INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND APPLICATIONS
VOLUME XVI, ISSUE VI, JUNE 22, WWW.IJCEA.COM, ISSN 2321-3469

NETWORK INTRUSION DETECTION SYSTEM USING MACHINE LEARNING

Prof. Vrushali V. Kondhalkar, Swapnil Shende, Abhishek Patwari, Pranav Ovhal
Dept. of Computer Engineering, JSCOE, Pune, Maharashtra, India

ABSTRACT

“Network Intrusion Detection System Based on Machine Learning Algorithms” is a software that monitors network of computers for malicious activities that are aimed at stealing sensitive confidential information or corrupting /hacking network protocols. Techniques used in Today’s NIDS are not able to deal with the Dynamic Complex types of security Cyber Attacks on Computer Networks. Performance of an intrusion Detection is mainly depending on accuracy. Accuracy for Intrusion detection must able to decrease false alarms and to increase the detection rate of alarms. To improve the performance, different techniques have been used in recent works. Analyzing huge network traffic data is the main work of intrusion detection system. A well-organized classification methodology is required to overcome this problem. NSL-KDD knowledge discovery data set is used, their accuracy and misclassification rate get calculated.

***Keywords:* NIDS, Machine Learning, KDD Dataset.**

1.INTRODUCTION

“Network Intrusion Detection System” is detecting abnormal malicious activities in the network or system by intruders using “machine learning Technique” . In today’s world one of the most severe threat to computer security or network security is the illegal intrusion into a computer system. As the network applications are growing rapidly, new sort of network attacks are rising continuously. For controlling suspicious activities our system should be prepared. Once an attack is identified or abnormal behavior is observed, the message can be displayed so that the one who is using the system can be notified about the intrusion. Network Intrusion detection systems (NIDS) take network based approach for recognizing and deflecting attacks. In either case, these products look for attack signatures that usually indicate malicious or suspicious intent.

When an IDS looks for these patterns in network traffic then it is network based. When an IDS looks for attack signatures in log files, then it is host based. Various algorithms have been developed to identify various types of network intrusions; however there is no heuristic to confirm the accuracy of their results. The exact effectiveness of a network intrusion detection system's ability to identify malicious sources cannot be reported unless a concise measurement of performance is available. This study defines the behavior pattern of machine learning for identifying intruders. This is very useful for preventing intrusions according to the associated individual type of attack. This model can identify the intrusion and will be able to display the message whether it is intrusion or not. In case of intrusion the message "anomaly" will be displayed or else "normal" will be displayed.

2. Existing Work

With today's world trending towards being reliant on computers and automation, it is very challenging for people to make networks and systems secure for everyday use. The number of security threats to organizations are increasing in greater amount with the growth of online market and their services. There are numerous solutions to fight network security threats. Network intrusion detection system are placed along firewalls in networks to fight security threats. They scan the network for all the incoming and outgoing traffics and analyze the packets to detect whether they are malicious are not. Machine learning algorithms helps the systems to learn the signature of known attacks. Some of the best intrusion detection system on the market are:-

1. Snort
2. Bro
3. Suricata.

Most of the intrusion detection system only detects the attacks and make the intruders being known to the systems such are called as passive intrusion detection system and on the other hand there are active and reactive type of network intrusion detection system.

3. Proposed System

"To identify all abnormal patterns and traffic using monitoring, detecting and responding to unauthorized activities within the system". However, regarding its huge and unbalanced dataset, NIDS encounters total data processing problem.

There are so many intrusion detection system techniques generates a huge number of alerts where most of them are real and some of are not (i.e., false alert) or are redundant alerts. The false alerts create a serious problem to intrusion detection systems. Alerts are defined based on source/destination IP and source/destination Ports. However, one cannot know which of those IP/ports bring a threat to the network. It is difficult for the security analyst to identify attacks and take remedial action for this threat. So it is necessary to assist in categorizing the degree of threat, by using Machine learning technique.

4. System Architecture

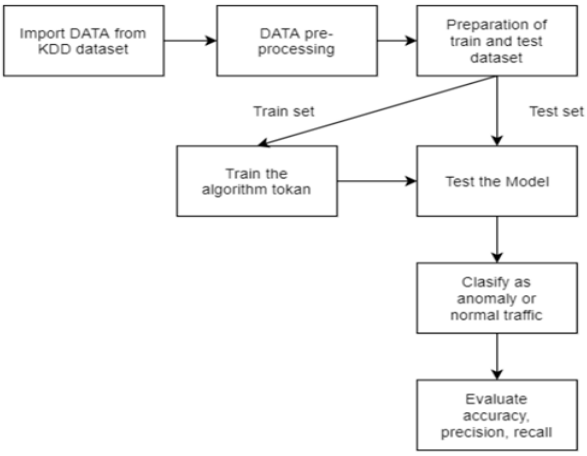

Import DATA from KDD dataset
DATA pre-processing
Preparation of train and test dataset
Train set
Test set
Train the algorithm tokan
Test the Model
Clasify as anomaly or normal traffic
Evaluate accuracy, precision, recall

Figure 4.1: System Architecture

As you can see in fig 4.1, there is the step to import data from KDD dataset. This is the first step to begin with. After we have imported the data from KDD dataset we have to perform pre processing on the dataset. In this step we remove the unwanted features from the data set like number of outbouds. This are the outgoing commands from the system and doesn't affect the intrusion that's why they are not needed. After pre processing next comes scaling and encoding to scale down the numerical and categorical features into same so that they can be processed further.

Now we have data sets splitted into two categories training and testing. First we train the model using machine learning algorithms using training data set and we compares it's output with test dataset's output. So after the training of model is done we give a unknown data as an input and the model classifies if it is anomaly or normal. As a result we get the accuracy of all the algorithms and we can decide which algorithm is best suited based on its accuracy percentage.

5. Algorithms

K-nearest neighbour:

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

The following function is used to train KNN Classifier model.

```
KNN_Classifier = KNeighborsClassifier(n_jobs=-1)
KNN_Classifier.fit(X_train,Y_train);
```

Decision Tree:

Decision Tree algorithm is one of the most popular methods of Classification and Regression. But it mostly uses for Classification method. The decision tree is a graph-like Structure where internal nodes represent features of dataset, branches represent decision rule and each leaf node represent results.

In our System, the working Decision tree Algorithm will be we gives training data to decision tree classifier to train the model. Based on model our system will train and then we supplied testing data as an input and accuracy of model is calculated. From the result its seen as accuracy of training data is 100% and Testing Accuracy is 99.47%. As compared to other algorithm, the results show that Decision tree algorithm gives an overall better result as compared to KNN.

The following function is used to train decision tree model

```
DTC_Classifier = tree.DecisionTreeClassifier(criterion='entropy', random_state=0)
DTC_Classifier.fit(X_train, Y_train)
```

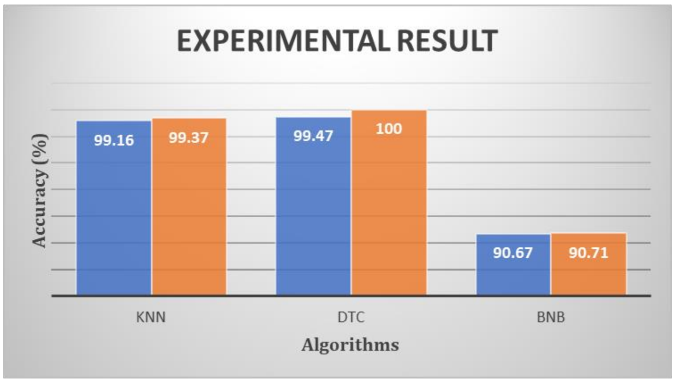
Naïve Bayes Classifier:

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Bayes Theorem:-

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

6. Experimental Result



From the above chart, there are accuracies of different type of Machine Learning algorithms that we have used in our model. We have implemented the decision tree, KNN algorithms and Naïve Bayes Classifier. Testing accuracy is shown in Blue Grid whereas training accuracy shown in orange grid. From experiment result chart we observe that the testing accuracy of KNN , Decision tree and Naïve Bayes Classifier algorithm is 99.16% , 99.47% , 90.67% respectively.

7. Applications:

1) Anomaly Detection

2) Network Traffic Processing

3) Thread Reporting

4) Thread Classification

5) Prevention System

8. Conclusions

We have successfully implemented Network Intrusion Detection System using machine learning algorithms such as Decision tree classifier, K-Nearest Neighbour and Naïve Bayes Classifier. In this project we have observed that Decision tree classifier gives the best accuracy among all. But it is difficult to develop an Network intrusion detection system with 100 percent success.

This system can only detect the intrusion and it will display the message as “anomaly” in that case but it is unable to remove the intruder from the system. This system can work only for standard dataset. It will not be able to work in real time application.

9. Future Scope

A more challenging problem being addressed right now is analysis and correlation. No matter how good the NIDS analyst, Identifying anomalies are difficult to detect, especially on large networks. So NIDS they examine strange packets and look to group them using sophisticated statistical analysis.

It’s much easy to discover and categorize patterns when you have all the relevant data in front of you, without the noise that generally follows. Furthermore, This NIDS similar to these will be used to fine tune filters and rules in order to reduce false positives, over time providing a kind of NIDS feedback system, based on administrator input and response.

REFERENCES

[1] Hurley, T.; Perdomo, J.E.; Perez-Pons, HMM-Based Intrusion Detection System for Software Defined Networking. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning

[2] Dey, S.K.; Rahman, M.M. Effects of Machine Learning Approach in Flow-Based Anomaly Detection on Software-Defined Networking. Symmetry 2020.

[3] Shone, N.; Ngoc, T.N.; Phai, V.D.; Shi, Q. A Deep Learning Approach to Network Intrusion Detection. IEEE Trans. Emerg. Top.Comput. Intell. 2018.

[4] Y. Liao and V. R. Vemuri, "Use of k-nearest neighbor classifier for intrusion detection," Computers & Security, vol. 21, no. 5, pp. 439–448, 2002.

[5] Ngo, D.M.; Pham-Quoc, C.; Thinh, T.N. Heterogeneous Hardware-based Network Intrusion Detection System with Multiple Approaches for SDN. Mob. Netw. Appl. 2020,

a. 25, 1178–1192.

[6] Jeng-Shyang, PAN, Yu-Long, QIA, Sheng-He SUN. A fast K nearest neighbors classification algorithm. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences. 2004; 87(4):961−3.

[7] Complete reference by www.geeksforgeeks.org

[8] www.tutorialspoint.com