DDoS Attack Prediction System

Dr. R. Manikandan[1], S. Dharshini[2], N. Pavithra[3]

Computer Science and Engineering, Government College of Engineering, Thanjavur

Abstract: Distributed Denial of Service attack (DDoS) is one of the most dangerous attack in the field of network security. It is an attack which coordinated stream of request is launched against from many locations at a same time. These attacks are increasing day by day and have become more sophisticated. So, it has become difficult to detect these attacks and secure online services. Thus Prediction take place with Random forest algorithm with accuracy of 98%.

I. INTRODUCTION

DDoS attack is crucial for network security. DDoS attacks are a type of cyberattack that can cause significant damage to businesses and organizations, resulting in financial losses, reputational damage, and legal liabilities. These attacks involve overwhelming a network or a website with a massive amount of traffic, making it impossible for legitimate users to access the network or website. Predicting DDoS attacks is essential for ensuring business continuity and maintaining the trust of customers and stakeholders. With the increasing frequency and complexity of cyber attacks, predictive models that can effectively detect and prevent DDoS attacks are essential for organizations to protect their assets and maintain their operations. These attacks are increasing day by day and have become more sophisticated. So, it has become difficult to detect these attacks and secure online services. The system presents a technique for predicting the DDoS attack using Random Forest algorithm on the UNSW_NB_15 dataset. The UNSW_NB_15 dataset is a publicly available dataset that contains different types of network traffic, including normal and attack traffic. The proposed model uses a feature selection technique to select relevant features from the dataset, and these features are used to train the Random Forest model. The model's effectiveness is evaluated using different performance metrics, including accuracy, precision and recall. This study contributes to the development of effective DDoS attack prediction models and highlights the effectiveness of the Random Forest algorithm in predicting such attack.

II. RELATED WORKS

Arun Nagaraja et al.,[1] proposed a hybrid model deep learning model for intrusion detection. They combined two deep learning models for the classification of CNN+ LSTM from the RNN model. The dataset was used in this work is KDD. They found an 85.14% average accuracy for the proposed. They used k mean cluster model for feature similarity detection and naïve Bayes model used for classification.

Ashutosh Nath Rimal and Dr. Raja Praveen [2] have proposed a DDoS attack detection system using ML algorithms and packet analysis in a smart way. In this case the classification

algorithm being used are Naïve Bayes and SVM algorithm out of which SVM was found to give maximum accuracy. This system has achieved with a maximum accuracy of 99.68%, if

the recommended combination of feature selection and classification algorithm is chosen. The user is left with the choice for both feature selection and classification algorithm.

Bakker, J. et [3] This paper has shown how statistical classification can be deployed using SDN to detect DDoS attacks. Three classifiers were selected in a off-line environment to be integrated with nmeta2. These were then evaluated on a physical network testbed by replaying a DDoS scenario. While statistical classification can be deployed using SDN to classify traffic, careful consideration must be made to pick classifiers that result in the smaller possible packet processing overhead.

Guiomar et al [4] has implemented a Network Intrusion Detection System (NIDS) using OPNET simulation. This was based on misuse detection and network traffic was imported using an ACE module into OPNET. A NMAP port scanner was used to simulate a flood attack, and the proposed IDS was tested in a controlled environment; the result was

satisfactory with around 93% accuracy rate.

Jing Wu, et al [5] have suggested a method for detecting DDoS Attacks in Software-Defined Networks through Feature Selection Methods and Machine Learning Models. In this study, The SDN-based detection systems developed for DDoS attacks were analyzed by using machine learning systems. In the first proposed approach, by analyzing flow data, algorithms with 98.3% accuracy ensure the detection of attacks without discriminating the type of traffic.With 97.7% sensitivity, KNN algorithms can perform this control by facilitating the charge of the controller.

Khuphiran, P et al [6] the application of machine learning algorithm for the problem of DDoS attack detection has been addressed. Two algorithms, Support Vector Machine (SVM) and Deep Feed Forward (DFF) have been evaluated to demonstrate the feasibility of applying these algorithms. The experiments have been conducted to compare the performance of these algorithms. It has been found that DFF can classify the data with a higher accuracy.

Kimmi kumari and M.Mrunalini [7] have proposed a system for detecting DDoS attack by using various machine learning algorithms. The major goal of this paper's work is to distinguish between normal and attacks scenarios by analyzing the throughput of the data packets respectively.

Larriva-Novo et al. [8] proposed two benchmark datasets, especially UGR16 and UNSWNB15, and the most used dataset KDD99 were used for evaluation. The pre-processing strategy is evaluated based on scalar and standardization capabilities. These pre-processing models are applied through various attribute arrangements. These attributes depend on the classification of the four sets of highlights: basic associated highlights, content quality, fact attributes, and finally the creation of highlights based on traffic and traffic quality based on associated titles Collection. The goal of this inspection is to evaluate this arrangement by using different information pre-processing methods to obtain the most accurate model.

Liu et al [9] have suggested a method for detecting DDoS attack using deep learning techniques. To increase the validity and effectiveness of feature extraction, a convolutional neural network (CNN) modeling approach for intrusion detection was applied. The convolution kernel was chosen and convolved with the data to extract local correlation. The new approach can raise classification accuracy for jobs involving intrusion detection and

recognition.

Maede Zolanvari et al.[10] proposed a recurrent neural network model for classification intrusion detection. They compared other deep learning models with RNN. Finally, they

found RNN is the best model for intrusion detection by using the KDD dataset. It was a multiple classification problem. They used advanced deep learning LSTM for multiple classification problems. They found good results with 89% average accuracy for the proposed work.

III.PROPOSED MODEL

In this research, we design a web page for predicting the DDoS attack by extracting some important features in dataset.

The proposed system are expected to meet the following requirements:
1) Random forest can overcome the issue faced in existing system by combining multiple decision trees and using voting system to classify network traffic.
2) Random forest is a decision tree-based ensemble method that can handle large dataset with high dimensionality and noise. It can also handle missing values and outliers, and is less sensitive to hyper parameters than SVM.
3) So, this system using the random forest algorithm for predicting the ddos attack by extracting some important features in the dataset.

The main contribution is to generate the best model for data utilization, as well as, model optimization; and which performs best for model learning. After getting the results, we performed performance measures in terms of precision, recall, and f1 score. In this research work, we used a well known supervised learning models which is Random Forest Classifer.

IV.RESULTS AND DISCUSSION

This section contains all the obtained results of our proposed models. All the results are shown step by step in the form of figures, as well as, results explanation. We briefly describe and evaluate the performance of our suggested model with several closest rivals and existing research studies.

A. Dataset

A Dataset is a collection of numbers or values that relate to a particular subject. In this project the UNSW_NB_15 dataset is used, that contains the different features about ddos attack including ID number, protocol, duration, attack_cat which represents the severity of DDoS attack.
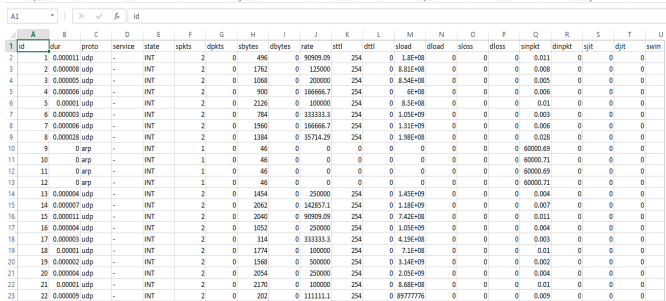
Fig. 1 UNSW_NB_15 dataset

B. Language And Tools

Python language is considered a suitable programming language both for simulations and real-world programming.

It is considered the most powerful high level language for model learning . We used a jupyter notebook as a tool. This tool is open-source and browser-based which has evolved to become a robust tool for researchers to share documentation and code. This tool functions as a virtual lab notebook.

C. Data Preprocessing

It is very important and time to clean the information from irrelevant data and convert it to quality in statistical techniques to clean data and replace those values which are not important for our experimental analysis. This module consists of two steps. First one is to check whether the missing value is present or not. And the second one is to replace into corresponding values such that replace HTTP. Because the DNS request are generally very small and fit well within UDP protocol and the HTTP requests are work using TCP because across the network accurately and in the proper sequence.


D. Data Visualization

The present of data where the information will understandable in the form of image or diagram. It is important to understand easily the information. This is the initial step, where we are selecting our target for the proposed algorithm. Also, this step is used for selecting the test class. This step is very important to understand data in a much better way. Through this method is able to select our target class for classification. The visualization of showed total number of Normal = 37,000, Generic = 18871, Exploits = 11,132, Fuzzers = 6,062, DoS = 4,089, Reconnaissance = 3,496, Analysis = 677, Backdoor = 583, Shellcode =378, and Worms = 44 attacks
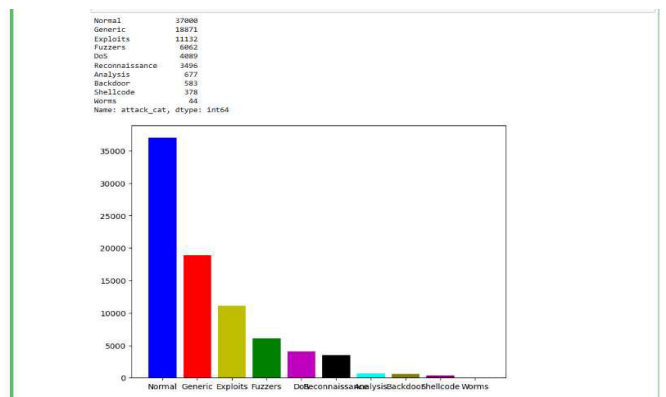
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x62497f0278>
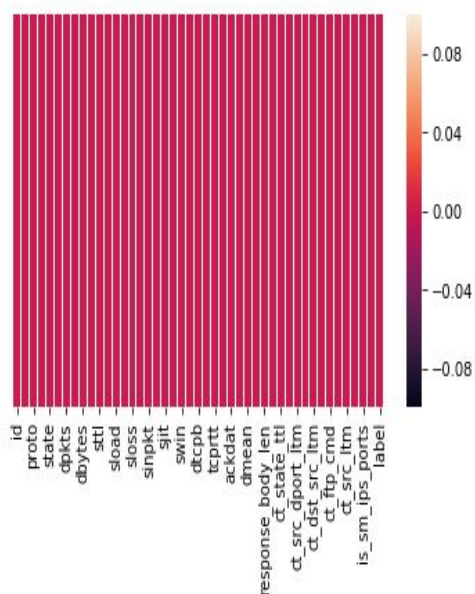
E. Train The Model

In this module, train the training dataset using random forest algorithm. It consists of two steps they are classification results and confusion matrix. In classification results it create combination of decision tree by selecting those features to predict the ddos attack. In confusion matrix denotes the overall number of actual and predicted labels for particular algorithm. It is used to calculate the accuracy of representation just like arranging true and prescient marks.

F. Random Forest Model

The Random forest, which is one of the most popular and powerful machine learning classification algorithm.

It can apply for both classification and regression problems.

It is based on ensemble learning, which integrates multiple classifiers to solve a complex issue and increases the model's performance.

```python
#prediction
predictions=model.predict(X_test)
print(predictions)
print(accuracy_score(y_test,predictions)*100)
print("Confusion Matrix:",confusion_matrix(y_test,predictions))
print("Precison :",precision_score(y_test,predictions,average='macro'))
print("Recall :",recall_score(y_test,predictions,average='macro'))
print("F1 Score :",f1_score(y_test,predictions,average='macro'))
```

```
['Normal' 'Normal' 'Normal' ... 'Normal' 'Normal' 'Normal']
97.7914334387929
Confusion Matrix: [[ 1445   283]
 [   80 14628]]
Precison : 0.964280853281382
Recall : 0.9153938175495322
F1 Score : 0.9380775255520706
```

V. CONCLUSIONS

The system predict the ddos attack using random forest algorithm. In this system it extract some important features from the UNSW_NB_!5 dataset for predicting the ddos attack. Based on the analysis, it can be concluded the random forest algorithm is an effective approach for predicting ddos attack. The model achieved a high accuracy rate indicating that it can be effectively distinguish between normal traffic and malicious traffic. The accuracy achieved is 97%.

REFERENCES

[1]   Arun Nagaraja et al.,[1] "Hybrid model deep learning model for intrusion detection".

[2]   Ashutosh Nath Rimal and Dr.Raja Praveen., "DDoS attack Detection System" using SVM and Naïve Bayes.

[3]   Bakker, J. et "DDoS Attack Detection System" using Software Defined Networking (SDN).

[4]   Guiomar et al., "Network Intrusion Detection System (NIDS)" using OPNET simulation.

[5]   Jing Wu, et al., "DDoS Attack Detection System" using KNN algorithm.

[6]   Khuphiran, P et al., "DDoS Attack Detection System" using Support Vector Machine (SVM) and Deep Feed Forward (DFF).

[7]   Kimmi kumari and M.Mrunalini., "DDoS Attack Detection System" using various machine learning algorithm.

[8]   Larriva-Novo et al., "proposed two benchmark datasets, especially UGR16 and UNSWNB15, and the most used dataset KDD99 were used for evaluation".

[9]   Liu et al., "DDoS Attack Detection System in deep learning" using CNN algorithm.

[10] Maede Zolanvari et al., "recurrent neural network model for classification intrusion detection".

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)