Network Intrusion Identification System using Artificial Neural Network

Anuj Kulkarni, Himansshu Mital, Prajakta Memane, Atul Kathole

Department of Computer Engineering,
Zeal College of Engineering and Research, Pune, Maharashtra, India

Abstract: In today's world the anomaly-based network intrusion detection techniques have become very efficient in identifying the both known and unknown types of network intrusions, in this paper we have implemented the anomaly-based technique with the help of feed forward neural network in deep learning, the dataset used for the model building was CICIDS2017 Dataset which is updated with attacks type of today's world. We were very successful in detection of intrusion types in 14 groups with very high accuracy.

***Keywords*– Anomaly Based Network Intrusion Identification System; Artificial Intelligence; Deep Learning; Sequential Model.**

I. INTRODUCTION

An Intrusion Identification System is a software application or a service that monitors the network traffic and based on the traffic it detects the network attacks. Identifying the intrusion is the art of detecting malicious, wrong or non-benign activity. Intrusion Identification System that operates on a host to detect malicious activity are called host based intrusion identification systems. Similarly, Intrusion Identification systems that operate on network data flows are called network based intrusion identification systems. We used network-based intrusion identification system in this paper. Basically, there are two types of intrusion identification systems: Signature based Intrusion identification system and Anomaly based intrusion identification system. Signature based Intrusion Identification System are fast and works well but have one major disadvantage. It does not detect the unknown attacks. So, with the modification of attacks this method may fail to detect the attacks. On the other side, we have another type that is anomaly-based intrusion identification system. This system can be used to identify the unknown attacks or some kind of new attacks. Anomalies show very crucial and rare events in the anomaly identification. For an instant, if any computer is facing some irregular traffic pattern then we can say that some unauthorized actions have been performed which can be a malicious data transmission to the destination

In anomaly-based network IDS, given system is trained with "benign" as well as "malicious" network traffic for generating the Model. When the generated Model for particular system is available, it is used to classify the malicious type based on its learning. Most important a self-learning system have to be use. It is an effective model which has characteristic to adapt to generalize systems behaviour and survive with changing environments.

This system is used in order to detect various on-going attacks based on anomalies and have the aim of catching the hackers before they do damage to our work[8]. This system inspects the data within the packets, analyses them, understands them, feed it to the trained and tested model and after learning various features and attributes based on the fed data, it detects the attack if it shows any kind of anomalies defined in our system.

Another critical problem in anomaly-based network intrusion identification is possibility of having labelled data for training and model validation. In normal behaviour, availability of labels is high but in intrusions labels are not highly available. In such cases, unsupervised and semi supervised techniques which are used to detect anomalies are highly recommended.

Hence, motivated by problems mentioned above, we propose an anomaly-based NIDS using deep learning to identify fourteen types of network intrusions in our system. With the help of deep learning, we expect to handle issues in anomaly-based network intrusion identification, like high intrusion identification rate, capability to adapt to dynamic network environments.

Using Artificial Intelligence in Intrusion identification system helped us to improve the accuracy of the system and also it helped to analyse huge volume of dataset in dynamic nature in comparatively less time. It efficiently recognizes the intrusion patterns and detects the attacks accuracy with very low false-positive rate and providing higher true-positive rate.

So, anomaly-based network intrusion identification using artificial intelligence allows us to detect unknown attacks with help of anomalies found in the network and detect them accurately and in very fast manner with the help of artificial intelligence.

Deep Learning Techniques:

The goal of deep learning states that it is used to calculate hierarchical features or representations of the observational data. It extracts the higher-level features or factors from lower level features. Also, it focuses on extracting and learning a better feature representation from a huge amount of labelled data. It helps model to be pre-trained in a supervised manner. The strategy of layer-wise supervised training allows adequate training of deep networks and gives auspicious solutions for different challenging learning problems, considerably bettering upon the current. In an area which is near a good local minimum, we initialized the weights, which made increase to internal distributed representations that were high level abstractions of the input, thus it brought a better generalization[8]. There are various Deep learning techniques used for pre-training and they are chosen depending on many various domains as network packets are available in sequential format. In our system, we have used feed-forward network for training the model.

II. LITERATURE SURVEY

Now a days, networks are more susceptible to intrusions, attacks. The heterogeneity of the attacks give rise to the continuously growing and changeable nature between network attacks. Anomaly Based Intrusion Identification technique is an appreciated methodology to expose various known as well as unknown and new attacks in intrusion identification systems. These identification becomes more accurate and precise with the help of artificial intelligence.

Alexet al [1] considered the signature-based identification methods in which performance is improved by utilizing an artificial neural network classifier for identification of shellcode patterns in network traffic. The research presented in given paper illustrates an offline approach to detecting shellcode patterns within the data.

Nguyen et al [2] considered the anomaly-based identification methods in which performance gets improved because of the adaptability of system in dynamic network environment. The authors have used deep learning methodologies to implement anomaly-based NIDS. The system detects the intrusion and classifies them into five groups with high accuracy. The time consumed by stacked auto-encoder is much more than the stacked restricted Boltzmann machine.

III. PROPOSED METHODOLOGY

1. IMPLEMENTATION

The system is implemented in two phases: In first phase we focused on packet capturing, processing data, data scaling, data transformation and fine tuning of data whereas in second phase we focused on training the supervised model.

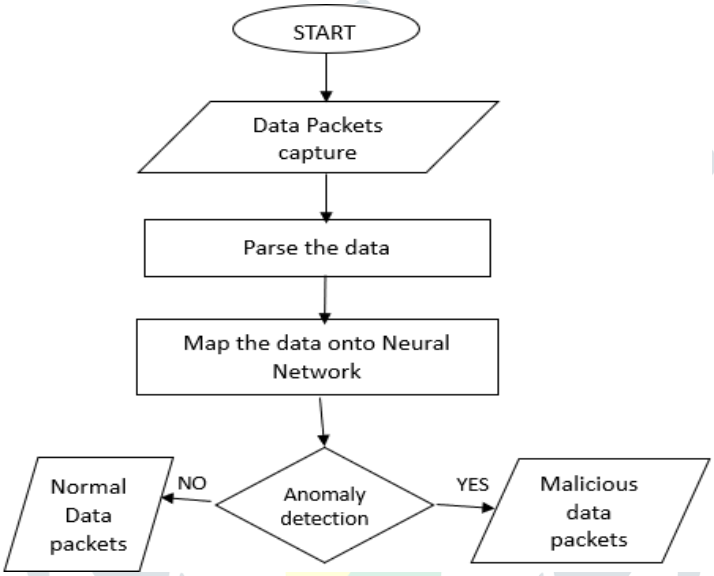

Figure 1: Process flow chart of proposed methodology

First Phase

In the first phase we used the CIC Flow Meter to collect the Network traffic. The acquired packets were then processed through data cleaning and transformation processes where the additional attributes such as type of attack were added for better classification during training and also the most of repetitive being packets were removed as they were common in all network traffic and were non-malicious.

After successful data cleaning and transformation, we applied standard scalar for scaling various attributes, we also identified and removed various impurities in data. The rows with missing values were eliminated from the dataset. The attributes such as destination port which had unique values in itself were not scaled. Also, as attributes like source port, source and destination IP addresses were not significant in identification of network intrusion. Hence, these attributes were eliminated from dataset[8]. The labelled data which was externally added to classify the packet type was in string format, so each label was given a unique number, so as to make data ready for neural network training.

Second Phase

In the second phase of the implementation, the process of training the ANN started. Here, the obtained dataset from previous phase was used for training the feed forward neural network. The Sequential Network Model is very simplest form of Neural network and works as name suggests in forward direction. We had implemented FFN in sequential pattern for better and efficient result. We focused on training the neural network using obtained data and making it adaptive for new attack patterns.

2. SEQUENTIAL MODEL

It is a directed graph which is acyclic in nature. It implies that there are zero feedback connections and loop present in the network. Multiple hidden layers can be present in any general network. Each node in layer is called as Neuron. It is the primary processing unit of a Neural Network.

Neuron

An Artificial Neuron is the basic unit of a neural network. It works in two stages. It takes the weighted inputs and makes sum of it. Further it applies those weighted sums to an activation function. It helps to make it normalize. The activation functions can be broadly classified into linear and nonlinear functions. Each input of a neuron is associated with same weights. These are the some factors which the network has to study during its training phase.

Activation Function

At the output of a neuron activation function helps as a decision-making body. It helps neuron to understand linear and non - linear decision boundaries which are hinge on the activation function. It normalizes the output of neuron which helps in avoiding it to become very large after various layers. It is due to the cascading effect. Practically Sigmoid, Tanh, Rectified Linear Unit (ReLU) are the three broadly used activation functions.

Input Layer

This is the initial layer of a neural network. It helps in providing the Input data and features to the network.

Output Layer

Predictions are taken out in this layer. The activation function to be used in this layer is dissimilar for distinct problems.

Hidden Layer

A feed forward network is used to apply a series of functions to the input. Benefits of having multiple hidden layers are they can figure out complex functions by cascading simpler functions. The most commonly used hidden unit as the activation function is a Rectified Linear Unit (ReLU). Depth of the neural network states the total number of non visible layers present in the network.

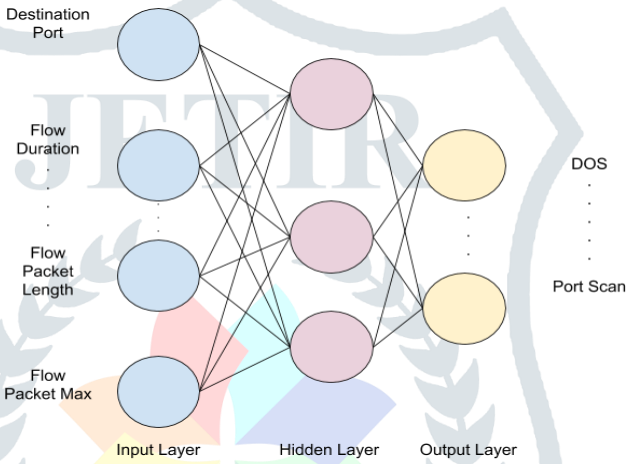

Figure 2: Sequential Model

IV. RESULTS

In our project we have been successful in identifying the network intrusions and also classify them in 15 groups or classes ('BENIGN', 'FTP-Patator', 'SSH-Patator', 'DoS GoldenEye', 'DoS Hulk', 'DoS Slowhttptest', 'DoS slowloris', 'Heartbleed', 'Web Attack One', 'Web Attack Two', 'Web Attack Three', 'Infiltration', 'Bot', 'PortScan', 'DDoS') with a high accuracy rate. For the training we had used a dataset containing 615072 records and for testing 205025 records were used, out the 205025 records the model was efficiently able to identify the 203102 records correctly as intrusion but some classes were misinterpreted for another malicious class whereas the 8 samples were not correctly classified, hence the overall aim of identification of intrusion was achieved with high accuracy but achieving a specific type of intrusion correctly was less accurate but above average.

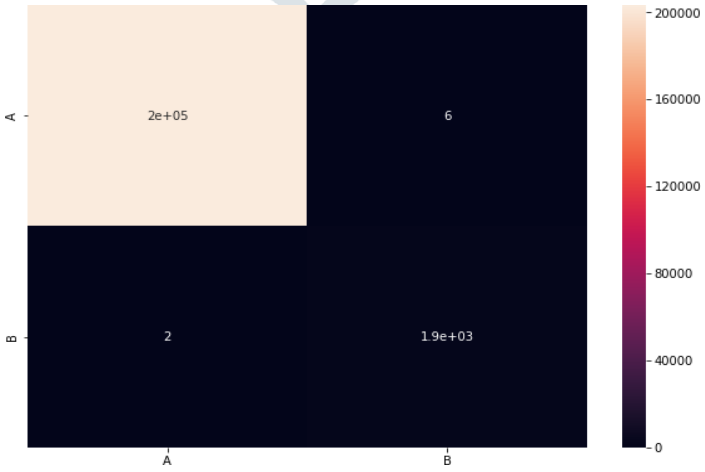

Figure 3: Confusion Matrix for Binary Classification

As we are classifying the intrusion in 15 groups as stated previously, we have built the confusion matrix for the multiclass classification ,as seen in figure 4 we have classes 0-14 ,the diagonal represents the number of records which are correctly classified and the non-diagonal values represent the misclassified values, one can clearly observe that in some places such as class 9 which is web attack type two, is misinterpreted for class 8 some times which is web attack one as these attack have very resemblance in their attack pattern ,similarly it can be observed for various DoS types such as DDoS, Dos slowloris, DoS Hulk, DoS GoldenEye and other DoS types may sometimes get misinterpreted one another, but the overall aim of identification of an intrusion is achieved vey correct as you can observe that almost all of benign records are correctly classified which very important in intrusion identification.

Figure 4: Confusion Matrix for Multiclass Classification

V. CONCLUSION

In this paper, we have researched, implemented and evaluated application of deep learning to anomaly-based NIDS. The results show that our system was able to detect the intrusion into 15 types and was very successful in identification of malicious records at very high accuracy of 94 percent, we have also observed that the various DoS types and Web Attack types were sometimes misinterpreted for each other due to very resemblance in their attack pattern, another observation was that many times the pattern of DDoS and Port Scan was found to be very similar as both use different port access for execution, but increase in the dataset for training model for identification of these types solves the issues at great extent. Hence, we can conclude that overall identification of malicious behavior of network is very accurately identified by our system.

VI. FUTURE SCOPE

In future we plan to implement the latest techniques of ensemble learning, and also try to implement the unsupervised learning model for intrusion identification in parallel with existing supervised model. This will improve the precision of intrusion identification highly. Also, we need to make system efficient enough that it is able to notify the network administrator about the identification of intrusion. Another aspect where we can improve our system is constantly generating the new network patterns and feeding to the neural network for continuous training.

REFERENCES

[1] Alex Shenfield, David Day, Aladdin Ayesh, 2018, April. Intelligent intrusion identification systems using artificial neural networks. In Information and Communication Technology Express (ICTE), 2018 The Korean Institute of Communications and Information Sciences (KICS)

[2] Nguyen Thanh Van, Tran Ngoc Thinh, Le Thanh Sach, 2017, July. An anomaly-based Network Intrusion Identification System using Deep Learning. In 2017 International Conference on System Science and Engineering (ICSSe) (pp. 210-214). IEEE.

[3] Peter Norvig, 2009, Artificial Intelligence: A modern approach.

[4] Geoffrey Hinton, Ruslan Salakhutdinov. Deep Boltzmann Machine.

[5] Y. Li, R. Ma and R. Jiao, 2015. A Hybrid Malicious Code Identification method based on Deep Learning. In International Journal of Security and Its Applications (IJSIA) (pp. 205-216).

[6] Md. Zahangir Alom, V. Bontupalli, and Tarek M. Taha, 2015. Intrusion Identification using DBN. In National Aerospace and Electronics Conference (NAECON), USA, 2015.

[7] M. Ahmed, A.N. Mahmood, J. Hu, 2015. A survey of network anomaly identification techniques. In Journal of Network and Computer Applications. (p. 13),2015.

[8] A B Kathole **,** "Optimization of Vehicular Adhoc Network Using Cloud Computing", 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing,IEEE.

[9] U. Fiore, F. Palmieri, A. Castiglione, A. Santis, 2013. Network anomaly identification with the RBM. In Neurocomputing, Elsevier B.V, (p. 11),2013.

[10] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Purtogal, January 2018.