International Journal of Engineering Sciences Paradigms and Researches [Volume 47, Issue: Special), March 2018]
www.ijesonline.com ISSN (Online): 2319-6564

Finding Obtrusions in Networks using Deep Learning Techniques

Sagram Keshari Sahoo
Raajdhani Engineering College, Bhubaneswar
sangramsahoo@rec.ac.in

Abstract- *A network obtrusion is an unlawful action on a digital network. Network intrusions almost invariably result in the theft of important network resources and jeopardize the security of the network and/or data. We presented a deep learning technique for blockage recognition using the NSL-KDD dataset. The identification technique uses the logistic regression method after pre-processing a dataset. A network obtrusion detection system is a system that detects these kinds of obtrusions, which happen when a hacker tries to get access to your system. Since its main objective is to ascertain whether a hacker or cracker is trying to access the system, it watches packets on a network cable.*

Keyword- Intrusion, Logistic regression, Dataset, Network.

The present study aimed at obtaining the accuracy in obtrusion identification system. Obtrusion identification can identify unknown attacks from network and has been an effective mean of network security. Nowadays, the existing systems/methods for detecting the obtrusion are based on traditional machine learning technique/models such as KNN, SVM, etc. However, these methods obtain great features, but they get low accuracy and mostly depend upon manual design of traffic features, which has been absolute in the age of big data. To reduce this accuracy and feature engineering problem we proposed the BAT model along with the Logistic regression algorithm where BAT model is combination of BLSTM and attention mechanism. The proposed model has high accuracy and better performance compared to other models.

Index Terms- Network traffic, Obtrusion identification, NSL-KDD, Deep learning.

I. INTRODUCTION

With the growth of internet technology, internet is providing many services for people which includes several activities such as online payments, internet banking, social networks etc. Government and private organizations store confidential data over network. These activities are becoming part of daily life which mostly imply on internet. Increase in usage of network leads to network viruses and malicious attacks, and the problems which are caused by internet can be solved by obtrusion identification system, which has gained the more focus by society and government department in network security.

Obtrusion Identification System (IDS) have become essential component of almost every security infrastructure predominantly because they provide a wall of Défense and restrain external attacks effectively, where other traditional security cannot perform well. Traffic types in the network are growing day by day with explosive growth of internet

business, which brings great challenge to obtrusion identification.Obtrusion Identification Systems (IDS) monitor the activities and events occurring in the systems and decide if these are intrusive actions or normal usage of the system. It scans the network or system for suspicious activity if any harmful activity detected then it reports to administrator. Based onanalysis technique IDS are classified as misuse and anomaly identification. The misuse method is very accurate in detectingknown attacks based on their signatures that are stored in the database whereas the anomaly identification approach automatically constructs a normal behaviour of the systems. This latter method can detect new attacks but, it can alsogenerate many false alarms. Obtrusion identification can be considered as a classification problem. Obtrusion detection accuracy may be greatly enhanced by boosting classifier performance in efficiently recognising malicious traffics.

II. RELATED WORK

[1] They suggested a machine learning technique for obtrusiondetection using a Nave Bayesian classifier in this research, which entails identifying irregular packets based on the system's experience. The dataset is created by analysing incoming packets and categorising them based on the value of attribute values. The dataset is used to determine if the next incoming packets are normal or aberrant. If suspicious packets are discovered, they can be reported. [2] The goal of this study is to lower the rate of false positive alarms, increase the rate of false negative alarms, and enhance the identification rate. They suggested a hybrid machine learning approach for network obtrusion identification based on a mix of K-means clustering and support vector machine classification, and they employed the NSL-KDD dataset to achieve this purpose. The classification was done with a support vector machine, and various measures were taken on the dataset to boost the classification results. The collected findings suggest that the proposed hybrid machine learning approach has achieved a positive identification rate and reduced the false alarm rate after training and testing. [3] To identify suspicious behaviour, they deployed artificial intelligence, which operates as a virtual analyst in tandem with IDS to defend against the threat environment and take necessary measures with the analyst's approval. They have concluded that data setsare more important which are used for training the machine learning algorithm and by experimental results they have found that k-Nearest Neighbours (KNN) technique’s performance is best if its k value has been chosen correctly. It gives good result in both evaluation and real-life scenario.

[4] Here they have studied different techniques of machine learning and deep learning for obtrusion identification such as Support Vector Machine (SVM), KNN, Decision tree, ANN, Random Forest and so on. After the survey they have concluded that both machine learning and deep learning techniques have their own advantages, but most important thing considered is dataset used. The machine learning and the deep learning methods will work better if we have proper datasets for training and testing the models. Whereas the deep learning is useful in handling the large datasets compared to machine learning but to obtain the interpretability some fine tuning is required. [5] In this paper they have proposed investigation of obtrusion identification as well as prevention from different network attacks. Before creating the training module, the suggested system performed data pre-processing, data normalisation, feature extraction, and feature selection. After feature extraction, any supervise classifier for a training module is applied. In the testing phase, the same process was followed according to the classification algorithm, and the classification accuracy for all assaults was evaluated. The suggested system was then tested using different supervise and unsupervised algorithms in the Weka 3.7 open-source environment. After the experiment they have concluded that the proposed implementation shows classification accuracy for selective input data set with various soft computing as well as machine learning algorithms.Also, they found that the average accuracy for entire system Naive Bayes provides highest classification accuracy while SVM introduces lower accuracy than other classical algorithms.

[6] The paper explores the pre-processing technique, model comparison for training as well as testing and evolution technique. Here they have proposed to develop a NIDS using machine learning which will learn from past data and oppose the intruder or else it will alert the administrator through the notification. [7] In this paper they have approached three level of stages such as Feed-Forward Neural Organization, Recurrent Neural Organizations (RNNS), and Recurrent Neural Organizations (RNNS) for obtrusion identification. They discovered that each technique to developing an obtrusion detection system has its own and a point clear from the comparisons among the various ways based on the testing findings. As a result, deciding which approach to use to create an obtrusion identification system is complicated. [8] In this paper, they performed analysis of the benchmark dataset NSL-KDD and CIDD-001 and proposed hybrid feature selection and ranking methods before applying self-learning (machine / deep learning) classification algorithmic approaches such as SVM, Naïve Bayes, k-NN, Neural Networks, DNN and DAE for getting optimal results. They looked at the performance of IDS using certain well-known performance indicator measures including Accuracy, Precision, Recall, and F1-Score. They discovered that on the NSLKDD dataset, k-NN, SVM, NN, and DNN classifiers perform almost 100 percent accuracy in terms of performance assessment metrics, but on the CIDDS-001 dataset, k-NN and Nave Bayes classifiers perform approximately 99 percent accuracy. [9] The paper provides a complete study about IDS, types of IDS, types of attacks, different types of tools and techniques of deep learning which can be used for obtrusion identification. They have developed IDS tool which can detect and prevent the obtrusion from the intruder. Here they have concluded that Obtrusion Identification Systems are not a only result to all security concerns, An IDS is not a replacement for a good security policy and Human intervention is required in obtrusion identification system.[10] They have proposed IDS which uses machine learning algorithm like Naïve Bayes and k-means clustering algorithm and implemented on it to detect the obtrusion. To evaluate these algorithms, they have used KDD cup99 dataset. Finally, the dataset is analysed by Weka tool, which is written in java and contains classification, cluster, association rules & visualization.

In this present research paper, further we are going to discuss about proposed work and methods used in the experiment and results, conclusion is discussed in detail.

III. PROPOSED WORK

Many different Machine Learning techniques are there for network anomaly identification and most of them obtains some outstanding feature, but these methods usually get low accuracy. To solve this problem, we have proposed IDS which uses deep neural network which is one of the deep learning techniques. In this experiment we used BAT-MC model and logistic regression algorithm. Where Binary Addition Tree (BAT) algorithm is used to get high accuracy and high efficiency. We have used the NSL-KDD dataset and logistic algorithm is applied on train the data and using test data result is predicted to obtain the desired accuracy in obtrusion identification.

BAT-MC model:

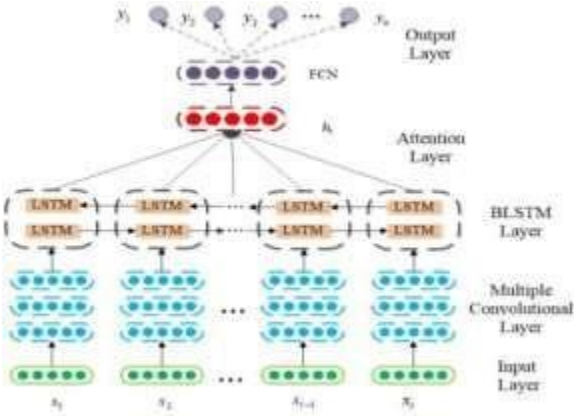

Figure 1: System architecture of BAT-MC model

As shown in Figure 1, from bottom to top, the BAT-MC model is made up of five layers: input, multiple convolutional layers, BSLTM layer, attention layer, and output layer. The BAT-MC paradigm turns each traffic byte into a one-hot data format at the input layer. Each byte of traffic is represented as an n-dimensional vector. We undertake normalisation procedures once the traffic byte is translated to a numerical representation. We use multiple convolutional layers to transform numerical input into traffic pictures. The convolutional technique is employed as a feature extractor for data packets having an image representation. At the BLSTM layer, the BLSTM model, which connects the forward and backward LSTM, is used to extract features from the traffic bytes of each packet. BLSTM can learn the sequential

qualities inside the traffic bytes since it is suitable to the structure of network traffic. In the attention layer, the attention mechanism is used to examine the important degree of packet vectors in order to get fine-grained characteristics that are more significant for malicious traffic detection. The features formed by the attention mechanism are then imported into a fully linked layer for feature fusion at the output layer, yielding the basic properties that describe network traffic behaviour correctly. The fused characteristics are sent into a classifier to get the final recognition results.

NSL-KDD dataset:

Network Security Laboratory (NSL) -KDD is a data collection designed to address some of the issues with the KDD'99 data set. The NSL-KDD data set provides a number of advantages, including the fact that it does not include redundant records in the train set, which means the classifiers will not be biased toward more frequent records. Because the suggested test sets contain no duplicate records, the learners' performance is not influenced by approaches that have higher identification rates on frequent records. The percentage of records in the original KDD data set is inversely related to the number of records picked from each difficulty level group. As a result, the classification rates of different machine learning algorithms vary more widely, making an accurate evaluation of different learning strategies more efficient. Because the train and test sets include a decent amount of records, it is feasible to perform the experiments on the entire set rather than a small subset at random. As a result, the assessment outcomes of various research projects will be consistent and comparable.

Logistic regression:

Under the Supervised Learning approach, one of the most prominent Machine Learning algorithms is logistic regression. It's a method for predicting a categorical dependent variable from a set of independent factors. A categorical dependent variable's output is predicted using logistic regression. As a result, the result must be a discrete or categorical value. It can be yes or no, 0 or 1, true or false, and so on, but instead of giving precise values like 0 and 1, it delivers probabilistic values that are somewhere between 0 and 1. The only difference between Logistic Regression and Linear Regression is how they are employed. For regression issues, Linear Regression is employed, whereas for classification difficulties, Logistic Regression is used.

In the proposed work we have used the logistic regression in pre-processing, test and predicting the data. Where in data preprocessing step we will process the dataset so that we can use it in our code efficiently and data is split into two parts as training data and test data. After the splitting logistic algorithm is applied to train data and using the test data, which will predict the result.

IV. RESULT AND DISCUSSION

Figure 2: View of Employee module

Employee module: Using employee module as shown in figure 2 user who is working in company can register with application and login to application and view profile as shown in figure 3. Employee who wants to check status or want to know if there is any anomaly is network connection or packets which are sent by him or any other employee inside company can request admin to check from the data set and update to employee. Employee can view anomaly packets inside the network after responding from admin.

Figure 3: Employee registration window

Admin module: Admin a module who looks after network related issues inside a company who will login with application, they can log dataset of networking packets which are part of the company and whenever they gets requests to check anomaly identification request admin will upload data set, pre-process data into test and train and then create model and predict from the test data and predicted result is displayed to admin and sent requested user. Figure 4 shows the admin module registration.

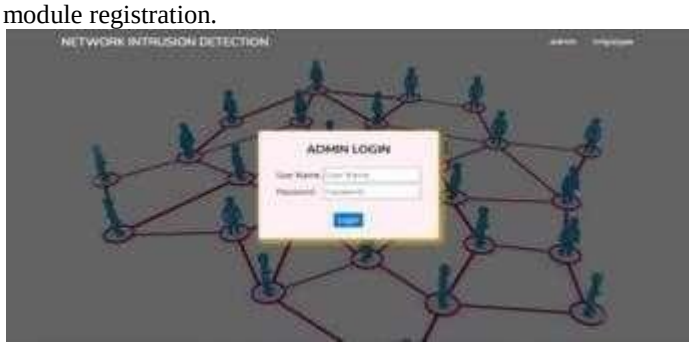

Figure 4: Admin details

Data Pre-processing:

Figure 5: NSL KDD dataset

Here the NSL-KDD dataset used for training and testing. The dataset is pre-processed before using with pre-process techniques.

Table 1. Evaluation metrics

Sl.No	Metrics	Values
1	Accuracy	0.8462
2	Precision	0.9884
3	Recall	0.8185
4	F1	0.8955

Table 1 shows the evaluation parameters for considered NSL_KDD Dataset with accuracy 84.6% for intrusion detection on network with deep learning method.

V. CONCLUSION

After the experiment the outcome is analysed to check the accuracy of the obtrusion identification. The problem of manual design features is well avoided with this technique. The BAT-MC model delivers good accuracy on the NSL-KDD dataset, according to experimental results. These comparisons reveal that BAT-MC model results are quite promising when compared to other existing deep learning-based approaches when compared to certain conventional classifier. As a result, we feel that the suggested technique is a useful tool for identifying obtrusions.We concludes that deep learning techniques for the obtrusion identification can obtain high accuracy compared to the machine learning techniques.

REFERENCES

[1] Ajit Kalekar, Niranjan Kshatriya, Sneha Chakranarayan, Snehal Wadekar, “Real Time Obtrusion Identification System using Machine Learning,” IJERT., Vol. 3 Issue 2, February – 2014.

[2] Hatim Mohamad Tahir, Wael Hasan, Abas Md Said3, Nur Haryani Zakaria, NorlizaKatuk, Nur Farzana Kabir, Mohd Hasbullah Omar, Osman Ghazali, and Noor Izzah Yahya, “Hybrid Machine Learning Technique for Obtrusion Identification System,”ICOCI 2015 11-13 August 2015.

[3] Syam Akhil Repalle, Venkata Ratnam Kolluru, “Obtrusion Identification System using AI and Machine Learning Algorithm,” IRJET., Volume: 04 Issue: 12 | Dec-2017.

[4] Hemavati*, Dr Aparna R, “A Survey on Obtrusion Identification System using Machine Learning and Deep Learning,” IJSRCSEIT., Volume: 05, Issue: 2, ISSN: 2456-3307, 2019.

[5] Abhishek Vaidya, Vikrant Karawande, Kunal Gaikwad, Parshwa Shah, Anand Dhawale, “Obtrusion Identification System using Soft Computing and Machine Learning Approach,” IRJET., Volume: 07 Issue: 04 | Apr 2020.

[6] Jayesh Zala, Aditya Panchal, Advait Thakkar, Bhagirath Prajapati, Priyanka Puvar, “Obtrusion Identification System using Machine Learning” IJSRCSEIT., Volume: 06, Issue: 3, ISSN: 2456-3307, 2020.

[7] R. Thenmalar, B. Madhavi, K. Naveenkumar, A. Neelakandasankar “Obtrusion Identification using Deep Learning,” Volume 8 Issue IX Sep 2020.

[8] Azam Rashid, Muhammad Jawaid Siddique, Shahid Munir Ahmed, “Machine and Deep Learning Based Comparative Analysis Using Hybrid Approaches for Obtrusion Identification System,” IEEE 2020.

[9] S. Santosh Kumar, M. Kannan, B. Vignesh, Mr. S. Rajarajan, “Obtrusion Identification System using Deep Learning,” IJERT., ISSN: 2278-0181, ICRADL – 2021.

[10] Jayasri P, Atchaya A, Sanfeeya Parveen M, Ramprasath J, “Obtrusion Identification System in Software Defined Networks using Machine Learning Approach,” ISSN: 2349-6495(P) | 2456-1908(O) Vol-8, Issue-4; Apr 2021.