ICSSR Sponsored - International Conference on Need of Digital Transition and CSR For Innovative Sustenance in the 2023 Economy
© 2023 IJSRST | Volume 10 | Issue 11 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
International Journal of Scientific Research in Science and Technology

Network Security Intrusion Detection System - A Machine Learning Approach

Dr Julie M David, Amalendu Anilkumar, Mayoogha Ashok

ABSTRACT

An Network Security Intrusion Detection is a monitoring system that detect suspicious or malicious activities and issues in network. Here we use the Machine Learning methodology for Network Security Intrusion Detection. The main benefit of Network Security Intrusion Detection System, is the ability to detect and stop the malicious activities in the network, suspicious network traffic , issues and variety of attacks on the network. By using Machine Learning Approach on Network Security Intrusion Detection we can easily detect all various intrusions on the network, speedup the detection and increase the accuracy of detection.

I. INTRODUCTION

Information Technology is one of the most developing field. Computer Networks is an important part of IT. It is widely used in various industries such as business, media etc and it become a part of day to day life of people. For effective use of network we need a perfect networking system. With every innovation there will be problems which can cause the efficient working of it. Same as here, in networks can be affected by other than normal traffic, there can be suspicious network traffic, malicious activities on network and various attacks. Because of these problems it is difficult to build a reliable network. The mentioned problems are a threat to the network security. So, for the smooth working of the network it is essential to detect and remove the threats to the network security. It is difficult to detect all attacks or traffics that can affect network security. So network security intrusion detection have a vital role in detecting and removing the threats to the security of networks. By implementation of network security intrusion detection system, it will monitor the activities in the network and detect the intrusion as quick as possible and remove the intrusion and makes the network more reliable. Here we are using Machine Learning approach for network security intrusion detection.[7] Machine learning focuses on the development of computer programs that can access data and use it for self-study. [3] The algorithms are divided into several categories, including supervised , unsupervised ,reinforcement learning and deep learning . Machine Learning helps to develop the network security intrusion detection system easily and it make the system more efficient and accurate. This paper shows in section 2, gives some related works about the topic. In section 3, gives the methods of machine learning in detecting intrusion in network security and using the suitable dataset shows its efficiency of the system. In section 4, it shows the method of detecting intrusion and in section 5, shows the result of applying the method using its corresponding dataset and in section 6, the paper will conclude with the findings in speed, accuracy and efficiency of network security intrusion detection system.

II. RELATED WORKS

Yin Luo 's paper discussed about the support vector machine (SVM) algorithm parameters were improved by the adaptive particle swarm optimization (APSO) algorithm and the APSO-SVM algorithm, which obtains for intrusion detection.[5]The experiments carried out using KDD'99 CUP dataset and thee results showed that the proposed method greatly reduced the running time of the algorithm and improved the performance and machine learning method is effective on IDS, which contributes to the further realization of network security. The accuracy of the APSO-SVM algorithm was the highest which is reaching 97.687%.

Niveditha S Naganhalli and Dr Sujatha Terdal 's paper describes to protect the network from distributed denial of service and malicious traffic which gives good accuracy .The decision tree and Navie Bayes are the methodologies applied. As an important application of machine learning ,an accurate intrusion detection model is built by choosing an effective classification approach. The algorithms are tested using the KDD data-set. [6]Effective classifier is identified by comparing the performances based on the accuracy and confusion matrix.It can be concluded from the result that the C4.5 decision trees classifier performs better than the other classifiers for the considered data-set and parameters. It has the accuracy of 99%.

Sayed Atir Razza, and Sania Shamim 's paper is a review paper which describes to ensure the confidentiality ,integrity and availability of network and network security .[10]The IDS is quick and effectively by means of machine learning methods, reviewing machine learning algorithms that can be used to detect network anomalies, to check which dataset(Darpa98, KDD99, CAIDA,NSL-KDD, ISCX 2012, CICISD2017) will be best enough by comparing other datasets. the anomaly-base detection method is being used intensively to detect and prevent network attacks. The paper concludes with that the Random forest and ID3 is the much better approach in network anomaly detection using machine learning as they give good efficiency, less error rate and it can overcome over-fitting issues.

Gulshan Kumar and Hamed Alquhtani's paper describes to enable flexibility in developing tools that can effectively analyze and detect malicious traffic and attacks in network. [9]The antiintrusion techniques can be divided into six categories; namely, intrusion prevention, intrusion detection, intrusion pre-emption, intrusion deterrence, intrusion deflection and intrusion countermeasure. It is the comprehensive review of ML techniques for detecting intrusion detection in SDN. The paper concluded that the application of machine learning techniques in detecting intrusion in SDN faces many challenges , so the development of ML-based intrusion detection in the SDN give better output.

Md Nasimuzzaman Chowdhury and Ken Ferens, Mike Ferens's paper discuss with the importance of implementing ML in IDS for effective analysis tool to detect any anomalous events occurred in the network traffic flow. [8]overall efficiency of the proposed method is dignified by evaluating the detection accuracy, false positive rate, false negative rate and time taken to detect the intrusion. The paper conclude with that detecting the intrusion gives 98.76% which is higher detection accuracy and lower false positive rate of 0.09% and false negative rate of 1.15%, and the 88.03% which is normal SVM based scheme achieved a detection accuracy and false positive rate of 4.2% and false negative rate of 7.77%.

III. METHODOLOGY

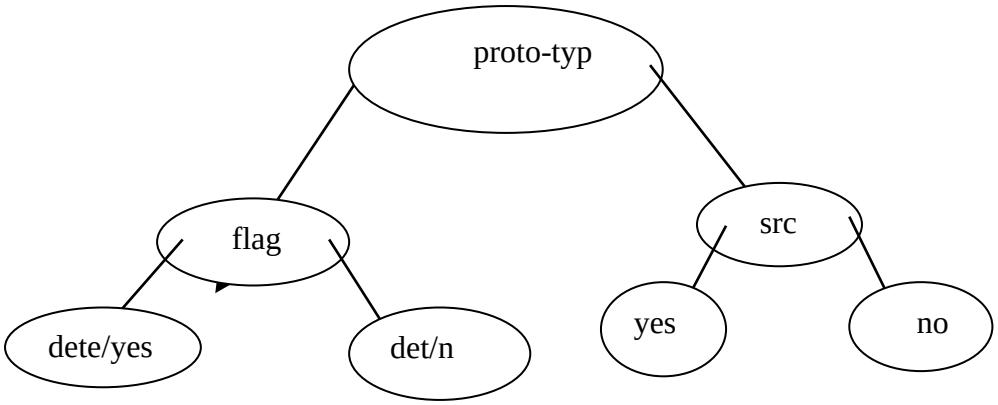
The decision tree approach in Machine Learning is used to build a decision tree to check the malicious traffic and attacks on the network depending on an available dataset to empower it to group the new cases accurately

and efficiently.[2]Diverse Machine learning methods have been successfully deployed to address such wide ranging problems in computer security. [4]Machine learning algorithm helps to study the high dimensional network traffic and identify abnormal flow in traffic with high accuracy. A decision tree is a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered. The decision tree model is trained and tested on a set of data that contains the desired categorization. A decision tree is defined as a tree-like diagram that contains a tree trunk representing the internal nodes to express a test of a particular characteristic, the branches representing the result of the test, and the leaves representing the nodes as a class mark. The purpose of classification in the approach of the decision tree is to frame the data so that it contains both the Decision node or Root node and Leaf node. Decision nodes are used to make any decision and may have further multiple branches. Leaf nodes are the output of those decisions that we taken and it does not contain any further branches. The decision tree is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions**.** There are many approaches available for the construction of the decision tree, the CRT (Classification and Regression Trees) is of good and simple method. Here to build a tree we make use of CART for our Intrusion Detection System. The CART algorithm is a combination of classification tree algorithm and regression tree algorithm. In classification tree algorithm the target variable is always fixed or categorical, moreover it is a Yes/No types. The algorithm is then used to identify the "class" within which a target variable would presumably fall into. A regression tree refers to an algorithm where the target variable is and therefore the algorithm is employed to predict the value of it and also it is a continuous data types. The experiments carried out using the KDDCUP99 dataset. The competition task was to build a network intrusion detector, a predictive model capable of distinguishing between "bad'' connections, called intrusions or attacks, and good'' normal connections. This data set is capable of giving the best results. Decision trees can examine information and recognize critical qualities in the system .This in turn increases the value of the some security frameworks by checking the arrangement of intrusion identification information. It can perceive patterns and examples that promote to check, the advancement of attack signatures and different activities of checking. Distinguishes the decision tree from other methods that the decision tree gives the rich arrangement of rules.

IV. DETECTION USING DECISION TREE

The purpose of Decision tree approach of machine learning is to build a decision tree to validate the incoming traffic depending on an available data set to empower it to group the new cases accurately. The data set is given as the input to the working model. The working model of detection using decision tree consists of three stages: the preprocessing stage, normalization stage and decision tree building. In preprocessing stage the dataset deals with the strings and numbers. The input dataset is usually consisting of strings and number. As the value of string cannot compared directly, for this we are required to digitize the string by making use of string manipulation operation. This is done by using the preprocessing stage. The processed data set may does not contain any uniformity. There may be smaller number of column sections that may play critical role. Therefore, its needed to perform the normalization of processed data before it is given to detection algorithm. In decision tree mode , we build the decision tree by making use of the training dataset and then obtain the output of the checking dataset.Decision tree mode is obtained by making use of the function by applying the Classification and Regression trees methodology.

V. GRAPH



VI. RESULT ANALYSIS

ACCURACY	PRECISION	RECALL	F-SCORE	F PR	T PR
0.957	0.863	0.960	0.908	0.042	0.944

The experimental analysis is done by evaluating our intrusion detection system on premier data set KDDCUP99.The implementation is done by using python. This research work requires an extensive number of legitimate test information. Information accumulation can be gotten through some capturing devices and we can check the Accuracy, Precision, Recall, F-Score, FPR and TPR. In this Investigation, we utilize KDDCUP99 dataset for our test. To ensure that all types of attacks are covered, the dataset is randomly divided based percentage as training and test data sets. This makes best for our Intrusion detection system. And the result of these form are as follows, Accuracy may be of correct predictions and incorrect predictions. Whether we want to know about the correct or incorrect predictions we choose the confusion matrix for it. Confusion matrix helps us to display the performance of a model or how has made its prediction in Machine learning. Then we get the accurate value or the correct value. And the corrected value or the accurate value of the dataset KDDCUP99 full dataset is 0.957. Precision and Recall are performances metrics used for pattern recognition and classification in machine learning. Precision helps us to visualize the reliability of the machine learning model in classifying the model as positive. The precision is obtained by dividing true positives to a total number of classified positive samples it may either correctly predicted or incorrectly predicted. Ie

Precision P = TP/TP+FP

TP – True Positive

FP –False Positive

Recall is the Total Positive Rate (TPR) it may have the ability to detect the positive samples. Recall is also known as sensitivity. The fraction example classified as positive, among the total number of positive examples. Precision and Recall is always used to calculate only for getting the positive values. Unlike the Precision, Recall is independent of the number negative samples classifications .The Precision of the KDDCUP99 full dataset is 0.863 and also the Recall of the KDDCUP99 full dataset is 0.960.

ie,

Recall = TP/TP+ FN

FN – False Negative

The F-Score is also known as f1-score matrix which is used to calculate or evaluate the performance of machine learning model and model accuracy in dataset. It is way of combining precision and recall. Here the F-Score of the KDDCUP99 full dataset is 0.908.

ie,

F1 Score = (2 × Precision× Recall)/(Precision + Recall)

FPR or also called it as "FALL OUT" is the number of false positive rates in an accuracy metric that can be measured on subset of machine learning models.

ie ,

FPR = FP/(FP + TN)

TN – True Negative

Here the false positive rate is measured as 0.042.TPR is the number of true positive values get by the machine learning model.ie, the value of this is 0.944 for the given KDDCUP99 fully dataset.

VII. CONCLUSION

There may be intrusions or attacks generally present in the network. The decision tree is employed to help the system administrator to conclude about approaching traffic and attacks.That is the regardless of whether the incoming information is malicious or not by building a model that isolates noxious and non-vindictive traffic. In this work here we built a system based on the decision tree, different strategies are also contrasted with this approach. [1]Both 20% dataset as well as the full data set is tried, and the experiment results demonstrate that our framework is sufficiently powerful. In the future, we will take part in the investigation of the IDS for different sorts of attacks.

VIII. REFERENCES

[1]. Shilpashree S, S C Lingareddy, Nayana G Bhat, Sunil Kumar G, Decision Tree: Machine Learning For Intrusion Detection, International Journal Of Innovative Technology And Exploring Engineering

[2]. Vitaly Ford and Ambareen Siraj, Applications of Machine Learning in Cyber Security, Computer Science Department, Tennessee Tech University Cookeville, TN, 38505, USA

[3]. Rami Ahmad 1,2,* , Raniyah Wazirali 3,* and Tarik Abu-Ain, Machine Learning for Wireless Sensor Networks Security: An Overview of Challenges and Issues, Institute of Networked and Embedded Systems, University of Klagenfurt, 9020 Klagenfurt, Austria ,Ubiquitous Sensing Systems Lab, University of Klagenfurt-Silicon Austria Labs, 9020 Klagenfurt, Austria , College of Computing and Informatics, Saudi Electronic University, Riyadh 11673, Saudi Arabia.

[4]. Prachi, Heena Malhotra, Prabha Sharma , Intrusion Detection using Machine Learning and Feature Selection, The NorthCap University, Gurgaon, The NorthCap University, Gurgaon, India.

[5]. Yin Luo, Research on Network Security Intrusion Detection System Based on Machine Learning, Sichuan TOP IT Vocational Institute, China No. 2000, Xiqu Avenue, High-tech District, Chengdu, Sichuan 611743, China.

[6]. Nivedita S Naganhalli, Dr Sujata Terda, Network Intrusion Detection Using Supervised Machine Learning Technique, International Journal Of Scientific & Technology Research Volume 8, Issue 09, September 2019.

[7]. Surya Lakshmisri, Department of Information Technology, Machine Learning On Network Security, International Engineering Journal For Research & Development

[8]. Md Nasimuzzaman Chowdhury and Ken Ferens, Mike Ferens1, Network Intrusion Detection Using Machine Learning, Department of Electrical and Computer Engineering University of Manitoba Winnipeg, Manitoba, Canada 1 Gourdie-Fraser, Inc., Traverse City, Michigan, US.

[9]. Gulshan Kumar1,* and Hamed Alqahtani, Machine Learning Techniques for Intrusion Detection Systems in SDN-Recent Advances, Challenges and Future Directions, Shaheed Bhagat Singh State University, Firozpur, 152024, India 2 King Khalid University, Abha, 61421, Saudi Arabia.

[10]. Syed Atir Raza and Sania Shamim, Network Anomaly Detection Using Machine Learning | A Review Paper, ST department University of management and technology, Lahore Pakitan