SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875930

SCALABLE INTRUSION DETECTION USING DNN

B. L. Malleswari
Professor, Department of Electronics and Communication Engineering,
Sridevi Women's Engineering College, Hyderabad, India, blmalleswari@gmail.com

K. Anusha
U.G Student, Department of Electronics and Communication Engineering,
Sridevi Women's Engineering College, Hyderabad, India

C. Aakanksha
U.G Student, Department of Electronics and Communication Engineering,
Sridevi Women's Engineering College, Hyderabad, India

M. Haritha
U.G Student, Department of Electronics and Communication Engineering,
Sridevi Women's Engineering College, Hyderabad, India

Abstract: In this Paper, Intelligent Intrusion Detection System using deep neural networks is proposed. Because of increase in technologies in our day to day life number of security breaches happening in various parts of the world is also increasing at considerable rate, now this makes us to realise that there is a severe need of strong intrusion detection systems to overcome the problems related to security breaches, in order to do that we propose a hybrid intrusion detection system that can effectively work to detect various malicious activities happening at host level and as well as network level. Different misused detections have been used previously to identify the intruders and their suspect activities performed at host level and network level but many of them failed as they are not accurate, in this paper we proposed a unique model of using Deep neural networks to solve the security breaches problems. We used various set of data sets like NSL and KDD Cup 99 to effectively detect the intrusions happening at different networks. The proposed system meets the higher level of accuracy. It compares the SVM and Random Forest algorithms and has a greater accuracy.
Keywords: Machine Learning, Deep learning, Neural Network, SVM, Random Forest.

1.Introduction

An Intrusion detection system is developed such that it monitors the network traffic and immediately alerts if any suspicious activity is identified. Malicious Activities if left neglected can lead to a destruction of entire systems. In order to not meet that and decrease the rate of cyber-attacks and security breaches a strong intrusion detection is required.

Based on the behaviour of Intrusion, systems are broadly classified as NIDS, HIDS and protocol base IDS. Intrusions which happen at network level are Network level intrusion detection systems and the Intrusions which happens at host level are called Host level Intrusion detection

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875930

system [1]. NIDS and HIDS are the two IDS which are mainly focused when trying to implement an effective intrusion system. NIDS deal with the data packets or network packets whereas HIDS deals with the system applications, operations and log files.

Various misused detections were used to implement a strong intrusion detection system, but they all failed due to the low accuracy rates and were not suitable for complex data sets.
Deep learning models have recently become a popular algorithm to effectively analyse the intrusions and give higher accuracy than previously used methods.

2. Method

All the intrusion detection systems previously Implemented are based on various Machine learning techniques, but they were not successful in identifying and could not deal with all types of data sets. Almost every technique used in previous IDS was compatible with only simple kind of data sets but could not have better performance with complex data sets. Besides Machine learning techniques various data mining tools have also been used, and a deep analysis on all of techniques are done and it was confirmed that most of them has a false accuracy rates and poor performances. An attacker points at anonymous login and always uses a greater security by their side by using complex data sets which cannot be identified by the IDS, they aim at using strong data sets hence many of the data mining tools have been showed as a poor performer in finding out the intrusion detections [2]. An old and unique algorithm known as Ad boost algorithm was also used to detect intrusions. This Algorithm uses decision stumps as a weak classifier and combine them for continuous features and categorial features so that it becomes strong classifier but this algorithm also failed due to their low computational rates and high false rates.

2.1 The Purpose of Research

The proposed method promotes a scalable and hybrid model. Here we use a deep neural network and analyse the previously used misused detections SVM (support vector machines) and Random Forest algorithms and compare them with proposed DNN models [3]. The proposed DNN model has a greater accuracy in identifying the various intrusions happening at host level and network level where as the misused detection has very low accuracy rates.
The proposed neural networks have many hidden layers so a deeper tasks and complex data sets can be solved using the proposed DNN model.
The misused detections SVM and Random Forest are the best techniques for linear and regression problems but both of them has a poor performance than the DNN. The deep neural network consists of many hidden layers it has an input layer, hidden layer, neurons and an output layer. These hidden layers consist of neurons which performs tasks and makes the process so smooth it automatically trains the data sets and generates the model there by giving us accurate results.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875930

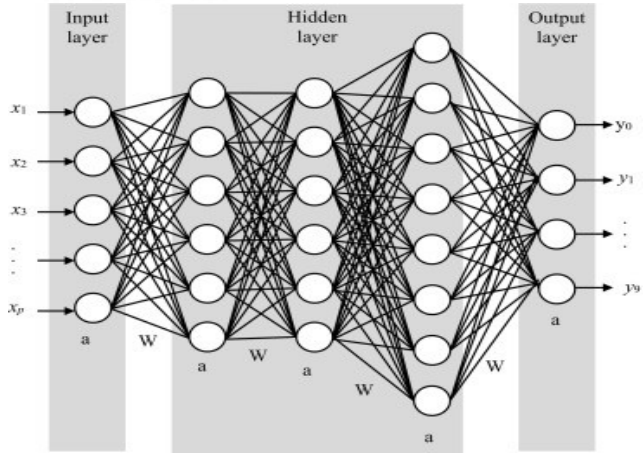

Fig 1 DNN Architecture

The proposed method uses data sets which are very popular to find any kind of intrusions. NSL and KDD CUP 99 Data sets are used in proposed model [3][4]. Testing results showed that their 5 classes of classification showed 80% of accuracy. The flow of execution deals with uploading data sets, training data sets, generating model, run SVM, run random forest and run DNN and finally shows the accuracy graph.

2.2. System Design

The system is designed in such a way that it collects publicly available data sets and compares the classical machine learning models with the proposed deep learning models. The misused detections used in this paper are Support Vector Machine and Random Forest algorithms [4].

In order to execute the three different models specified in this paper we need to import certain modules in python such that the required calculations, pre-processing of the data, plotting the graphs takes place without any errors [5]. The different python libraries like Pandas, NumPy, Sklearn, Matplotlib need to be imported.

SVM is the supervised machine learning technique, it is best used for linear and regression problems; it starts its work by dividing the given data sets into two classes and separating them using hyperplanes, margin and support vectors. Margin is the distance between the hyperplane and the support vectors, in SVM support vectors are nothing but the data points [5][6].

Random Forest is the second misused detection proposed in this paper. Random Forest is also a supervise machine learning technique which is best used for classification and regression problems. It works on the principle of ensemble learning that is combining two simple classifiers to solve a complex problem. Random Forest forms the decision tress, it takes the average instead relying on the decision trees. The more no. of decision tress represents a greater accuracy. It selects K points from the data sets and form decision trees.

Deep neural network is an artificial neural network which consists of multiple layers between the input layer and output layer [5]. We are using the concept of deep learning because complex tasks and difficult problems are easily solved using deep neural networks as it contains neurons, these neurons are responsible for predicting accurate output as it behaves exactly like huma brain and analyses the problem, it trains the data and use these trained data sets to test the models [6]. Deep neural network is a feed forwarding model where data moves from input layer to output layer without looping back. DNN creates a virtual neuron and assigns random

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875930
numerical values(weights), then these weights and input values are multiplied to give result between 0 and 1.

2.3 DNN Module

The layers of DNN helps to extract the features which is very important in IDS. Each layer of DNN Estimates some features an pass them to the next layers and finally the last layer performs the classification. Hidden layer contain 41 neurons for KDDCup99 dataset and 41 neurons for NSLKDD data set. An output layer contains 1 neuron for binary classification for all the types of datasets. 5 neurons are required for multi-class classification for KDDCup99 and 5 neurons for NSL-KDD data set.

The DNN is trained using the backpropagation mechanism [13]. Generally, the units in input to hidden layer and hidden to output layer are fully connected. The DNN is composed of various components, a brief description of each component is given below. Fully connected layer: This layer is called as fully connected layer since the units in this layer have connection to every other unit in the succeeding layer. Generally, the fully connected layers map the data into high dimensions. The output will be more accurate, when the dimension of data is more. It uses ReLU as the non-linear activation function. Dropout (0.01) and Batch Normalization was used in between fully connected layers to obviate overfitting and speedup the DNN model training [10]. A dropout removes neurons with their connections randomly. In our alternative architectures, the DNNs could easily overfit the training data without regularization even when trained on large number samples. Classification: The last layer is a fully connected layer which uses sigmoid activation function for Binary classification and SoftMax activation function for multi-class classification. The prediction loss function for sigmoid is defined using Binary cross entropy and the prediction loss for SoftMax is defined using the Categorical cross entropy as follows: The prediction loss for Binary classification is estimated using Binary cross entropy given by,

loss (pd, ed) = − 1 N X N i=1 [edi log pdi+(1−edi) log(1−pdi)] (11)

where pd is a vector of predicted probability for all samples in testing dataset, ed is a vector of expected class label, values are either 0 or 1. The prediction loss for multi-class classification is estimated using Categorical cross entropy given by,

loss (pd, ed) = −X x pd(x) log(ed(x)) (12)

where ed is true probability distribution, pd is predicted probability distribution. We have used adam as an optimizer to minimize the loss of Binary cross entropy and Categorical cross entropy.

3. Research Results

Publicly available data sets are used to measure the performance of intrusion detection. Machine learning and DNNs are used as a baseline method. By using NSL and KDDCUP99 datasets the accuracy was shown as 50% in misused detections and 80% in DNNs [7].

The data sets are first collected then tested and trained after that model is generated, the test datasets are used to evaluate the trained machine learning and DNN models and then evaluated to give the accuracy graph [8][9].

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875930

In all the cases it is observed that the DNN is more efficient for intrusion detection system when compared with the classical machine learning algorithms [10].

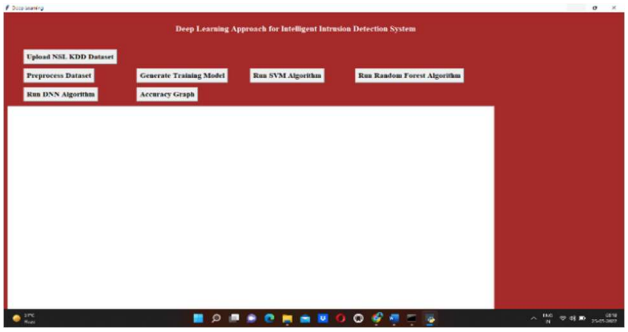

Fig 2 Display window

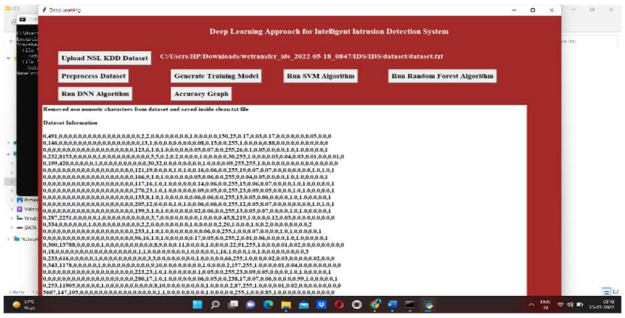

Fig 3 Pre-processing data

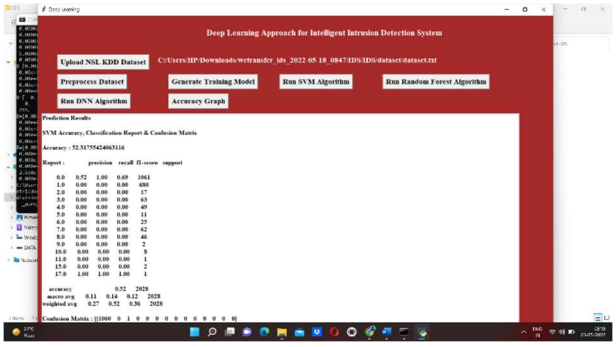

Fig 4 Run SVM

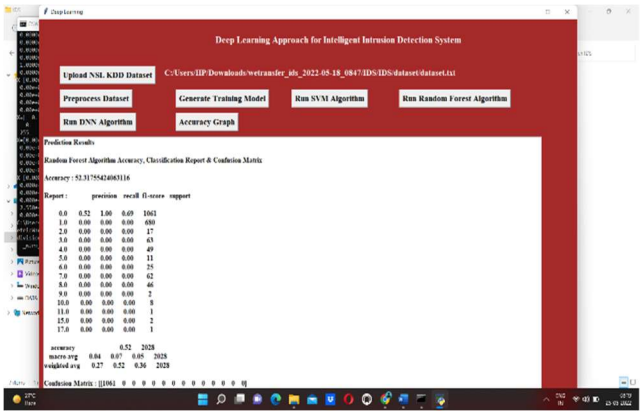

Fig 5 Run Random Forest

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875930

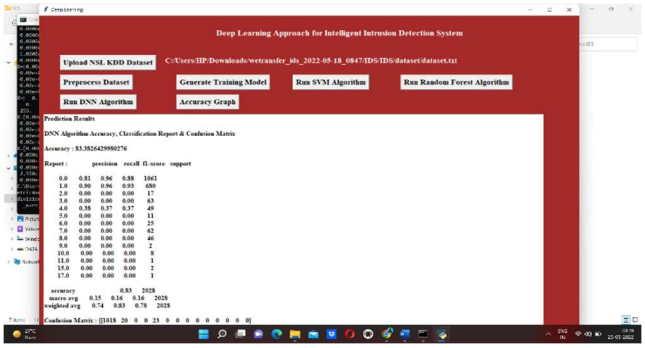

Fig 6 Run DNN

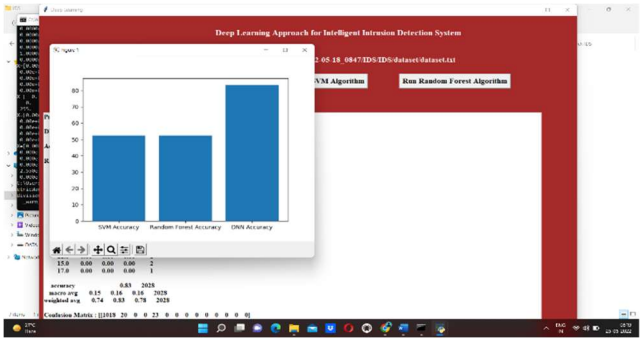

Fig 7 Accuracy graph

4. Conclusions

In this paper a hybrid scalable intrusion detection system is proposed which gives us the best accuracy rate in finding various malicious activities happens at host level and network level. [15][16]. In this paper three algorithms have been used SVM, Random Forest.

Deep neural network is the proposed algorithm which performs complex tasks with greater accuracy as it has hidden layers and neurons which acts exactly like human brain and predicts the output by self-training the data sets [20]. Through analysis and by performing accuracy tests we have observed that DNN has showed a better performance and it can also perform better with complex data sets.

In future the proposed model can be enhanced by adding more and more modules and there by monitoring the DNS.

Overall performance can be improved by using complex data sets and training them in order solve the more complex cyber security problems.

References

[1] Mukherjee. Herberlein L.T., &Levitt, K. N. (1994) Network intrusion detection system IEEE Network, pp 1-2

[2] U. Fiore, F. Palmieri. A. Castiglione, and A De Santis, “Network Anomaly Detection with the Restricted Boltzmann Machine” Neurocomputing, vol 122, pp 3-4

[3] M. A. Salama, A. Darwish, and A. E. Hassanien, “Hybrid Intelligent Intrusion Detection Scheme”

[4] S. Thaseen and C. A, Kumar “An Analysis of Supervised Tree Based Classifiers for

DOI: 10.5281/zenodo.6875930
Intrusion Detection for Pattern Recognition".
[5] Venkatraman, S., Alazab, M. "Use of Data Visualisation for Zero-Day Malware Detection," Security and Communication Networks, vol. 2018, Article ID 1728303, 13 pages, 2018.
[6] Mishra, P., Varadharajan, V., Tupakula, U., & Pilli, E. S. (2018). A detailed investigation and analysis of using machine learning techniques for intrusion detection. IEEE Communications Surveys & Tutorials
[7] Staudemeyer, R. C. (2015). Applying long short-term memory recurrent neural networks to intrusion detection.
[8] Hubballi, N., Biswas, S., & Nandi, S. (2011, January). Sequencegram:n-gram modeling of system calls for program based anomaly detection.In Communication Systems and Networks (COMSNETS), 2011 Third International Conference on (pp. 1-10). IEEE.
[9] H. Kayacik, A.N. Zincir-Heywood, and M.I. Heywood, "Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets." Proceedings of the third annual conference on privacy,security and trust 2005, PST 2005, DBLP
[10] Subba, B., Biswas, S., & Karmakar, S. (2017, November). Host based intrusion detection system using frequency analysis of n-gram terms. In Region 10 Conference, TENCON 2017-2017 IEEE (pp. 2006-2011).
[11] H. G. Kayacik, A. N. Zincir-Heywood, and M. I. Heywood, ''Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets,'' Proc. 3rd Annu. Conf. Privacy, Secur. Trust, 2005.
[12] J. Zhang, M. Zulkernine, and A. Haque, ''Random-forests-based network intrusion detection systems,'' IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 38, no. 5, pp. 649–659, Sep. 2008.
[13] H. Leung and S. Haykin, ''The complex backpropagation algorithm,'' IEEE Trans. Signal Process., vol. 39, no. 9.
[14] K. Simonyan, A. Vedaldi, and A. Zisserman. (2013). ''Deep inside convolutional networks: Visualising image classification models and saliency maps.'' [Online]. Available: https://arxiv.org/abs/1312.6034
[15] M. Alazab, S. Venkatraman, P. Watters, and M. Alazab, ''Zero-day malware detection based on supervised learning algorithms of API call signatures,'' in Proc. 9th Australas. Data Mining Conf., vol. 121, 2011, pp. 171–182.
[16] A. Alazab, M. Hobbs, J. Abawajy, and M. Alazab, ''Using feature selection for intrusion detection system,'' in Proc. Int. Symp. Commun. Inf. Technol. (ISCIT), Oct. 2012, pp. 296–301.
[17] A. Saracino, D. Sgandurra, G. Dini, and F. Martinelli, ''Madam: Effective and efficient behavior-based android malware detection and prevention,'' IEEE Trans. Dependable Secure Comput., vol. 15, no. 1, pp. 83–97, Jan. 2018.
[18] S. Naseer et al., ''Enhanced network anomaly detection based on deep neural networks,'' IEEE Access, vol. 6, pp. 48231–48246, 2018.
[19] N. R. Sabar, X. Yi, and A. Song, ''A bi-objective hyper-heuristic support vector machines for big data cyber security.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875930
[20] W. Wang et al., ‘‘HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection,’’ IEEE Access, vol. 6, pp. 1792–1806, 2018.