International Journal of Engineering & Technology, 7 (4.6) (2018) 181-184

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Novel Intrusion Detection System Using Artificial Neural Networks and Feature Subset Selection

L. Haripriya[1]*, M.A. Jabbar[2], B. Seetharamulu[3]

[1] *M. Tech Scholar, Centre for Data Science, Vardhaman College of Engineering, Hyderabad, Telangana*
[2] *Professor, Centre for Data Science, Vardhaman College of Engineering, Hyderabad, Telangana*
[3] *Professor, Department of CSE, Vardhaman College of Engineering, Hyderabad, Telangana*
**Corresponding author E-mail:haripriya.lakkireddy@gmail.com*

Abstract

The growth of internet and network technologies has been increasing day by day.With the increase of these technologies, attacks and intrusions are also increasing. The prevention of these attacks has become an task. Intrusion Detection System (IDS) provides prevention against these attacks. Data Mining and Machine Learning techniques are used for IDS to reduce error rate and to improve accuracy and detection rate. In this paper, we proposed a novel Artificial Neural Network (ANN) classifier using Back propagation algorithm to model IDS. ANN is widely used supervised classifier for IDS. The performance of our model is evaluated by conducting experiments on KYOTO data set which is refined version of KDD99 data set. Empirical results show that proposed model is efficient with high detection rate and accuracy.

Keywords: *ANN; Back propagation; IDS; Machine Learning; KYOTO*

1. Introduction

An Intrusion is an unauthorized access or malicious utilization of a computer resource. Intrusion is used to reduce factors of a resource like integrity, confidentiality and availability [1]. An Intruder existing in the real world attempts for gaining the access to unauthorized data and performs damage to the malicious activities present.

Intrusion Detection System (IDS) is used to detect all these kinds of malicious activities happening on the network and indicates the network administrator to get the data secured against these attacks [2]. The growth of IDS has improved the network security and protecting the data of an organization. Hence IDS is a security system that observes network traffic and computer system. An IDS provides security of firewall. A firewall safeguards an organization by identifying malicious activities from the internet and IDS detects if any one attempts to break firewall security or trying to have access and it immediately alerts the administrator to take action [3]. Hence IDS are the security systems detecting various activities that attack on the network and keeps our systems safe.

The concept of Artificial Neural network (ANN) is taken from the Biology subject where a neural network plays an important role in our human body. ANN is a computing system containing large collection of units which are interconnected in some manner that allows communication between the units [4] [5]. These units are called nodes or neurons which are simple processors operate in parallel. In ANN, the input layers are connected to the hidden layers which are associated with some weights that allow processing. These hidden layers are connected to the output layers for getting the results.

In this paper we propose a novel IDS using ANN to achieve high accuracy in classification of attacks. Section 2 discusses the Literature Review and Section 3 discusses about Related Work. Explanation of our proposed model is discussed in Section 4. Section 5 analyzes the experimental results. Finally, we conclude in Section6.

2. Literature Review

2.1. Intrusion Detection System

Now a day the usage of internet has been increasing day by day. It is playing an important role in many applications such as education, business, health care and in many fields. Every individual is using internet. The main issue comes here the data which we get through internet has to be secured. This security of data over network is done by Intrusion Detection System (IDS). Traditional approaches like firewalls and some mechanisms of authentication have been used for security of data were considered as first degree of protection and second degree used is IDS.

The Intruders perform different techniques to crack the system data like password hacking, peer to peer attack, sniffing attack, DOS attack etc. Hence a security system is required to protect the details of our organization from Internet and users from our organization. IDS detect the attacks by identifying the intrusions and prevent system from getting damaged.

2.2. Types of Attacks

IDS play a major role in identifying different types of attacks. The main aim of IDS is finding intrusion which is considered as classi-

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

fication problem. IDS is divided into various attacks such as DOS, probe, U2R, R2L [6] [7].

1. Denial of Service (DOS)
This means to shut down a system or a network by making inaccessible to its users. Some of the attacks of this type are Back, Land, Mail blood Smurf and others.

2. Probe
It is an attempt to gain access to a computer and its files through a known or weak point in the computer system. The attacks of this type are Mscan, Nmap, Saint, Satan and Ipsweep.

3. User to Root attack (U2R)
In this attack the person tries to exploit vulnerability for gaining root access. Some attacks of this type are Eject, Ps, Perl, Ffbconfig and others.

4. Remote to Login (R2L)
It is an attempt in which the user gets an unauthorized access from a remote machine. Some of the R2L attacks are Guest, Phf, Sendmail, Named and others.

2.3. Artificial Neural Networks

The concept of Artificial Neural network (ANN) is taken from the Biology subject where a neural network plays an important role in our human body. Neural network is interconnection of neurons which are present in millions of number. With these neurons, parallel processing is done in our body and hence it is the best example of parallel processing [9].

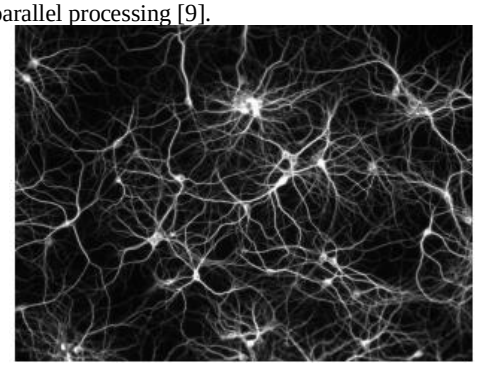


Fig. 1: Neural network in Human body [9]

Similarly ANN is a computing system containing large collection of units which are interconnected in some manner that allows communication between the units. These units are called nodes or neurons which are simple processors operate in parallel. In ANN, the input layers are connected to the hidden layers which are associated with some weights that allow processing. These hidden layers are connected to the output layers for getting the results.

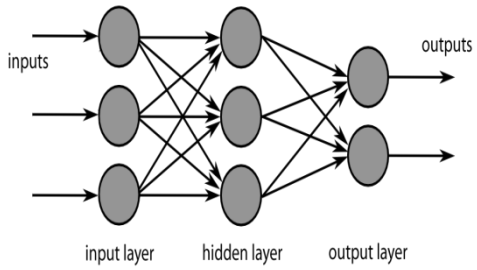


Fig.2: A simple neural network showing layers [9]

2.4. Feature Selection

Feature selection is the pre processing step used often in data mining. It removes the irrelevant features present in the data set and thus increases the accuracy [8]. It is very effective in dimensionality reduction feature. It identifies the features that are useful for class prediction. Feature selection methods are classified as (a) filter method, (b) wrapper method and (c) embedded method.

3. Related Work

Nabila Farnaaz [8] and M.A.Jabbar used random Forest classifier as a model for IDS, which gives better performance when compared to other classifiers for classification of attacks. Experiments are conducted on NSL-KDD data set in ARFF format. The proposed model is efficient with low false alarm rate and high detection rate. The performance measure for Random Forest is recorded as follows:

Table 1: Performance Measure For Random Forest [8]

Attack	Accuracy	DR	FAR	MCC
DOS	99.67	99.84	0.00527	0.99
Probe	99.67	99.82	0.00502	0.99
R2L	99.67	9.82	0.00505	0.99
U2R	99.67	99.84	0.00552	0.99

K. Kanaka Vardhini [10] and Dr.T.Sitamahalakshmi proposed an enhanced heuristic algorithm known as Ant Colony Optimization (ACO) for increasing success rate in IDS. The experiments are conducted on KDD data set. The proposed model has sensitivity and specificity factors for increasing success rate. The proposed ACO increased the accuracy when compared to conventional methods.

AODE (Average One Dependence Estimators) is one of the recent improvements of Naive Bayes algorithm. AODE solves the problem of independence by averaging the models which is generated by one dependence estimator and it is suitable for incremental learning. An Intelligent Network IDS using AODE was proposed by Amreen Sultana and M.A.Jabbar [1] for detecting different types of attacks. The evaluation of the model was performed on NSL-KDD data set. The performance of our model is as follows:

Table 2: Performance Of Aode [1]

S.No	Attack	Accuracy	DR	FAR	MCC
1	DOS	97.19	98.63	4.44	0.943
2	Probe	96.48	98.19	5.45	0.927
3	U2R& R2L	96.25	98.65	6.48	0.925

4. Proposed Work

This section discusses about Back propagation algorithm and our proposed method for IDS.

4.1. Back Propagation Algorithm

Back propagation is a supervised learning algorithm used for multilayer feed forward networks. In supervised learning both the inputs and outputs are provided for training of network. Now the network processes the inputs by comparing it with the resulting outputs over the desired outputs [11].The errors occurred through this process are back propagated through the system, making the system to adjust its weights by controlling the network. Back propagation algorithm is based on error correction learning rule.

The error back propagation consists of two passes: forward pass and backward pass.In forward passing, associate input vector is applied to the network and it propagates through the complete network layer by layer that the outputs are created. In forward pass the conjugation weights of network are fastened. In backward pass, the conjugation weights of network are adjusted per the error correction rule. The actual target response is deducted from the desired target response in order to produce an error signal. This error signal is back propagated towards the network and hence it is called as "error-back propagation".

4.2. Proposed Approach

Description of our proposed algorithm is below
Algorithm: A Novel IDS using Artificial Neural Networks
Input: KYOTO data set
Output: Classification of types of attacks
Step 1: Load KYOTO data set
Step 2: Apply Feature Selection technique - ReliefF Attribute Eval
Step 3: Partition the data set into training and testing
Step 4: Data set is given to ANN for training
Step 5: The test data is given to ANN for classification
Step 6: Calculate Accuracy, Precision, Recall and F-measure

For our experiment, we used Back propagation algorithm and it is implemented on R tool. R is an open source software widely used for execution. R is mostly used software for computing statistics and graphics [12] . R was developed by Ross Ihaka and Robert Gentleman in 1990's. We used R 3.4.2 version for our experiment. It is an integrated software containing facilities like storing and handling data. In R tool there is a package available called 'neuralnet' for ANN. By applying Back propagation algorithm in ANN using R tool provides an accuracy of 95.46% before feature selection.

We used weka tool for Feature Selection. ReliefF Attribute Eval is the Feature Selection technique used to run the experiment.

4.3. ReliefF Attribue Eval

RELIEF algorithm was developed by Kira and Rendell in 1992 to solve classification problems containing high dependant variables such as parity problems. But RELIEF algorithm is very sensitive to noise and it cannot handle incomplete data and regression problems [13]. ReliefF is an extension of RELIEF algorithm and it is capable of handling incomplete data, multi class problems and regression problems. ReliefF is used in inductive logic programming (ILP) for estimating the utility of literals during construction of theory.

In Weka, we used ReliefF for attribute ranking. The list of ranking attributes for KYOTO dataset is given below

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 Label):
ReliefF Ranking Filter
Instances sampled: all

0.938748	2	Service
0.833555	10	Dst-host-srv-count
0.693342	4	Destination-bytes
0.258589	9	Dst-host-count
0.253662	5	Count
0.234621	6	Same-srv-rate
0.174168	3	Source-bytes
0.046205	1	Duration
0.03755	8	Srv-serror-rate
0.020506	11	Dst-host-same-src-port-rate
0.018509	13	Dst-host-srv-serror-rate
0.003329	7	Serror-rate
0.000133	12	Dst-host-serror-rate

Ranked attributes:

Selected attributes: 2,10,4,9,5,6,3,1,8,11,13,7,12 : 13

The ranked attributes are shown according to the selection of attributes. Dst_host_srv_serror_rate is the least ranked attribute shown in ranked attributes . To improve the accuracy we removed the least ranked attribute (LRA) . After removing LRA , the accuracy improved is 98.66%.

5. Experimental Results

The experiments were carried out using R tool. We used KYOTO data set which is refined version of KDD CUP 99 data set. KYOTO data set contains of 15 attributes in which the last attribute contains class label. Accuracy, Precision, Recall and F-measure are the performance measures derived from the confusion matrix.

	Classified as Normal	Classified as Attack
Normal	TP	FP
Attack	FN	TN

where
TP - True Positive
FP - False Positive
FN - False Negative
TN - True Negative

$$\text{Accuracy} = \frac{\text{No. of samples correctly classified in test data}}{\text{Total no. of samples in test data}}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

We conducted all our experiments using R tool. The performance of our proposed model is shown in following tables.

when hidden layers = 3

Table 3: Performance Of Our Model When C=3

S.No	Accuracy	Precision	Recall	F-measure
1	98.66	98.95	99.30	99.12

when hidden layers = 4

Table 4: Performance Of Our Model When C=4

S.No	Accuracy	Precision	Recall	F-measure
1	98.4	98.57	99.28	98.92

when hidden layers = 5

Table 5: Performance Of Our Model When C=5

S.No	Accuracy	Precision	Recall	F-measure
1	97.86	99.64	97.53	98.56

The following figure shows comparison of our approach with different models

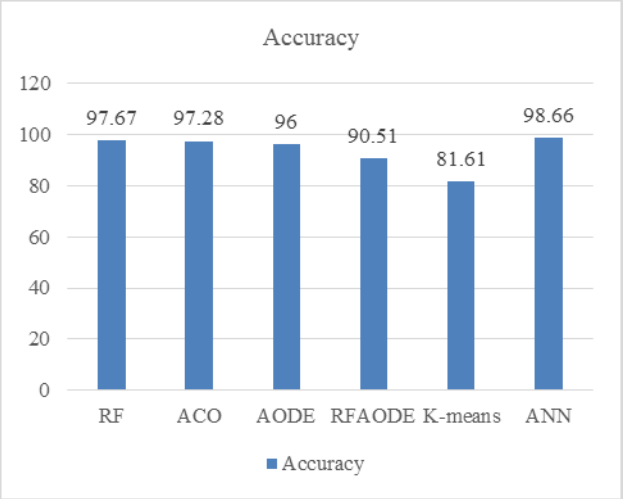


Fig .4: Comparison of our approach with different models

6. Conclusion

This paper deals with Back propagation algorithm in ANN to classify the attacks as either normal or attack (presence of an attack). We applied Feature Selection on the data set to remove duplicate and irrelevant features. We applied ReliefF attribute evaluation in Weka for feature selection.Our proposed approach is evaluated and compared using KYOTO data set. The experimental results showed that Accuracy, Precision, Recall and F-measure are increased by our proposed method when compared with different models. In future, we will work on feature selection techniques to improve the accuracy of the model.

References

[1] Amreen Sultana et,al," Intelligent Network Intrusion Detection System using Data Mining Techniques", 2nd International Conference on Applied and Theoretical Computing and Communication Technology (icATcct),978-1-5090-2399-8/16.

[2] Mr. MohitTiwari, Raj Kumar, AkashBharti, Jai Kishan, "INTRUSION DETECTION SYSTEM", International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com, Volume 5, Issue 2 (March - April 2017), PP. 38-44.

[3] BambangSetiawan, SupenoDjanali, Tohari Ahmad," A Study on Intrusion Detection Using Centroid-Based Classification ", 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia.

[4] Vidushi Sharma et.al," A Comprehensive Study of Artificial Neural networks", International Journal of Advanced Research in Computer Science and Software Engineering Research paper, Volume 2,Issue 10,October 2012 ISSSN:2277 128X.

[5] Ms.sonali et.al, "Research Paper on Basic of Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 issue: 1 96-100.

[6] K.KanakaVardhini et.al, "Enhanced Intrusion Detection System using Data Reduction: An Ant Colony Optimization Approach", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 9 (2017) pp.1844-1847.

[7] Dr.L.Dhanabal et.al, "A study on NSL-KDD dataset for IDS based on Classification Algorithms", International journal of advanced research in computer and communication engineering, vol.4, issue6, June 2015.

[8] Nabila Farnaaz et.al, "Random Forest Modeling for Network Intrusion detection Sytem",12th International Multi-Conference on Information Processing-2016(IMCIP-2016), 213-217.

[9] MichielHermans, BenjaminSchrauwen, "Training and Analyzing Deep Recurrent Neural Networks", Ghent University, ELIS department, SintPietersnieuwstraat 41, 9000 Ghent, Belgium.

[10] M.A.Jabbar et.al, "RFAODE: A Novel Ensemble Intrusion Detection System", 7th International Conference on Advances in Computing and Communications, ICACC-2017, 22-24 August 2017, Cochin, India.

[11] Martin T. Hagan, Howard B. Demuth, Mark Beale, "Neural Network Design", China Machine Press, 2002.

[12] https://www.r-project.org/

[13] International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 10, October 2014 "Feature Selection using ReliefF Algorithm" R.P.L.DURGABAI.