ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

Intrusion Detection System Using Neural Network

[1] Abarna Anusia. S, [2] Saveetha.D
[1] M.Tech (ISCF), SRM University, Kattanakulathur, India
[2] Assistant Professor (O.G), IT Department, SRM University, Kattanakulathur, India

***Abstract*: Nowadays increase in internet usage and network technologies has led to extent increase in number of attacks and intrusions. Detection of these attacks and intrusion has become an important part of the security. Intrusion Detection System (IDS) is one type of software application that monitors a network or system for malicious activity or policy violation. In this paper we propose a Hybrid Intrusion Detection System that is combination of both Signature-based and Anomaly-based Intrusion Detection System. The proposed IDS uses a back propagation algorithm and feed forward Artificial Neural Network(ANN) to learn system's behavior. We use the KDD'99 data set in our experiments and obtain the expected results.**

***Key Words:* Intrusion Detection System(IDS); Artificial Neural Network(ANN); NSL-KDD dataset.**

I. INTRODUCTION

In the last few years, the essence of the computers and internet play a vital role in the world. The use of internet has become a trend these days, most of the people prefer to use internet either for personal or business purpose due to their portability and easy access. Online shopping, internet banking, social network are popular among internet user, which attracts the hackers and intruders towards them. As the usage of internet increasing, internet threats are also increasing. Some of the internet threats are stealing personal and financial informations, ransomware, DOS attack, etc. Several users become the victim of these threats by using malicious websites and applications which look legitimate. Providing an extra layer of security from the network side can protect the user from advanced threats, which a typical antivirus could not detect.

A. Intrusion Detection System (IDS)

Intrusion Detection Systems (IDS) are important security tools used to detect possible attacks, inappropriate, incorrect or abnormal activities and to alert the occurrence to network administrators. The main goal of an intrusion detection system is to detect the attacks efficiently and it is also equally important to detect attacks at a beginning stage in order to reduce their impacts. Intrusion detection system are either network-based or host-based.

Network based Intrusion Detection Systems (NIDS) are placing IDS key components in the network segment and monitoring the overall network traffic passing through this segment. A network IDS, using either a network tap, span port, or hub collects packets that traverse a given network. Using the captured data, the IDS system tries to identify unauthorised and anomalous behaviour. Well known example is SNORT.[2]

Host-based Intrusion Detection System (HIDS) involves an agent installed on the local nodes of the computer network. The installed agent uses a combination of signatures, rules, and heuristics to identify unauthorised activities. Examples are OSSEC (Open Source Host-based Intrusion Detection System), Verisys, etc.[2]

According to the operation, Intrusion Detection System (IDS) can be classified into Signature-based IDS and Anomaly-based IDS. In Signature based IDS, signatures of already known attack pattern are stored in the database and matched with captured data, if the match is positive it is declared as an attack otherwise it discard the data. Each intrusion leaves a footprint behind and these footprints are stored in database as signatures. For each incoming packet there is a verification against stored signatures and the IDS check whether the packet consist malicious data or not. The disadvantage of signatures based IDS is that signature database must be continually updated and maintained and it unable to find unknown attacks.

In Anomaly based IDS, the detection involves defining or learning "normal" activity and looking for dedication from this baseline. Protocol Anomaly involves looking for deviations from a standard protocol. Network Anomaly involves watching or learning the normal traffic levels. Behavioural Anomaly involves learning normal user behaviour. The main advantage of an anomaly-based system is its ability to detect unknown attacks when they appear. The major drawback is to defining its rule set.

B. Artificial Neural Networks(ANNs)

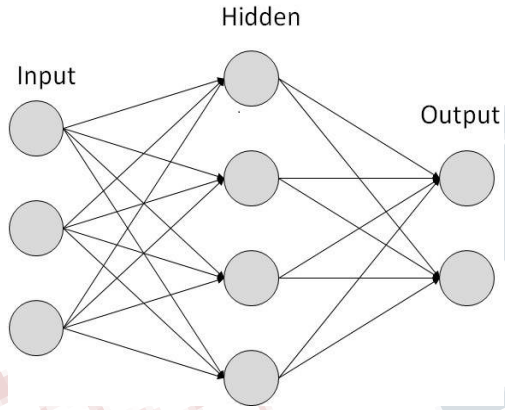
The Artificial Neural Network (ANN) enables system to think and act intelligently. The neural network is based on the working of human brain by making right connections, can be imitated using silicon and wires as living neurone and dendrites. The ANNs are composed of multiple

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

nodes, which imitated biological neurone of human brain and they are connected by links. Each link is associated with weight. ANNs are capable of learning, by altering weight values. There are two types of ANNs, they are Feed Forward ANN and Feed Back ANN.

It contains interconnected neurons. The connection between them is realised via dendrites which transport the electric signals. It is composed of multiple layer containing interconnected neurons which are activated after receiving signals, these last ones are summed then performed through an activation function and the obtained result activates the next neurons and so on this process continues until the last layer. This layer is called the output one and it returns the final result. It has the ability to learn, react and respond to problems belonging in what he learned. To achieve this, the ANN should be accomplished through training and test phases.



C. Metrics

In order to evaluate the IDS effectiveness, it is necessary to calculate the following indicators:

•True Positive(TP): The IDS gives an alert against a real threat.

•False Positive(FP): The IDS gives an alert for an event which does not correspond to a real attack.

•True Negative(TN): The IDS does not gives an alert for a normal event.

•False Negative(FN): The IDS does not detect a real attack.

D. Description of KDD Dataset

KDD is a publicly available dataset of attacks for IDS. It contains 41 features and one more attribute for class. In fact, there are 41 parameters in KDD set, their values depend on the attack's type which is indicated in the 42nd feature. The parameters are divided into four categories:

•Basic attributes.

•Attributes which are related to content.

•Attributes based on the time using windows of two-second time.

•Time-based attributes using windows of 100 connections time.

The details of the 41 attributes are described in table I.

I. TABLE I CLASSWISE DETAIL OF KDD DATA SET ATTRIBUTES

	Basic attributes
A1	Connection duration(nb of seconds)
A2	Protocol type, ex, tcp, udp, etc.
A3	Network service(destination) ex. http,telnet
A4	Connection status(normal or error)
A5	Number of data(in bytes) from the source to the destination
A6	Number of data(in bytes) from the destination to the source
A7	1 if the connection is from/to the same host; 0 otherwise
A8	Number of "erroneous" fragments
A9	Number of urgent packets
	Attributes related to content
A10	Number of indicator "hot"
A11	Number of unsuccessful login attempts
A12	1 if login success; 0 otherwise
A13	Number of conditions of "compromises"
A14	1 if the shell root is obtained; 0 otherwise
A15	1 if the command is "root su"; 0 otherwise
A16	Number of access to "root"
A17	Number of creation of files operations
A18	Number of shell prompts
A19	Number of operations related to file of access control
A20	Number of outbound commands in ftp session
A21	1 if login belongs on "hot" list; 0 otherwise
A22	1 if the login is "invited"; 0 otherwise
	Attributes based on the time using windows of two-second time
A23	Number of connections for the same host
A24	Number of connection for the same service

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

A25	% of connections for the same host with "SYN" error
A26	% of connections for the same service with "SYN" error
A27	% of connections for the same host with "REJ" error
A28	% of connections for the same service with "REJ" error
A29	% of connections for the same host using the same service
A30	% of connections for the same host using different service
A31	% of connections for the same service using different host
	Time-based attributes using windows of 100 connections time
A32	Number of connections for the same host
A33	Number of connections for the same host using the same service
A34	% of connections for the same host using the same service
A35	% of connections for the same host using different service
A36	% of connections for the same host having the "src" port
A37	% of connections for the same host and same service using different host
A38	% of connections for the same host with "SYN" error
A39	% of connections for the same host and same service with "SYN" error
A40	% of connections for the same host with "REJ" error
A41	% of connections for the same host and same service with "REJ" error

E. back propagation algorithm

Step 1: Normalise the inputs and outputs with respect to their **Step 1:** Normalise the inputs and outputs with respect to their maximum values. it is proved that the neural networks work better if input and outputs lie between 0-1 for each training pair, assume there are 'l' given by {I}1 (l*1) and 'n' outputs {O}o (n*1) in a normalised form.

Step 2: Assume the number of neurons in the hidden layer to lie between 1<m<2l.

Step 3: [V] Represents the weights of synapses connecting input neurons and hidden neurons and hidden neurons and [w] represents weights of synapses connecting hidden neurons and output neurons. Initialise the weights to small random values usually from -1 to 1. for general problems, λ can be assumed as 1 and the threshold values can be taken as zero.

[V]o={random weights}
[W]o = {random, weights}
[ΔV]o =[ΔW]o=[o]

Step 4: for the training data, present one set of inputs and outputs, Present the pattern to the input layer {I}, as inputs to the input layer may be evaluated as

$$\{O\}r = \{I\}I \text{ for } (l*1)$$

Step 5: Compute the inputs to the hidden layer by multiplying corresponding weights of synapses as

{I}H ={V}T{O}I where I and V sets are m*1 and and o is l*1

Step 6 : Let the hidden layer units evaluate the output using the sigmoidal function as

$$\{O\}H = \{- - -1/(1+e(-IHi))- - -\}$$

Step 7 : Compute the inputs to the output layer by multiplying corresponding weights of synapses as

$$\{I\}o = \{w\}T\{O\}H$$

where i is set of n*1,w is set of n*m matrix and O is set of m*1 matrix.

Step 8 : Let the output layer units evaluate the output using sigmoidal function as

$$\{O\}O = \{- - - ((1/ (1+e(-IOj))\}$$

the above is the network output.

Step 9 : Calculate the error and the difference between the network output and the desired output as for the ith training set as

$$EP = \sum((Tj - Ooj)2)/n$$

Step 10 : Find {d} as

$$\{d\} = \{(TK-Ook)Ook(1-Ook)\} \text{ of matrix } n*1$$

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

Step 11 : Find {Y} matrix as [Y] = {O}H <d> where Y is matrix of m*n, o is matrix of m*1, d is matrix of 1*n

Step 12 : Find [ΔW]t+1 =α[Δw]t + η[Y] where each matrix is of m*n

Step 13 : Find {e} = [w] {d} where e is matrix of m*1, w is matrix of m*n, and d is matrix of n*1.

$$\{d^*\}= \{ei(OHi)(1-OHi)\}$$

where d is matrix of m*1

Step 14: For finding X matrix as [x] = {O}r<d*> = {I}I<d*> where x = matrix of 1*m, O,I = single value, d =matrix of 1*m

Step 15: Find [ΔV]t+1 = α[ΔV]t + η[X] all matrix of 1*m

Step 16: To Find the updated weights and threshold[Θ]

$$[V]t+1 = [V]t +[\Delta V]t+1 \; [W]t+1 = [W]t +[\Delta W]t+1$$
$$[\Theta]ot+1 = [\Theta]ot+[\Delta\Theta]ot+1$$

$$[\Theta]Ht+1 = [\Theta]Ht+[\Delta\Theta]Ht+1$$

Step 17: Find error rate as

$$\text{Error rate} = ((\sum EP)/nset)$$

Step 18 : Repeat steps 4-18 until the convergences in the error rate is less than the tolerance value.

F. Problem Statement

The Network Intrusion Detection Systems (NIDS) is either signature-based or anomaly-based. The signature-based IDS only detect the known attack that is available in the database. The anomaly-based IDS detect by analysing and observing the normal behaviour of network system. The drawback is it lead to high rate of false alarm.

II. LITERATURE SURVEY

The Neural Network Intrusion Detection System (NNIDS) is able to successfully recognise learned malicious activities in a network environment and it was tested for the SYN flood attack, UDP flood attack, nMap scanning attack and for non-malicious communication was proposed by Ádám et al. [2].

In some KDD-features have either no role, or a minimum impact in attack detection. Consequently, the large number of parameters that must be selected to produce a neural network detecting attacks must be optimised in order to decrease error and increase the IDSs performance was proposed by El Farissi et al. [3]. The KDD99 includes a wide variety of intrusions simulated in a military network environment. It proposes six datasets: Dataset1: it contains 4898431 patterns, including the classes “Normal”, “DOS”, “Probe”, “R2L” and “U2R”. Dataset2: This base includes patterns belonging on the same five classes. Dataset3: it contains 10 percent of dataset1 content. And three unlabelled datasets.

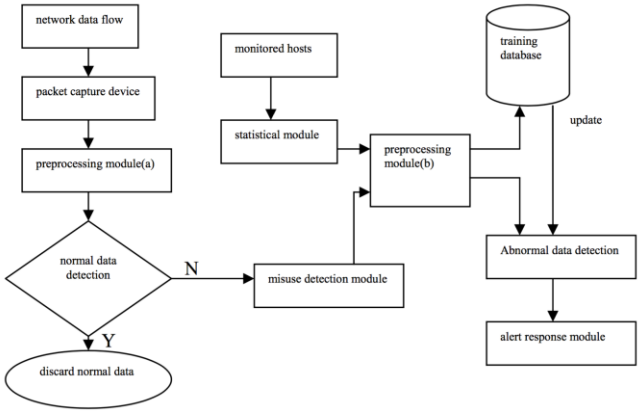
Anomaly-based IDSs need to be able to learn the dynamically changing behaviour of users or systems. In this paper, we are experimenting with packet behaviour as parameters in anomaly intrusion detection. There are several methods to assist IDSs to learn system’s behaviour. The proposed IDS uses a back propagation artificial neural network (ANN) to learn system’s behaviour was proposed by Al-Janabi et al. [4].

The five different types of NNs are described: multilayer perceptrons (MLP), radial basis function (RBF), self-organising feature map (SOFM), adaptive resonance theory (ART) and principal component analysis (PCA). An intrusion detection system combined with Genetic Algorithm (GA) and Back Propagation (BP) network is presented. Finally, a discussion of the future NN technologies, which guarantee to enhance the detection efficiency of IDS is provided was proposed by Fu, Y et al. [6].

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 normalise number of instance of each class before training. The system then behaves intelligently to classify test data into attack and non-attack classes. The aim of the feature reduced system is to achieve same degree of performance as a normal system. 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

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017



was proposed by Manzoor et al. [7].

III. PROPOSED SYSTEM

The proposed model is the Network Intrusion Detection System (NIDS) using simple Artificial Neural Network (ANN). It is the combination of both signature-based and anomaly-based intrusion detection system. The hybrid system detect the known attacks using signature based IDS by identifying a match and unknown attacks are detect by checking if there is any deviation from normal behaviour to increase detection rate and decrease false alarm rate.

IV. SYSTEM ARCHITECTURE

The intrusion detection system model presented in this paper adopts signature detection and anomaly detection. The system is composed of eight modules, its structure is shown in above figure and their descriptions as follows.

1.Network packet capture device. It captures data packet from the network and decode information.

2.Preprocessing module(a). The goal of this module is to preprocess data based on the patterns of normal data and to do protocol analysis. Initially it converts binary data obtained from network to network packet information in ASCII format. It processes network packet information using data preprocessing program.

3.Normal data detection module. To meet the demand of high-speed network, one-class classifier is used to identify normal data. Here, normal data are separated by using BP algorithms with anomaly detection method. Finally the abnormal data are separated and transferred to the misuse detection module.

4.Misuse detection module. It detects known attacks. First, it makes alarm when an attack is found. Then the attack packet information was detected and abnormal data are together transferred to preprocessing module(b)

5.Statistical module. It is in charge of data feature statistics based on destination host and services. Statistics based on destination hosts include: the number of connection same as the connected destination hosts at a time and Statistics based on service include: the number of connection same as the connected services at a time.

6.Preprocessing module(b). In this module, data from misuse detection module and statistical module are implemented integrated processing, and finished connection record items. Then connection record containing attack information are transferred to training database of neural network, and the rest of connection records are sent to abnormal data detection module.

7.Abnormal data detection module. It uses the same algorithm as the normal data detection module. The mainly difference between them are input data and training data. The former extracts only network connection feature and takes normal data as training data, but the latter process feature statistics based on destination host and services, except for network connection feature and content feature, and attack data as its training data.

8.Alert response module. A response and alert mechanism is implemented to give an alarm for intrusion or attack detected and takes measure according to the response regulation which the users define.

V.CONCLUSION

If this NIDS is implemented then it will provide the network security. This model is able to detect known and unknown threats. Although, NIDS produce high accuracy, it can be further improved by fine tuning various parameters of the proposed IDS model.

ACKNOWLEDGMENT

Thank all the pioneers who make a great contribution to the research of the intrusion detection system and neural network, then I can apply this idea to my project to enrich the network security, at last I will thank my guide Saveetha for giving me the help.

REFERENCES

1. Subba, B., Biswas, S., & Karmakar, S. (2016, March). A Neural Network based system for Intrusion Detection and attack classification. In Communication

(NCC), 2016 Twenty Second National Conference on (pp. 1-6). IEEE.

2. Ádám, N., Madoš, B., Baláž, A., & Pavlik, T. (2017, January). Artificial neural network based IDS. In Applied Machine Intelligence and Informatics (SAMI), 2017 IEEE 15th International Symposium on (pp. 000159-000164). IEEE.

3. El Farissi, I., Saber, M., Chadli, S., Emharraf, M., & Belkasmi, M. G. (2016, October). The analysis performance of an Intrusion Detection Systems based on Neural Network. In Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on (pp. 145-151). IEEE.

4. Al-Janabi, S. T. F., & Saeed, H. A. (2011, December). A neural network based anomaly intrusion detection system. In Developments in E-systems Engineering (DeSE), 2011 (pp. 221-226). IEEE.

5. Kumar, S., Viinikainen, A., & Hamalainen, T. (2016, December). Machine learning classification model for Network based Intrusion Detection System. In Internet Technology and Secured Transactions (ICITST), 2016 11th International Conference for (pp. 242-249). IEEE.

6. Fu, Y., Zhu, Y., & Yu, H. (2009, September). Study of neural network technologies in intrusion detection systems. In Wireless Communications, Networking and Mobile Computing, 2009. WiCom'09. 5th International Conference on (pp. 1-4). IEEE.

7. Manzoor, I., & Kumar, N. (2017). A feature reduced intrusion detection system using ANN classifier. Expert Systems with Applications, 88, 249-257.

8. Esmaily, J., Moradinezhad, R., & Ghasemi, J. (2015, May). Intrusion detection system based on Multi-Layer Perceptron Neural Networks and Decision Tree. In Information and Knowledge Technology (IKT), 2015 7th Conference on (pp. 1-5). IEEE.