Detection of DDoS attacks using Machine Learning

Ms. J Anusha1, Mr.B. Akhil Krishna2 ,U. Mr. Munindhar3, Ms.K Vaishnavi4, Ms.B.Srinidhi5

1. Assistant Professor of ECE Dept. & (2,3,4,5) Students,

Department of Electronics & Communication Engineering, KITS(S), Huzurabad, Karimangar, T.S..

### Mail ID: j.anushaece@gmail.com

## ABSTRACT

Cyber-attacks by sending large data packets that deplete computer network service resources by using multiple computers when attacking are called Distributed Denial of Service (DDoS) attacks. Total Data Packet and important information in the form of log files sent by the attacker can be observed and captured through the port mirroring of the computer network service. The classification system is required to distinguish network traffic into two conditions, first normal condition, and second attack condition.

In this project, we have used various machine learning algorithms such as support vector machine (SVM), Random Forest, XGBOOST, ADABOOST, KNN and Naïve Bayes. The random forest model resulted in superior performance compared to conventional algorithms. To train these algorithms, we have used Canadian Institute for Cyber security (CIC) dataset which contains 10 different attacks of IOT environment and benign (normal) class. The Random Forest classification is one of the methods that can be used to process numeric attribute as input and determine two decisions of access that occur on the computer network service that is " normal " access or access under " attack " by DDoS as output.

The aim of the paper consists of a machine learning techniques to train a model which can be used to detect and classify the type of DDoS attackwith greater accuracy than that of each individual machine learning technique used.

**Keywords:** DDoS Attack, IOT Environment, SDN, Forest Fire, LoRa.

## I.INTRODUCTION

Under the innovative architecture environment of SDN, deep packet analysis available through the full network view. It supports quick response and update of traffic policies andrules. The SDN has the capability of perceived control of the global visualization view, lexible and schedulable rapid deployment capability, and service open intelligent scheduling capability. While ensuring network services and reducing deployment costs, the software defined network enhances the quality of user experience and facilitates the promotion of the whole network deployment. Researchers aimed at traditional network architecture proposed a lot of DDoS attack detection methods. Software-Defined Networking (SDN) introduces an innovative architecture to decouple the control and data plane, which otherwise are intermingled in traditional networks. Basically, SDN divides a network into three layers: application layer, control layer and data layer. SDN switches in the data plane are deprived of the ability of thinking and are managed by a centralized controller in the control plane. The advantage of this revolution is obvious as there is an ease of management. However, the controller can easily become a single point of failure. That is, once the controller is down, all the SDN switches attached might stop working because they lose connection to the controller. The link between the control and data plane is defined by the Open Flow protocol, which consults with the controller about the decision regarding how to process certain packets.

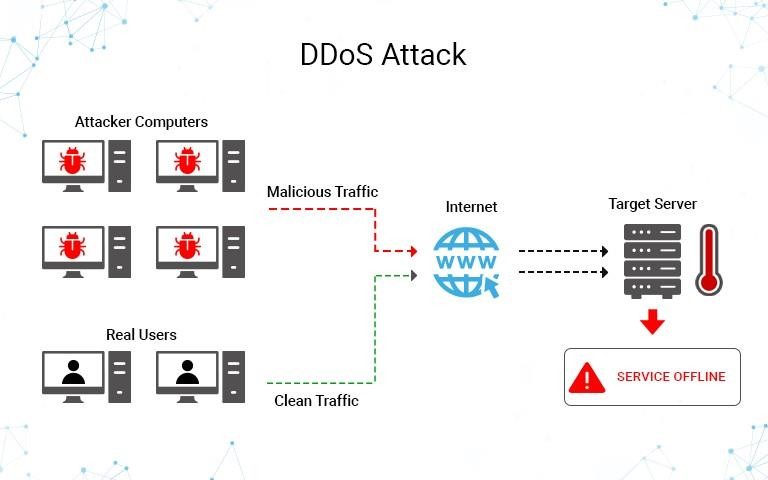


Fig: 1 DDos Attack

IoT enables machine-to-machine communications and data exchange to broaden the range of coverage area. Through sensors identification and ubiquitous computing, IoT tries to involve a diverse range of devices and merge different networks altogether to achieve localization, monitor, management, etc. In outdoor deployments, the number of devices under the network has

an exponential growth due to a high demand for environment monitoring and data collection. Recently, new technologies—such as LoRa, DASH7 and Narrowband (NB-IoT)—promise to provide low-power and long-range connectivity solutions for IoT applications, and they also meet the key requirements of low cost, long battery life, extended coverage area, support for a massive number of connected devices (scalability), security and privacy. However, the massive number of IoT devices connected to the same network increases the attack vector, which raises new security issues for IoT networks.

## II.PROBLEM OVERVIEW

With the continuous development of network technology, the ceaseless expansion of network business needs, and rapid growth of the Internet economy in the Internet age, the services of network with important business and industry information have been spread tothe production and life of current society. The emergence of DDoS attacks can lead to abnormalities in the related network services, causing huge economic losses and even causing other catastrophic consequences. DDoS attacks are one of the serious network security threats facing the Internet. It is a key research topic in the security field to detect DDoS attacks accurately and quickly. SDN is an emerging network innovation architecturethat separates the network data plane and the control plane, which has the characteristics ofnetwork programmable, centralized management control, and interface opening. Network attackers attack network bandwidth, system resources, and application resources, to achievethe effect of denial of service attacks.

The difficulties of DDoS attack detection are as follows:

1. The attack traffic characteristicsnot being easy to identify;
2. The lack of collaboration between the coherent network nodes;
3. The change of the attack tool being strengthened, with the threshold of its use decreasing; (4)The widely used address fraud making it difficult to trace the source of the attack; (5)The duration time of attack being short and response time being limited.

In the traditional network architecture, the main methods of DDoS attack detection technology can be divided into attack detection based on traffic characteristics and attack detection based on traffic anomaly. The former mainly collects all kinds of characteristics information related to the attack and establishes a characteristics database of DDoS attack. By comparing and analyzing the data information of the current network data packet and characteristics database, we can judge whether it is attacked by DDoS or not. The main implementation methods are characteristics match, model reasoning, state transition, and expert systems. The latter is mainly to establish traffic model and analysis of abnormal flow changes, to determine whether the traffic is abnormal or not, so as to detect whether the server was attacked.

## III. LITERATURE SURVEY

Lin and Wang [5] proposed a DDoS attack detection and defense mechanism based on SDN, but the method used three Open Flow management tools with sFlow standard to perform anomaly detection, so the deployment and operation are complex. Yang et al. [6] dished a method in which the flow information and the IP entropycharacteristic information are combined, which is detected by a single flow information and IP entropy characteristic information, which has a higher and more accurate detection effect. Although information entropy is flexible and convenient, it still needs to be combined with other technologies in determining the threshold and multi element weight distribution.

Saied et al. [7] advanced that based on analysis the characteristics of each protocol of TCP/UDP/ICMP through the training ANN algorithm to detect DDoS attacks, the method needs to distinguish packet protocol, which is complex and inefficient.

In the SOM algorithm is used to detect DDoS attacks by extracting the flow statistics related to DDoS attacks. This method has the characteristics of low consumption and high detection rate. The key point lies in the extraction of time interval. The disadvantage of this method is that the detection has a certain hysteresis and the attack behavior is not timely and accurately found.

In the authors proposed a framework for detection and mitigation of DDoS attacks in a large- scale network, but it is not suitable for small-scale deployment.

In a DDoS attack detection mechanism based on a legitimate source and destination IP address database is proposed. Based on the nonparametric cumulative algorithm CUSUM, it analyses the abnormal characteristics of the source IP address and the destination IP address when the DDoS attack occurs and effectively checks the DDoS attack, but the method needs to adjust and determine the threshold.

Due to the high false positive rate of information entropy, the SOM algorithm needs to determine the number of neurons in advance. Therefore, we summarize the characteristicsof several DDoS attacks, then collect the switch flow table information, extract the six tuple characteristic values matrix, and establish their SVM classification model. The algorithm can process multidimensional data and map the low-dimensional nonlinear separable data into the high- dimensional feature space to make it linearly separable and able to be classified with high accuracy. At present, the algorithm is widely used in anomaly detectionand classification. Dao et al. [12] define a table in the controller to track the packets by IP address during a DDoS attack. All the new packets are regarded as suspicious packets andassigned a small timeout value in the flow entry. The number of packets using that connection is also compared with a minimum value to determine if it is a normal request oran attack. From the simulation, this method effectively reduces flow entries in the switch, and the bandwidth of controller-switch channel is still available during DDoS attacks. However, this mechanism consumes a huge amount of resources on the controller if the attacker modifies source address. Mousavi and St-Hilaire [13] propose to use entropy for DDoS detection due to its ability to measure randomness, where two essential components are time period and threshold. Although it may improve detection accuracy in the real network, the proposed techniques only address detection without providingcountermeasures. Similarly, Dong et al. [14] suggest a statistical tool, called Sequential Probability Ratio Test (SPRT), to improve existing false positive and false negative issues. It predefines two boundaries (A and B, B < A) related to the probabilities of false positive (a) and false negative (b) (it is suggested A = b/ (1 − a), B = (1 − b)/a), and the decision is made from the log-probability ratio. The evaluation of the DARPA Intrusion Detection Data Sets shows its promptness and accuracy. However, the proposed method is evaluated using onlymathematical results without simulations, where random variables can be introduced.

Yan et al. [15] propose a “Multislot” strategy to process requests in each time slot so that legitimate users can communicate to each other properly during DDoS attacks.

**IV. ATTACK DETECTION**

## a. low rate ddos attack detection

Low-rate Distributed Denial-of - Service attacks (low-rate DDoS) represent a new threat to cyberspace, as attackers send a vast amount of similar traffic-like attack packets to bypass legitimate flows. Zhang et al. [4] Proposed a congestion-participation (CPR) metric and a CPR-based approach for detecting and filtering DDoS attacks at low rates. They found that low-rate DDoS flows actively induce congestion in the network while normal TCP flows actively prevent congestion in the network. The proposed method was conceived to distinguish flows of attack from legitimate flows. However, the testing of its effectiveness requires more experiments and analyzes using real datasets. Du and Abe [6] have proposed an entropy metric for the size of an IP packet to detect both long-term low- rate DDoS attacks and short-term high-rate DDoS attacks. Based on the assumption that many applications have typical packet sizes depending on requests for and responses to data and acknowledgments, they claimed that the distribution of the packet size changes under attacks; this can be used to identify attacks to some degree. However, because the proposed method relies heavily on the packets in the observation window, this approach is constrained in its scalability, and it takes a long detection period to achieve a high probability of detection while suffering from a low-rate DDoS attack. Jadhav and Patil[5] suggested an efficient, unbiased method of entropy-based detection of DDoS attacks at lowlevels. This approach is a major improvement over conventional metric entropy. The distance value between regular traffic and attack traffic, however, is very small and therefore the false positive rate is increasing.

## b. high rate ddos attack detection

comprehensive research on the protection schemes for the spoofed DDoS attacks has been performed. Each scheme has its advantages and its limitations. The attacker forges TCP / IP header fields for a spoofed DDoS attack to launch. Any TCP / IP header field can however be forged, but its Time-To-Live (TTL) field cannot be forged. Hence, TTL valueis used to identify the spoofed IP packets. The challenge in this computation is since the header contains only final TTL value. All Operating Systems (OSs) have different initial TTL value and for a particular IP address its OS may get changed with time. The approach suffers from false positive if the legitimate packet is coming from the unlisted OS and false negative if the attacker correctly predicts the hop-count between source and victim. To differentiate attack traffic from the legitimate traffic. A path fingerprint approach is proposed, where a unique path fingerprint is embedded in each packet. Path fingerprint represents the route traverses by a packet to reach its destination. The incorrect path fingerprint declares the packet as spoofed. The scheme is not able to detect subnet spoofing, as the packet reaches the same subnet and it requires calculation at the intermediate nodes. The TCP/IP header fields’ values such as TTL, IP Don’t Fragment (DF), window size, and total length are used to identify the OS of a packet. A fingerprint is created using these values and attached with the packet at the source side. If the fingerprint matches at the receiver side, then the packet is declared as legitimate; otherwise, it is treated as a spoofed packet.

## 

## V.EXISTING METHODOLOGIES

In the SDN architecture, the Open Flow switch forwards the main network data at a high speed. The SDN controller is responsible for the forwarding and management of the forwarding decision and the collection of traffic information of switches. In the SDN switch, the core data

structure of the forwarding policy management control is the flow table. The SDN manages the relevant network traffic by searching the flow table entries, where the flow entry can forward the packet to one or more interfaces. Each entry includes the header field, the counters, and the actions. The packet forwarding of the switch is based on the flow table. Each flow table is composed of multiple flow entries. The flow table entries form the rules for data forwarding. Figure 1 shows the flow table entry structure diagram. The flow diagram of the attack detection consists mainly of the flow state collection, the extraction characteristic values, and the classifier judgment, as shown In Figure 2. The flow state collection periodically sends a flow table request to the Open flow switch and sends the flow table information replied from the switch to the flow state collection. The characteristic values extraction is mainly responsible for extracting the characteristic values related to the DDoS attack from the switch flow table and composingthe six-tuple characteristic values matrix. Six-tuple characteristic values information is classified by using an SVM-based algorithm [13] to distinguish between normal traffic andattacking abnormal traffic.

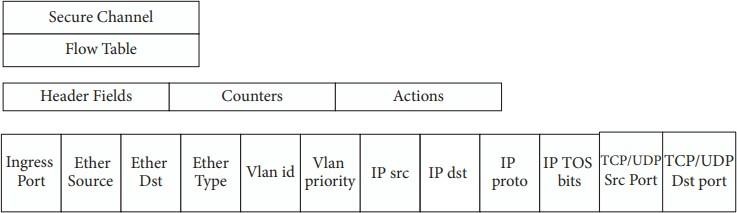
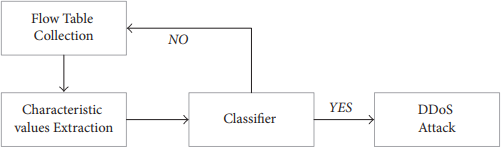


Figure 2: Flow table structure. Figure 3: Attack detection process

## Proposed methodology

DDoS attacks are a common threat to the network, although the attacker typically doesnot aim to steal any data. Basically, DDoS attacks aim at consuming system resources untilthe target is not available to offer its services. DDoS attacks could be divided into three categories: application layer attack, protocol attack and volumetric attack. For volumetric attack, an attacker can deplete the available resources of the victim or bandwidth towards the target. Not only the data plane in the SDN, but also the controller and southbound interface, could suffer from this kind of attack as well; this is because a client host can trigger inquiry from the data plane to control plane. Although there have been a lot of discussions about DDoS attacks in the SDN and IoT networks, the large number of IoT gadgets is still a good chance to launch attacks, as well as the communication link betweencontrollers and switches in SDN. Additionally, more validations in the real network are required. Moreover, the programmability and centralized control in the SDN give users more options to probe into this threat. In this paper, volumetric attack is implemented

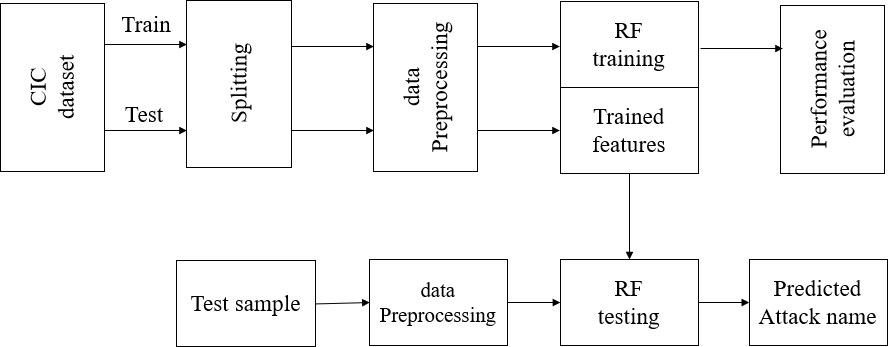


Figure 4 Proposed block diagram

The block diagram of proposed method. Initially, CIC dataset is spitted into 80% for training and

20% for testing. Then, dataset preprocessing operation is performed to normalize the entire dataset. Further, random forest classifier is used for prediction of DDoS attack from test sample. The performance evaluation is carried out to show supremacy of proposed method.

# Performance evaluation cic dataset:

CICDDoS2019 contains benign and the most up-to-date common DDoS attacks, which resembles the true real-world data (PCAPs). It also includes the results of the network traffic analysis using CICFlowMeter-V3 with labeled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack (CSV files). Generating realistic background traffic was our top priority in building this dataset. We have used our proposed B-Profile system to profile the abstract behaviour of human interactions and generates naturalistic benign background traffic in the proposed testbed (Figure 2). For this dataset, we built the abstract behaviour of 25 users based on the HTTP, HTTPS, FTP, SSH and email protocols.

# 1.Preprocessing:

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data preprocessing task.

1. **Need of Data Preprocessing:** A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency ofa machine learning model.

* Getting the dataset
* Importing libraries
* Importing datasets
* Finding Missing Data
* Encoding Categorical Data
* Splitting dataset into training and test set
* Feature scaling

# .Feature Selection:

Feature Selection is one of the key machine learning principles that greatly impacts the model 's efficiency. The data characteristics you use to train your machine learning models will have a huge impact on the performance you can achieve. Collection of features and Data Cleaning are the most important step in developing your layout. Feature Selection is the process where you select those features that most contribute to your predictive variableor output you are interested in, automatically or manually. With irrelevant features in yourdata, the model 's accuracy can be reduced and your model learned based on irrelevant features.

# 4..Model Training:

Run Algorithms: using this module we will input 80% train data to Random Forest, XGBOOST, ADABOOST, SVM, Naïve Bayes and KNN algorithms to train a model andthis model will be applied on test data to calculate prediction accuracy.**Comparison Graph:** using this module we will display comparison table and graphof all algorithms.

**5.Predict Attack from Test Data:** using this module we will upload test data andthen machine learning models will predict attack from that test data. Test Data you can find inside test folder and this test data contains all features without any class label and this label will be predicted by machine learning model

Use Case Diagram

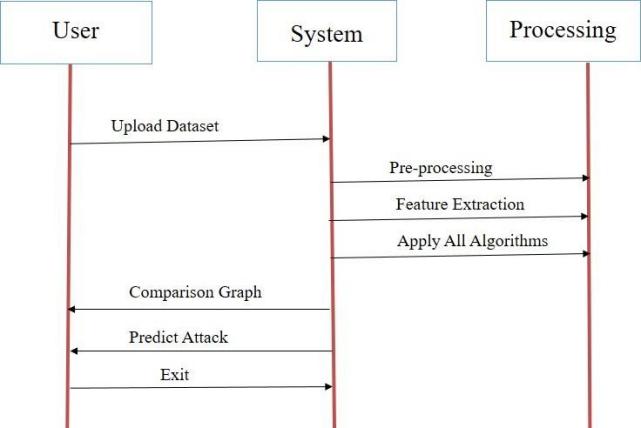
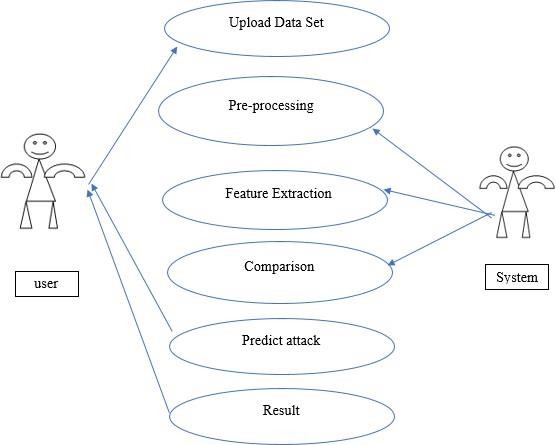


Figure 5.1: Use case diagram Figure 5.2: Sequence diagram

## VI.RESULTS

Test Data you can find inside test folder and this test data contains all featureswithout any class label and this label will be predicted by machine learning model.

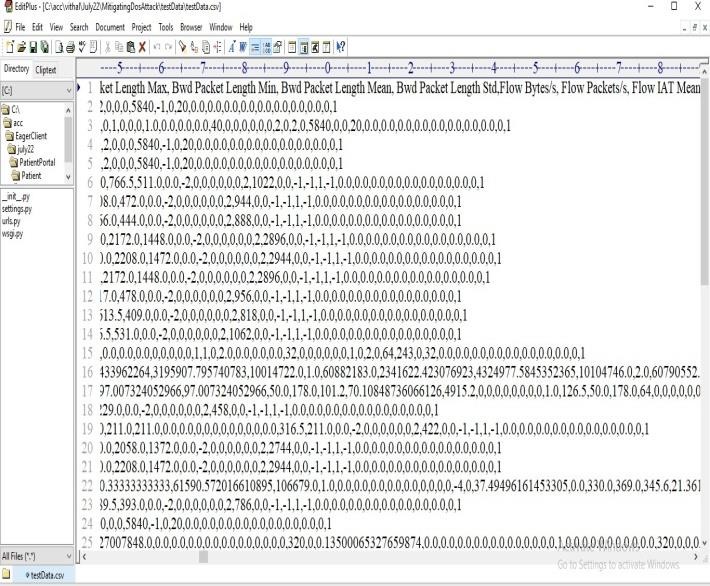


Figure 6.1: Test Data

In above TEST DATA screen there is no class label or attack name and this will be predicted by ML model.

**SCREEN SHOTS:** To run project double click on ‘run.bat’ file to get below screen

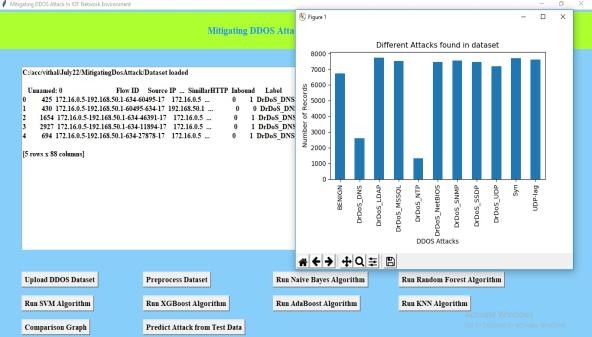


Figure 6.2: Graphical user interface (GUI) Application Figure 6.3: Different Attacks found in dataset

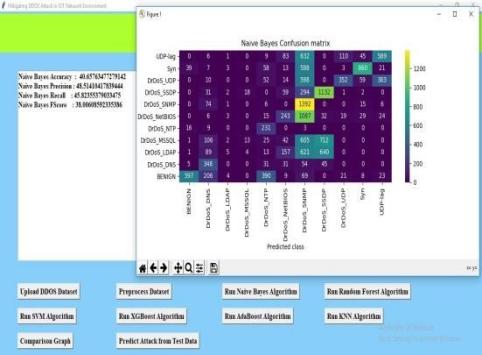




Figure 6.4: Preprocessing Dataset Output Figure 6.5: Naïve Bayes Confusion matrix In above screen click on ‘Upload DDOS Dataset’ button to upload dataset and get below output

In above screen dataset loaded and we can see dataset contains both numeric and non- numeric data and in above graph x-axis represents attack names and y-axis represents count of those records. Now close above graph and then click on ‘Preprocess Dataset’ button to process dataset and get below screen.

In above screen we can see all dataset values converted to numeric format and dataset contains more than 70000 records and each record contains 87 features and then we have split dataset into train and test and for training application using 56685 records for trainingand 14172 for testing. Now train and test data is ready and now click on ‘Run Naïve Bayes’

In above screen with Naïve Bayes we got 40% accuracy and in confusion matrix graph x- axis represents PREDICTED classes and y-axis represents TRUE classes and prediction count in same row and column names are the correct prediction and count in different row and column names are the incorrect prediction and we can see Naïve Bayes predicted so many wrong prediction and close above graph and then click on ‘Run Random Forest Algorithm’ button to get below output. In above screen with Random Forest we got more than 96% accuracy and in graph also we can see lots of predictions are correct. Now close above graph and then click on ‘Run SVMAlgorithm’ button to get below output

In above screen with SVM we got 67% accuracy and now close above graph and then clickon ‘Run XGBOOST Algorithm’ button to get below output

VII.CONFUSION MATRIX

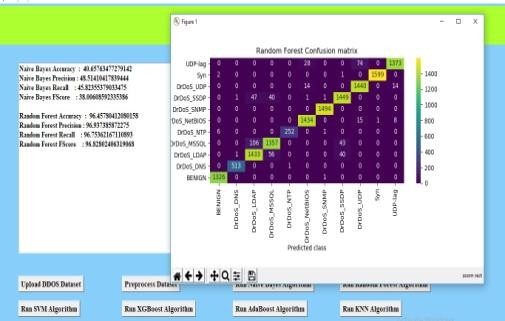
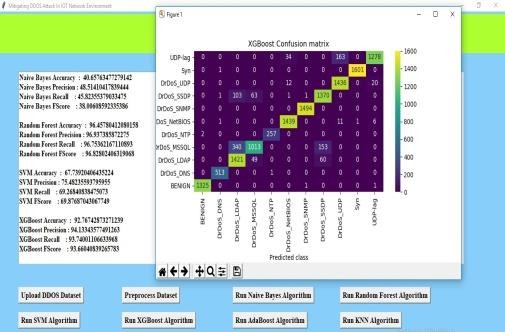
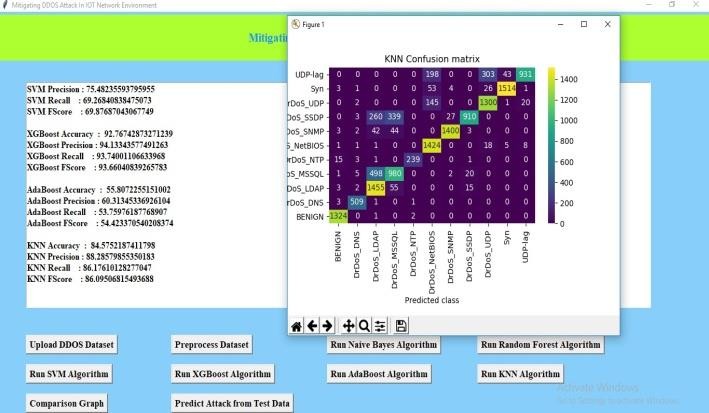
  

Figure 7.1: Random Forest Confusion matrix Figure 7.2: SVM Confusion matrix Figure 7.3: XGBOOST Confusion matrix

In above screen with XGBOOST we got 92% accuracy and now close above graph and then click on ‘Run ADA BOOST Algorithm’ button to get below output



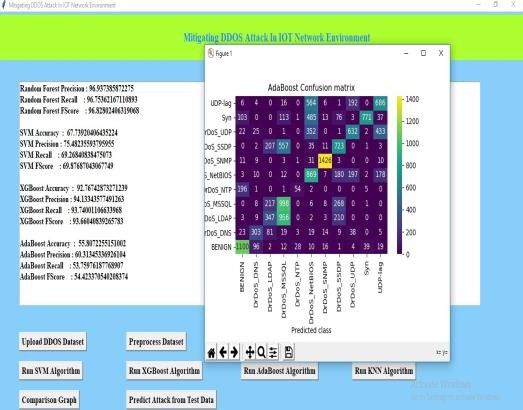


Figure 7.4: ADABOOST Confusion matrix Figure 7.5: KNN confusion matrix





Figure 7.6: Comparison Table Figure 7.7: Test Data Output In above

In above screen with ADABOOST we got 55% accuracy and now close above graph andthen click on ‘Run KNN Algorithm’ button to get below output In above screen with KNN we got 84% accuracy and now close above graph and then clickon ‘Comparison Graph’ button to get below ggraph and comparison table we can see Random Forest got high accuracy and in above graph different colour bar represents different metrics such as accuracy, precision, recall and FSCORE. Now click on ‘Predict Attack from Test Data’ button to upload test data and get below output

In above screen selecting and uploading TEST DATA file and then click on ‘Open’ button to get output. In above screens with each different test records different attacks and benign (normal) classes are predicted.

## VIII.CONCLUSION AND FUTURE SCOPE

With the development of network technology, DDoS attacks have increasingly become an important security risk that endangers the network. It uses common protocols and serviceswhen attacking, so it is difficult to detect through traditional methods. Based on the idea ofrational thinking, DDoS attack detection can be simulated as a classification problem that distinguishes between "rational" and "irrational" network flow states. This article analyzes the common TCP flood attacks, UDP flood attacks, and ICMP flood attacks in detail. Define the characteristics of data stream information entropy to characterize attack behavior. A DDoS attack detection method based on random forest classification model isproposed. Establish classification models for the above three types of typical attack methods. Through training and learning, it is finally predicted whether the network traffic is normal.

# A.Future Scope

Now-a-days to detect cyber-attack are using static and dynamic analysis of request data. Static analysis is based on signature which we will match existing attack signature with new request packet data to identify packet is normal or contains attack signature. Dynamicanalysis will use dynamic execution of program to detect malware/attack, but dynamic analysis is time consuming. To overcome from this problem and to increase detection accuracy with old and new malware attacks, we are using machine learning algorithms andevaluating prediction performance of various machine learning algorithms such as Support Vector Machine (SVM), Random Forest, Decision Tree, Naïve Bayes, Logistic Regression, KNearest Neighbours and Deep Learning Algorithms such as Convolution Neural Networks (CNN) and LSTM (Long Short-Term Memory). Among those, various models Deep learning CNN resulted in superior performance compared to other models.

## 

## REFERENCES:

[1]. H. Zhang, Z. Cai, Q. Liu, Q. Xiao, Y. Li, and C. F. Cheang, “A surveyon security-aware network measurement in SDN,” Security and Communication Networks, Article ID 2459154, 2018.

[2]. J. Cao, M. Xu, Q. Li, K. Sun, Y. Yang, and J. Zheng, “Disrupting SDN via the data plane: a low-rate fow table overow attack,” in Proceedings of the 13th EAI International Conference onSecurity and Privacy in Communication Networks, Niagara Falls, Canada, October 2017. 8 Security and Communication Networks

[3]. Z. Cai, Z. Wang, K. Zheng, and J. Cao, “A distributed TCAM coprocessor architecture for integrated longest prefx matching, policy fltering, and content fltering,” IEEE Transactions on Computers, vol. 62, no. 3, pp. 417–427, 2013.

[4]. Y. Li, Z. Cai, and H. Xu, “LLMP: exploiting LLDP for latency measurement in sofware-defneddata center networks,” Journal of Computer Science and Technology, vol. 33, no. 2, pp. 277– 285, 2018.

[5]. H. Lin and P. Wang, “Implementation of an SDN-based security defense mechanism against DDoS attacks,” in Proceedings of the 2016 Joint International Conference on Economics and Management Engineering (ICEME 2016) and International Conference on Economics and Business Management (EBM 2016), Pennsylvania, Penn, USA, 2016.

[6]. J. G. Yang, X. T. Wang, and L. Q. Liu, “Based on trafc and IP entropy characteristics of DDoSattack detection method,” Application Research of Computers, vol. 33, no. 4, pp. 1145–1149, 2016.

[7]. A. Saied, R. E. Overill, and T. Radzik, “Detection of known and unknown DDoS attacks usingartifcial neural networks,” Neurocomputing, vol. 172, pp. 385–393, 2016.

[8]. R. Braga, E. Mota, and A. Passito, “Lightweight DDoS fooding attack detection using NOX/OpenFlow,” in Proceedings of the 35th Annual IEEE Conference on Local Computer Networks (LCN ’10), pp. 408–415, Denver, Colo, USA, October 2010.

[9]. N. Z. Bawany, J. A. Shamsi, and K. Salah, “DDoS attack detection and mitigation using SDN: methods, practices, and solutions,” Arabian Journal for Science and Engineering, vol. 42, no. 2, pp. 425– 441, 2017.

[10]. X. Wang, M. Chen, C. Xing, and T. Zhang, “Defending DDoS attacks in sofware-defned networking based on legitimate source and destination IP address database,” IEICE Transaction on Information and Systems, vol. E99D, no. 4, pp. 850–859, 2016.

[11]. J. Xia, Z. Cai, G. Hu, and M. Xu, “An active defense solution for ARP Spoo ng in OpenFlow network,” Chinese Journal of Electronics, vol. 3, 2018.

[12]. Dao, N.N.; Park, J.; Park, M.; Cho, S. A feasible method to combat against DDoS attack in SDN network. In Proceedings of the 2015 International Conference on Information Networking (ICOIN), Siem Reap, Cambodia, 12– 14 January 2015; pp. 309–311. doi:10.1109/ICOIN.2015.7057902.

[13]. Mousavi, S.M.; St-Hilaire, M. Early detection of DDoS attacks against SDN controllers. In Proceedings of the 2015 International Conference on Computing, Networking and Communications (ICNC), Anaheim, CA, USA, 16–19 February 2015; pp. 77–81. doi:10.1109/ICCNC.2015.7069319.

[14]. Dong, P.; Du, X.; Zhang, H.; Xu, T. A detection method for a novel DDoS attack against SDN controllers by vast new low-traffic flows. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–6. doi:10.1109/ICC.2016.7510992.