A FEATURE REDUCES INTRUSION DETECTION USING ANN CLASSIFIER

S. YUVASRI, Department of Computer Application,

Mr.J.Jayapandiyan ,MCA,M.Phil., Associate Professor,

Krishnasamy College of Engineering and Technology,

Cuddalore.

Abstract: Rapid increase in internet and network technologies has led to considerable increase in number of attacks and intrusions. Detection and prevention of these attacks has become an important part of security. Intrusion detection system is one of the important ways to achieve high security in computer networks and used to thwart different attacks. Intrusion detection systems have curse of dimensionality which tends to increase time complexity and decrease resource utilization. As a result, it is desirable that important features of data must be analyzed by intrusion detection system to reduce dimensionality. This work proposes an intelligent system which first performs feature ranking on the basis of information gain and correlation. Feature reduction is then done by combining ranks obtained from both information gain and correlation using a novel approach to identify useful and useless features. These reduced features are then fed to a feed forward neural network for training and testing on KDD99 dataset. Pre-processing of KDD-99 dataset has been done to normalize number of instances of each class before training. The system is tested on five different test datasets and both individual and average results of all datasets are reported. Comparison of proposed method with and without feature reduction is done in terms of various performance metrics. Comparisons with recent and relevant approaches are also tabled. Results obtained for proposed method are really encouraging.

1.Introduction: Intrusion was detected by authentication, and decryption techniques and firewalls etc. These are known as first line of defense in computer security. This enables evaluation of computer programs installed on the host to detect known vulnerabilities. After evaluation, the, attacker can by pass them easily and first line defense mechanism is not flexible and powerful enough to thwart different kinds of this type of detection system has the same problem as antivirus software which needs periodic updation to detect a new type of attacks. Anomaly Detection system creates a normal profile by analyzing and observing the normal behavior of network system known as normal and patterns which deviate from normal profile are called outliers, anomalies, aberrations. A significant deviation from the normal profile is considered an attack. In anomaly detection system there is no need of prior knowledge of signatures. Anomaly detection can be divided into static and dynamic. The static detector works on a principle that only a fixed part of system which does not change, is monitored e.g. operating system software. Dynamic anomaly detector addresses network traffic data or audit records. Dynamic detector sets a threshold to separate normal consumption from anomalous consumption of resources. This method can detect an seen as well as a new attack which is an advantage over misuse detection systems but may lead to high rate of false alarm. Another drawback is, if an attacker knows that he or she is being profiled, they can slowly change the profile to train the anomaly detection system of intruder's $^{\text{malicious}}$ behavior as normal. Such systems can be further categorized into Network Intrusion Detection System (NIDS) and Host Intrusion Detection System (HIDS). Network-based intrusion detection monitors and analyses network traffic to differentiate and detect normal usage patterns from attack patterns. If a malicious pattern is detected, it is said to be an intrusion. Host-based intrusion detection analyzes log files for attack signatures. HIDS analyze host based audit sources such as audit trails, system logs and application logs to detect attacks. Detection System ensembles both misuse and anomaly based detection systems.

2.Objective: The Neural Network (NN) approach to Intrusion Detection, we first have to expose NN to normal data and to attacks to automatically adjust coefficients of the NN during the training phase. Performance tests are then conducted with real network traffic and attacks. . In order to apply this approach to Intrusion Detection, we would have to introduce data representing attacks and non-attacks to the Neural Network to adjust automatically coefficients of this Network during the training phase. In other words, it will be necessary to collect data representing normal and abnormal behavior to train the Neural Network. After training is accomplished, a certain number of performance tests with real network traffic and attacks were be conducted.

3. Proposed approach: The proposed methodology for Intrusion Detection in Computer Networks is based on using Artificial Neural Network (ANN) for detecting the Normal and Abnormal conditions of the given parameters, which leads to various attacks. The neural network approach for this purpose has two phases; training and testing. During the training phase, neural network is trained to capture the underlying relationship between the chosen inputs and outputs. After training, the networks are tested with a test data set, which was not used for training. Once the networks are trained and tested, they are ready for detecting the intrusions at different operating conditions. The following issues are to be addressed while developing an ANN for Intrusion Detection.

3.1 DATA COLLECTION: There are two ways to build IDS, one is to create our own simulation network, and collect relevant data and the other one is by using previously collected datasets. Issues like privacy, security, and completeness greatly restrict people from generating data. The advantage of using previously collected datasets is that the results can be compared with others in the literature. Some of the popularly used IDS datasets [15] are DARPA 1998 data set, DARPA 1999 data set and KDD Cup 1999 data set which are available in the MIT Lincoln Labs. In this work, we use KDD Cup 1999 data.

3.2: Data preprocession: Before training the neural network, the dataset should be preprocessed to remove the redundancy present in the data and the non-numerical attributes should be represented in numerical form suitably.

3.3: Data Normalization: During training of the neural network, higher valued input variables may tend to suppress the influence of smaller ones. Also, if the raw data is directly applied to the network, there is a risk of the simulated neurons reaching the saturated conditions. If the neurons get saturated, then the changes in the input value will produce a very small change or no change in the output value.

3.4: SELECTION OF NETWORK STRUCTURE: To make a Neural Network to perform some specific task, one must choose number of input neurons, output neurons, hidden neurons and how the neurons are connected to one another. For the best network performance, an optimal number of hidden-units must be properly determined using the trial and error procedure. The hidden layer neurons have tangent hyperbolic function as the activation function and the output have linear activation function. Once the appropriate structures of the network are selected, the ANN model is trained to capture the underlying relationship between the input and output using the training data. The output layer to the hidden layer to update the weight matrix. After training, the networks are tested with the test.

3.5: Network training and testing: In Back propagation algorithm is used to train the network, which propagates the error from the output layer to the hidden layer to update the weight matrix. After training, the networks are tested with the test data set to assess the generalization capability of the developed network.

4. Hybrid instruion detection:

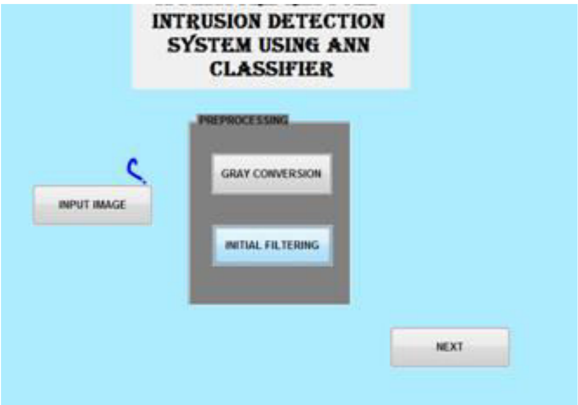

Fig.no:1

Artificial Neural Networks can be viewed as parallel and distributed processing systems which consists of a huge number of simple and massively connected processors. The ANN architecture is the most popular paradigm of artificial neural networks in use today.

5.Future Scope: Instrusion Detection has been proposed to distinguish the abnormal and normal network behaviors. The KDD taining and testing for dataset and no misuse data.

6.Conculsion: Intrusion detection system that works on reduced number of features. The system extracts features using concepts of information gain and correlation. Features are first raked using information gain and correlation and combined thereafter using a suitably designed mechanism. The method uses pre-processing to eliminate redundant and irrelevant data from the dataset in order to improve resource utilization and reduce time complexity. A classification system was designed using ANN which was trained on compact dataset and tested on five different subsets of KDD99 dataset. It can be seen from results that the method outperforms other methods for attack and non-attack classes. Overall, the method has reported an increased detection rate and decreased false alarm rate. The system was put to test against contemporary techniques and results were found to be encouraging. The implication of the proposed system is demonstration of the fact that feature reduction can be an important phenomenon to reduce dimensionality and training time of the system. The performance of the feature reduced system is actually better than system without feature reduction thereby influencing design of systems with far less time complexities. The proposed method of Intrusion detection system can be used to provide security in network, organizational and social areas where security is prime importance. The study can also inspire researchers from field of data science, big data to utilize their work to propose more challenging solutions for the current research problem.

7.References:

[1] A. Zhong and C.F. Jia. 2004, "Study on the applications of hidden Markov models to computer intrusion detection," in Proceedings of

the Fifth World Congress on Intelligent Control and Automation WCICA, Vol. 5, pp. 4352-4356.

[2] M.Analoui, A.Mizaei, and P.Kabiri. 2005, "Intrusion detection using multivariate analysis of variance algorithms," in Third International Conference on Systems, Signals & Devices SSD05, Vol.

[3] V. Boloan cane, N. Sanchenez Marano,
A. Aloanso betanzos 2011, "Feature selection and classification in multiple class
Dataset.

[4] P. Barmejo,OSSA, L. Gamez &
J. puertaj 2012 fastwrapper feature subset
Knowledge based system.

[5] Network anomoley detection methods
System and tools.IEEE communication.