International Journal of Scientific Research in Computer Science, Engineering and Information Technology

© 2017 IJSRCSEIT | Volume 2 | Issue 5 | ISSN : 2456-3307

A Performance Evaluation of Intrusion Detection system to get better detection rate using ANN Technique

Aakanksha Kori[1], Harsh Mathur[2]

[1]Research Scholar, Department of Computer Science and Engineering, IES College of Technology, Bhopal, Madhya Pradesh, India
[2]Assistant Professor, Department of Computer Science and Engineering, IES College of Technology, Bhopal, Madhya Pradesh, India

ABSTRACT

Intrusion Detection System (IDS) is a Detection System that works for detecting malicious attacks. This can be defined as software for security management. Many researchers have proposed the Intrusion Detection System with different techniques to achieve the best accuracy. This paper outlines an investigation on the unsupervised neural network models and choice of one of them for evaluation and implementation. In this paper, the performance of intrusion detection is compared with various neural network classifiers. In the proposed research the two algorithms used are Back-propagation algorithm and Growing Self organization Map algorithm. After implementing these algorithms, we have proposed a comparative analysis between them and choose the best accuracy rate among them. Here, it has been proved that, the ANN procedure is validated against a simulated IoT network. The experimental results demonstrate far better accuracy and when use in implementation of application software, it can successfully detect various attacks.

Keywords: Intrusion Detection System, BP Algorithm, GSOM Algorithm

I. INTRODUCTION

Computer security is used frequently, but the content of a computer is vulnerable to few risks unless the computer is connected to other computers on a network. As the use of computer networks, especially the Internet, has become pervasive, the concept of computer security has expanded to denote issues pertaining to the networked use of computers and their resources. The major technical areas of computer security are usually represented by the initials confidentiality, integrity, and authentication or availability. Confidentiality means that information cannot be access by unauthorized parties. Confidentiality is also known as secrecy or privacy; breaches of confidentiality range from the embarrassing to the disastrous. Integrity means that information is protected against unauthorized changes that are not detectable to authorized users; many incidents of hacking compromise the integrity of databases and other resources. Authentication means that users are who they claim to be. Availability means that resources are accessible by authorized parties; "denial of service" attacks, which are sometimes the topic of national news, are attacks against availability. Other important concerns of computer security professionals are access control and no repudiation.

The main goal of intrusion detection is to detect unauthorized use, misuse and abuse of computer systems by both system insiders and external intruders. Among automated intrusion detection systems, a particular system for network intrusion detection, known as a network-based intrusion detection system (IDS), monitors any number of hosts on a network by scrutinizing the audit trails of multiple hosts and network traffic. It is usually comprised of two main components: an anomaly detector and a misuse detector [1][2]. The anomaly detector establishes the profiles of normal activities of users, systems, system resources, network traffic and/or services and detects intrusions by identifying significant deviations from the normal behavior patterns observed from profiles. The misuse detector defines suspicious misuse signatures based on known system vulnerabilities and a security policy. This component probes whether these misuse signatures are present or not in the auditing trails. Currently many network-based IDS's have been developed using diverse approaches Nevertheless, there still remain unresolved problems to build an effective network-based IDS [6]. As one approach of providing the solutions of these problems, the previous work [8] identified a set of general requirements for successful network-based IDS and three design goals to satisfy these requirements: being distributed, self organizing and lightweight. In addition, Kim and Bentley (1999a) introduced a number of remarkable features of human immune systems that satisfy these three design goals. It is anticipated that the adoption of these features should help the construction of an effective network based IDS This paper proposes the use GSOM algorithm for developing an effective network-based IDS to improve detection rate in IDS System.

II. IDS FUNCTIONS

Functions of IDS are

- Monitoring users and system activity.
- Auditing system configuration for vulnerabilities and misconfigurations.
- Assessing the integrity of critical system and data files.
- Recognizing known attack patterns in system activity.
- Identifying abnormal activity through statistical analysis.
- Managing audit trails and highlighting user violation of policy or normal activity.
- Correcting system configuration errors
- Installing and operating traps to record information about intruders.

III. CONCEPT OF NEURAL NETWORK

The concept of neural network is highly inspired by the recognition mechanism of the human brain [1, 20]. The human brain is a complex, nonlinear and parallel computer, whereas the digital computer is entirely the opposite, it is a simple, linear and serial computer. The capability to arrange neurons to perform computation is again and again quicker than a contemporary information processing system existing these days. Human vision is a good example for understanding this difference.

There is no universally accepted definition of neural network, but there are some architectural and fundamental elements that are the same for all neural networks. First of all, a neural network is a network with many simple processors, which are known as the neurons.

They have the task to receive input data from other neurons or external sources and use this to compute new data as output for the neural network or input data to the neurons of the next layer. Communication channels, better known as the weights, carry the received or computed data. The weights, which connect two neurons posses certain, values and will be adjusted upon network training [19]. The adjustment of the weights is processed in parallel, meaning that many neurons can process their computations simultaneously. The magnitude of the adjustment of the neurons depends on the training data and is carried out with a so called training.

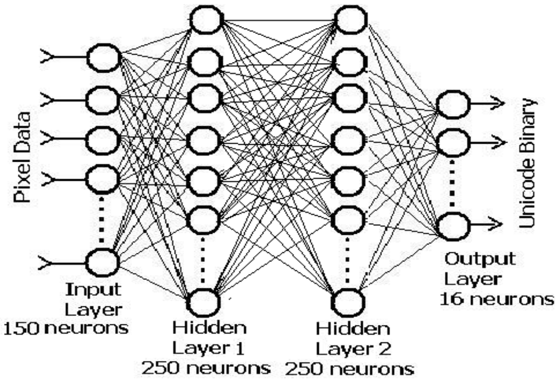

Figure 3.1. Structure of a simple Fully-Connected Neural Network with four layers

Network model is shown (Fig. 5.1). In this project, four layers were used where the layers are organized as follows; the first layer is the input layer that receives data from a source. The 4th is the output layer that sends computed records out of the neural network. The second and third layer is called hidden layer, whose input and output signals remain within the neural network, see 5.1. In the particular example, the network is fully connected, which means that every neuron in one layer is connected with all neurons in the preceding layer and so on. Although it is not a rule and a neural network does not need to be fully connected. It is used in artificial intelligence, have traditionally been viewed as simplified models of neural processing in the brain. Roughly, the overall task of a neural network is to predict or make approximately correct results for a given condition. Neural networks are trained with training data and the elements (e.g., neurons and weights) of the network will be adjusted on the basis of this training data. When further training does not change the network significantly or a given criterion is fulfilled the network is ready to produce results. Test data can be put into the network, be processed and the network will come up with a result.

IV. LITERATURE SURVEY

A lot of research works have been carried out in the literature for intrusion detection and some of them have motivated us to take up this research. Brief reviews of some of those recent significant researches are presented below:

Elike Hodo, Xavier Bellekens, Andrew Hamilton,Pierre-Louis Dubouilh al[1] This paper focuses on the type of regular and threat styles on an iot network. The ANN procedure is validated against a simulated IoT network. The experimental results demonstrate 99.4% accuracy and can successfully detect various DDoS/DoS attacks

Tich Phu oc Tran [2] has applied Machine Learning techniques to solve Intrusion Detection problems within computer networks. Due to complex and dynamic nature of computer networks and hacking techniques, identifying malicious activities remains a challenging task for security experts, that is, defense systems that were currently available suffer from low detection capability and high number of false alarms.

Ye Yuan et [3] proposed a method of evidence assignment in combination with Dempster-Shafer theory to identify network attack data. In this method, extracted features were identified by a multi generalized regression neural network classifier, which determined the basic probability assignment.

Snehal A [4] proposed the decision tree based algorithm to build multiclass intrusion detection system. Support Vector Machines was the classifiers which were initially designed for binary classification.

Shun J and Malki H. A.[5] presented a neural network-based intrusion detection method for the internet-based attacks on a computer network.

Aida O. Ali [6] described a relative study between the performances of recent nine artificial neural networks (ANNs) based classifiers was assessed centred on a particular set of features. The outcomes showed that; the Multilayer perceptrons (MLPS) based classifier yielded the best results; about 99.63% true positive attacks were detected.

Pohsiang Tsai [7] suggested a Machine Learning (ML) framework in which various types of intrusions would be detected with different classifiers, containing different attribute selections and learning algorithms. Appropriate voting techniques were used to combine the outputs of these classifiers. The pattern-learning abilities of the IS has been modeled and described by **Timmis, Neal, and Hunt (2008) and Dasgupta, Cao, and Yang (2003)** who successfully applied their AISs to recognition and classification tasks.

Also **Byoung-Doo [8]** in 2006 built IDS deals well various mutated attacks, as well as well-known attacks by using Time Delay Neural Network classifier that discriminates between normal and abnormal packet flows. It seems that the area here the notion of AIS has been most widespread is in the area of computer security.

A. H. M. Rezaul Karim [9] proposed collaborative IDS for MANET using Bayesian method using a set of very useful features which guarantee the effectiveness of the IDS [12].

L. Khan and et al. [10] proposed a method with a scalable solution for detecting network based anomalies [13]. They used Support Vector Machines (SVM) for classification. They used the Dynamically Growing Self-Organizing Tree (DGSOT) algorithm for clustering.

Tsong and et al.[11] introduced a three-tier architecture of intrusion detection system which consists of a blacklist, a whitelist and a multiclass support vector machine classifier.They designed a three-tier IDS based on the KDD'99 benchmark dataset.

Weiming Hu and et al.[12] proposed an intrusion detection algorithm based on the AdaBoost algorithm. The discrete AdaBoost algorithm was selected to learn the classifier.

Hu Zhengbing1 and et al.[13] proposed an algorithm to use the known signature to find the signature of the related attack quickly. They used nine different-sized databases,

Amit Kumar Choudhary and et al.[14] proposed a neural network approach to improve the alert throughput of a network and making it attack prohibitive using IDS. For evolving and testing intrusion the KDD CUP 99 dataset were used.

Stefano Zanero and et al.[15] proposed a novel architecture which implements a network-based anomaly detection system using unsupervised learning algorithms. They described how the pattern recognition features of a Self Organizing Map algorithm can be used for Intrusion Detection purposes on the payload of TCP network World Journal of Science and Technology 2012, 2(3):127-133 131 packets.

Liberios VOKOROKOS[16] presented intrusion detections systems and design architecture of intrusion detection based on neural network self organizing map. Result of the designed architecture is simulation in real conditions.

V. PROBLEM IDENTIFICATION IN PREVIOUS WORK

Intrusion Detection Systems (IDS) has become a general component in safety infrastructures as they allow network administrators to detect violations of policy. These policy violations are ranging from outside attackers trying to gain unauthorized access to insiders abusing their access.

Present day IDS have a number of sizable drawbacks-modern-day IDS are generally tuned to locate recognized carrier level network attacks. This leaves them vulnerable to original and novel malicious attacks. Data overload: Another aspect which does not relate directly to misuse detection but is extremely important is how much data an analyst can efficiently analyze. That amount of data he needs to look at seems to be flourishing rapidly. Depending on the IDS tools employed by a company and its size there is the possibility for logs to reach millions of records per day. False positives: A common place grievance is the quantity of false• positives an IDS will generate. A false positive occurs when normal attack is mistakenly classified as malicious and treated accordingly.

False negatives: This is the case in which an IDS does no longer generate an alert when an intrusion is definitely taking location. (Classification of malicious traffic as normal).

Limitations of Current State of Art:

Modern-day safety systems are going through challenges like:

- More time for execution
- Less efficient on the larger size datasets
- Slow and provide low accuracy
- Do not detect unknown or unseen attack.

VI. PROPOSED ALGORITHM

Growing Self organization Map Algorithm

A growing self-organizing map (GSOM) is a growing variant of the popular self-organizing map (SOM). The GSOM is developed to address the issue of identifying a suitable map size in the SOM (self organizing map). It starts with a least number of nodes (usually 4) and grows new nodes on the boundary based on the heuristic. Using the value called Spread Factor (SF), the data analyst has an ability to control the growth of the GSOM.

Learning Algorithm of the GSOM:

The GSOM process is as follows:

Initialization phase:

- Initialize the Weight vectors of the starting nodes (commonly 4) with random numbers between zero and 1.
- Calculate the growth threshold ( ) for the given data set of dimension according to the spread factor ( ) using the formula

Growing Phase:

- Present input to the network.
- Decide the weight vector this is closest to the input vector mapped to the current feature map (winner), using Euclidean distance. This step can be summarized as: find such that where , are the input and weight vectors respectively, is the position vector for nodes and is the set of natural numbers.
- The Weight vector alteration is implemented most effective to the neighborhood of the winner and the winner itself. The neighborhood is a set of neurons around the winner, but in the GSOM the starting neighborhood selected for weight adaptation is smaller compared to the SOM (localized weight adaptation). The amount of adaptation(Learning rate) is likewise decreased exponentially over the repeatations.. Even within the neighborhood, weights that are closer to the winner are adapted more than those further away. The adaptation of weight can be explained by where the Learning Rate , is a series of +ve parameters converging to zero as ….are the weight(wt.) vectors of the node previous and afterwards of the adaptation and is the neighbourhood of the winning neuron at the th iteration. The reducing value of in the

GSOM based on the number of nodes present in the map at time .

- Increase the error value of the winner (error value is the difference between the input vector and the weight vectors).
- When (where is the total error of node and is the growth threshold). Grow nodes if (i) is a boundary node. Assign weights to neighbours if is a non-boundary node.
- Initialize the new node weight vectors to match the neighbouring node weights.
- Initialize the learning rate ( ) to its beginning value.
- Repeat steps 2 – 7 until all inputs have been presented and node growth is reduced to a minimum level.

Smoothing Phase

Reduce learning rate and fix a small starting neighbourhood. Find winner and adapt the weights of the winner and neighbours in the same way as in growing phase.

VII. RESULT AND ANALYSIS

Because the goal of this work is to study and enhance the learning capabilities of the techniques for intrusions detection, the Back- Propagation Algorithm is compared. The use the full set of samples sampled from the KDD Cup98 dataset and witch contain 5000 sample. The original data set contain 5 million records which specify various attacks in which 1% sample consisting of about 5000 records was used in our experiment.

Now, we compare the aforementioned clustering algorithms on the whole data set with 2500 data set The Computation time for the clustering algorithms with 100 clusters are shown in Table:

Experiment 1):

In this experiment (a), we investigate computation time of GSOM and Back-Propagation Algorithm.

In the training phase, GSOM and the Back-Propagation Algorithm was used to cluster the training data. After training, each cluster was labeled according to the majority type of data in this cluster.

For instance, if more than 50% of the connections in cluster were intrusions, the cluster and its centroid weight vector would be labeled as intrusion.

GSOM Algorithm perform significantly better ($p < 5\%$) than the others in terms of computation time with much less run time. Comparing the results for 100 clusters is shown in table (6.1).

Table 6.1 Clustering results with 100 clusters with time efficiency

Cluster	Algorithm	
	Back-Propagation algorithm	Growing Self Organization Map Algorithm
	Time (ms)	Time (ms)
20	27	15
40	34	27
60	42	30
80	57	40
100	67	60

GSOM Algorithm algorithms perform significantly better ($p < 5\%$) than the others in terms of computation time. Comparing the results for 100 clusters, we observe that Back-Propagation take more execution time than GSOM Algorithm.

This experiment is run on individual clusters for an individual cluster on KDDCUP99 Data set.

This data set contain only numeric value not categorical valued. GSOM Algorithm is fast than Back-Propagation algorithm. Times are calculated in Millisecond (ms).

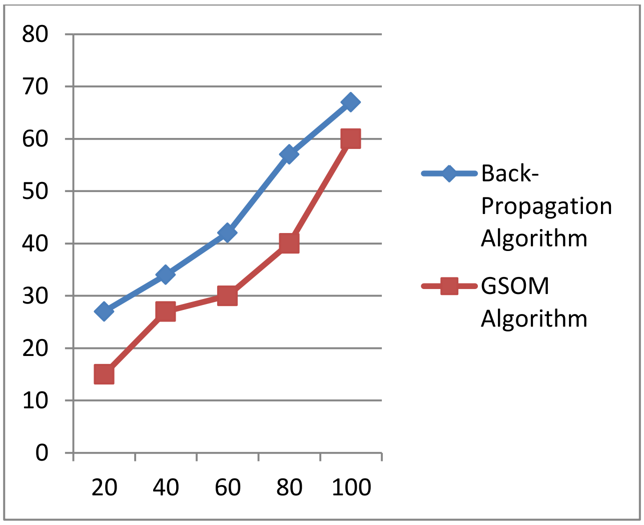

Figure 6.1 Clustering results with 100 clusters with time efficiency

Since our aim is to detect network intrusion using clustering algorithms [10], we now analyze the unsupervised intrusion detection accuracies or times for detect the unseen or new attack.

Experiment (2):

Now we find the detection rate of GSOM and Back-Propagation Algorithm. To evaluate the accuracy of a system, we use indicator: Detection Rate (DR)

DR equals the number of intrusions divided by the total number of intrusions in the data set

$$\text{DR} = \frac{\text{the number of intrusions}}{\text{The total number of intrusions in the data set}}$$

We partitioned 5000 instances of KDDCUP-99 dataset using the Back-Propagation and Self Organizing Map algorithm with different initial values of k. The Detection rate of GSOM and Back-Propagation is

Table: 6.2 Summary Detection rate results with 100 clusters shown below:

Cluster	Algorithm	
(Group of neurons or nodes)	**Back propagation algorithm (in percentage)**	**Growing Self Organization Map Algorithm(in percentage)**
20	93	99
40	74	93
60	60	84
80	44	63
100	38	52

Figure below Shows the Graph for detection rate of the GSOM

Back-Propagation Algorithm with 100 clusters respectively. It can be seen that for 100 clusters, the GSOM Algorithm has high detection rate than Back-Propagation. GSOM Algorithm clusters performs extremely well, it can detect more than 92% attacks. Overall, GSOM Algorithm better ones and stable across different Number of clusters.

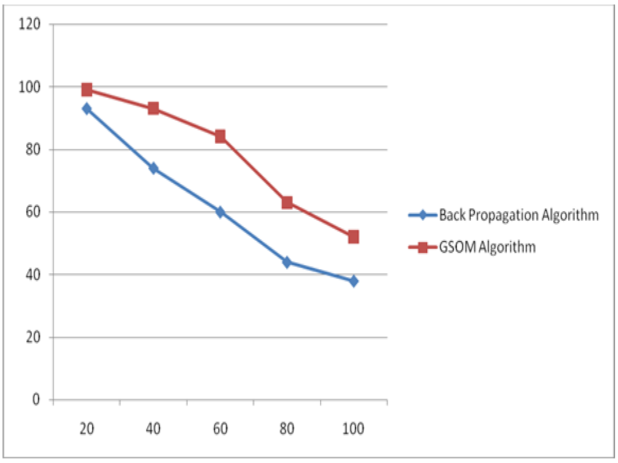

Figure 6.2 Graph for detection rate of the GSOM

VIII. CONCLUSION AND FUTURE WORK

The feasibility of unsupervised intrusion detection using centroid-based clustering algorithms is investigated in this study. Considering the dynamic nature of network traffic intrusions, unsupervised intrusion detection is more appropriate for anomaly detection than classification-based intrusion detection methods.

An empirical study consisting of centroid based clustering and Artificial Neural Network Immune System Technique is performed with a case study of data obtained from KDDCUP99 Data set. A comparative analysis and evaluation of the clustering algorithms yielded reasonable intrusion detection rates.

Promising clustering and detection results encourage us to proceed our **future work** in several directions. Identifying the precise attack category associated with a cluster and the discriminating features that are unique to a given cluster can do a further detailed analysis of individual clusters

.

In addition, feature selection/weighting for clustering will be investigated. This will eventually enhance our understanding and detection of new attack categories. Sophisticated self-labeling techniques, taking into consideration of additional network security domain knowledge, can be developed to improve the performance of clustering-based intrusion detection.

IX. REFERENCES

[1]. Elike Hodo, Xavier Bellekens, Andrew Hamilton, Pierre-Louis Dubouilh," Threat analysis of IoT networks Using Artificial Neural Network Intrusion Detection System" Systems (ICICES), pp. 232"236, 2013.

[2]. Tich Phu oc Tran, "Intrusion Detection: A Brief Introduction and History", Security & Privacy " Supplement " IEEE Computer Magazine, pp. 27-30, 2002.

[3]. Shun J and Malki H. A., "Limiting Uncertainty in Intrusion Response", 2001

[4]. IEEE Man Systems and Cybernetics Information Assurance Workshop, pp. 142-147, 2001.

[5]. Tsong Song Hwang , Tsung-Ju Lee, Yuh-Jye Lee, "A Neural Network Component for an Intrusion Detection System", Proc. 1992 IEEE Computer Society Symposium on Research in Computer Security and Privacy, pp. 240-250, 1992.

[6]. Aida.O.Ali et,"A Three-tier IDS via Data-Mining Approach",MineNet’07, pp.212-217, 2007.

[7]. Pohsiang Tsai, "On Spectral Clustering: Analysis and an Algorithm," Advances in Neural Information Processing Systems, Volume. 14, pp. 849-856, 2002.

[8]. Timmis, Neal, and Hunt, "Anomaly Detection Based on Unsupervised Niche

[9]. Clustering with Application to Network Intrusion Detection," IEEE Congress on Evolutionary Computation, Volume.10, pp. 128-138, 2008.

[10]. Byoung-Doo, "Detecting Attacks on Networks", IEEE Computer, Volume. 30, No 12. Pp. 16-17, 1997.

[11]. R A., H. M. Rezaul Karim, "Improving Intrusion Detection Performance using Keyword Selection and Neural Networks", Computer Networks, Volume. 34, No. 4, pp. 597-603, 2000.

[12]. Tsong, "Intrusion and Misuse Detection in Large-Scale Systems", IEEE Computer Graphics and Applications, Volume. 22, No. 1, pp.38-48, 2002.

[13]. Weiming Hu, "Application of Neural Networks to Intrusion Detection Classification and detection of computer intrusions", pp. 110-117, 2005.

[14]. Hu Zhengbing1, "DoS intrusion detection using generalized grey self-organizing maps," 2007 IEEE International Conference on Grey Systems and Intelligent Services, pp. 1548-1551, 2007.

[15]. Amit Kumar Choudhary, "Host-based intrusion detection using self-organizing maps," Proceedings of the 2002 IEEE World Congress on Computational Intelligence, Honolulu, HI, pp. 1714-1719, 2002.

[16]. Stefano Zanero, "An Efficient Collaborative Intrusion Detection System for MANET Using Bayesian Approach", pp.187- 90. 2006.

[17]. Liberios VOKOROKOS, "A Comparison of Intrusion Detection Systems", Computers & Security, Volume. 20, pp. 676-683, 2001.

[18]. H. Günes Kayacık, "Network-based Intrusion Detection Using Ada Boost Algorithm", National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Science.pp.468-476, 2012.

[19]. Zhenwei YU, "Pattern Recognition and Machine Learning," Springer-Verlag New York, LLC, pp.121-130, 2006.

[20]. V. K. Pachghare, G. Vachtsevanos, and B. Litt, "One-class novelty detection for seizure analysis from intracranial EEG", J. Machine Learning Research (JMLR), vol. 7, pp. 1025"1044, 2006.

[21]. Mansour M. Alsulaiman, "An introduction to intrusion detection, Crossroads", Volume.2, Issue 4, pp. 3-7, 1996.