INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2367 - 2371
RECEIVED : 25.10.2017 PUBLISHED : 09.11.2017

Nov 09, 2017

Enhanced efficient clustering analysis using artificial neural networks

[1]S Archana M.Phil, Research Scholar, K.M.G College Of Arts & Science, Gudiyattam.

[2]P. Vinodhini Assistant Professor, PG & Research Department Of Computer Science & Applications, K.M.G College Of Arts & Science, Gudiyattam.

Abstract:

Intrusion Detection System (IDS) is an example of Misuse 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. In this paper it is projected that intrusion detection system with the amalgamation of k-means clustering and artificial neural network to improve the system. To obtain a better result benchmark dataset was split into training and testing part and then cluster the dataset into five different divisions. Elman Neural Network. After implementing these functions we have proposed a comparative analysis between them and choose the best accuracy rate among them. Here, it has been proved that, using the clustering technique a better accuracy rate can be found that improve the system with the best neural network functions which is the probabilistic neural network. It is also important to select efficient feature sets for better accuracy.

Keywords: IDS, Misuse detection system, Neural network.

1. INTRODUCTION

Intrusion Detection System is one of the common phenomena in this early age of internet security. IDS works for known and unknown attacks or it can be said that it can also be worked for novel attacks also [1]. But this paper worked for known attacks only. Intrusion Detection System monitors the network traffic and warns for the suspicious activities that enter into the system.

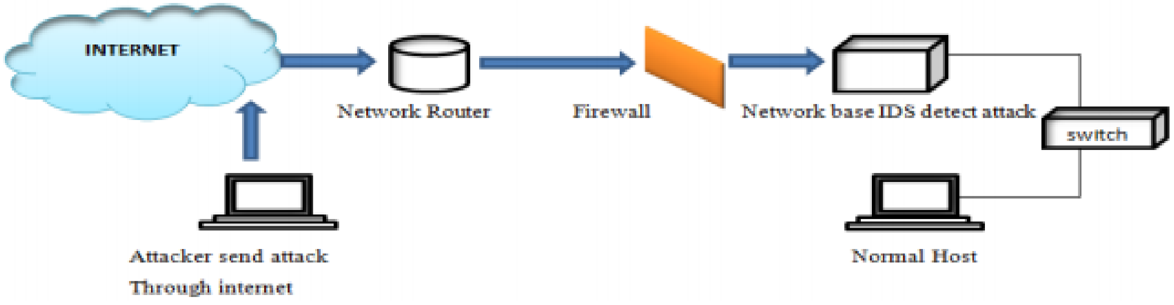


Fig.1.Basic figure

Network Based IDS and 2.Host Based IDS. Network based IDS detects network attacks as payload is analyzed and Host based IDS detects local attacks before they hit the network. IDS can be more understandable by its precision rate, the harmonic mean of sensitivity and the misclassification comparison rate. In data mining, clustering is an unsupervised technique that main advantage is splitting the dataset into different groups with its similarities . Among many clustering process, the k-means clustering is selected which is the common and simplest calculation technique. After doing the clustering part the dataset will be alienated into different cluster which will be trained by the different artificial neural network functions. But before training the ANN with the clustered data, feature selection process has been applied to reduce the irrelevant features to get a proper result. Intrusion detection system with artificial neural network is one of the widely used data mining techniques that proves how it is worthy while doing the complex terms in an easy way. And we have merged the IDS, clustering and ANN to get a better result for our proposed system. Though clustering and ANN have both some drawbacks, but the combination of these two can improve the performance of IDS. In k-means clustering the low capability to pass the local optimum and strong sensitivity are the main shortcomings [9]. On the other hand, lower detection precision for low number of attacks and weaker detection stability are the main disadvantages of artificial neural networks [1]. But when we cluster a dataset into several cluster and train individual neural networks with each cluster the precision for low frequent records improves up to maximum. This research work was tested on NSL-KDD dataset where there are five types of data. They are- Normal, Probe, Dos, U2R, and R2L. Low number of attacks means the remote to local (R2L) and user to root (U2R) data's. These data are called the low-slung frequent attacks.

2. RELATED WORK

IDS is the most used and developed system that can detect attack. Intellectual IDS can perform as a dynamic defensive system which is capable of adapting dynamically changing traffic pattern. Many researchers have worked on Intrusion Detection System to give the preeminent output through their system. It can be more categorized where researchers have mentioned the IDS system with the merging of the artificial neural network and clustering techniques. They showed the performance with other well-known methods such as decision tree, naïve byes, in terms of detection precision and detection stability. They mainly used the clustering techniques to generate different training subsets. And based on these subsets, different ANN modules are trained to formulate different base models. Finally they used an aggregation model to aggregate the result. They reduced the complexity of each of the sub training set and correspondingly they increase the performance of the detection. Moreover, many of them proposed a survey report on where they have cited the detection of accuracy from IDS using only the artificial neural network classifier.

However, some have provided the combined approach for anomaly detection using clustering and neural networks techniques [18]. Using the modified SOM or self-organizing map is used to create the network with the help of distances threshold, connection strength and neighborhood functions and k-means clustering algorithms groups the nodes in the network with the help of similarity measures. It disclosed when the learning rate increases the number of output nodes decreases. And many have provided, Intrusion Detection using Fuzzy Clustering and Artificial Neural Network where it has also proved the better accuracy for using these two techniques [4] where they proved the better accuracy rate for each of the attack in the system.

3. PROPOSED SYSTEM

This proposed system has 3 stages. The system is described in detail. Then we will discuss the modules of the system later. Primarily the dataset is divided into train and test set with ratio of 70:30 respectively and then cluster the train set with k-means clustering. Then train different ANN for

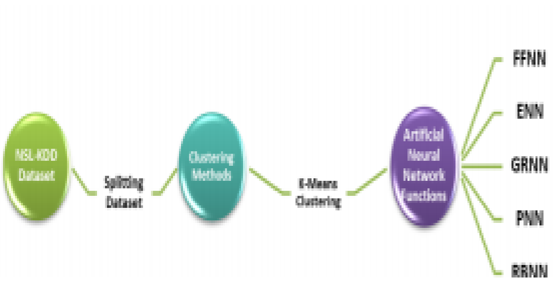


Fig.2.Proposed Model

different clusters, and cumulate the ANN on the last stage. So all in all it can be articulated that the main purpose is to prove the best accuracy using the different techniques that are being used in intrusion detection system. The more the techniques the more the efficient results have been established. From the above state of art it can be explained that each and every paper has been provided with different models where few of them have implemented too many techniques. But in this proposal two different techniques i.e. clustering and neural network are implemented with feature selections to improve the output so much prominent than others.

While working with the dataset we have used the NSL- KDD99 Dataset and improvise the dataset as an advantage of calculation. In the dataset the nominal data are changed with some numeric value to be more effective while calculating. There are four attack types and they are- Probe, Dos, U2R (User to Root) and R2L (Remote to Local) and a Normal class. These names are initiated with the numeric value from 1-5. This dataset consists of 41 features and the features in columns 2, 3, and 4 in the KDD99 dataset are the protocol type, the service type, and the flag, respectively. The value of the protocol type may be tcp, udp, or icmp, the service type could be one of the 66 different network services such as http and smtp and the flag has 11 possible values such as SF or S2. For instance, in the case of protocol type feature, 0 is assigned to tcp, 1 to udp, and 2 to the icmp symbol.

4. ANALYSIS

Artificial neural network invented by the idea of the human brain that deals with visual data and learned to distinguish objects. A neural network is a set of alarmed units following a particular topology. It creates connections between many types of processing elements where each parallel to a single neuron in a biological brain.

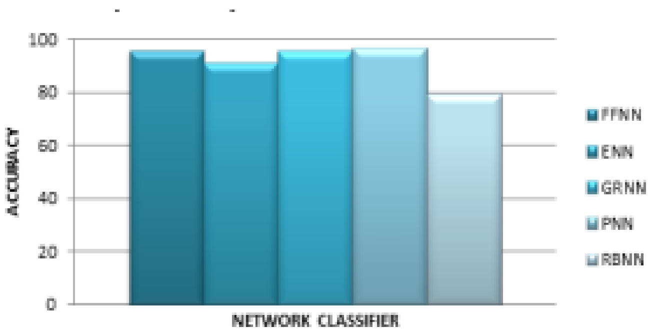
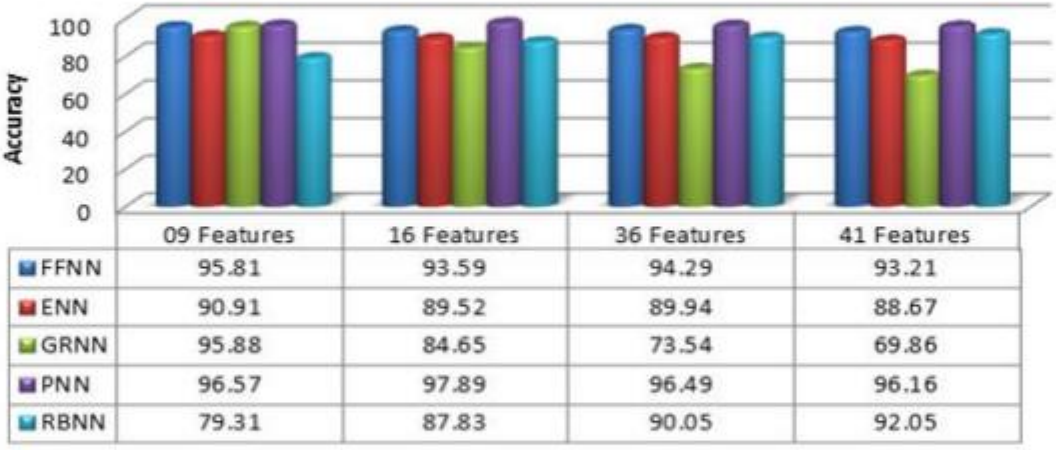


Fig.3.Accuracy

Neural networks have been extensively used in anomaly detection system as well as misuse detection. Its main advantage is it's a nonparametric model and it is easy to understand compared to statistical methods. There are many artificial neural network functions. For larger features sets like 36 and 41 features the performance of PNN is far better than other functions. It is here to be mentioned that for larger feature sets GRNN performance is lagged behind at a huge margin. So one thing that is common for all the dataset, is the accuracy for probabilistic neural network is preeminent than others.



	09 Features	16 Features	36 Features	41 Features
🟦 FFNN	95.81	93.59	94.29	93.21
🟥 ENN	90.91	89.52	89.94	88.67
🟩 GRNN	95.88	84.65	73.54	69.86
🟪 PNN	96.57	97.89	96.49	96.16
🟦 RBNN	79.31	87.83	90.05	92.05

Fig.4.Result Summary

In this paper it has been proposed that using the clustering techniques with neural network results is higher accuracy compare to other model and in detecting the attacks it is important to get the leading result. Here, it is seen that, imposing features selection over the clustering process has solved the redundancy complexions so far. Clustering is the mainstream for detecting attack of low frequent data. And as the k-means clustering is the most easiest and effective clustering so the output has verified this characteristics very flourishing. Though we have selected those features selection method from some previous research.

CONCLUSION

In this study, it has been proved that, the Probabilistic Neural Networks provide better accuracy over other Neural Network functions i.e. Feed Forward Neural Network, Elman Neural Network, Generalized Regression Neural Network and Radial Basis Neural Network. And the reduction of feature matrix makes the performance enhanced. So improving the accuracy efficient feature selection techniques can be applied to improve the accuracy. We have used the clustering technique for identifying the low frequent data for more specification. And thus prove a better accuracy rate with the proposed techniques.

REFERENCES

[1]. Bouzida Y., Cuppens F., 2006, Neural networks vs. decision trees for intrusion detection, In IEEE / IST Workshop on Monitoring, Attack Detection and Mitigation.

[2]. Shrivas A.K., Dewangan A. K., 2014 An Ensemble Model for Classification of Attacks with Feature Selection based on KDD99 and NSL-KDD Data Set, International Journal of Computer Applications (0975 – 8887) Vol 99 – No.15.

[3]. Elhamahmy M. E., Hesham N. E. and Imane A. S., 2010 A New Approach for Evaluating Intrusion Detection System, CiiT International Journal of Artificial Intelligent Systems and Machine Learning, Vol 2, No 11

[4]. Surana S. 2013 Intrusion Detection using Fuzzy Clustering and Artificial Neural Network, Advances in Neural Networks, Fuzzy Systems and Artificial Intelligence, ISBN- 978-960-474-379-7.

[5]. Osoba O., Kosko B., 2013 Noise-enhanced clustering and competitive learning algorithms, Neural Networks 37 (2013) 132–140.

[6]. Kumar V., Chauhan H., Panwar D., 2013 K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4.

[7]. Wu, Junjie, Advances in K-means Clustering, A Data Mining Thinking, Springer Press, ISBN: 9783642298073