Crucial data mining techniques are used for this purpose. It will lead to effective, adaptive and intelligent intrusion detection.

3. PROPOSED WORK

Traditional methods of web intrusion detection are instituted on the saved outlines of understood attacks. The main drawback of instituted methods is that they cannot notice unfamiliar intrusion. It is additionally capable of knowing new aggressions to a slight degree of resemblance to the learned ones; the neural webs are extensively trusted as an effectual method to adaptively categorize outlines. With the assistance of data excavating methods, these difficulties can be facilely overcome. Briefly, the data excavating methods have endowed the following benefits:

- **Improved variants detection**

This is exceptionally real for anomaly detection. Not manipulated to pre-defined signatures, the concern alongside variants is not as distant as beforehand, as every single deviation from a normal signature will be indulged as intrusion, encompassing those beforehand unfamiliar variants of intrusions.

- **Controlled fake alarms**

Even nevertheless these are fake positives, alongside a discovering procedure to understand recurring sequences of fake alarms, it is probable for us to filter those normal arrangement hobbies and retain the rate of fake alarms at a satisfactory level.

- **Reduced fake dismissals**

With data excavating methods, outlines (or signatures) of normal hobbies and atypical events (intrusions) can be crafted automatically. It is additionally probable to familiarize new kinds of aggressions across an incremental discovering process. As a consequence, supplementary and supplementary aggressions can be noticed correctly. This leads to a cut number of fake dismissals.

- **Improved efficiency**

One tremendously appealing feature of data excavating methods is the skill to remove most meaningful data out of large numbers of data. Later a pace of feature extraction and/or feature selection, the discovering procedure can be finished distant supplementary efficiently.

Classification is the most vital supervised discovering procedure utilized to find the constructions or outlines in a collection of labelled data. The Main Contributions of this Scrutiny are:

1. A comprehensive comparative discover of Neural Web methods for analysing colossal intrusion detection datasets is to be conducted. The work will examine the suitability of Manmade Neural Webs and empirically contrasted countless unsupervised discovering to recognize the unseen or unfamiliar attack in words of run-time efficiency.
2. The main aim of this work is to elucidate this supremacy of Neural Web technique. In this work, an endeavour is made to enhance the discovering skills of neural web and cut the number of era and resource demanded by discovering procedure by sampling the input data set to be discovering retaining Soft computing method.
3. Reduction of fake forecast worth and rises in assurance value.
4. Lower rate of fake detection and fake alarm.

4. METHODOLOGY

Neural Network

The neural web has a facile clarification as a form of input-output flawless, alongside the weights and thresholds (biases) the free parameters of the model. Such webs can flawless intentions of nearly arbitrary intricacy, alongside the number of layers, and the number of constituents in every single solitary layer, ascertaining the intention complexity. Vital subjects in MLP design encompass specification of the number of hidden layers and the number of constituents in these layers.

Normally sigmoid purpose is utilized for this ideal and is expressed as follows:

$$f(x)=1/(1+e^{-x})$$

Each synaptic link has a web weight. The web heaviness from constituent i to constituent j is expressed as w_{ij} and the output worth for constituent i is expressed as O_i. The output benefits input signal. Consequently, to change the output worth to a wanted worth, adjustment of these web weights are needed. In counselled method, we use back propagation discovering as discovering method. Back propagation discovering is a supervised discovering.

Training Phase

The multilayer back propagation algorithm is utilized to train the neural web for classifying the intrusions. Weights are initialized to random benefits amid +0.1 and -0.1 and acceded error is selected as 0.009.41 inputs are taken and 20 hidden layers are used for the training purpose. One output layer provides the output as shown in figure 1.

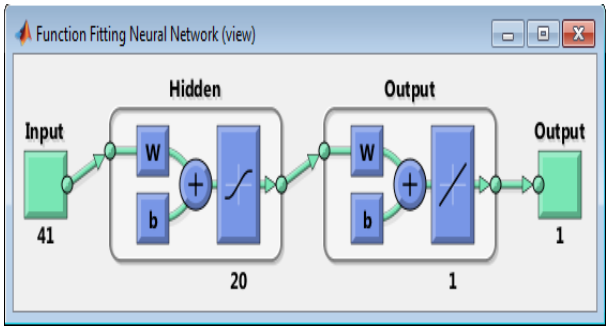

Figure 1: Architecture of Function Fitting Neural Network

Activation function
If a multilayer perceptron consists of a linear activation intention in all neurons, that is, a facile on-off mechanism to notice whether or not a neuron fires, subsequent it is facilely proved alongside linear algebra that every single number of layers can be cut to the average two-layer input-output model. What makes a multilayer perception disparate is that every single solitary neuron uses a nonlinear activation intention that was industrialized to flawless the frequency of deed potentials, or enkindling, of biological neurons in the brain. This intention is modelled in countless methods, but has to always be normalizable and differentiable.

The two main activation intentions utilized in present demands are both sigmoid, and are delineated by hyperbolic tangent that scopes from -1 to 1, and the last is equivalent in form but scopes from 0 to 1. Here y_i is the output of the ith node (neuron) and v_i is the weighted sum of the input synapses. Supplementary enumerated activation intentions encompass radial basis intentions that are utilized in one extra class of supervised neural web models. Most area activation intentions are the logistic and hyperbolic tangent sigmoid functions.

Weight Adjustment
The webs weights demand to be adjusted in order to minimize the difference or the error amid the output and the anticipated output. This is elucidated in the equations below. The error gesture at the output layer of the ith neuron at iteration n is given by

$$e_i(n) = X_i(n) - X'_i(n)$$

Where Xi represent the desired output and X'i represent the actual output. The error function over all neurons in output layer is given by Eq.

$$E_l(n) = \sum e_i^2(n)$$

The error function, over all input vectors in the training image, is

$$E = \sum E_l,\ E_l = (X', w)$$

where l indexes the picture blocks (inputs vector), X' is the vector of outputs, and w is the vector of all weights. In order to minimize the error purpose alongside respect to heaviness vector (w) it is vital to find an optimal resolution (w*) that gratify the condition:

$$E(w^*) \le E(w)$$

The necessary condition for the optimality is

$$\Delta E(w) = 0$$

where Δ is gradient operator, ΔE (w) is gradient vector (g) of error function is defined as follows:

$$\Delta E(w) = \partial E / \partial w$$

The resolution can be obtained employing a class of unconstrained optimization methods established on the believed of innate iterative descent. Starting alongside early estimate denoted w(0), produce a sequence of eight vectors w(1),w(2)… such that the error purpose is decreased for every single iteration:

$$E(w_{n+1}) \le E(w_n)$$

Back propagating Perceptron with Gradidant Decsent for IDS:

- Let $x_1,\ldots x_n$ be the traning set (current search space in our case of Intrusion Detection)
- Consider linear unit without threshold and continuous output o (–1,1)
- $o=w_0 + w_1 x_1 + \ldots + {}_{wn} x_n$
- Initialize each w_{ij} to some small random value
- Until the termination condition is met, Do

  For each training example $<(x_1,\ldots x_n),t>$ Do
- Input the instance $(x_1,\ldots,x_n)$ to the network and compute the network outputs o_k

  For each output unit k

  $\delta_k = o_k(1-o_k)(t_k-o_k)$

  For each hidden unit h

  $\delta_h = o_h(1-o_h) \sum_k w_{h,k} \delta_k$

  For each network weight w_{ij} Do

  $w_{i,j}=w_{i,j}+\Delta w_{i,j}$ where $\Delta w_{i,j}= \eta \delta_j x_{i,j}$ here η is learning rate.

  $\delta_{k \text{ and}} \delta_h$ are errors at k^{th} and h^{th} nodes.

This is shown in Fig. 2 on next page.

After Training of weight Layers Gradient descent can be utilized to find optimal solution:

- Gradient descent over entire network weight vector

  Easily generalized to arbitrary directed graphs
- Will find a local, not necessarily global error minimum in practice often works well (can be invoked multiple times with different initial weights)
- Often include weight momentum term

  $\Delta w_{i,j}(t)= \eta \delta_j x_{i,j} + \alpha \Delta w_{i,j}(t-1)$
- Minimizes error training examples it generalize well to unseen instances (i.e. avoids over-fitting).

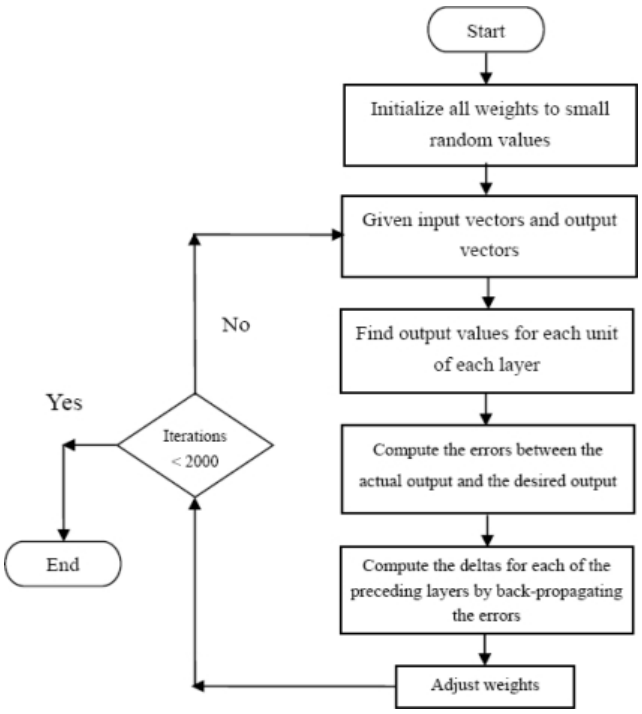

Figure 2: Neural Network Training Chart for IDS

5. RESULT AND ANALYSIS

IDS Data Set

With the colossal progress of computer webs rehearse and the huge development in the number of demands running on top of it, web protection is becoming increasingly supplementary important. All the computer arrangements tolerate from protection vulnerabilities that are both technically tough and frugally luxurious to be resolved by the manufacturers. Therefore, the deed of Intrusion Detection Arrangements (IDSs), as special-purpose mechanisms to notice anomalies and aggressions in the web, is becoming supplementary important. The scrutiny in the intrusion detection earth has been usually pondered on anomaly-based and misuse-based detection methods for a long time. As misuse-based detection is normally favored in company produce due to its predictability and elevated accuracy, in intellectual scrutiny anomaly detection is normally conceived as a supplementary prominent method due to its hypothetical probable for addressing novel attacks.

Conducting a methodical scrutiny of the present scrutiny trend in anomaly detection, one will encounter countless contraption discovering methods delineated to have a tremendously elevated detection rate of 98% as keeping the fake alarm rate at 1%. Though, afterward we stare at the state of the fine art IDS resolutions and company instruments, there is insufficient produce retaining anomaly detection methods, and practitioners yet contemplate that it is not a mature vision yet. To find the reason of this difference, we learned the features of the scrutiny finished in anomaly detection and trusted varied aspects such as discovering and detection methods, training data sets, assessing data sets, and evaluation methods.

The main vital deficiency in the KDD data set is the huge number of redundant records. Analysing KDD train and examination sets, we discovered that considering 78% and 75% of the records are duplicated in the train and examination set, respectively. This large number of redundant records in the train set reasons discovering algorithms to be biased towards the supplementary recurrent records, and consequently halt it from discovering infrequent records that are normally supplementary harmful to webs such as U2R attacks. The attendance of these recapped records in the examination set, on the supplementary hand, will cause the evaluation aftermath to be biased by the methods that have larger detection rates on the recurrent records.

In supplement, to scrutinize the difficulty level of the records in KDD data set, we retained 41 parameters to label the records of the finished KDD train and examination sets.

20 hidden layers have been used here and one output layer. The reason we came to be these statistics on both KDD train and examination sets is that in countless papers, random servings of the KDD train set are utilized as examination sets. As a consequence, they finish considering 98% association rate demanding tremendously facile contraption discovering methods. Even demanding the KDD examination set will consequence in owning a minimum association rate of 86% that makes the analogy of IDSs quite tough as they all vary in the scope of 86% to 100%.

KDD CUP 99 DATA SET DESCRIPTION:

KDD'99 has been the most wildly utilized data set for the evaluation of anomaly detection methods. This data set is instituted on the data grabbed in DARPA'98 IDS evaluation program[14].DARPA'98 is considering 4 gigabytes of compressed raw (binary) tcp dump data[15] of 7 weeks of web traffic, that can be processed into considering 5 million connection records, every single solitary alongside considering 100 bytes. The two weeks of examination data have considering 2 million connection records. KDD training dataset consists of considering 4,900,000 solitary connection vectors every single solitary of that encompasses 41 features and is labelled as whichever normal or an attack, alongside precisely one specific attack type. The simulated aggressions plummet in one of the pursuing four clusters:

1) **Denial of Service Attack (DoS)**: It is an attack in that the attacker makes a little computing or recollection resource too busy or too maximum to grasp legitimate demands, or denies legitimate users admission to a machine.

2) **User to Origin Attack (U2R)**: is a class of exploit in that the attacker starts out alongside admission to a normal user report on the arrangement (perhaps obtained by sniffing passwords, a lexicon attack, or communal engineering) and is able to exploit a little vulnerability to gain origin admission to the system.

3) **Remote to Native Attack (R2L)**: occurs after an attacker who has the skill to dispatch packets to a contraption above a web but who does not have a report on that contraption exploits a little vulnerability to gain innate admission as a user of that machine.

4) **Probing Attack**: is an endeavour to gather data concerning a web of computers for the seeming intention of circumventing its protection controls.

Fig 3 shows the error histogram for the data set we have used.70%of the data is used for training and 15% for validation and 15% for the testing purpose and Fig 4 shows the regression at training, validation, testing and overall level.

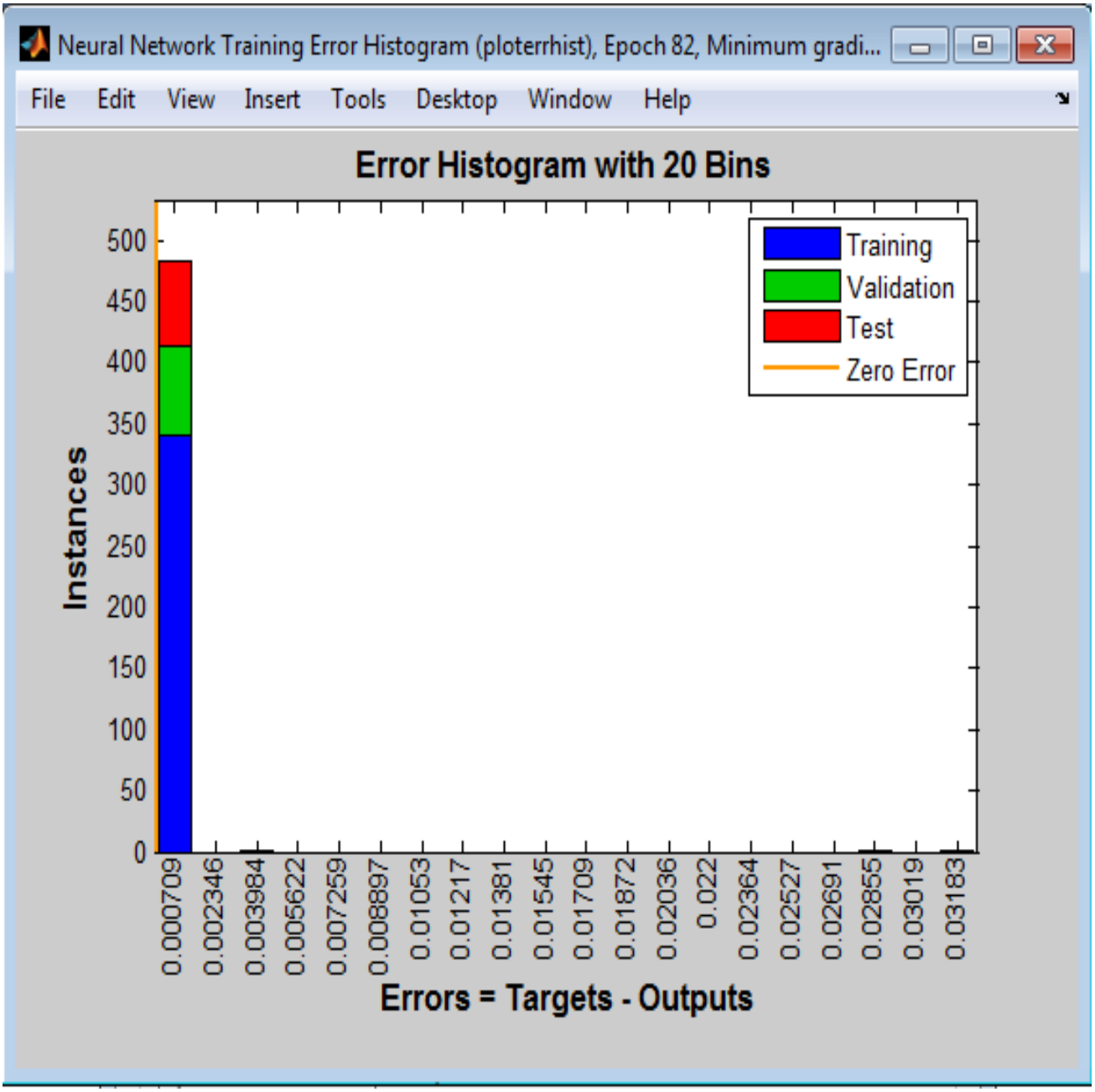

Figure 3: Calculated Errors histogram for the input dataset

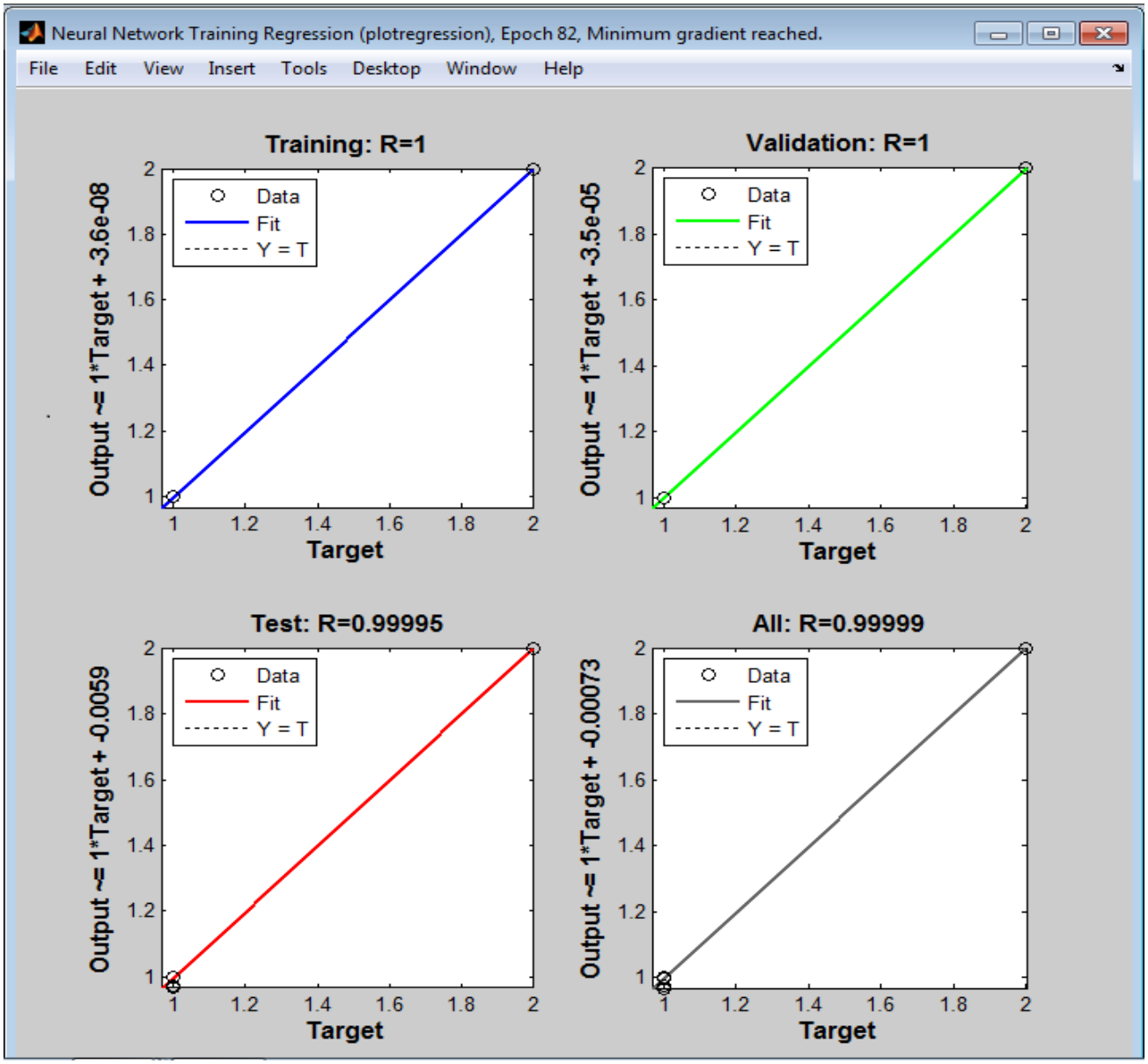

Figure 4: Success Rate at different Levels

6. CONCLUSION AND FUTURE SCOPE

There are additionally a number of unsolved subjects pondering the scrutiny of the audit trail. Signature scrutiny is clearly in the company span nowadays, but has been shown to be insufficient for noticing all attacks. Therefore, work is yet in progress to examination alongside new methods to both knowledge-based and deeds instituted intrusion detection. The detection of abuse-of-privilege aggressions (primarily associate attacks) is additionally the subject of ongoing work.

In hybrid method, we will focus on how supplementary than one above method are used. In the consecutive serving, we will endeavor to difference the centroid - instituted clustering and disparate ANN instituted methods.

7. ACKNOWLEDGMENT

We, the writers, would like to show the gratitude to the anonymous arbitrators for their respected annotations that enriched the content and the staging of this research and the paper.

8. REFERENCES

[1] Corchado, Emilio, and Álvaro Herrero. "Neural visualization of network traffic data for intrusion detection." Applied Soft Computing 11, no. 2 (2011): 2042-2056.

[2] Modi, Chirag, Dhiren Patel, Bhavesh Borisaniya, Hiren Patel, Avi Patel, and Muttukrishnan Rajarajan. "A survey of intrusion detection techniques in cloud." Journal of Network and Computer Applications 36, no. 1 (2013): 42-57.

[3] Casas, Pedro, Johan Mazel, and Philippe Owezarski. "Unsupervised network intrusion detection systems: Detecting the unknown without knowledge." Computer Communications 35, no. 7 (2012): 772-783.

[4] Casas, Pedro, Johan Mazel, and Philippe Owezarski. "Unsupervised network intrusion detection systems: Detecting the unknown without knowledge." Computer Communications 35, no. 7 (2012): 772-783.

[5] Robert Mitchell and Ing-Ray Chen. "A survey of intrusion detection techniques for cyber-physical systems." ACM Computing Surveys (CSUR) 46, no. 4 (2014): 55.

[6] Amin Dastanpour Suhaimi Ibrahim, and Reza Mashinchi. "Using Genetic Algorithm to Supporting Artificial Neural Network for Intrusion Detection System." In The International Conference on Computer Security and Digital Investigation (ComSec2014), pp. 1-13. The Society of Digital Information and Wireless Communication, 2014.

[7] Alina Oprea Zhou Li, Ting-Fang Yen, Sang Chin, and Sumayah Alrwais. "Detection of Early-Stage Enterprise Infection by Mining Large-Scale Log Data." arXiv preprint arXiv: 1411.5005 (2014).

[8] Liyuan Xiao Yetian Chen, and Carl K. Chang. "Bayesian Model Averaging of Bayesian Network Classifiers for Intrusion Detection." In Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International, pp. 128-133. IEEE, 2014.

[9] M .Govindarajan "Hybrid Intrusion Detection Using Ensemble of Classification Methods." IJ Computer Network and Information Security 2 (2014): 45-53.

[10] Weiming Hu Jun Gao, Yanguo Wang, Ou Wu, and Stephen Maybank. "Online adaboost-based parameterized methods for dynamic distributed network intrusion detection." Cybernetics, IEEE Transactions on 44, no. 1 (2014): 66-82.

[11] Mradul Dhakar and Akhilesh Tiwari. "A Novel Data Mining based Hybrid Intrusion Detection Framework." Journal of Information and Computing Science 9, no. 1 (2014): 037-048.

[12] Laheeb Mohammad Ibrahim, "Anomaly network intrusion detection system based on distributed time-delay neural network (DTDNN)." Journal of Engineering Science and Technology 5, no. 4 (2010): 457-471.

[13] Susan C Lee and David V. Heinbuch. "Training a neural-network based intrusion detector to recognize novel attacks." Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 31, no. 4 (2001): 294-299.

[14] KDD-CUP-99 Task Description; http://kdd.ics.uci.edu/databases/kddcup99/task.html

[15] H.G.Kayacık,A.N.Zincir-Heywood,M.I. Heywood, "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets", May 2005.

[16] Manoranjan Pradhan, Sateesh Kumar Pradhan, Sudhir Kumar Sahu, "Anomaly Detection Using Artificial Neural Network"International Journal of Engineering Sciences & Emerging Technologies, April 2012, Volume 2, Issue 1, pp: 29-36.