DEVELOPMENT OF ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM BASED NETWORK INTRUSION DETECTION SYSTEM

NITIN S. KHACHANE[1], PROF.DEEPAK SINGH TOMAR[2], PROF. AMIT SINHAL [3]

1Department of CSE, TIT Bhopal, RGPV Bhopal, MP, India

2Department of CSE, TIT Bhopal, RGPV Bhopal, MP, India

3Department of CSE, TIT Bhopal, RGPV Bhopal, MP, India

Corresponding Author: Email- khachane.nitin001@gmail.com

***Abstract*— Intruders' computers, who are spread across the Internet, have become a major threat in our world. Many researchers proposed a number of techniques such as (firewall, encryption) to prevent such penetration and protect the infrastructure of computers as well as information, but with this, the intruders managed to penetrate the computers. IDS which are increasingly a key part of system defense are used to identify abnormal activities in a computer system. So IDS has taken much of the attention of researchers, IDS monitors the resources computer and sends a report on the activities strange patterns the proposed system. We are going to design Adaptive Neuro-Fuzzy based system for effectively identifying the intrusion activities within a network. The proposed Adaptive Neuro-Fuzzy Inference based system will be able to detect an intrusion behavior of the networks. The experiments and evaluations of the proposed network intrusion detection system will be performed with the NSL-KDD dataset.**

***Keywords*—NIDS, Anomaly based IDS, NSL-KDD dataset and ANFIS.**

I. INTRODUCTION

Due to the increase use of computer networks in many aspects of life, the numbers of vulnerabilities also are increasing causing the network resources unavailable and violate the system confidentiality, integrity and availability. Intrusions pose a serious security threat for the stability and the security of information in the network environment. A network intrusion attack encompasses a wide range of activities. It includes attempting to destabilize the network, gaining unauthorized access to files with privileges, or mishandling and misusing of software. The possibilities and opportunities are limitless; unfortunately, so too are the risks and chances of malicious intrusions. It is very important that the security mechanisms of a system are designed so as to prevent unauthorized access to system resources and data. We need to search for new mechanisms and architecture to protect computers. So, Adaptive Neuro-Fuzzy Inference System based Network Intrusion Detection System may be the solution for this. Intrusion Detection Systems (IDS's) are security tools that, like other measures such as antivirus software, firewalls, and access control schemes, are intended to strengthen the security of information and communication systems (Teodoro, 2009).

An Intrusion Detection System is an important component of the computer and information security framework. Its main goal is to differentiate between normal activities of the system and behaviors that can be classified as intrusive. The main goal of intrusion detection is to build a system that could automatically scan the network activity and detect such intrusion attacks. An IDS is used to detect several types of malicious behaviors that can compromise the security and trust of a computer system or network. There are two main intrusion detection approaches: anomaly intrusion detection system and misuse intrusion detection system .The anomaly detection focuses on the unusual activities of patterns and uses the normal behavior patterns to identify an intrusion. The misuse detection recognizes known attack patterns and uses well-defined patterns of the attack. On the other hand, IDS's may be categorized according to the host system into two types:

i. Host-based IDS (HIDS)

ii. Network-based IDS(NIDS)

The first operates at the host level and monitors a single host machine using the audit trails of the host operating system, whereas the other operates at the network level and monitors any number of hosts on the network [1].

A. NSL-KDD Dataset

The NSL-KDD dataset is a common benchmark dataset usually used by many researchers for evaluation of Intrusion detection techniques. This dataset has been obtained from the KDD Cup 99 dataset [2].The dataset has 41 features for each connection record plus one class label. Some features are derived features, which are useful in distinguishing normal connection from attacks. These features had all Forms of continuous and symbolic with extensively varying ranges falling in four categories:

• In a connection, the first category consists of the intrinsic features which comprises of the fundamental features of each individual TCP connections. Some of the features for each individual TCP connections are duration of the connection, the type of the protocol (TCP, UDP, etc.) and network service (http, telnet, etc.).

• The content features suggested by domain knowledge are used to assess the payload of the original TCP packets, such as the number of failed login attempts.

• Within a connection, the same host features observe the recognized connections that have the same destination host as present connection in past two seconds and the statistics related to the protocol behavior, Service, etc. are estimated.

• The similar same service features scrutinize the connections that have the same service as the current connection in past two seconds.

There are 4 main categories of attacks in the NSL-KDD CUP dataset [3]. A brief description of each class in the subsequent sections

1. Denial-of-service Attack (DOS): It is a class of attacks where an attacker makes some Computing or memory resource too busy or too full to respond to requests.

2. Probing (Probe): It is a class of attacks where an attacker scans a network to get some information about potential vulnerabilities in the network.

3. User to Root Attacks (U2R): It is a class of attacks where an attacker gets an access to a normal user account on the system to get a root user access to the system later.

4. Remote to Local Attacks (R2L): It is a class of attacks where an attacker sends some packets to a system over a network remotely, and then it gets some information about the potential vulnerabilities in this system.

II. Related Work

Intrusion detections are rules dependent .If the behavior of the packets flowing in the network is new, and then the system cannot take any decision. So they purely work in the basis of the initial rules provided.
i) It cannot create its own rule depending on the current situation.
ii) It requires manual energy to monitor the Inflowing packets and analyze their behavior. Intrusion detection has emerged as a significant field of research, because it is not theoretically possible to set up a system with no vulnerabilities [5]. One main confrontation in intrusion detection is that we have to find out the concealed attacks from a large quantity of routine communication activities [6], Fuzzy Logic [7], Genetic Algorithm [8], and Data mining refers to a set of techniques that use the process of extracting previously unknown but potentially useful data from large stores of past data. Data mining method excels at processing large system logs (audit data).However they are less useful for stream analysis of network traffic [9]. Segmentation technique refers to allowing extraction of patterns of unknown attacks. This is done by matching patterns extracted from a simple audit set with those referred to warehoused unknown attacks [10] and more have been extensively employed to detect intrusion activities both known and unknown from large quantity of complex and dynamic datasets. Generating rules is vital for IDSs to differentiate standard behaviors from strange behavior by examining the dataset which is a list of tasks created by the operating system that are registered into a file in historical sorted order [11].

III. Adaptive Neuro-Fuzzy Inference System

The ANFIS is a fuzzy Sugeno model put in the framework of adaptive systems to facilitate learning and adaptation (Jang, 1993). Such framework makes the ANFIS modeling more systematic and less reliant on expert knowledge. To present the ANFIS architecture, two fuzzy if-then rules based on a first order Sugeno model are considered:
Rule 1: If (x is $A1$) and (y is $B1$) then ($f1 = p1x + q1y + r1$)
Rule 2: If (x is $A2$) and (y is $B2$) then ($f2 = p2x + q2y + r2$)

where x and y are the inputs, Ai and Bi are the fuzzy sets, fi are the outputs within the fuzzy region specified by the fuzzy rule, pi, qi and ri are the design parameters that are determined During the training process. The ANFIS architecture to implement these two rules is shown in Fig. 1, in which a circle indicates a fixed node, whereas a square indicates an adaptive node.In the first layer, all the nodes are adaptive nodes. The outputs of layer 1 are the fuzzy membership grade of the inputs, which are given by:

$$O_i^1 = \mu_{A_i}(x) \quad i = 1, 2 \qquad (1)$$

$$O_i^1 = \mu_{B_{i-2}}(y) \quad i = 3, 4 \qquad (2)$$

Where $\mu_{A_i}(x),\ \mu_{B_{i-2}}(y)$ can adopt any fuzzy membership function. For example, if the bell shaped membership

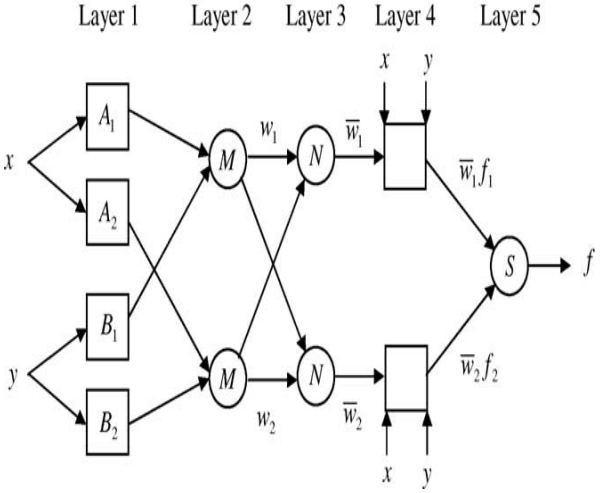


Fig.1. ANFIS Architecture

Function is employed; $\mu Ai\ (x)$ is given by:

$$\mu_{A_i}(x) = \frac{1}{1 + \left\{ \left(\frac{x - c_i}{a_i} \right)^2 \right\}^{b_i}} \quad (3)$$

Where *ai*, *bi* and *ci* are the parameters of the membership function, governing the bell shaped functions accordingly. In the second layer, the nodes are fixed nodes. They are labeled with *M*, indicating that they perform as a simple multiplier. The outputs of this layer can be represented as:

$$O2i = wi = \mu Ai\,(x)\,\mu Bi\,(y) \quad (4)$$
$$i = 1, 2$$

Which are called firing strengths of the rule. In the third layer, the nodes are also fixed nodes. They are labeled with *N*, indicating that they play a normalization role to the firing strengths from the previous layer. The outputs of this layer can be represented as:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \quad (5)$$

This is so-called normalized firing strength. In the fourth layer, the nodes are adaptive nodes. The output of each node in this layer is simply the product of the normalized firing strength and a first order polynomial. Thus, the outputs of this layer are given by:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad i = 1, 2 \quad (6)$$

In the fifth layer, there is only one single fixed node labeled with *S*. This node performs the summation of all incoming signals. Hence, the overall output of the model is given by:

$$O_i^5 = \sum_{i=1}^{2} \bar{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{w_1 + w_2} \quad (7)$$

It can be observed that there are two adaptive layers in this ANFIS architecture, namely the first layer and the fourth layer. In the first layer, there are three modifiable parameters {*ai*, *bi*, *ci*}, which are related to the input membership functions. These parameters are the so-called premise parameters. In the fourth layer, there are also three modifiable parameters {*pi*, *qi*, *ri*}, pertaining to the first order polynomial. These parameters are so-called consequent parameters (Jang, 1993) [12].

IV. Proposed Work

Presently, it is unfeasible for several computer systems to affirm security to network intrusions with computers increasingly getting connected to Internet. So, in recent years, Intrusion Detection System (IDS) has become one of the hottest research areas in Computer Security. It is an important detection technology and is used as a counter measure to preserve data integrity and system availability during an intrusion. An Intrusion Detection System is a system for detecting intrusions in the network. For intrusion detection, a wide variety of techniques have been applied specifically, data mining techniques, artificial intelligence technique and soft computing techniques. Most of the data mining techniques like association rule mining, clustering and classification have been applied on intrusion detection. By taking into consideration these, we have developed Adaptive Neuro-Fuzzy Inference System based Network Intrusion Detection System. The different steps involved in the proposed system for anomaly-based network intrusion detection (shown in figure 1) are described as follows:

A) Classification of Data (Input)
B) Training System
C) Decision System
D) Output

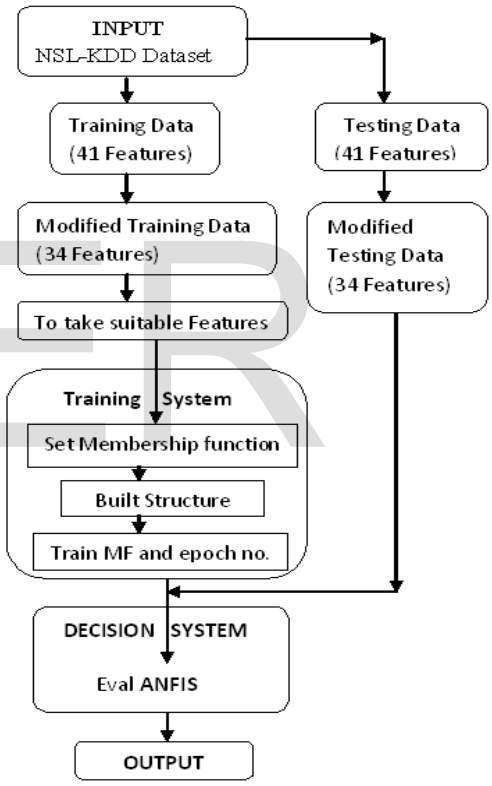


Fig.2. Steps of Proposed System

A) Classification of Data

In first step of proposed system, NSL-KDD dataset we have taken for analyzing the intrusion detection behavior using the proposed system. Detail analysis of this dataset is in introduction section. The NSL-KDD dataset mainly categories into four types of attacks with 41 attributes that have both continuous and symbolic attributes. The proposed system is designed only for the continuous attributes because the major attributes in NSL-KDD datasets are continuous in nature. Therefore, we have taken only the continuous attributes for instance, 34 attributes from the input dataset by removing discrete attributes.

After that we are taking only suitable features which are selected by using LDA & GA [13] for training system.

B) Training System

In Training System, The ANFIS procedure is shown Fig.2, in the first step initializing the fuzzy system using genfis1 command, in the second step learning process start and the number of epochs is set. In the third step learning process start by using anfis command and last in the fourth step perform fuzzy inference calculation. In the training phase, the membership functions and the weights will be adjusted such that the required minimum error is satisfied or if the number of epochs reached. At the end of training, the trained ANFIS network would have learned the input output map

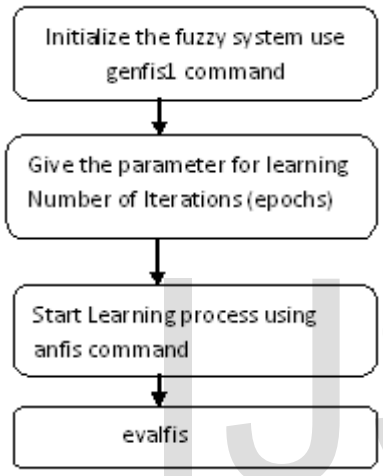


Fig.3. ANFIS Procedure

C) Decision System

In Decision System, It evaluates fuzzy inference system on testing dataset records and generates the decision.

D) Output

Decision means attack type name is displayed.

V. Experiment Result

The training dataset contains four types of attacks, which are given to the proposed system using suitable attributes for training system. Table No1 shows no of record of each attack type for training proposed system using ANFIS.

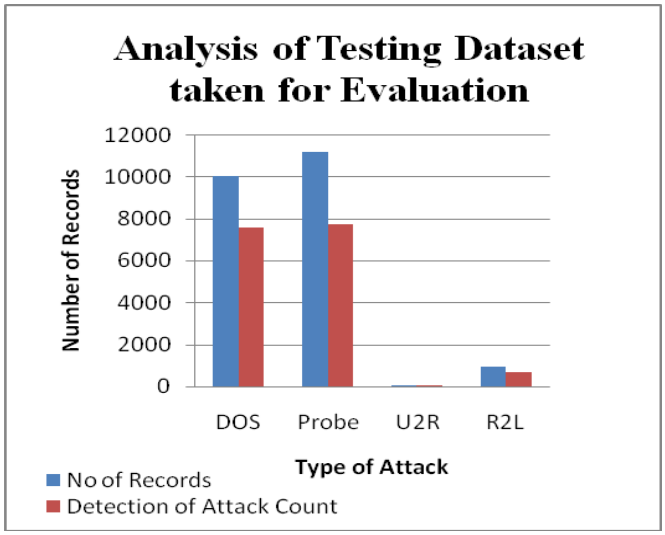
Table No1: Training dataset taken for experimentation

Types Of Attack	No of Records
DOS	6681
Probe	1122
U2R	12
R2L	262

In the testing phase, the testing dataset as input is given to the proposed system; it performs fuzzy inference calculation then which classifies the type attack i.e. decision. Through proposed system the evaluated values in Table No2 as follows-

Table No2: Testing dataset taken for Evaluation

Types Of Attack	No of Records	Detection Of Attack Count	Percentage of Attack Detection
DOS	10082	7599	75.37
Probe	11224	7753	69.07
U2R	39	34	87.18
R2L	937	659	70.33



From above analysis of obtained result the overall performance of the proposed system is achieved near about 70% for three types of attack and it is more than 85% accuracy for U2R type of attacks.

VI. CONCLUSION

We have developed an anomaly based network intrusion detection system in detecting the intrusion behavior within a Network using Adaptive Neuro-Fuzzy Inference System. A Neuro-fuzzy decision-making module was designed to build the system more accurate for attack detection, using the fuzzy inference approach. We have used NSL-KDD dataset for evaluating the performance of the proposed system and results showed that the proposed method is effective in detecting various intrusions in computer networks. Based on this experiment, it is also concluded the ANFIS technique could be used as a very accurate, reliable and fast method for detection of attacks .One reason may because of the combination the abilities of both quantities and qualities of neural network and fuzzy logic to improve the detection rather than the other methods.

REFERENCES

[1] B.V. Dasarathy, "Intrusion Detection", Information Fusion, Vol.4, No.4, pp.243-245, 2003.

[2] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani A Detailed Analysis of the KDD CUP 99 Data Set In proceedings of the 2009 IEEE symposium computational intelligence in security and defense applications(CISDA 2009).

[3] NSL-KDD dataset http://iscx.ca/NSL-KDD.

[4] KDD Cup 1999 Data: http://kdd.ics.uci.edu/databases/ kddcup99.

[5] Susan M. Bridges and Rayford B.Vaughn, "Fuzzy Data Mining And Genetic Algorithms Applied To Intrusion Detection", In Proceedings of the National Information Systems Security Conference (NISSC), Baltimore, MD, pp.16- 19, October 2000.

[6] Jian Pei, Upadhyaya, S.J., Farooq, F., Govindaraju, V, "Data mining for intrusion detection: techniques, applications and systems ", in Proceedings of the 20th International Conference on Data Engineering, pp: 877 - 87, 2004.

[7] R. Shanmugavadivu , Dr.N.Nagarajan "NETWORK INTRUSION DETECTION SYSTEM USING FUZZY LOGIC" Indian Journal of Computer Science and Engineering (IJCSE) Vol. 2 No. 1 ISSN : 0976-5166 pp. 101-111 ,

[8] Satya Keerthi N V L, P Lakshmiprasanna, B Minny Priscilla, M V B T Santhi. Intrusion Detection System Using Genetic Algorithm, International Journal of P2P Network Trends and Technology-Volume1Issue2 pp.1-7, 2011

[9] Yu Y, and Huang Hao, "An Ensemble Approach to Intrusion Detection Based on Improved Multi-Objective Genetic Algorithm", Journal of Software, Vol.18, No.6, pp.1369-1378, June 2007.

[10] W. Lee, S. Stolfo, and K. Mok, "A Data Mining Framework for Building Intrusion Detection Model", In Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA, pp. 120-132, 1999.

[11] Dewan Md. Farid and Mohammad Zahidur Rahman, "Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm", Journal of Computers, Vol.5, No.1, January, 2010

[12] Jang J-SR. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 1993;23(3):665–85.

[13] Hafiz Muhammad Imran, Azween Bin Abdullah, Muhammad Hussain, Sellappan Palaniappan, Iftikhar Ahmad "Intrusions Detection based on Optimum Features Subset and Efficient Dataset Selection" *2012* International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 6, December 2012, ISSN: 2277-3754.

[14]Abbasi, E., & Abouec, A. (2008). Stock price forecast by using neuro-fuzzy inference system. Proceedings of World Academy of cience, Engineering and Technology, 36, 320–323.