SECURING NETWORKS USING SOFTWARE DEFINED NETWORKS AND MACHINE LEARNING

By: **Fazeela Mughal**

2019-NUST-MS-CS-19 317739

Supervisor: **Dr Abdul Wahid**

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science (MSCS)

In

School of Electrical Engineering and Computer Science,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan

(April 2022)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Securing Networks using SDN and Machine learning" written by FAZEELA MUGHAL, (Registration No 00000317739), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________

Name of Advisor: Dr. Abdul Wahid

Date: 25-May-2022

HoD/Associate Dean: ______________

Date: ______________

Signature (Dean/Principal): ______________

Date: ______________

Approval

It is certified that the contents and form of the thesis entitled "Securing Networks using SDN and Machine learning" submitted by FAZEELA MUGHAL have been found satisfactory for the requirement of the degree

Advisor : Dr. Abdul Wahid

Signature: ____________________

Date: 25-May-2022

Committee Member 1: Dr. Muhammad Moazam Fraz

Signature:

26-May-2022

Committee Member 2: Prof. Hasan Ali Khattak

Signature: ____________________

Date: 26-May-2022

Signature: ____________________

Date: ____________________

DEDICATION

This thesis is dedicated to my family especially to my sister, who encouraged me to do MSCS from NUST. Also I would like to dedicate this to my supervisor Dr. Abdul Wahid who guided, appreciated & motivated me in the whole thesis. Moreover, I also want to give credit to my friends; Raisa Suleman, Haris Ahmed & Muhammad Luqman who helped me in resolving every query.

Certificate of Originality

I hereby declare that this submission titled "Securing Networks using SDN and Machine learning" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name:FAZEELA MUGHAL

Student Signature: ______________

ACKNOWLEDGEMENT

All praises be to ALLAH: Al-Muizz, Al-Kabeer, Al-Hadi and Al-Fattah

The successful completion of this thesis is accomplished by the devoted participation and cooperation of all guidance committee members. With gratitude and affection, I acknowledge active and guided guidance of my honorable supervisor Ass prof. Dr. Abdul Wahid. He supported me in hours of need and channelized my way in hard times. His motivation, guidance and supervision acted as the driving force that has enabled me to achieve my objective. I also admire and value participation of my respectable committee members; Associate Professor Dr. Moazam Fraz and Professor Dr. Hasan Ali Khattak for their time and advice. I am very grateful for the teachings of faculty members of Computer Science Department that has fueled my sense of continued determination over the years. I appreciate efforts of my all-family members, friends and class fellows who raised my morale and their motivation opened new ways for me. Their prayers and ALLAH's help had enabled me to be the best version of myself

Table of Contents

ABBREVIATIONS

DEFINATION	ABBREVIATIONS
Distributed Denial of Service	DDoS
Denial of Service	DoS
Machine Learning	ML
Intrusion Detection System	IDS
Machine Learning based Intrusion Detection System	ML-IDS
Network Intrusion Detection System	NIDS
Logistic Regression	LR
K-Nearest Neighbor	KNN
Decision Tree	DT
Random Forest	RF
Multi-Layer Perception	MLP
Support Vector Machine	SVM
eXtreme Gradient Boosting	XGB
User Datagram Protocol	UDP
Transmission Control Protocol	TCP
Packet Sender	PS

LIST OF TABLES

LIST OF FIGURES

Abstract

Machine Learning techniques are used in Networks to detect DoS and DDoS attacks and to resolves network security issues. As many researchers done their research either on real time datasets or synthetic datasets on different attacks however in our thesis, we aimed to check the performance of Machine learning algorithms, that which one is giving high accuracy in detection of DDoS attack. For this purpose, we have generated simulated datasets in Mininet and in Packet Sender tool. In addition, two well-known datasets has been chosen in which one of them is real time dataset that is ToN-IoT whereas the other one is synthetic dataset Mendeley DDoS. By applying Machine Learning techniques on these datasets, we investigate seven different algorithms: K-Nearest Neighbor, Decision Tree, Random Forest, Logistic Regression, Multi-Layer Perception, XG-Boost, Support Vector Machine and Ensemble Method, results are produced on the basis of accuracy rate. Results are computed on the basis of best features present in all datasets.

.

Chapter 1 Introduction

As the usage of internet in the new era is growing day-by-day therefore privacy and security issues are also considered. People prefer that their data should transfer to the destination with secure means. Privacy and security of any data should not be negligible in any system. Confidentiality and integrity of data is important, on internet it is also possible that one user might access the data of other user if there isn't any proper privacy mechanism implemented or we can say that any hacker can access or manipulate private data of other user over internet. Hackers attack on system for many reasons such as:

- To access information or resources.
- To manipulate information.
- To render a system unreliable or unusable.

There are many attacks which damage any computer network system either attacker attack from one particular host to any destination host (DoS) or from many hosts to one particular host (DDoS) [1] as shown in Figure 1.1. DoS attack is basically of two types; one is flooding the server in which large amount of packets are transferred from source host to destination server host and the other one is crashing the server by malicious traffic.

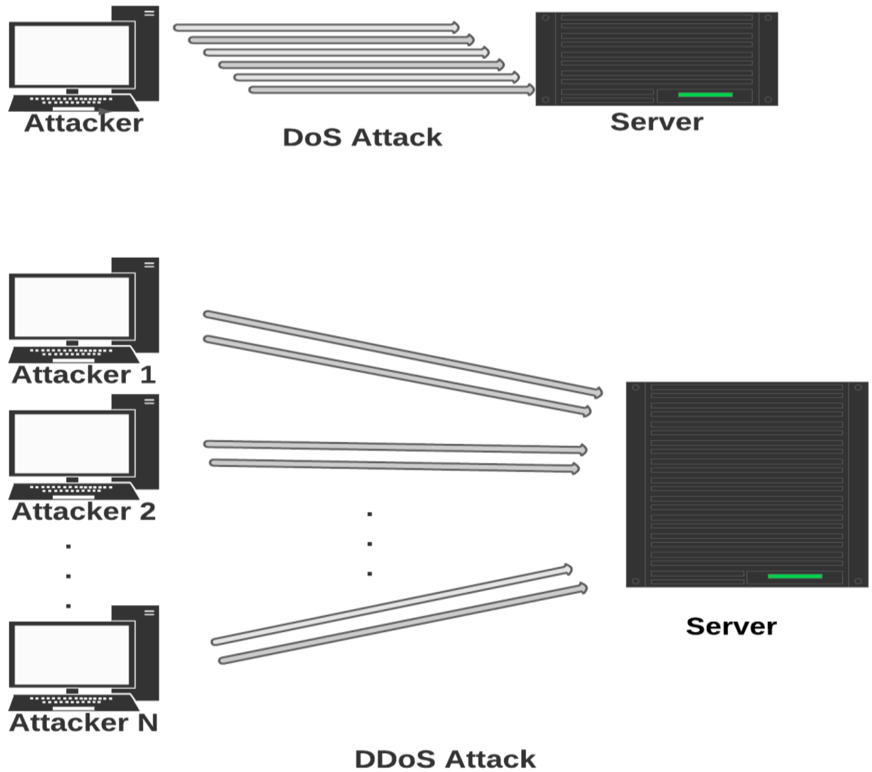


Figure 1.1 DoS and DDoS Attack

Distributed Denial of Service (DDoS) attack suspends the online services of a server either on temporary basis or permanently. There are three main categories of DDoS attack [2, 3, & 4]; Volume based attack, Application Layer attack and Protocol layer attack. These categories have sub category type of DDoS attack as shown in Figure 1.2.

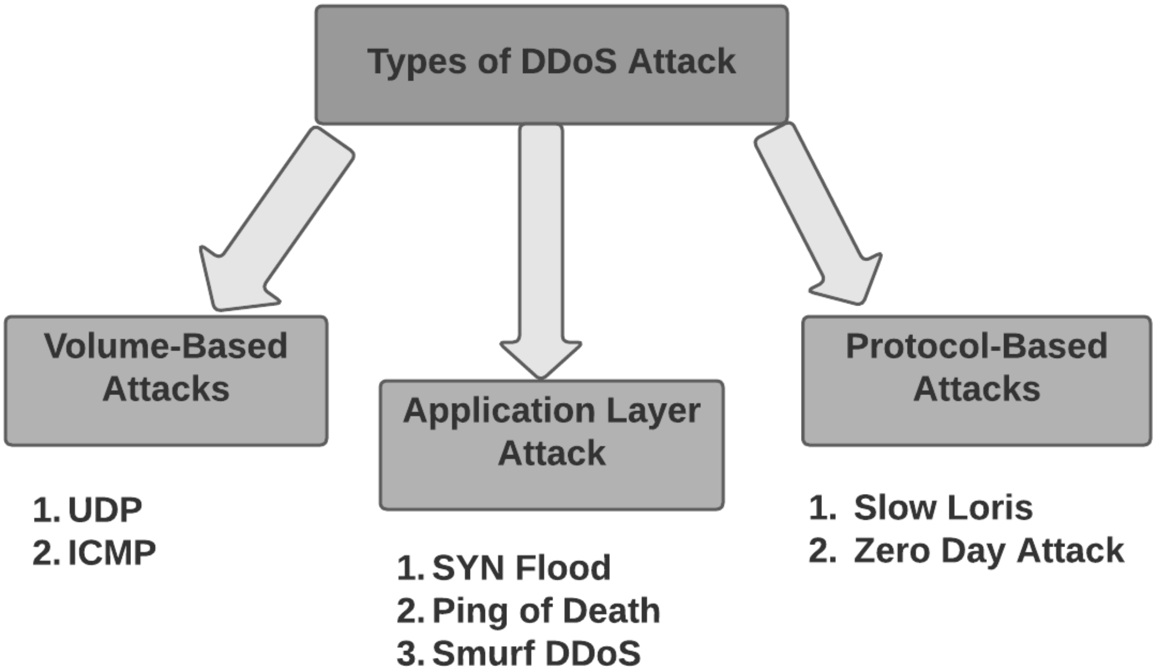


Figure 1.2 Types of DDoS Attacks

- **UDP Flood:** UDP Flooding is basically included in volume based flooding, many source hosts targets a single destination host and sends UDP packets as shown in Figure 1.3.
- **Ping of Death:** POD attack is done by many source hosts sending multiple malformed packets or many malicious pings to a destination host [5].
- **ICMP Flood:** It slows down the destination host by sending requests without waiting for any response from server [6] as shown in Figure 1.4.
- **Slow Loris:** It is an application layer attack which partially looks like HTTP request connection. It opens the connection as long as possible so that the other users might not connect and access the target web server [7].
- **SYN Flood:** The hosts sends many SYN requests to destination host and destination host sends back SYN-ACK and source hosts ignore the acknowledgement and sends requests again and again which overwhelms the server [8].
- **Zero-Day-Attack:** All DDoS new and unknown attacks are represented by zero day attack [9].

- **NTP Flood:** It is reflection based Network Time Protocol amplification attack which sends volumetric UDP packets to and overwhelms the server. In NTP Flood an attacker may be spoofed its IP address and send UDP packets [10].
- **HTTP Flood:** Source hosts do not use malformed packets and IP spoofing, it sends legitimate HTTP GET/POST request to a target server which slows down the services of a server [11].

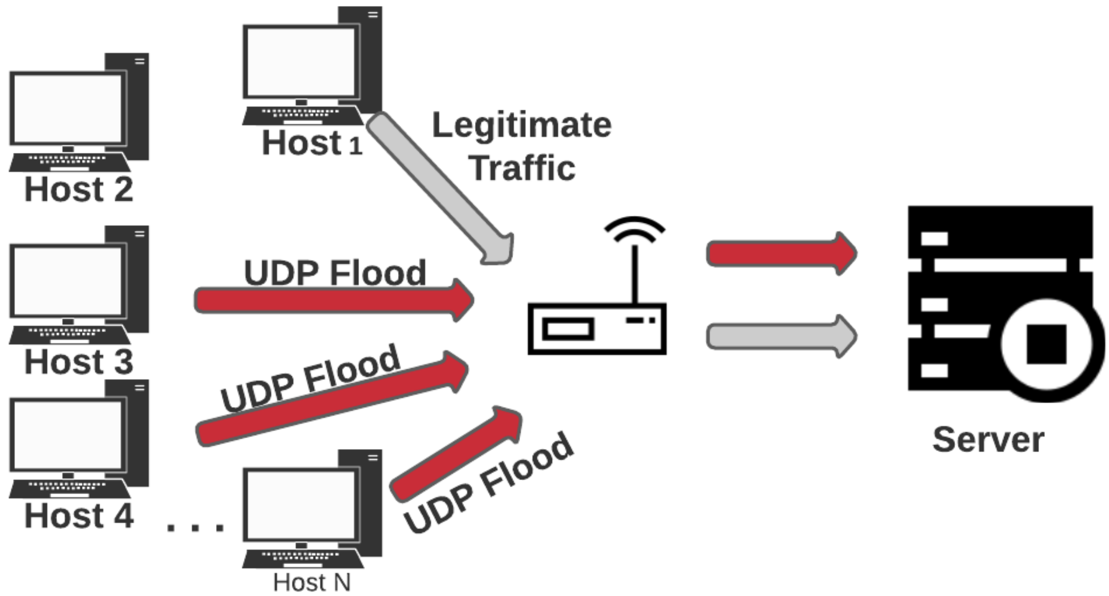


Figure 1.3 UDP Flood

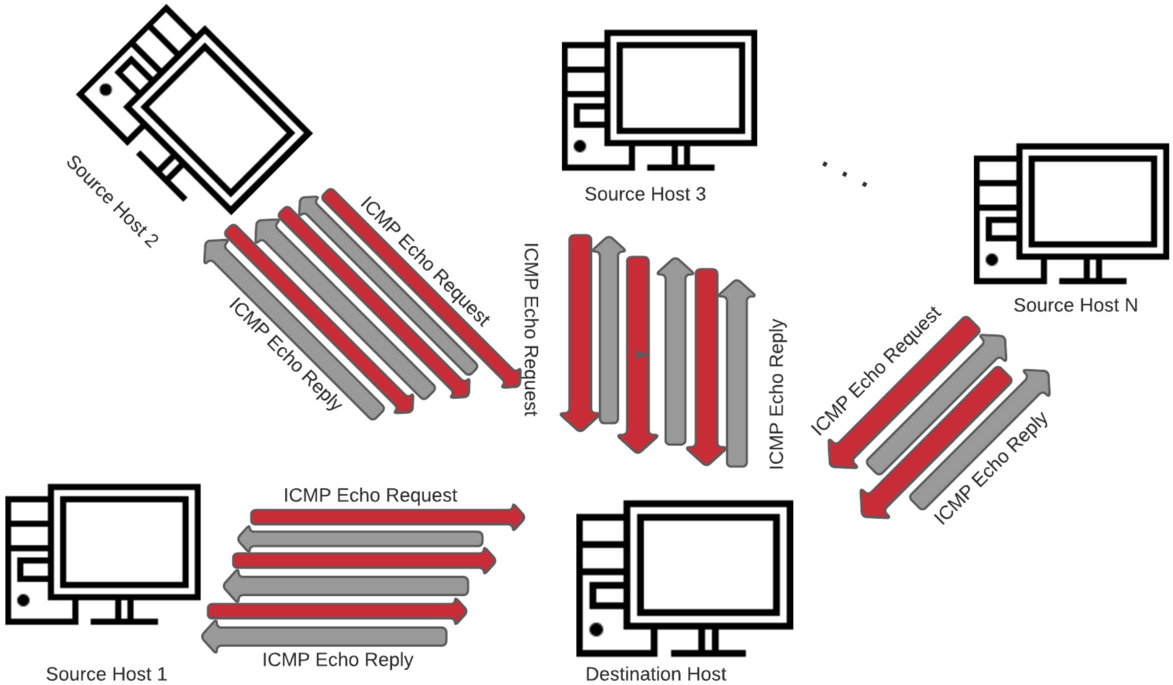


Figure 1.4 ICMP Flood

1.1 Motivation:

Denial of Service (DoS) or Distributed Denial of Service (DDoS) can easily target a system so it is necessary to detect these attacks on time before having major damage to a network system. DDoS is the most common and major attack which can damage any network by massive traffic as the traffic generated from different source hosts to one particular destination host. Intrusion Detection System (IDS) plays an important role in detecting these attacks by analyzing and identifying normal and abnormal traffic in a network [12]. These days Machine learning based IDS are securing networks by detecting attack and doing prevention.

In Networks ML-based IDS are used to detect malicious and abnormal traffic flow [13]. There are two methods to detect malicious traffic as shown in Figure 1.5. First one is signature based detection which detects the threat we know about whereas the second one is known as anomaly based detection whose methodology of detection is change in behavior of traffic

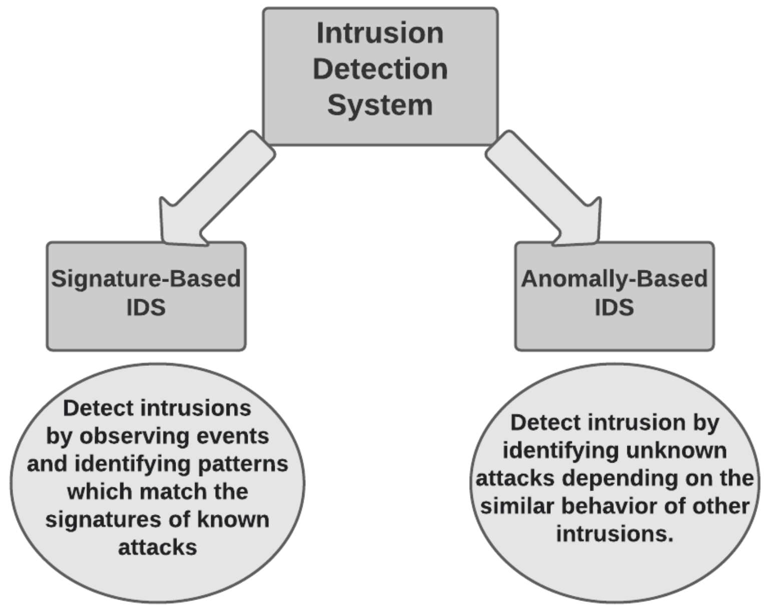


Figure 1.5 Types of IDS

1.2 Problem Statement:

As compare to traditional approaches ML-based IDS is efficient and more accurate for the detection of anomalies. There have been a lot researches on many attacks in networking and detection of anomalies is happened by ML-based IDS [14]. Many authors worked on publically available datasets [14] and few of them generated simulated datasets [16] and showed the results in the form of accuracy and said simulated datasets are better because publically available datasets are not updated regularly without comparing results on both types of datasets, also they concluded few algorithms are performing well.

1.3 Thesis Contribution:

In this thesis, our aim is to check the performance of algorithms in detection of DDoS attacks in different datasets, in addition, our aim is to generate simulated UDP DDoS dataset on Packet Sender tool for traditional networks and TCP, UDP and ICMP DDoS dataset on Mininet for Software Defined Networks to check the performance of algorithms. Different datasets are publically available in which real-time and simulated datasets are included. In our research, two well-known datasets were selected to check the algorithms' performance on both types of datasets. Necessary data pre-processing is done and seven algorithms has been tested and ensemble methods are applied on top three models which gave high accuracy. Here two datasets were selected and results are produced, there is one real-time dataset; ToN-IoT whereas the other one is simulated dataset; Mendeley DDoS [17].

Thesis report is divided into five major chapters listed below:-

- **Chapter 1: Introduction**
- **Chapter 2: Literature Review**
- **Chapter 3: Methodology**
- **Chapter 4: Results**
- **Chapter 5: Discussion**
- **Chapter 6: Conclusion & Future Work**
- **Chapter 7: References**

Chapter 2 Literature Review

In network systems, security became an important and challenging factor. In today's networking, cyber security plays an important role by providing security, integrity and confidentiality to users on internet. Intrusion can be in various forms on internet but the most common one is in the form of DoS and DDoS attack and for this Network Intrusion Detection Systems are used.

As in past there were some issues in detection of new attacks in a network system when intrusion is detected by Signature-based IDS, a research has been done to overcome the limitations of Signature-based IDS by the use of Anomaly-based IDS [19]. In many network systems, most uncertain traffic flow is evaluated by selecting limited features which helps in decision making and shows good performance which helps the controller to detect normal and malicious traffic [20].

Xie et al. [21] applied Machine learning techniques to find the optimal algorithms to classify the traffic, predict the Quality of Service or Quality of Experience, for routing optimization and for providing security and resource management. Different algorithms were applied to predict all of the above and for Traffic Classification the optimal algorithms were; Decision Tree, Random Forest, Deep NN, SVM KNN, for QoS/QoE; Decision Tree, KNN, Random Forest, Neural Network and for Resource Management the optimal algorithms were; Naive Bayes, Linear SVM**,** Radial SVM**,** Decision Tree and K-NN.

In article [22], authors tried to detect Low-Rate DDoS attack, for this he used CIC-DDoS Dataset and did evaluation on six models to detect DoS and DDoS attacks and got 97% overall accuracy for DDoS attacks but faced problems in finding LR-DDoS so they created simulated environment by using ONOS controller on Mininet Virtual Machine. In simulated environment which was having resemblance with real dataset they found out LR-DDoS

In Table 2.1, different attacks are identified also the type of dataset is mentioned either it is simulated or real-time based dataset.

Authors	Dataset Type	Attacks	Classifiers	Accuracy
[16]	Simulated Dataset	UDP Flood	1.SVM 2.Naïve Bayes 3.Logistic Regression 4.Decision Tree	97.5%
[19]	Real-time Dataset	U2R, R2L Probe	1.ANN	97%
[22]	Real-time Dataset	LR-DDoS	1.J48 2.Random Forest 4.MPL 5.SVM	97%
[23]	Simulated Dataset	DDoS	1.SVM	High accuracy.
[24]	Real-Time Dataset	TCP & SYN	1.Decision Fusion	97%
[25]	Real-time Dataset	Warmhole	1. K-mean 2.Decision Tree	KM-IDS achieved 70% to 90%
[26]	Simulated Dataset	DDoS	1.SVM 2.Naive Bayes 3.KNN 4.Self Organizing Map	97.14%

Table 2.1 Related Work for Traditional Networks

.

DDoS attacks are studied in detail in different network system such as in Traditional Networks, in Software Defined Networks which separates the forwarding plane to control plane by having centralized control plane as shown in Figure 2.1 In Table 2.2, there are some related work shown in which proposed solution and methodology is mentioned.

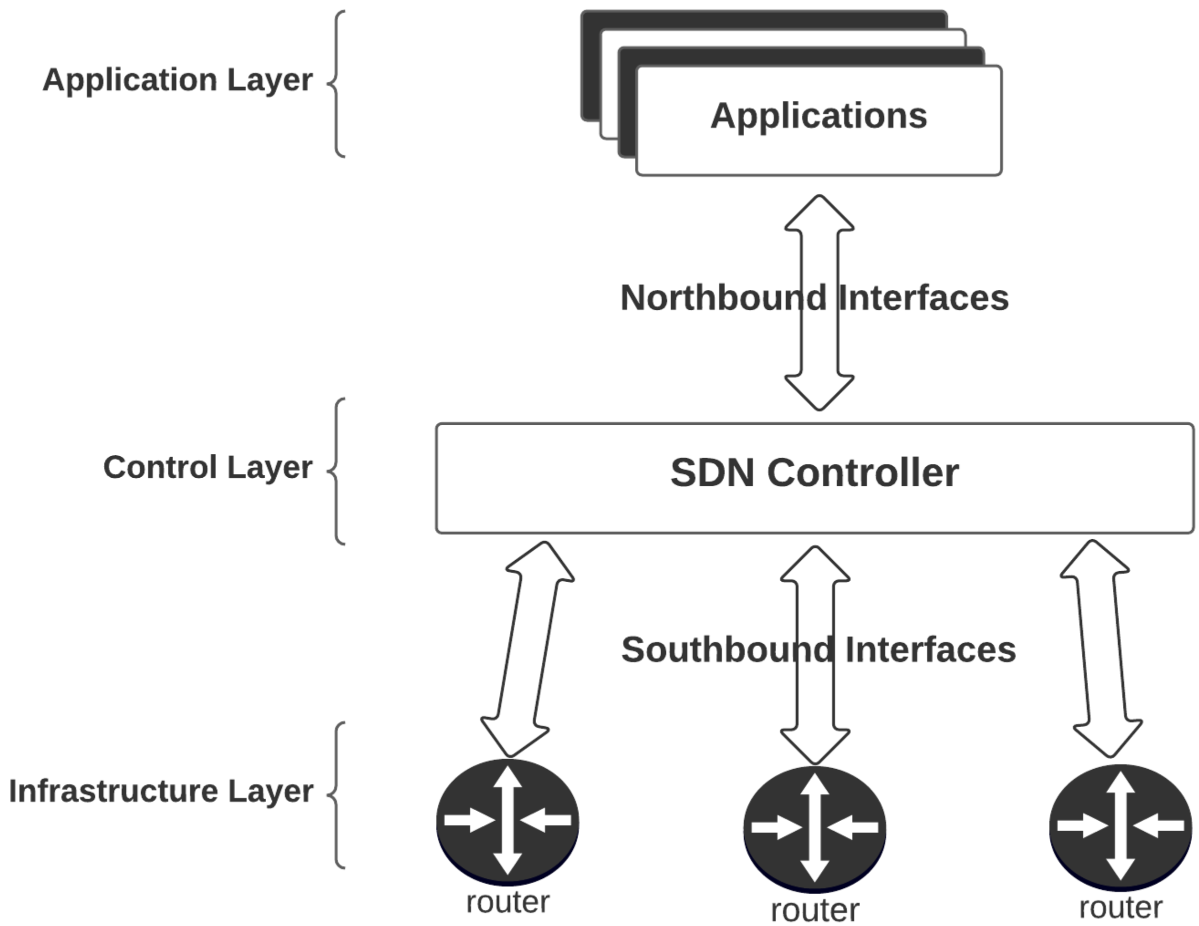


Fig 2.1 Software Defined Network Architecture

Sr #	Title	Year	Journal	Proposed Work	Methodology	Limitations & Future Work
[27]	**A survey on ML application for SDN security**	2019	International conference on applied Cryptography & network security.	In the given paper the author used Machine Learning techniques for the security of SDN. They also introduced the standard dataset, tools and test beds for research purpose.	Authors selected the papers and classify them in following categories: Survey. Proposal for framework. Experiments of existing tools. ML based **IDS** in SDN. **ML Techniques:** • **RBM** • **CNN** • **ANN** • **KNN** • **NEAT** • **Generic NN** • **Naive Bayes.**	It is beneficial to extend the analysis of Machine Learning techniques used in reviewed papers with a more detailed classification.
[28]	**Comparison for ML Algorithms For DDoS**	2020	Wiley Online Library	One of the most recent solutions to detect a DDoS	Six Algorithms were used to compare with each other for DDoS	The author said that there is a need to pay attention on the

	attack detection in SDN.			attack is using machine learning algorithms to classify the traffic. Authors also pointed out that the main features that identify malicious traffic compared to normal traffic. It will make it easier to build a DDoS protection system with a more compact data- set, focusing only on the data needed.	attack detection. • Naive Bayes. • Decision Tree. • Random Forest. • SVM • MLP • K-Nearest Neighbours. On the basis of processing time & accuracy author found out that Naive Bayes & Decision tree were the most suitable algorithms.	selection of data quality by comparing the results of detection between simulation dataset and real time based dataset.
[29]	**A Flexible SDN based Architecture for identifying &**	2020	IEEE Access 8	In the given paper, they designed and implemented modular and flexible	Achieved 95% accuracy rate by using six ML models: • J48. • Random	The aim of improving the performance with newer ML & Deep learning models/algorithm

	Mitigating Low Rate DDoS attacks using Machine Learning.			security architecture to detect and mitigate LR-DDoS attacks in SDN environments. The modularity of the design allowed one to easily replace any module without affecting the other modules of the architecture. They also deployed their architecture in real virtualized environment using mininet virtual machine & ONOS controller.	Tree. • REP Tree. • Random Forest. • MLP. • SVM. By Canadian Institute of Cybersecurity **CIC** -DoS dataset they evaluated their performance on ML models.	s. Also in terms of scalability it is important to include a selective testing mechanism of flows from the Intrusion Prevention System to Intrusion Detection System.
[21]	**A Survey of Machine Learning Techniques**	2018	IEEE Communications Surveys	In the given paper, authors delivered the comprehensive	The learning models researchers found out that best classifiers for:	The given article attempts to briefly explore how ML algorithms work

	Applied to SDN: Research Issues & Challenges.		& Tutorials 21(1).	analysis on the literature having ML techniques which were applied on SDN. For the perspective of QoS, traffic classification, QoE prediction, resource management, routing optimization, & security.	**QoS/QoE:** • Decision Tree • KNN • Random Forest. • Neural Network **Traffic Classification:** • Decision Tree. • Random Forest. • Deep NN • ML Classifier. • SVM • KNN • Semi Supervised Learning **Routing Optimization:** • Decision Tree. • Random Forest • Regression	and when they should be used to solve problems in SDN. The significant research challenges and future research directions in ML-based SDN, including high-quality training datasets, distributed multi-controller platform, improving network security, cross-layer network optimization, and incrementally deployed SDN.

					Tree • Neural Network **Resource Management:** • Naive Bayes. • Linear SVM • Radial SVM • Decision Tree • K-NN **Security:** • Decision Tree • Random Forest • HMM • SVM • Naive Bayes • Decision Table • Deep NN • Bayes Net • SOM	
[30]	**A Novel SDN**	2020	International Conference	In the given paper they	Two SDN datasets were created:	Limited types of attacks are used

	Dataset for Intrusion Detection in IoT Networks		on Network and Service Management	handled normal traffic and different types of traffic attacks (DoS, DDoS, Port Scanning, OS Fingerprinting & Fuzzing). For this purpose they introduced a novel dataset for IoT environments managed software defined network.	In our first SDN dataset number of IoT devices change time to time (Dynamic IoT environment). In second SDN, they test the performance of attack detection models trained using the first dataset in a dynamic IoT environment.	in this paper we can add more.
[20]	**Machine-learning based Threat-aware System in Software Defined Networks**	2017	International Conference on Computer Communications and Networks (ICCCN)	The given paper proposed threat aware system based on ML. This system is consisted on the following: Data pre-processing Predictive data modeling for	They developed a new method to deal with undecided data/alerts given the high resilience of SDN. With the help of **Utility Assessment** they achieve high accuracy.	The current proposed framework can be enhanced by using following additional advanced techniques: Multiple Classifiers. Contextual Knowledge

				Ml and anomaly detection Decision making for intrusion response in SDN.	The proposed system reacts to uncertainty in SDN by using **Reactive Routing.**	Advance Sophisticated Response System.
[31]	**Machine Learning Based Intrusion Detection System for Software Defined Networks**	2017	International Conference on Emerging Security Technologies (EST)	To detect Flow based anomaly attacks in the SDN environment, they proposed machine learning (Neural Network) based intrusion detection for SDN.	By using **Pattern Recognition** of neural networks they detect almost all possible anomaly attacks. For training data they used NSL-KDD Dataset and Achieve 97% accuracy rate.	
[23]	**An SVM Based DDoS Attack Detection Method for Ryu SDN Controller**	2019	International Conference on emerging Networking Experiments and Technologies	In this paper, they implement DDoS attack on Ryu SDN controller using Mininet Emulator. And for detecting	Following techniques are used to implement and detect DDoS attack on SDN: Python based Open Source Controller Ryu is used.	By improving feature correlation, traffic generation, and real-time performance we can extend the current work.

				DDoS attack SVM is used and after that they added flows in switches by doing this the percentage of DDoS attack is reduced by 36%.	Simulate DDoS attack using Mininet Emulator. SVM is used to detect DDoS attack To differentiate and train the model with normal and abnormal traffic Entropy is used.	

Table 2.2 Related Work for other Networks

By extensive literature review, we aimed that to work on real time datasets and on simulated datasets which are publically available to detect DDoS attacks and compare the accuracy results of algorithms. By doing so, optimal algorithms for each dataset can be found either they are tree based algorithms like Decision Tree, Random Forest or Regression Tree based like Logistic Regression. For doing this evaluation four datasets are used in this article to detect DDoS attack.

Chapter 3 Methodology

Over all implementation is divided into datasets, Analysis, Preprocessing, Feature Selection, Machine Learning classifiers and their hyper-parameter tuning. Flow diagram of implementation is given below:

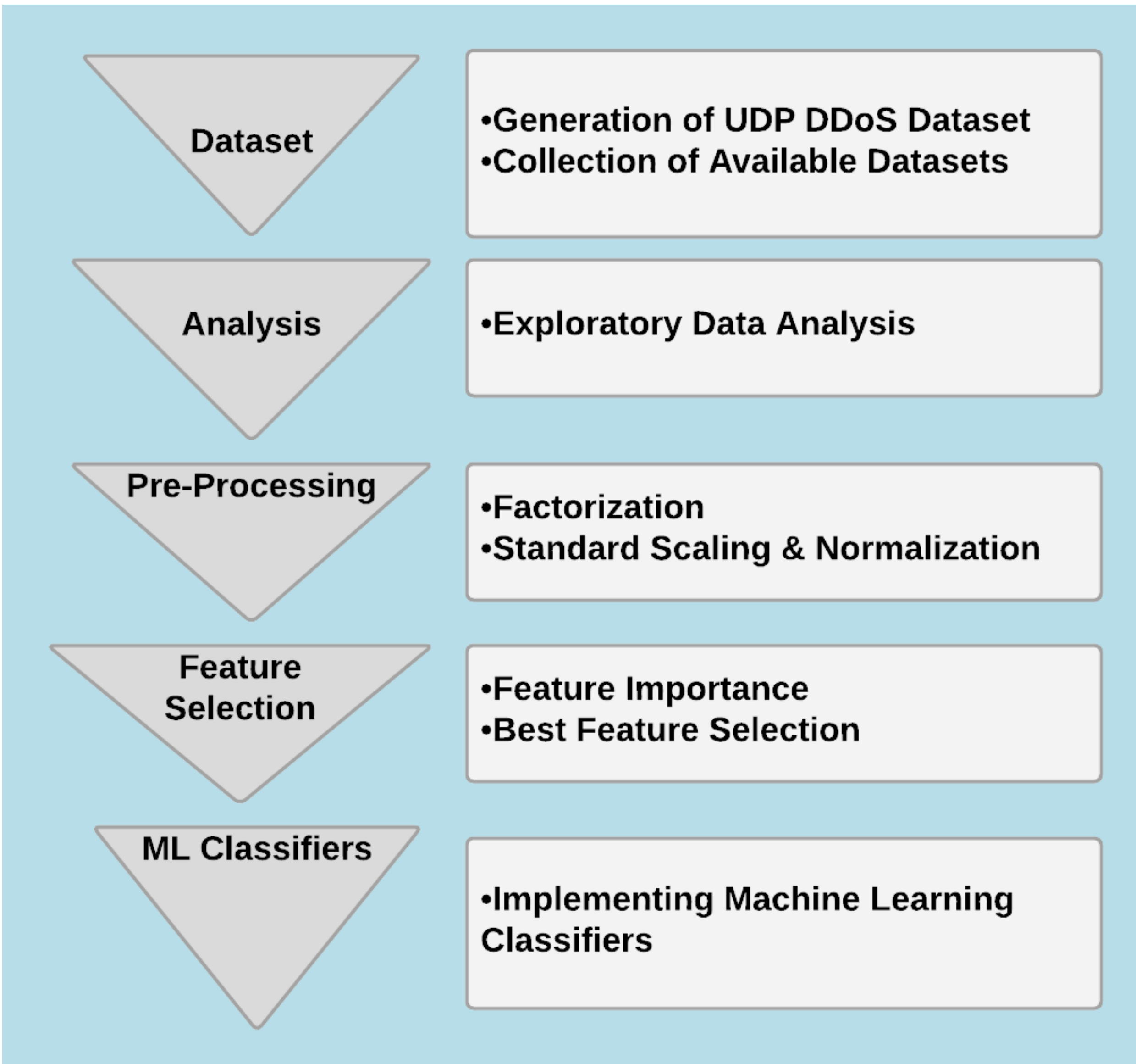


Figure 3.1 Implementation Flow

3.1 Datasets:

Different DDoS datasets are used in our research; we have generated UDP flood dataset in Packet Sender tool for traditional networks whereas for Software Defined Networks we generated TCP, UDP and ICMP DDoS attack dataset in mininet [32], in addition, we have also selected two well-known real-time and synthetic datasets from internet. Classification of datasets is shown in Figure 3.2.

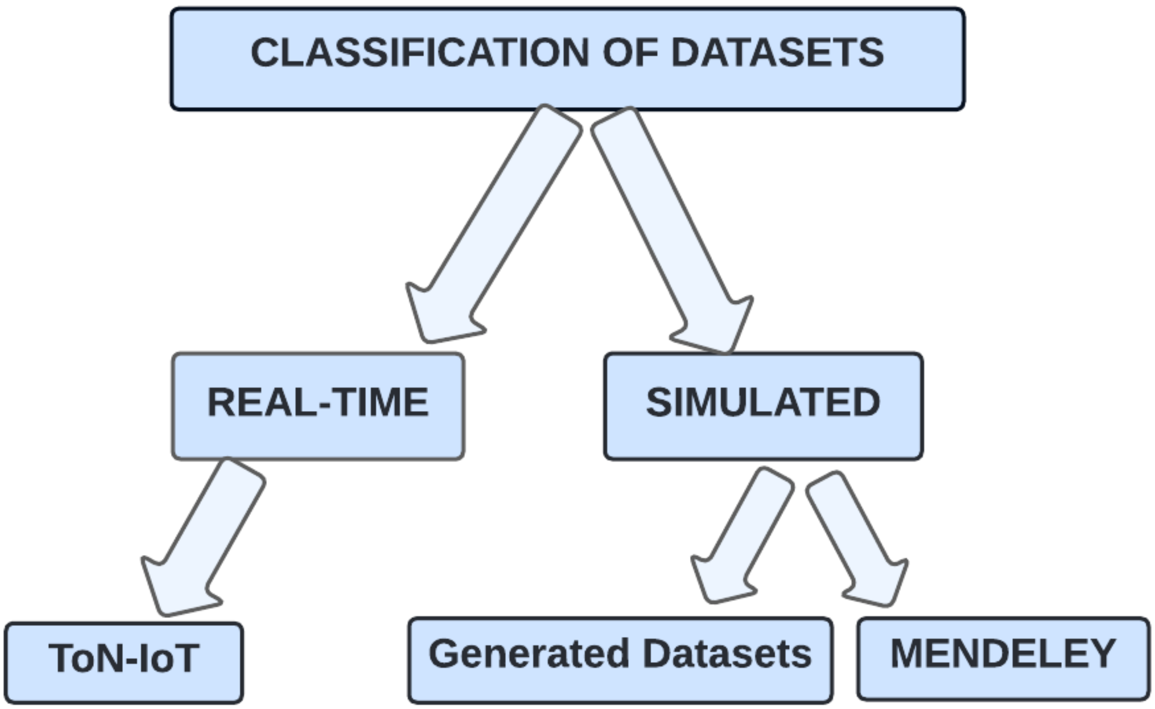


Figure 3.2 Classification of DDoS Datasets

3.1.1 Generation of DDoS Dataset in Packet Sender:

UDP traffic is generated using Packet Sender which is an open source utility that allows sending and receiving of TCP and UDP packets. Packet Sender operates at Network Layer (Layer-3), independent of switch configuration [33]. In Packet Sender tool we can define the limit of malicious traffic and time delay also protocol type can be decided either IPv4 or IPv6 as shown in Figure 3.3.

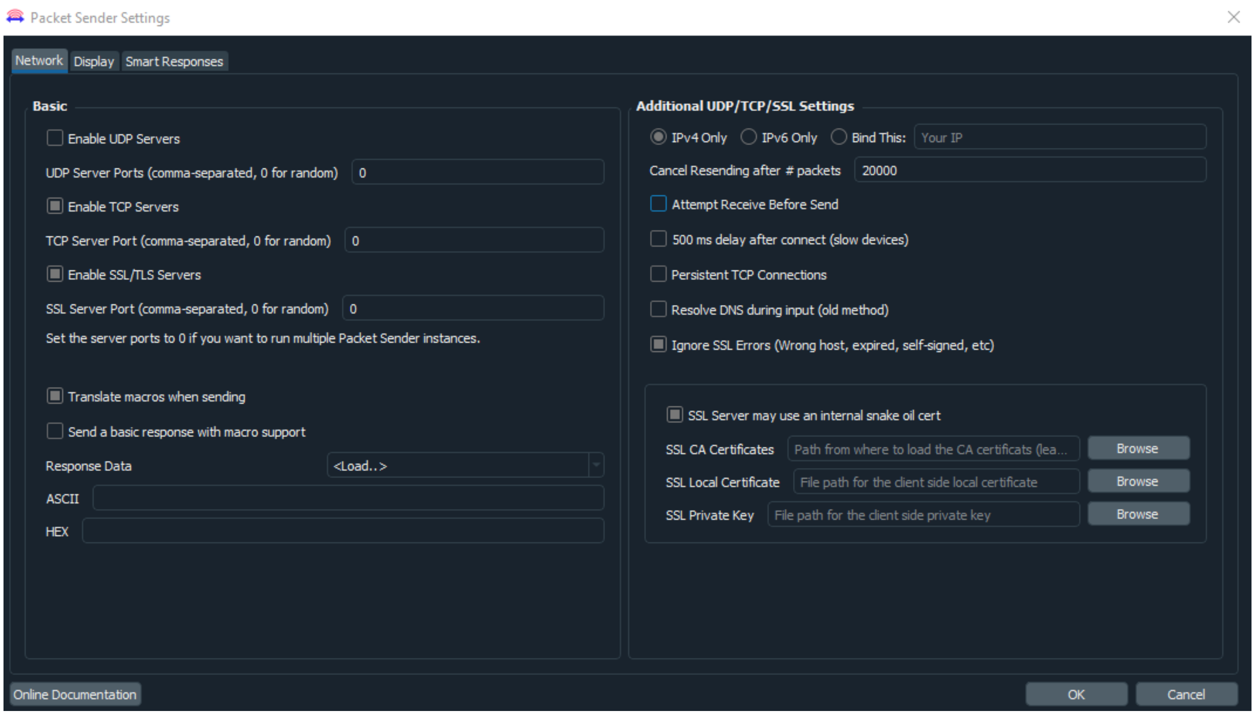


Fig 3.3 Protocol Type & Packet Limit

The dataset is generated in simulated environment (Figure 3.4) in Client-Server Architecture. Delivery of packets is on the basis of logical addressing scheme which ensures the host-to-host delivery.

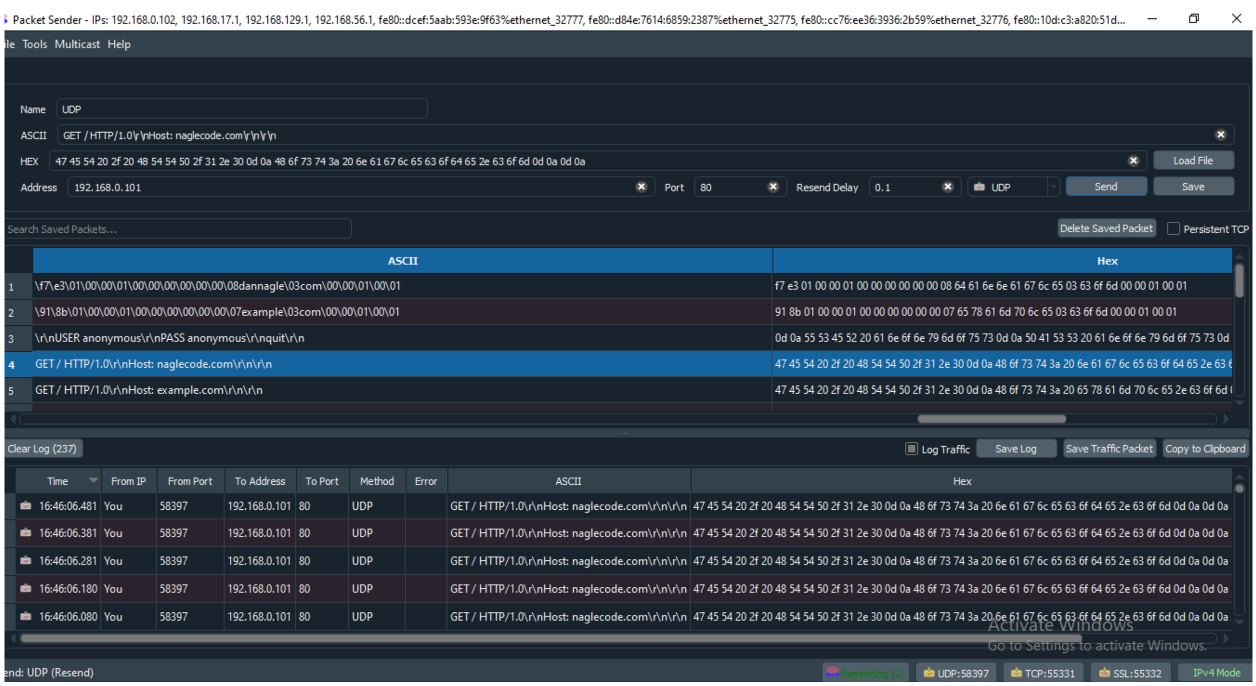


Fig 3.4 Simulated Environment of Packet Sender

In Packet Sender DDoS dataset there are nine features, which are shown in Table 3.1.

Sr #	Features	Description
1	Time	Time in seconds.
2	Source IP	The IP address of device sending the packet.
3	Destination IP	The IP address of device receiving the packet.
4	Source Port	The port of device sending the packet.
5	Destination Port	The port of device receiving the packet.
6	Method	Method represents the protocol type.
7	ASCII	ASCII value which represents Packet Size.
8	Hex Value	Hex value represents the byte count.
9	Attack, Non-Attack	Attack is represented by ‘1’ whereas normal traffic is represented by ‘0’.

Table 3.1 Description of Features of Generated DDoS Attack in Packet Sender

UDP traffic is generated on the basis of scenario shown by Figure 3.5. The destination victim IP is 192.168.0.101 whereas source IPs of hosts which are targeting destination IP are; 10.0.0.6, 10.0.0.2, 126.0.0.2, 126.0.0.5 and 192.168.1.106. Normal traffic flow is generated by the hosts which are in green having IPs; 126.0.0.3, 126.0.0.4, 126.0.0.6, 10.0.0.3, 10.0.0.7 and 192.168.1.104 towards Destination IP: 192.168.0.101.

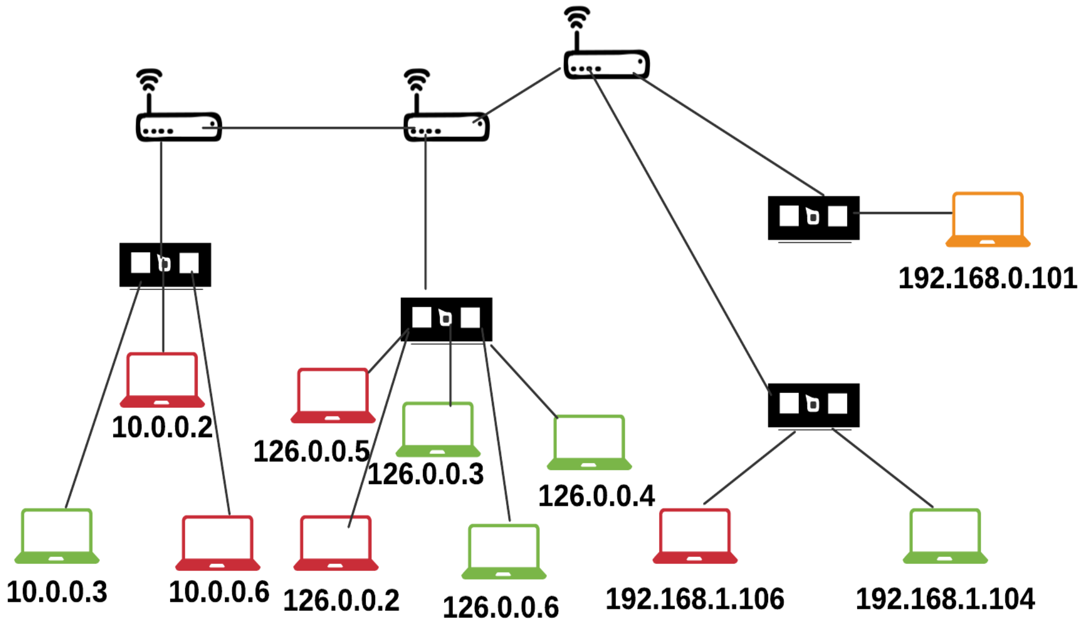


Figure 3.5 Topology of Generated DDoS Dataset in Packet Sender

3.1.1 Generation of DDoS Dataset in Mininet:

In mininet, we have generated TCP, UDP and ICMP DDoS attacks with normal traffic flow. There are three hosts; **h1**, **h2**, and **h3** which are attached to a switch **s0** and a central Ryu controller **c0** as shown in Figure 3.6**.** Host1 and host2 are attacking on host3 by using Scapy tool IP spoofing is done. Normal traffic flow is done by their IPs. Host1 has IP **10.0.0.1**, h2 has **10.0.0.2** whereas h3 has **10.0.0.3** and spoofed IPs is **10.0.0.23, 126.0.0.1, 126.0.0.2, 126.0.0.3, 192.168.0.1.** To check the configuration of topology following command (Fig 3.7) is used:

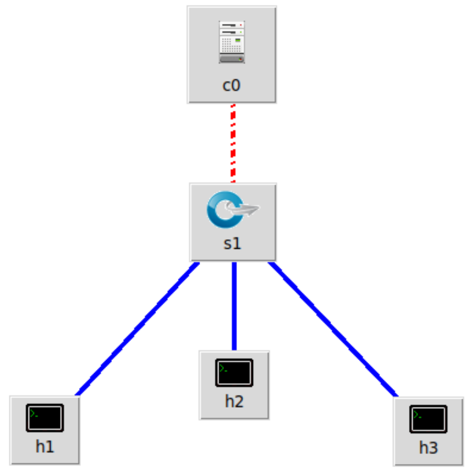


Figure 3.6 Topology of SDN

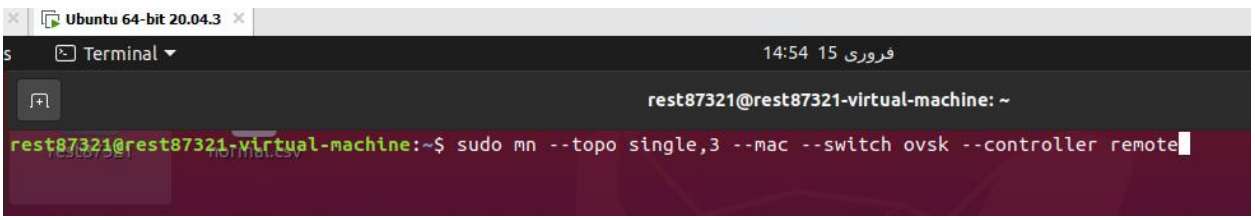


Fig 3.7 Command for Connecting Remote Controller

After this command you can see that (in Fig 3.8) Ryu remote controller is connected 127.0.0.1 with port 6653 and adding the links.

Fig 3.8 Adding Links

To check the reachability/connection used the command **pingall** h1 pings h2 and h3, h2 pings h1 and h3 whereas h3 pings h1 and h2 as shown in Fig 3.9.

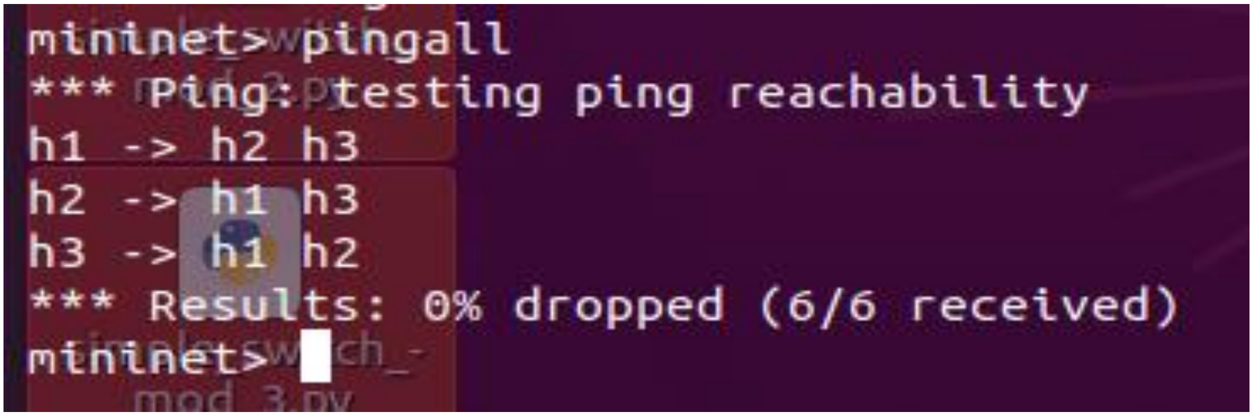


Fig 3.9 Pingall Command

Open another terminal by going to home then click on ryu folder again click on another sub-ryu folder now click on apps. In this terminal you can monitor packet flow (as shown in Fig 3.10)

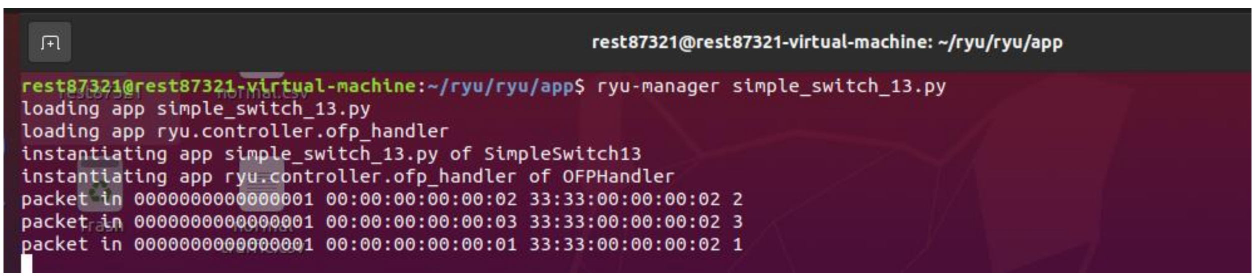


Fig 3.10 Ryu Manager

We cannot capture these packets by ryu manager therefore for packet capturing used Wireshark tool by running the command **sudo wireshark** in another terminal. By running this command wireshark window is opened as shown in Fig 3.11.

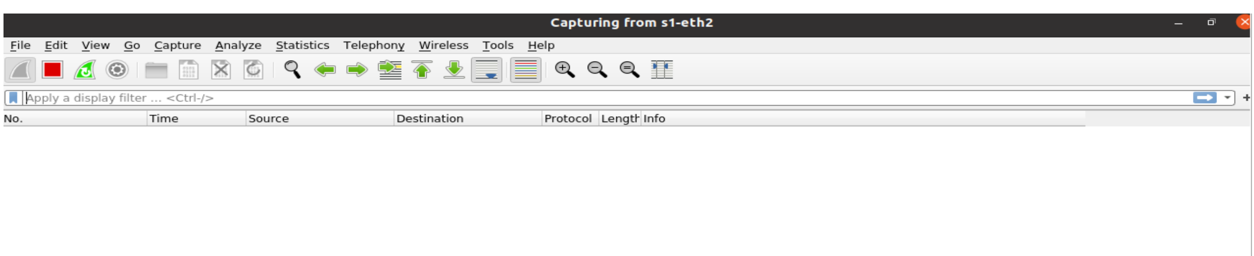


Fig 3.11 Wireshark

Now wireshark is capturing packets we run another pingall command which can be seen from Fig 3.12 that h3 sends ICMP request to h2 and h2 replied to h3.

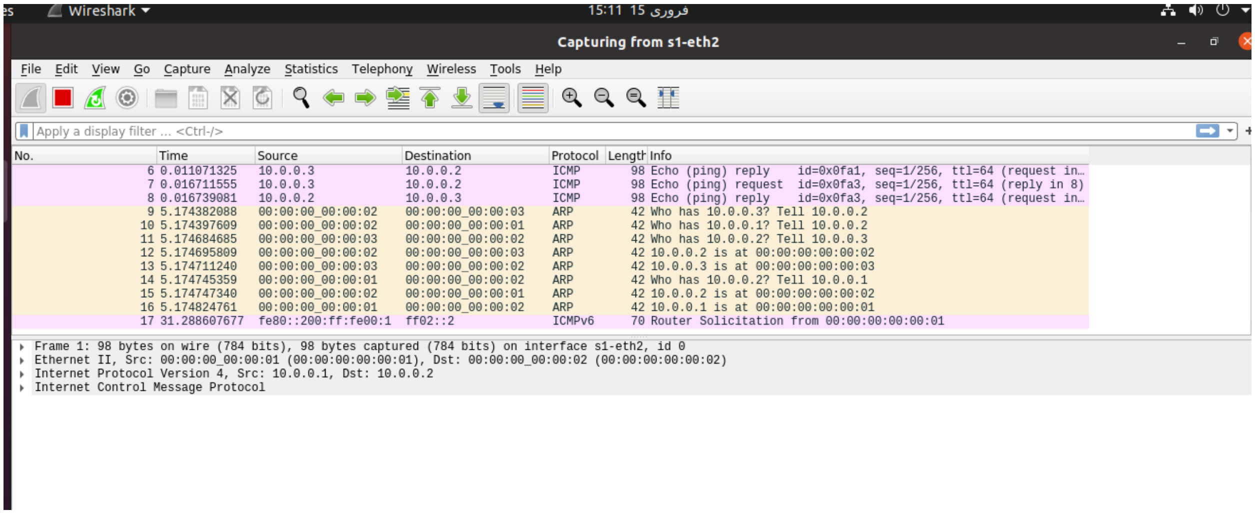


Fig 3.12 Packet Capture using Wireshark

In mininet xterm h1 command open Node: h1 window where we can run scapy command to send malicious packets to other host [32] as shown in Fig 3.13. We can open any host and send malicious traffic also we can set the packet limit and packet type like TCP, UDP and ICMP.

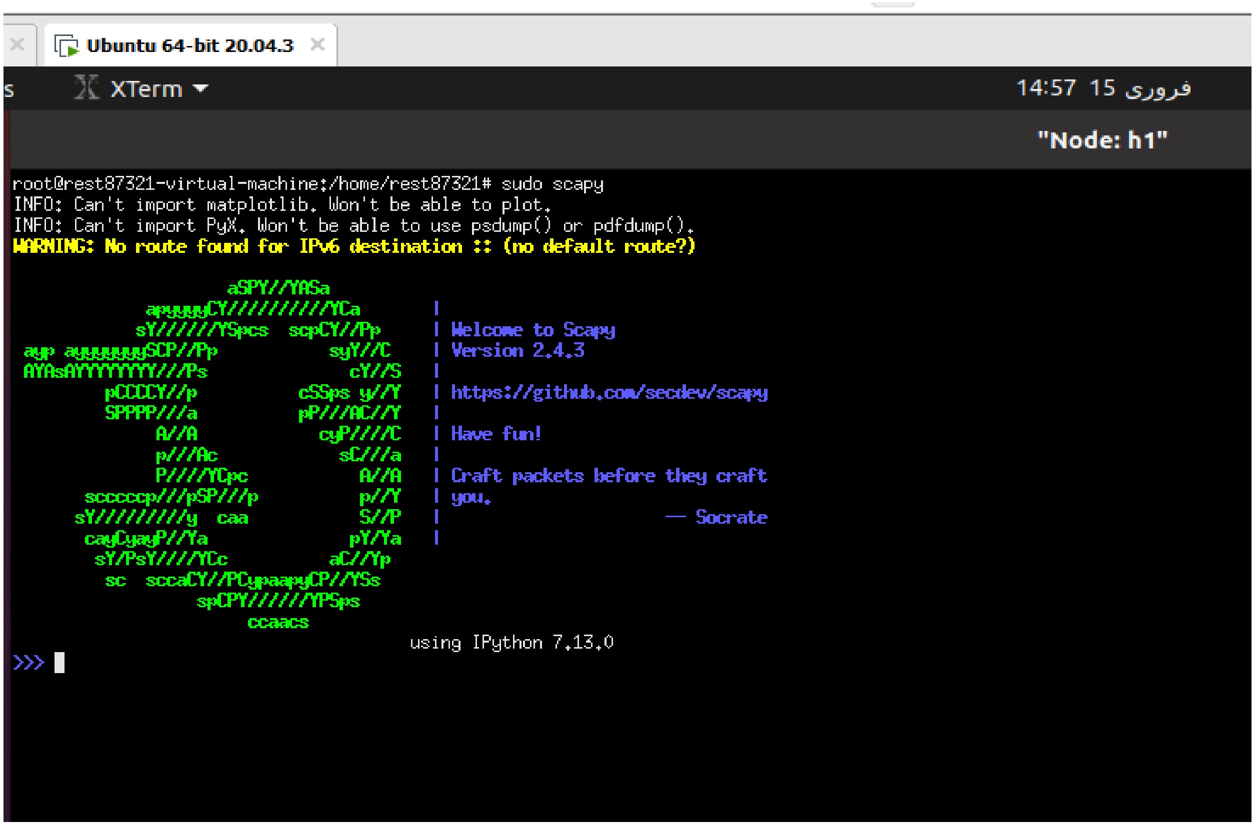


Fig 3.13 Scapy Tool

As shown in Fig 3.14 that we have sent 5 TCP packets at destination 10.0.0.3 from spoofed IP 126.0.0.3 and 9 UDP packets from IP 126.0.0.1 here dots are representing the packets.

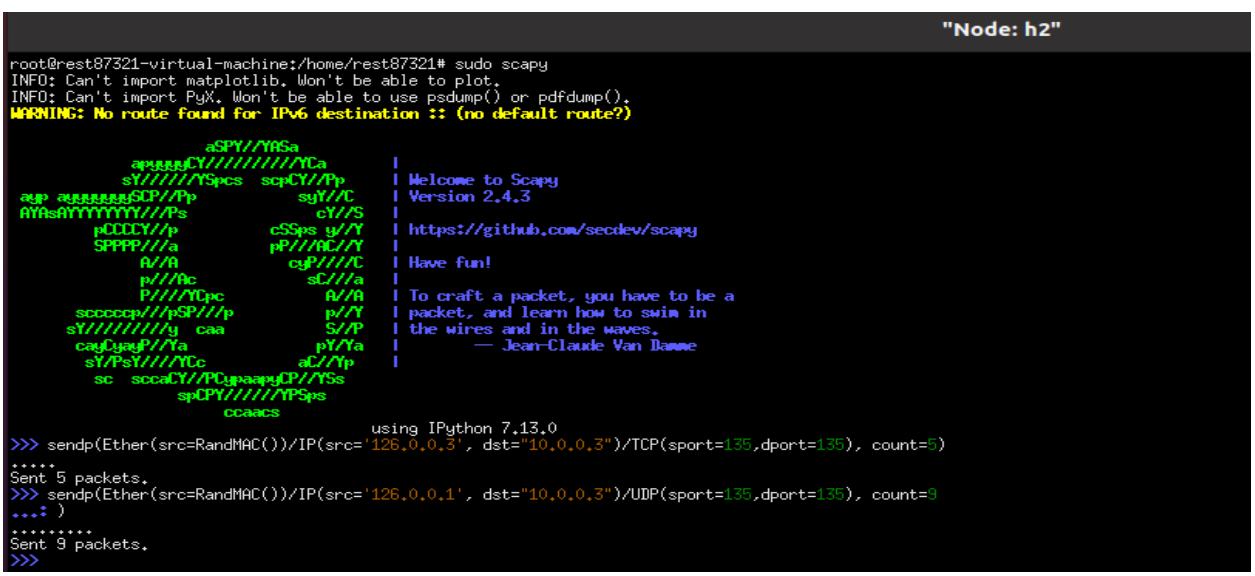


Fig 3.14 TCP and UDP Attack Command

As the packets are sending from source host to destination host Wireshark is capturing these packets which are shown in Fig 3.15.

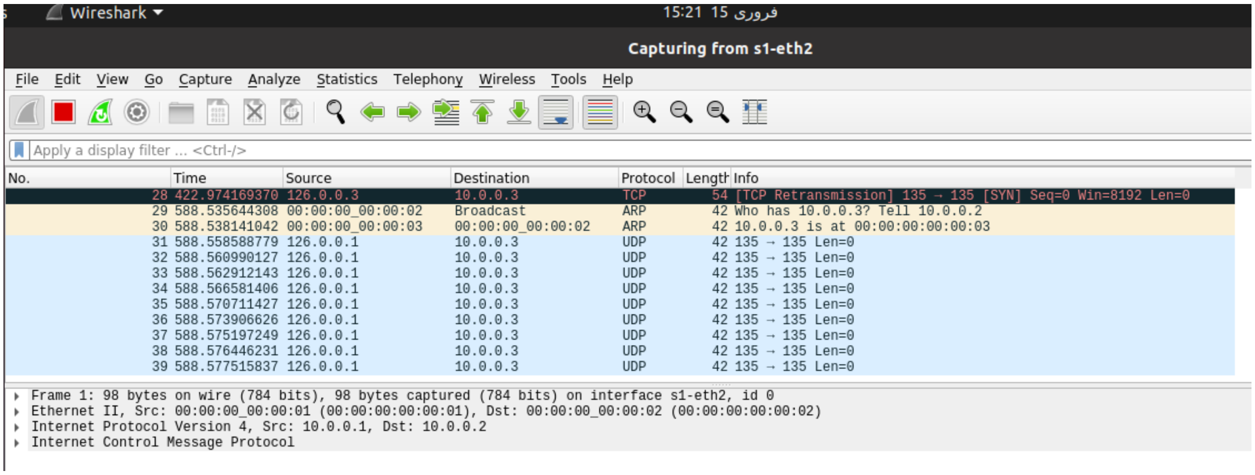


Fig 3.15 Packets in Wireshark

In our dataset, we have generated 153240 attacks and 98912 normal traffic. UDP attacks are 77163 where as TCP attacks are 39620 and ICMP attacks are 36457.

3.2.1 Available Datasets:

Following publically available datasets are used in our research:

1. **ToN-IoT:**

   It is collected from several heterogeneous sources from IIoT and IoT sensors and designed at UNSW Canberra at Australian Defence Force Academy. It was gathered in parallel manner to collect many cyber-attacks and normal traffic from a network system. This dataset has 127 features and by these features it can be seen that it is the updated one which comes after BoT-IoT and covers more attacks [34].

2. **Mendeley DDoS:**

   It is generated in simulated environment, it has 24 features and its results are high for Random forests its accuracy rate is 98.8% with minor false rate alarm depending on these features. It is simulated that's why authors said that results are high [35].

Features	Description
Time	Time in seconds
Source_IP	IP address of packet from where it was sent.
Destination_IP	IP address of packet to where it was received.
Frame Length	Length of Packet in Bytes.
Frame Number	Incremental Packet Count.
Source_Port	TCP source port of packet.
Destination_Port	TCP destination port of packet.
ACK	Acknowledgement flag of packet.
SYN	If packet is TCP then SYN flag is zero and if it is empty then it's not TCP packet.

TCP_Protocol	If packet belongs to transport layer IP it is TCP or UDP packet.
TTL	Value of packet's Time to live.
RST	Flag

Table 3.2 Description of Features of Mendeley DDoS Dataset

3.2 Tools and Technology:

Python language is used for the analysis of algorithms, also Jupyter Notebook, Anaconda and Packet Sender Tool is used in our research.

ML-based methods are used to detect DDoS attacks in a dataset [36]. As we have selected the above mentioned datasets which are different from each other on the basis of attack, non-attack, categories, some features have the value as a string whereas some are in the form of 0 (non-attack) and 1 (attack). ToN-IoT, Generated dataset and Mendeley DDoS Datasets are in the form of attack non- attack, and different categories of attacks are involved in it. Exploratory data analysis is done by Jupyter notebook in which loaded all required libraries, dataset and selected required columns. Flow of ML based Detection System is given in Figure 3.16.

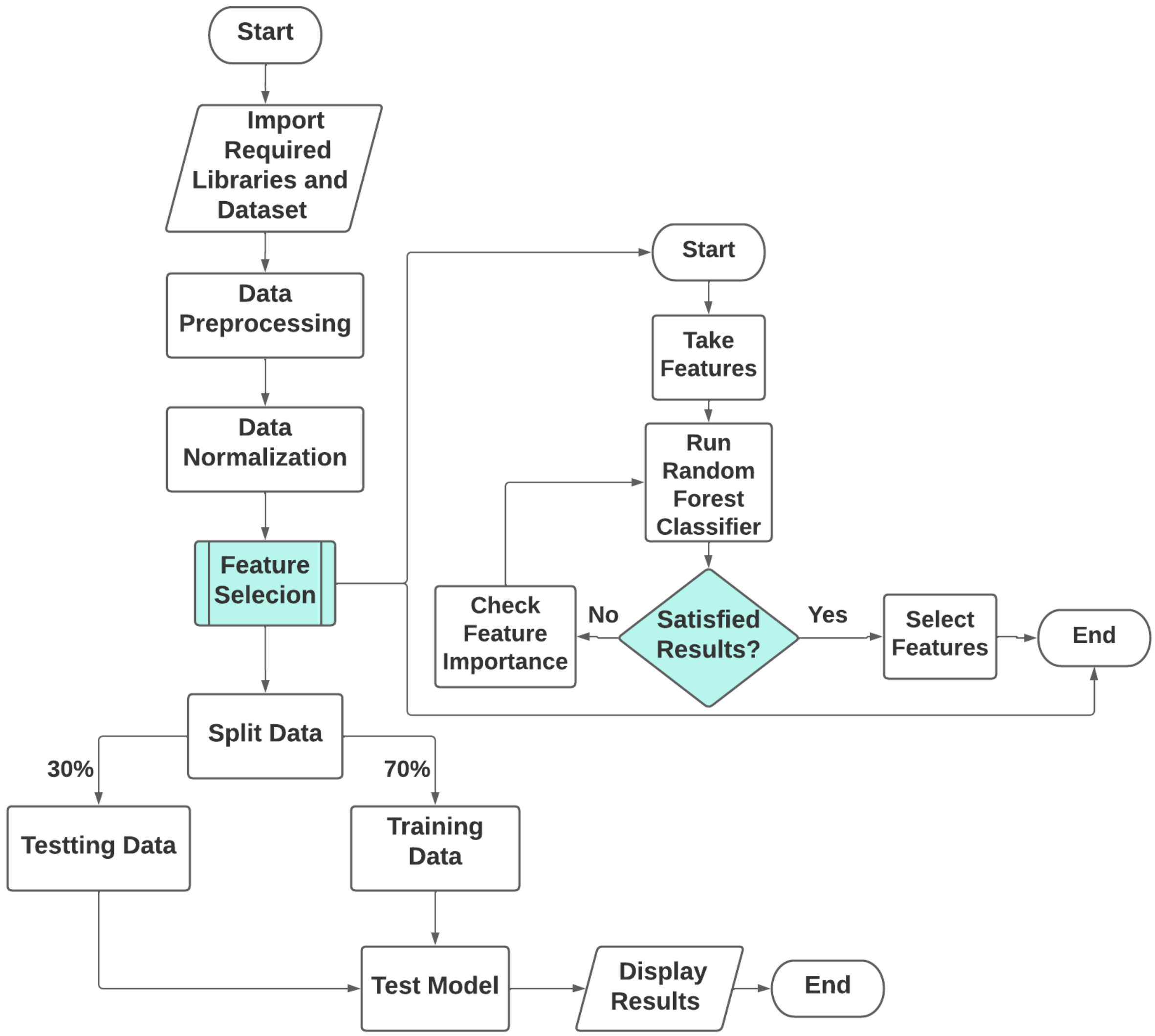


Figure 3.16 Flow of ML-based Detection System

3.3 Pre-Processing:

To enhance the performance of algorithms we only required certain features for our research therefore, data pre-processing is done on each dataset. Following data pre-processing steps were performed on datasets as shown in Figure 3.17:

i Useless columns were dropped in each dataset.

ii Blank cells were filled by 0 or -1 in some datasets.

iii As data is in mixed format (integer, objects) so it were converted into int 64 category.

iv Standard scalar were also applied.

v For data normalization label encoder is used.

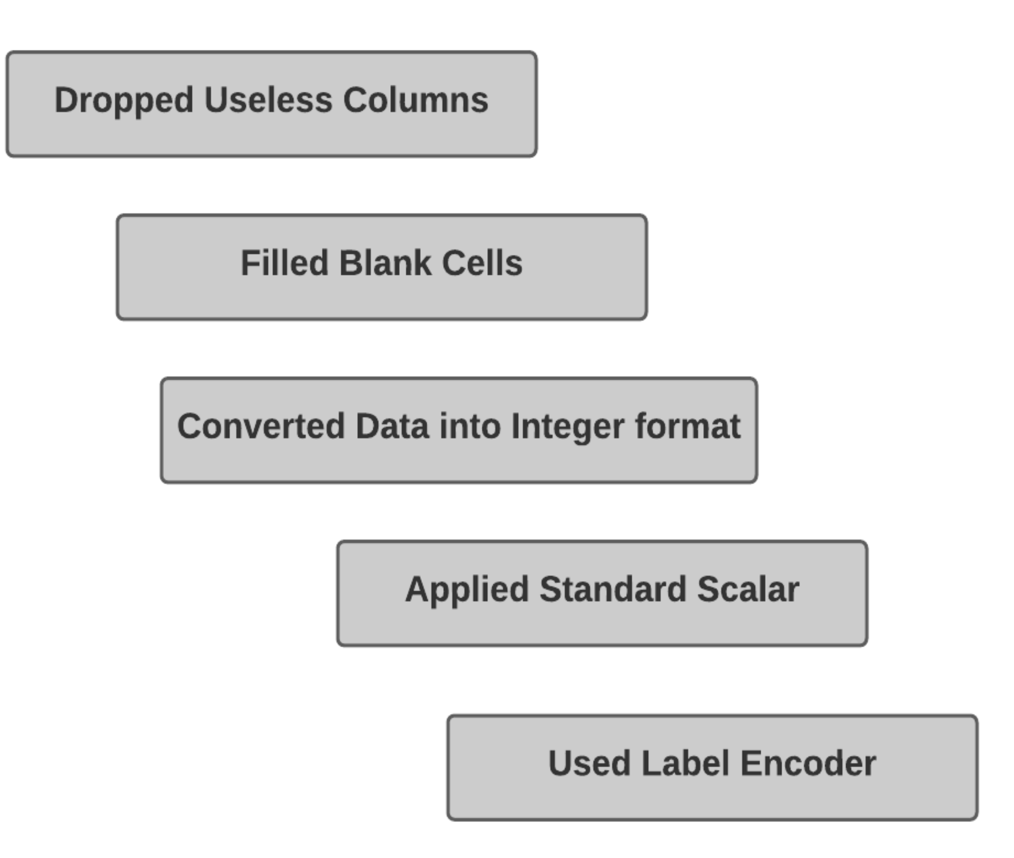


Figure 3.17 Pre-Processing Steps

Feature standardization and normalization can be done by the procedure mentioned in [37] Figure 3.18.

Sudo Code For Standard Scalar and Normalization

```
df scalar= StandardScalar()
scalar.fit()
normalized_df=scalar.transform(df)
normalized_df=Convert to Dataframe( normalized_df
le = Label Encoder()
for x in columns
            normalized_df[x] = le.fit_transform( normalized_df[x])
            best_feature =  normalized_df[all columns]
            target_features =  normalized_df[target column]
```

Figure 3.18 Sudo-Code for Standard Scalar and Normalization

3.4 ML Classifiers:

Seven different ML classifiers were used for learning different patterns. Following classifiers were used in our research:

- **Logistic Regression:** LR is a statistical analysis technique which uses to predict a value on the basis of prior known knowledge of dataset [41].
- **Decision Tree:** In DT the decision is taken by learning simple decision rules.[42]. As it can be seen from its name that for classification it uses tree structure. It gives best classification rates by making small subsets of dataset.
- **Support Vector Machine:** SVM is used for finding hyper-planes which distinguish data points [43].
- **Random Forest:** While growing the trees it adds extra randomness to the classifier. It searches and select best features when splitting any node. It produces good prediction and performs very well in both classification and regression tasks [44].
- **K- Nearest Neighbour:** It is also used for regression and classification tasks. Its learning methodology is simple; it determines the value of a point by analysing its nearest data points [45].

- **Multi-Layer Perception:** MLP can distinguish the data which is not linearly separable. It can find any abnormality by its gesture/behaviour, also it has the ability that how to do tasks on a particular given dataset [46].
- **XGBoost:** It's an implementation of Gradient Boosted Decision Trees .It is used for tabular or structured data, designed for efficient performance and learning speed [47].
- **Ensemble Methods:** It is a technique where various models are combined for better results [48].

3.5 Best Features Selection:

All columns were taken and run Random Forest classifier then results were checked and exclude certain features and order important features. Following total number of features were selected in each dataset which fits best in them for our research purpose.

3.5.1 Features of ToN-IoT:

In ToN-IoT, there are 127 features in total, after pre-processing the selected features are 50 which were useful in our research.

3.5.2 Features of Mendeley DDoS:

Mendeley DDoS is a simulated dataset it has 24 features in it and here we selected 20 features which are highly contributing in it. Feature selection is done by evaluating Random Forest classifier.

3.5.3 Feature Importance of Generated DDoS Dataset in Mininet:

There are total 8 features from which we have used 5 features in our research which are contributing highly. Selected features are; Time, Source, Destination, Protocol and Length.

3.5.4 Feature Importance of Generated DDoS Dataset in Packet Sender:

We generate it in simulated environment in packet sender tool; it has nine features in total and all features are important in detecting DDoS attack.

Chapter 4 Results

In this section, results were produced on the basis of feature selection. First of all, we have selected best features of each dataset by calculating feature importance of given dataset and then we provided the Grid Search CV our desired algorithms and possible hyper-parameters which returned us the best parameters for each specific algorithm, then all the algorithms were trained based on those hyper-parameters and produced results.

It can be seen from Table 4.1 that accuracy rate of different algorithms in every dataset has different. All datasets have high accuracy rate of decision tree and random forest and XGBoost whereas Logistic Regression have least accuracy rate among all datasets. Here the results are on the basis of individual best features of datasets.

Algorithms	Accuracy of ToN-IoT	Accuracy of Mendeley	Accuracy of SDN	Accuracy of Packet Sender
Logistic Regression	90.12%	64.06%	91.64%	62.93%
K-Nearest Neighbor	98.47%	93.71%	99.79%	97.48%
Multi-Layer Perception	95.26%	78.26%	90.52%	65.54%
Decision Tree	97.07%	94.45%	97.53%	100%
Random Forest	99.29%	99.98%	99.66%	100%
XG-Boost	98.44%	99.99%	99.79%	100%
Support Vector Machine	91.50%	73.82%	63.20%	62.98%
Ensemble	99.27%	99.90%	99.23%	100%

Table 4.1 Results of All Datasets

4.1.1 Results of ToN-IoT Dataset:

In this dataset, the performance of every algorithm is good however RF, XGB and KNN performed very well in less time whereas SVM takes a lot of time. Following are the confusion matrixes of each algorithm which shows accuracy rate and misclassification rate.

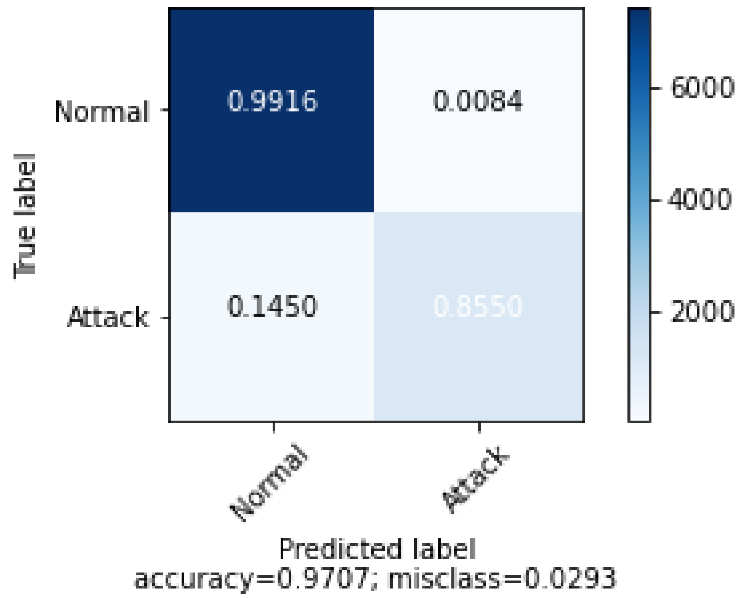


Figure 4.1.1 Confusion Matrix of Decision Tree of ToN-IoT

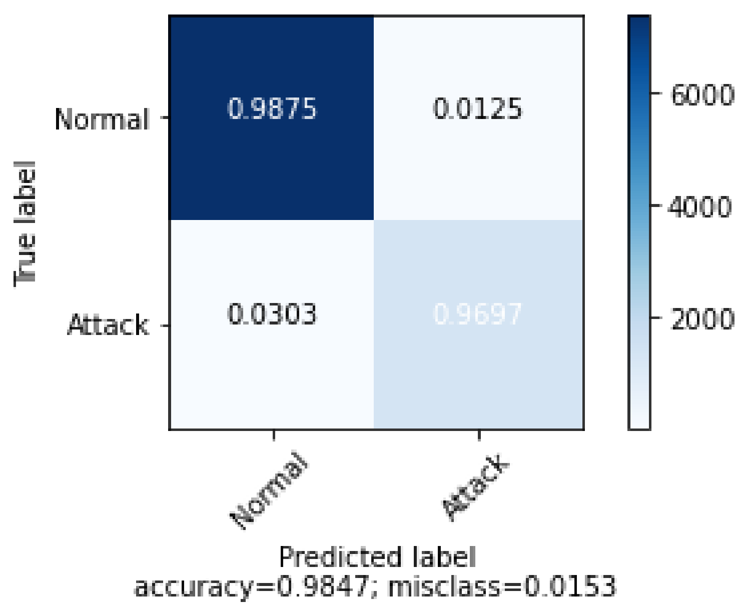


Figure 4.1.2 Confusion Matrix of K-Nearest Neighbour of ToN-IoT

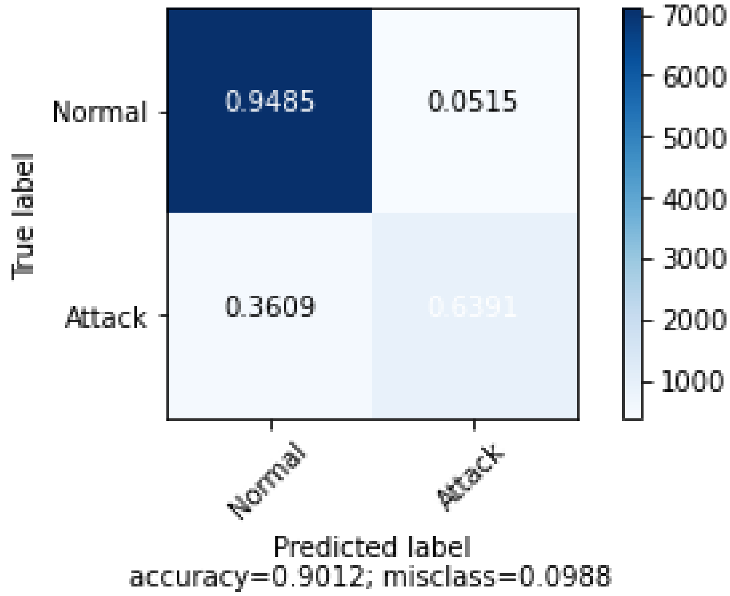


Figure 4.1.3 Confusion Matrix of Logistic Regression of ToN-IoT

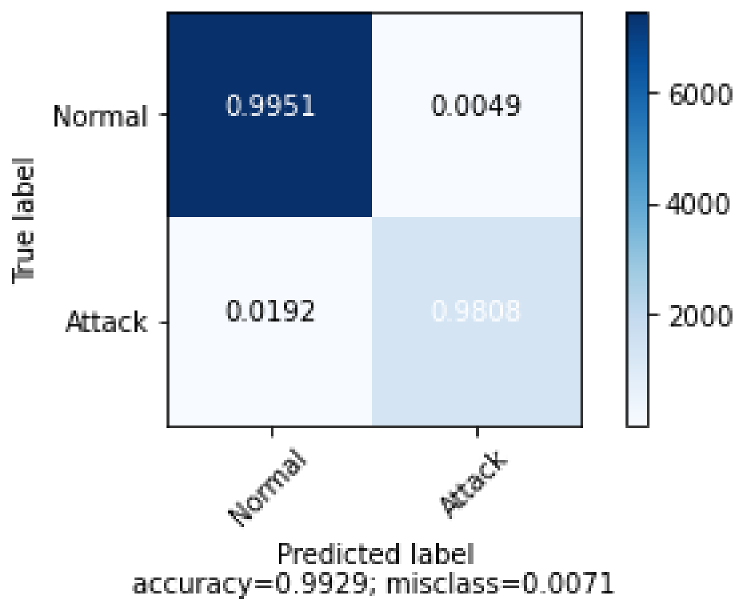


Figure 4.1.4 Confusion Matrix of Random Forest of ToN-IoT

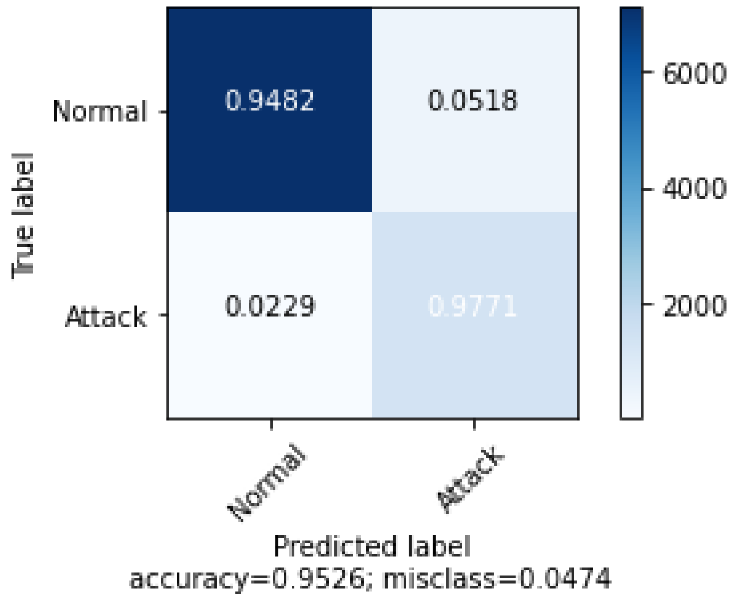


Figure 4.1.5 Confusion Matrix of Multi-Layer Perception of ToN-IoT

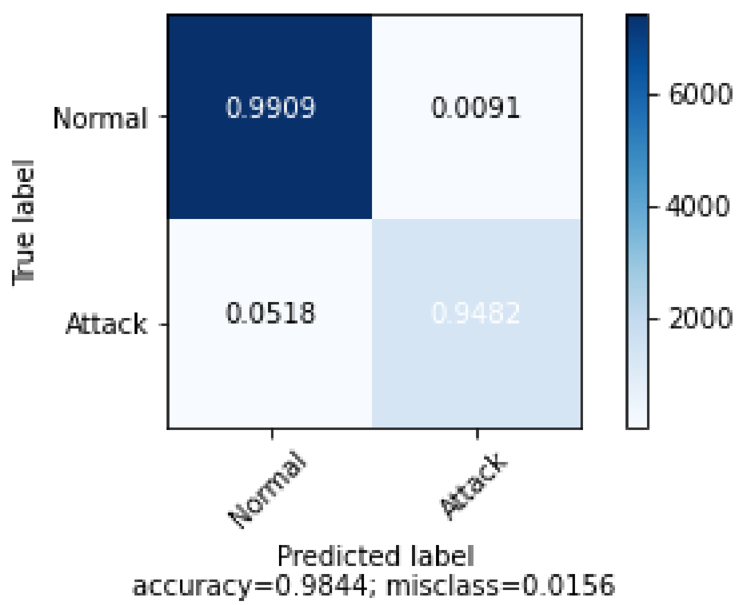


Figure 4.1.6 Confusion Matrix of XG-Boost of ToN-IoT

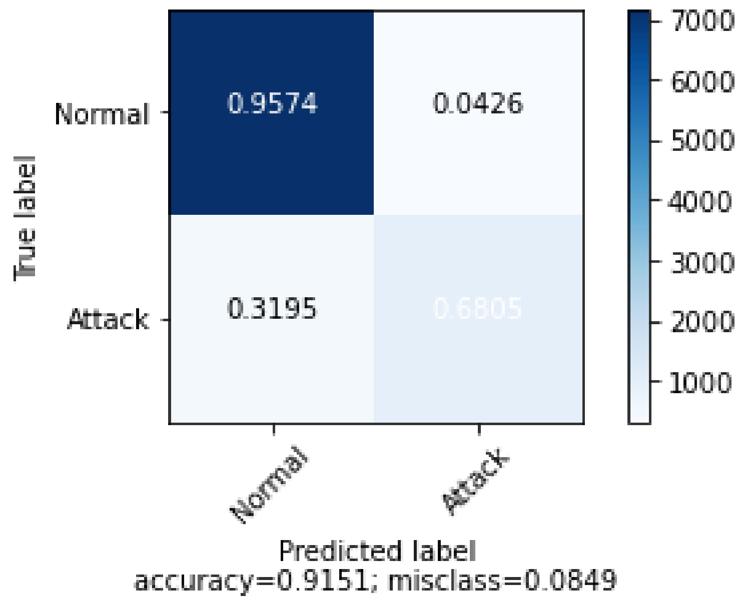


Figure 4.1.7 Confusion Matrix of Support Vector Machine of ToN-IoT

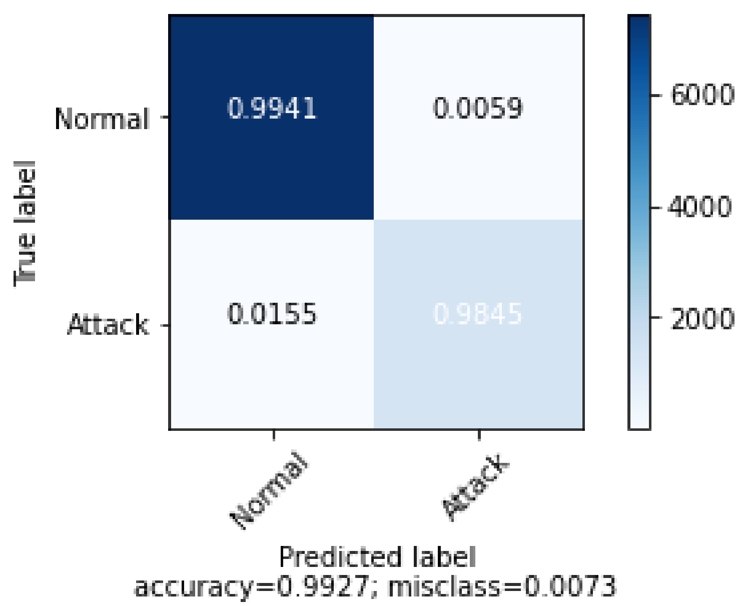
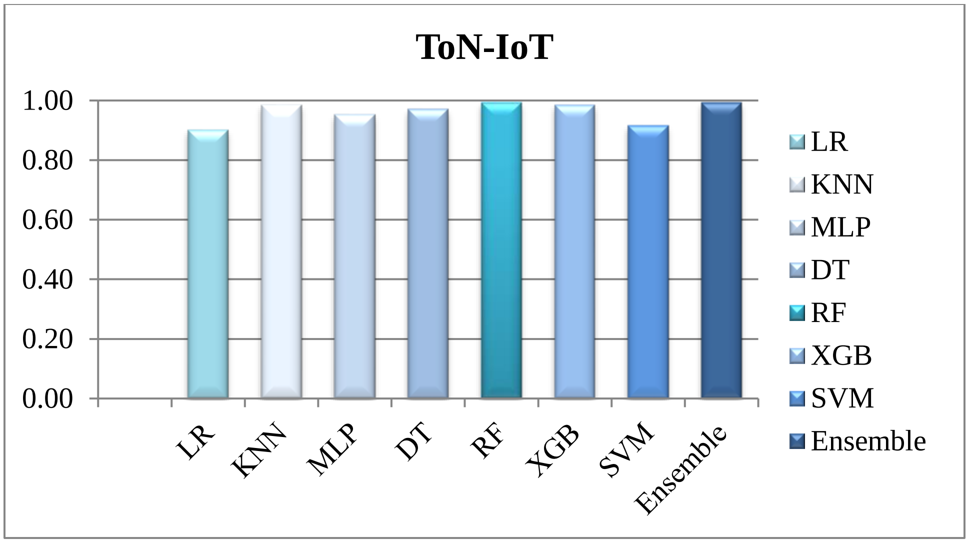


Figure 4.1.8 Confusion Matrix of Ensemble Method of ToN-IoT



Accuracy Bar Graph 4.1.1 of ToN-IoT

4.1.2 Results of Mendeley DDoS Dataset:

As this dataset is simulated so we achieved 94.45% accuracy rate of Decision Tree and for Random Forest 99.98% whereas the accuracy of XG-Boost is 99.99%. In this dataset other algorithms such as Logistic Regression achieved 64.06% accuracy rate whereas Multi-Layer Perception achieved 78.26% and Support Vector Machine have 73.82% accuracy rate, overall these algorithms performed well but in comparison of DT, RF and XGB have low accuracy rate. As shown in following confusion matrices:

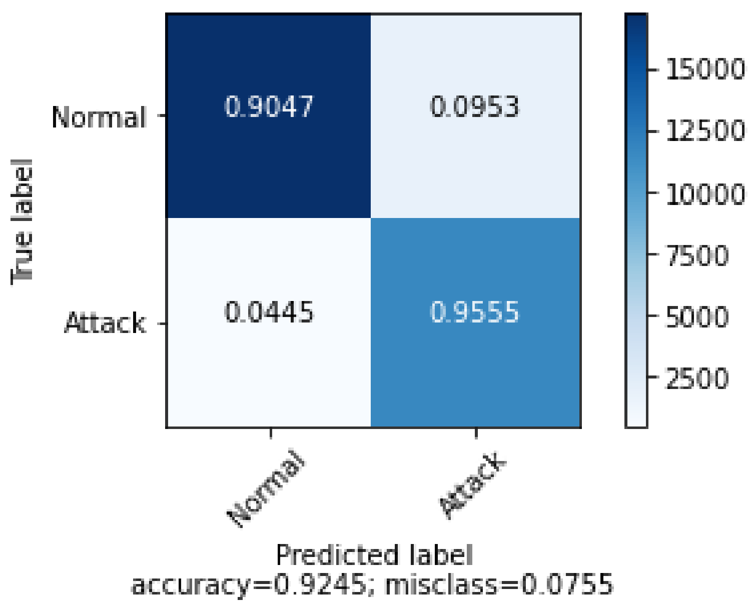


Figure 4.2.1 Confusion Matrix of Decision Tree of Mendeley

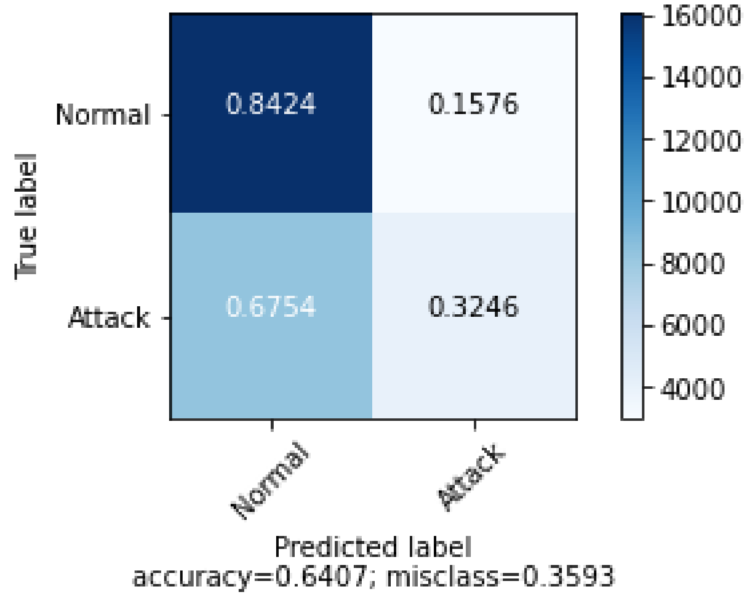


Figure 4.2.2 Confusion Matrix of Logistic Regression of Mendeley

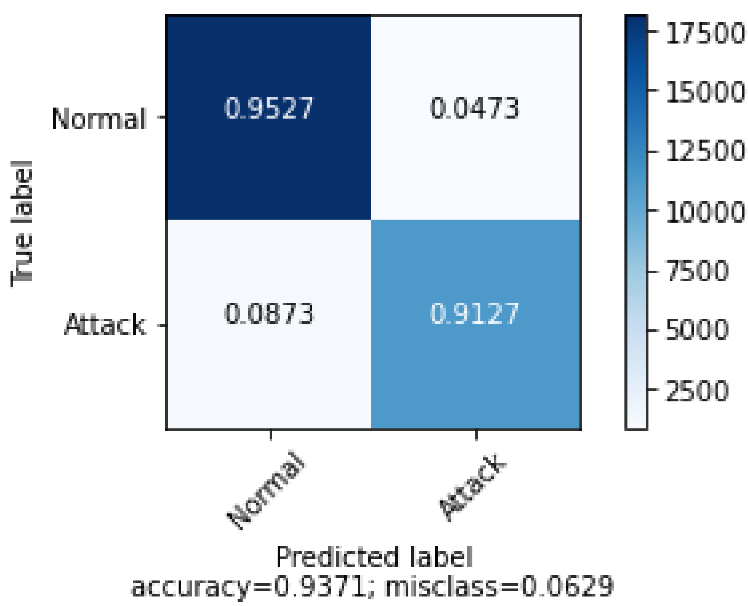


Figure 4.2.3 Confusion Matrix of K-Nearest Neighbour of Mendeley

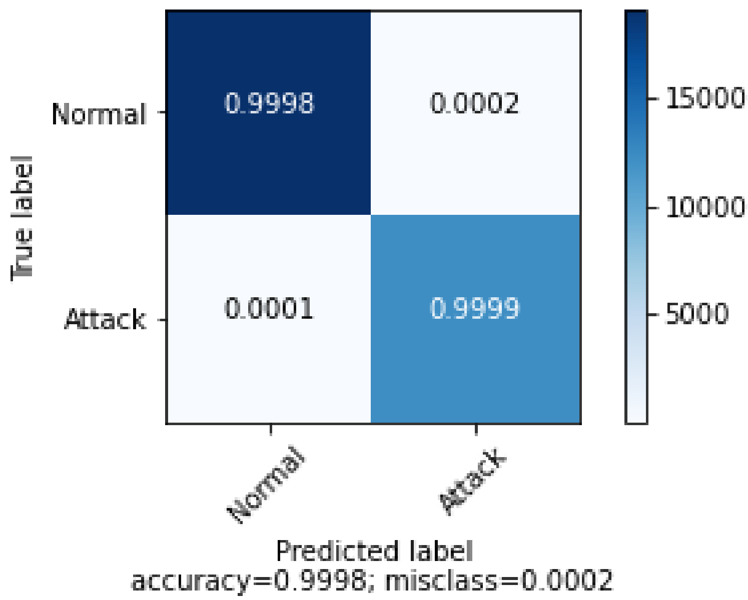


Figure 4.2.3 Confusion Matrix of Random Forest of Mendeley

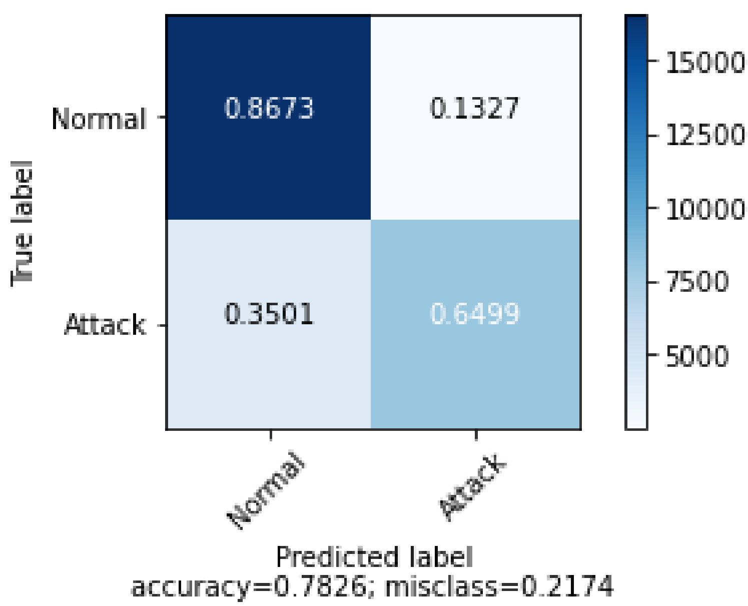


Figure 4.2.4 Confusion Matrix of Multi-Layer Perception of Mendeley

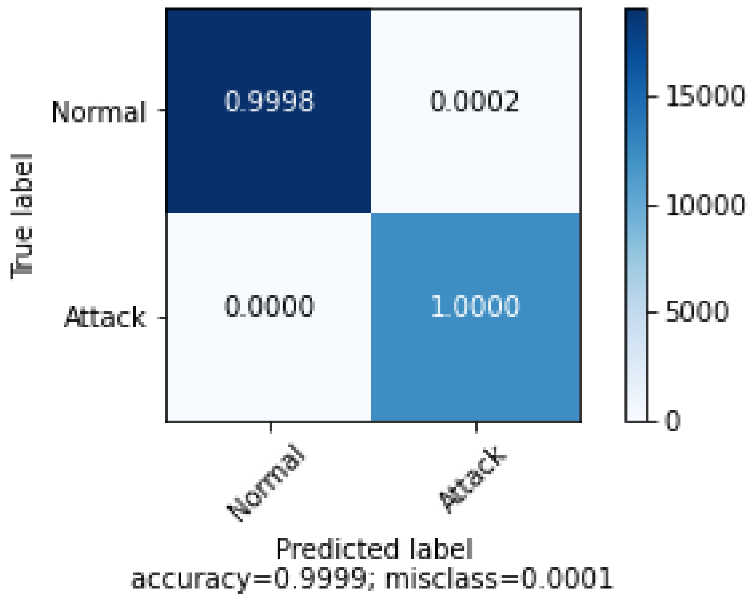


Figure 4.2.5 Confusion Matrix of XG-Boost of Mendeley

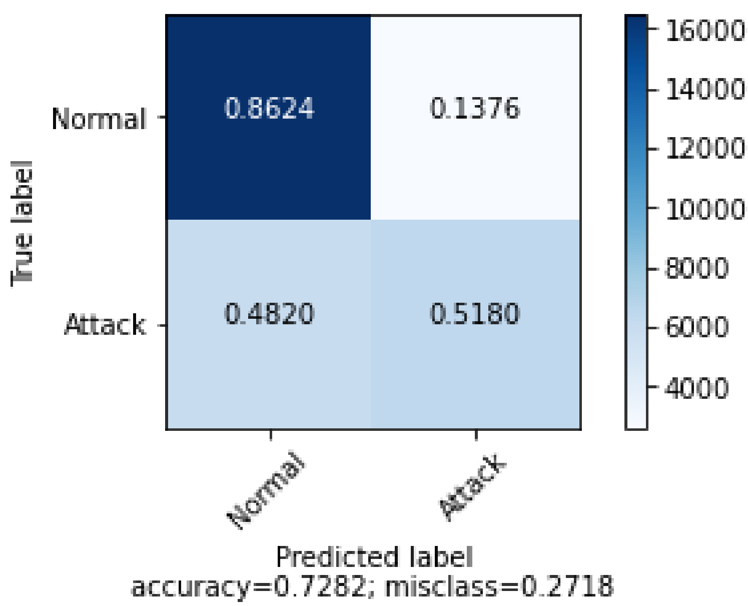


Figure 4.2.6 Confusion Matrix of Support Vector Machine of Mendeley

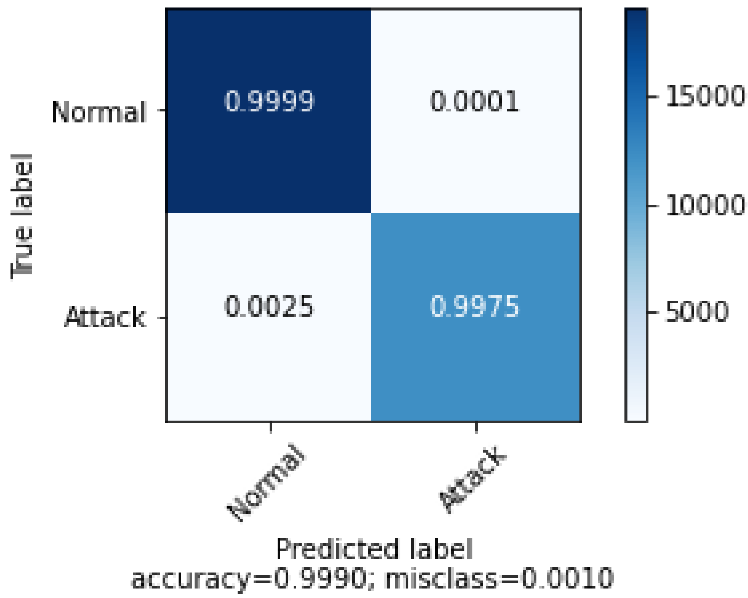
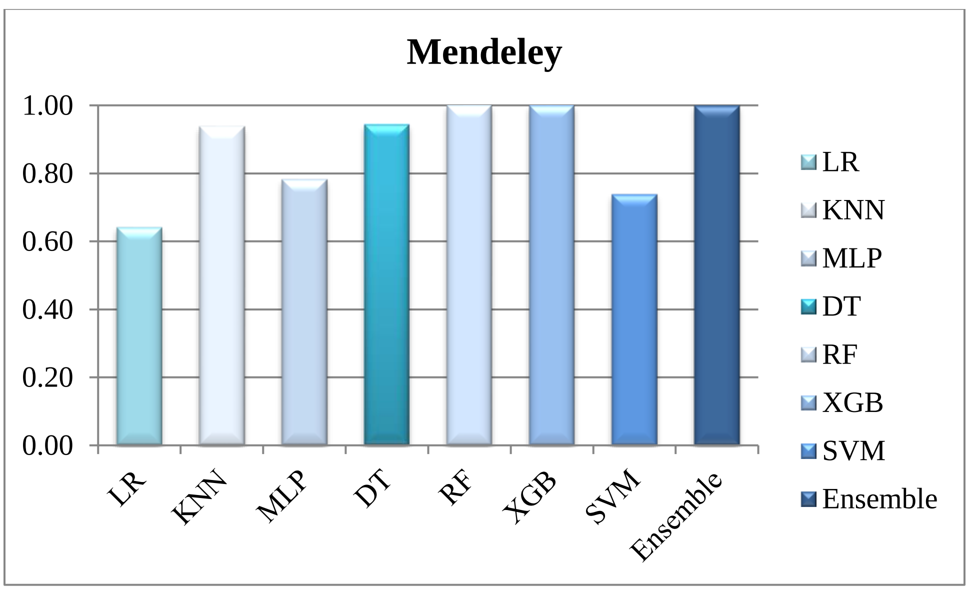


Figure 4.2.7 Confusion Matrix of Ensemble Method of Mendeley



Accuracy Bar Graph.4.1.2 Results of Mendeley

4.1.3 Results of Generated DDoS Dataset in Mininet:

The results are on the basis of malicious and normal traffic; every algorithm performs very well in detecting DDoS attack but four algorithms are giving high accuracy rate such as KNN, DT, RF and XGB also ensemble method is applied on DT, RF and XGB algorithms and as a result accuracy of ensemble method is 99.23%. Following are the normalized Confusion Matrices of algorithms:

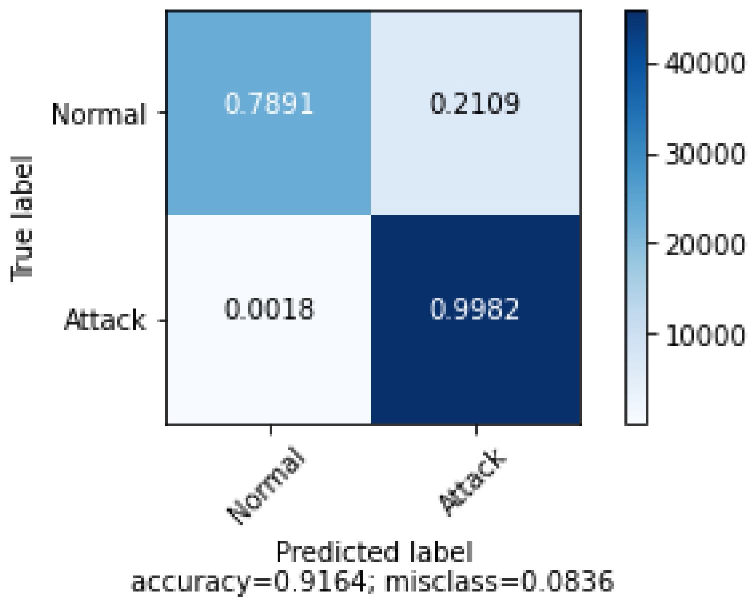


Figure 4.3.1Confusion Matrix of Logistic Regression of Mininet

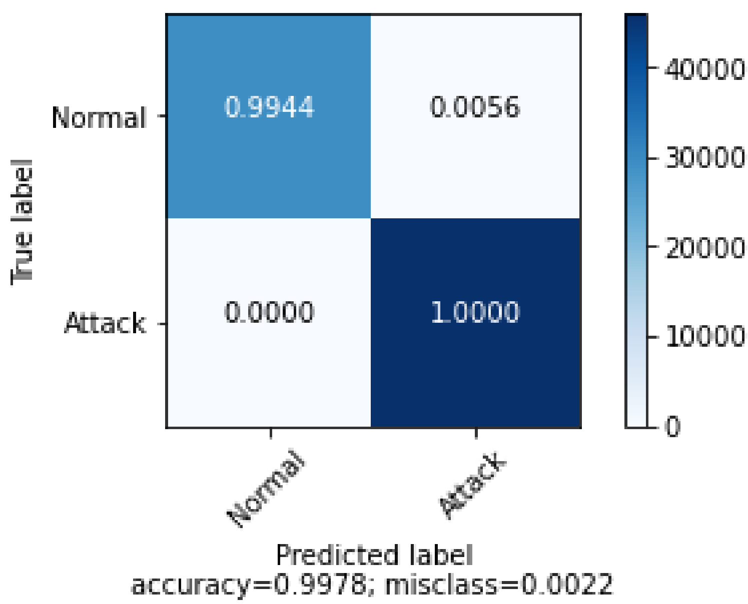


Figure 4.3.2 Confusion Matrix of K-Nearest Neighbor of Mininet

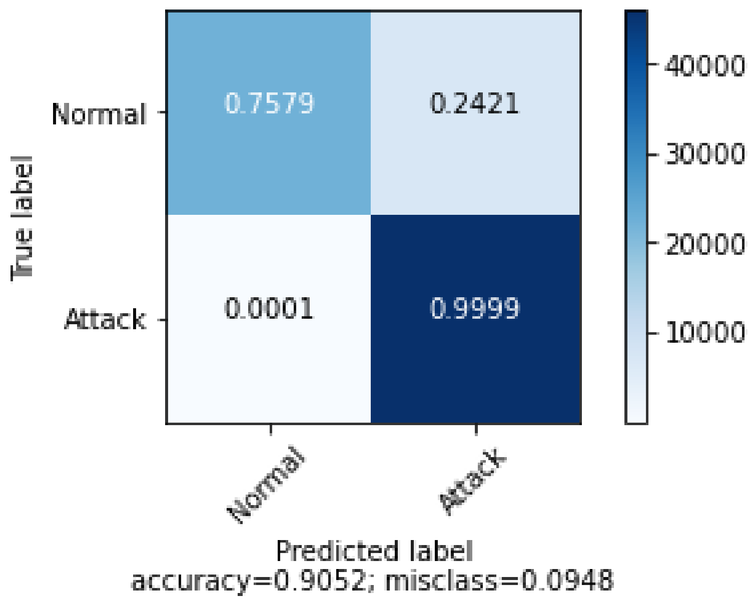


Figure 4.3.2 Confusion Matrix of Multi-Layer Perception of Mininet

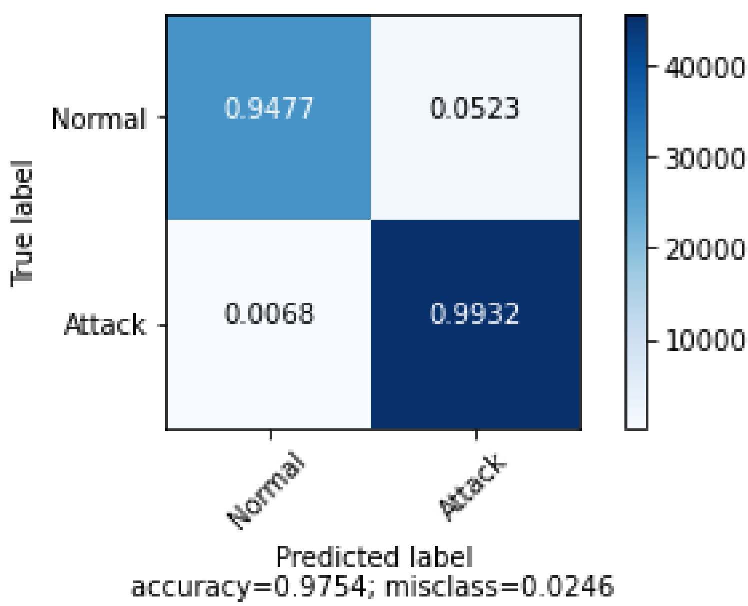


Figure 4.3.3 Confusion Matrix of Decision Tree of Mininet

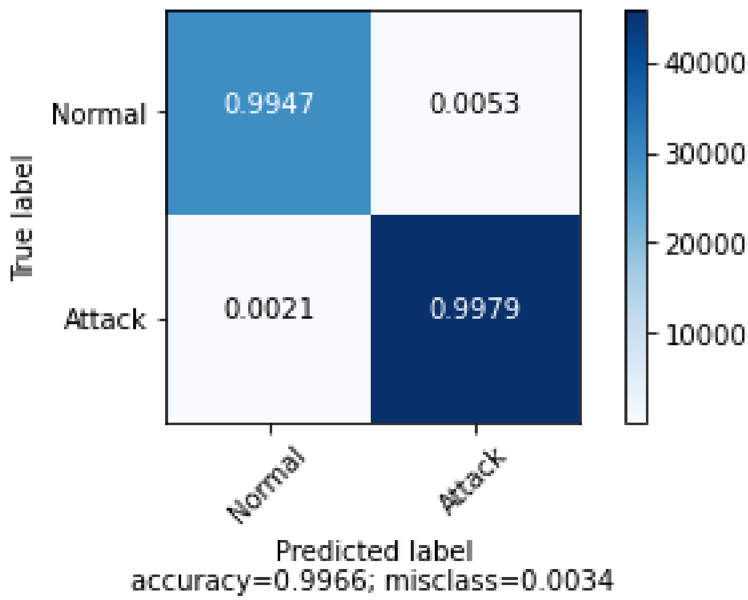


Figure 4.3.4 Confusion Matrix of Random Forest of Mininet

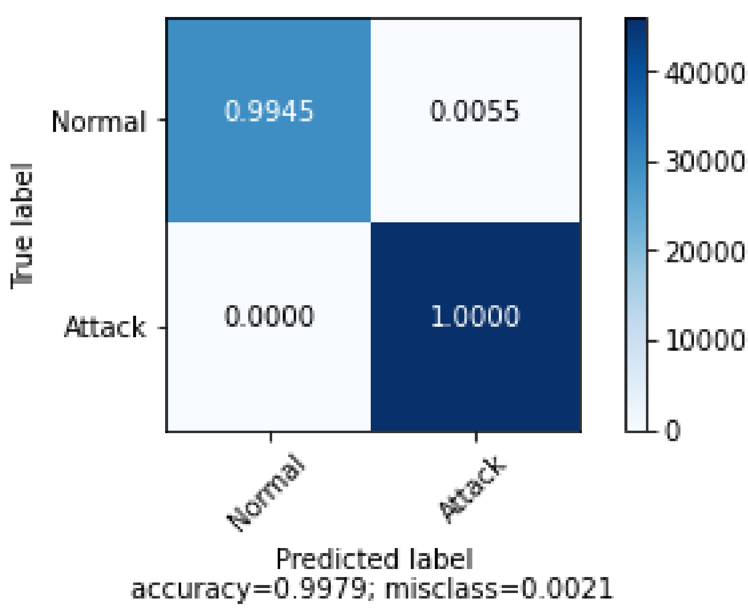


Figure 4.3.5 Confusion Matrix of XG-Boost of Mininet

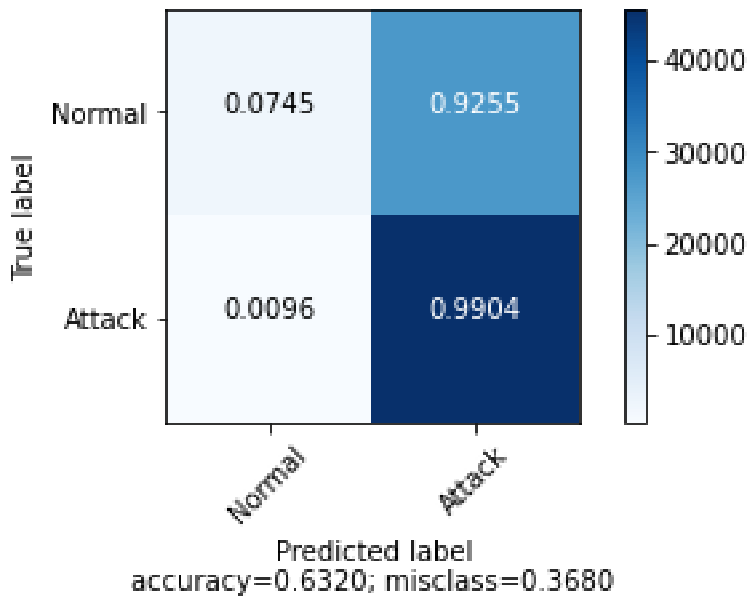


Figure 4.3.6 Confusion Matrix of Support Vector Machine of Mininet

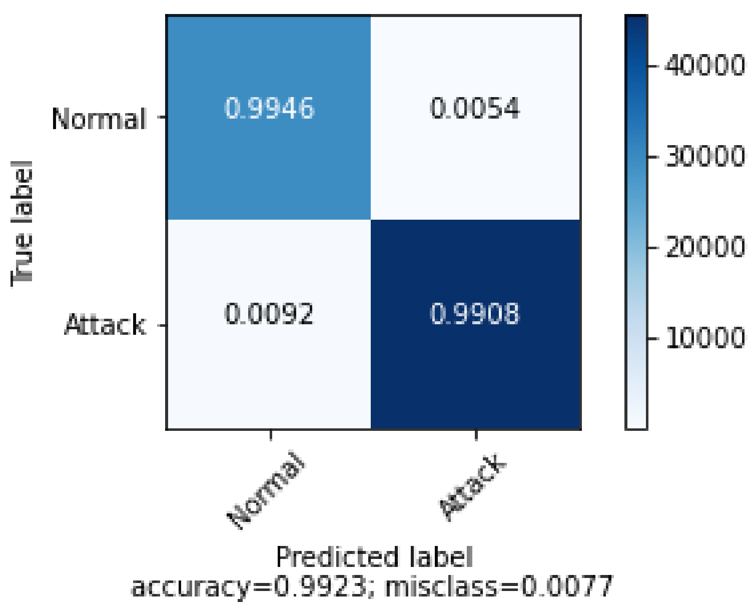
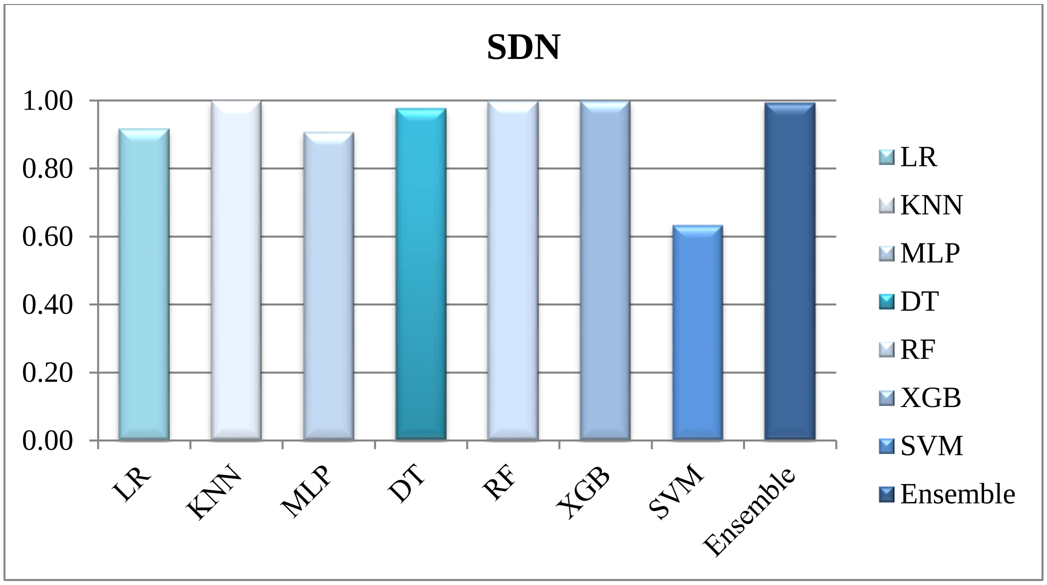


Figure 4.3.7 Confusion Matrix of Ensemble Method of Mininet



Accuracy Bar Graph 4.1.3 of SDN

4.1.4 Results of Generated DDoS Dataset in Packet Sender:

Five algorithms performed very well in this dataset; Decision Tree, Random Forest, XG-Boost and Ensemble Method by giving 100% and KNN gave 97.48% accuracy rate whereas Logistic Regression and SVM didn't perform well by giving accuracy rate of 62.93% and 62.98% respectively. Following are the Confusion Matrices of algorithms:

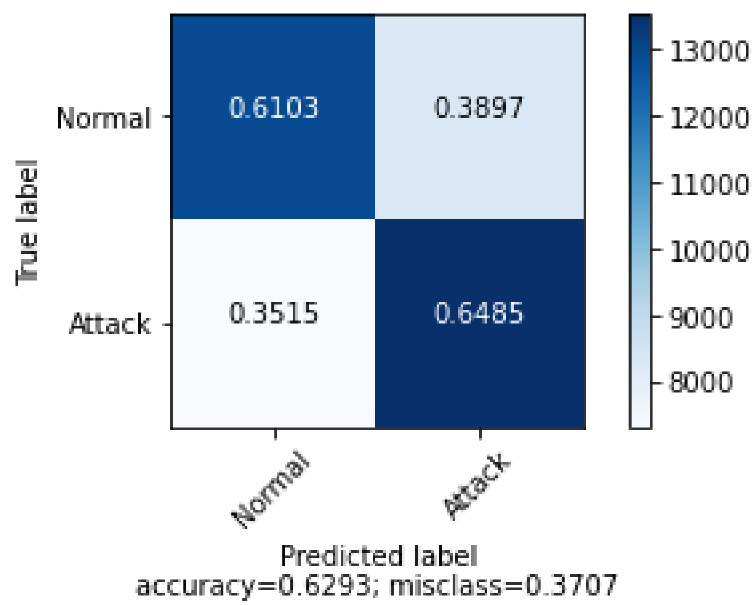


Figure 4.4.1 Confusion Matrix of Logistic Regression of PS

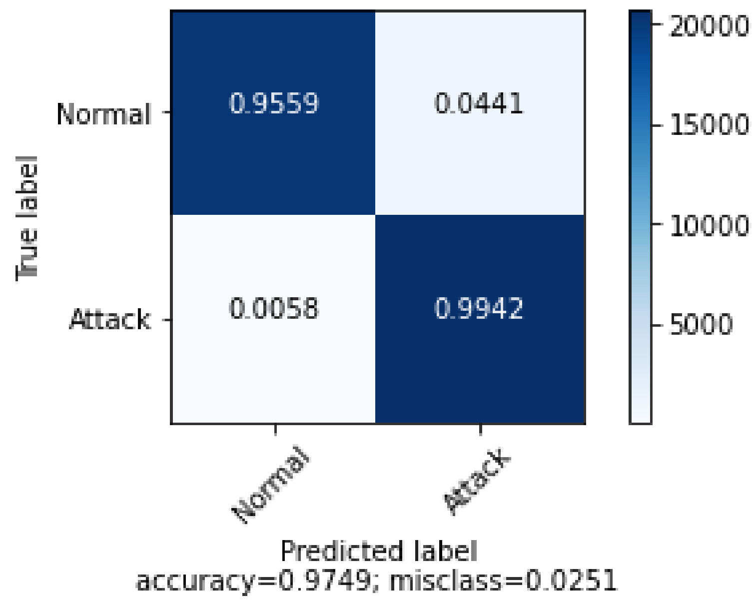


Figure 4.4.2 Confusion Matrix of K-Nearest Neighbour of PS

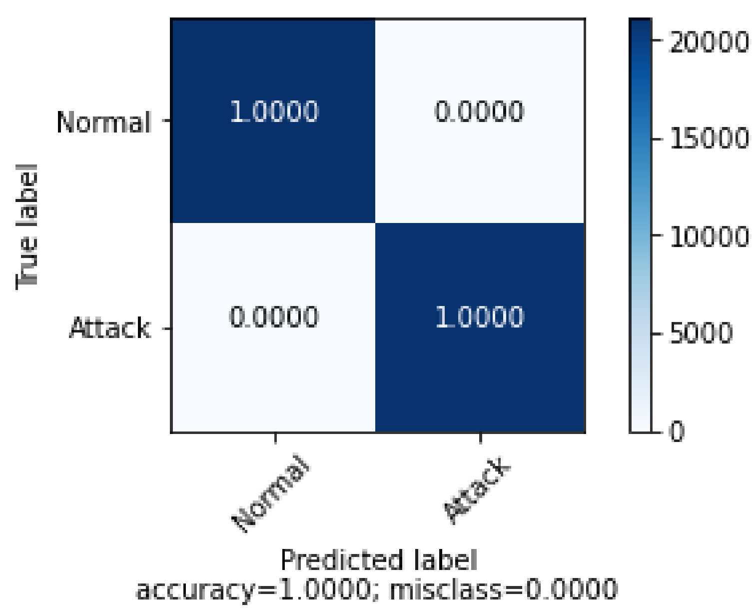


Figure 4.4.3 Confusion Matrix of Decision Tree of PS

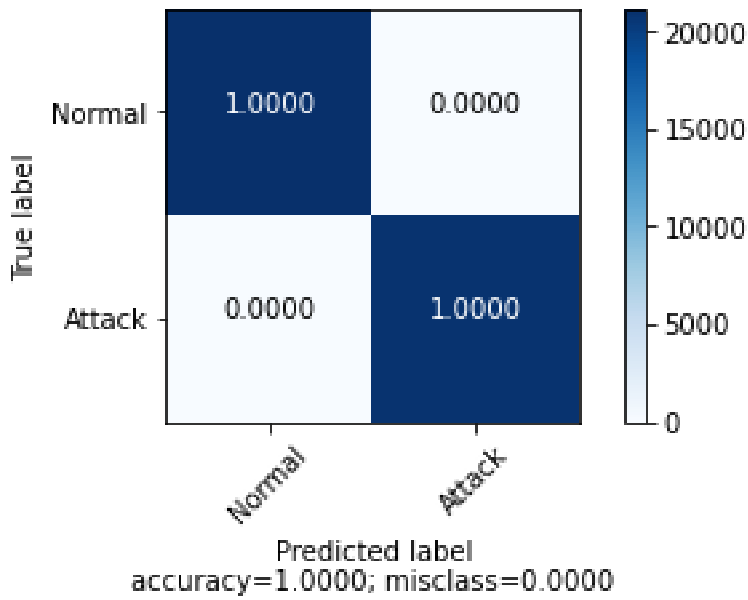


Figure 4.4.4 Confusion Matrix of Random Forest of PS

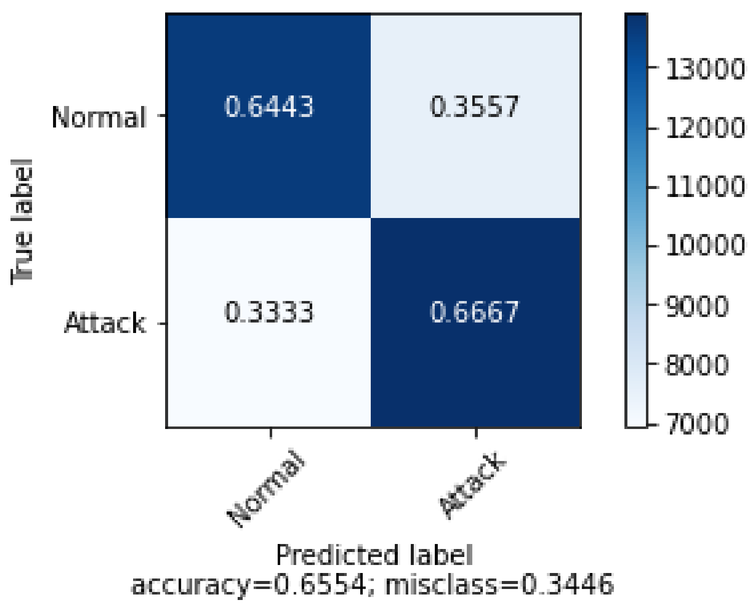


Figure 4.4.5 Confusion Matrix of Multi-Layer Perception of PS

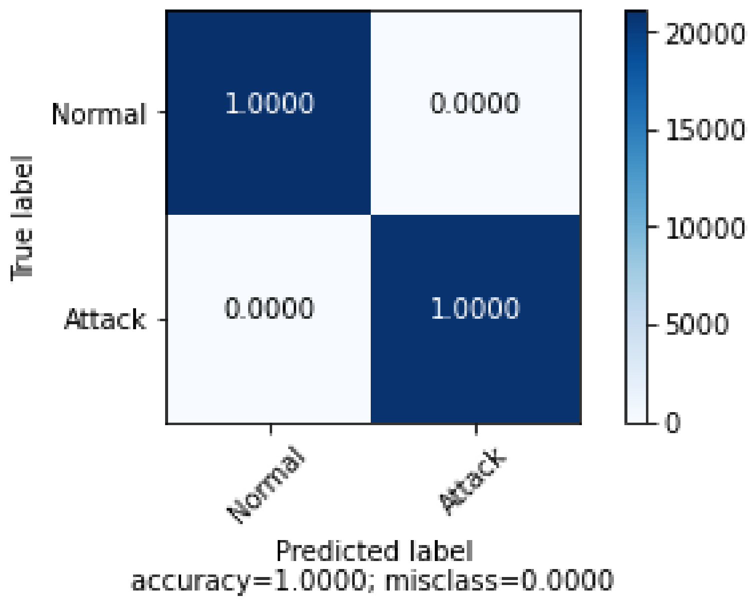


Figure 4.4.6 Confusion Matrix of XG-Boost of PS

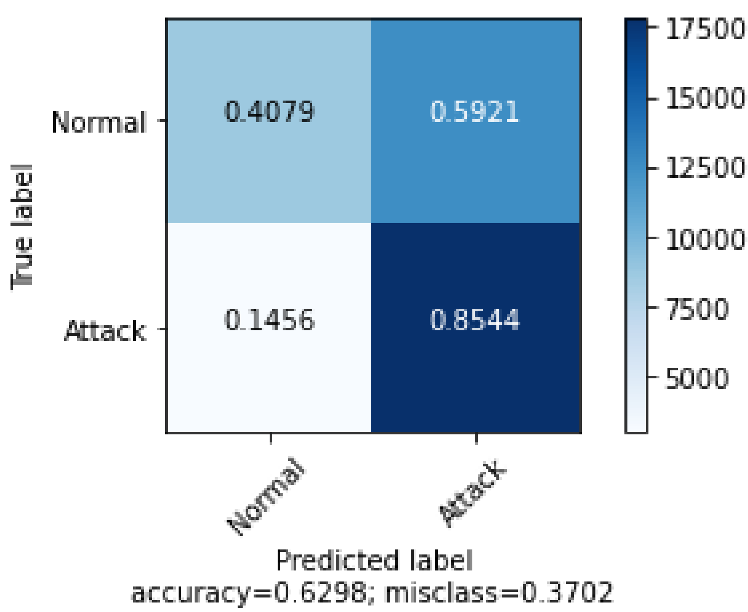


Figure 4.4.7 Confusion Matrix of Support Vector Machine of PS

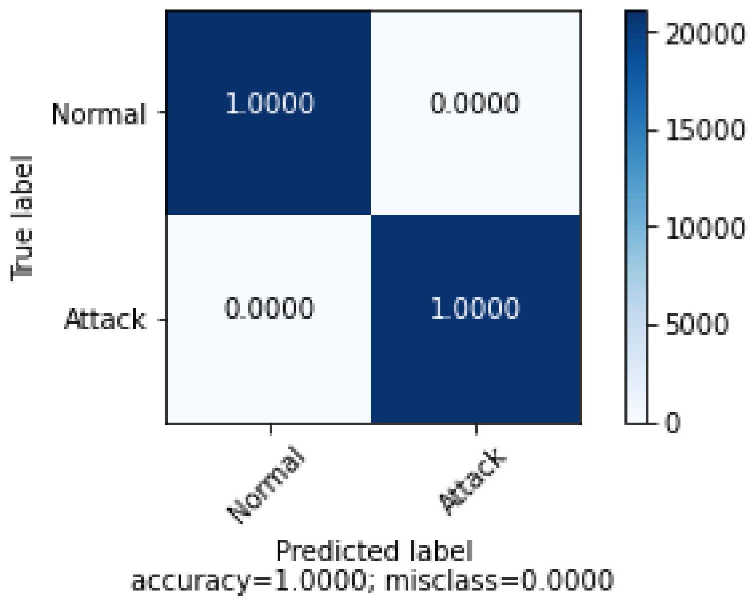
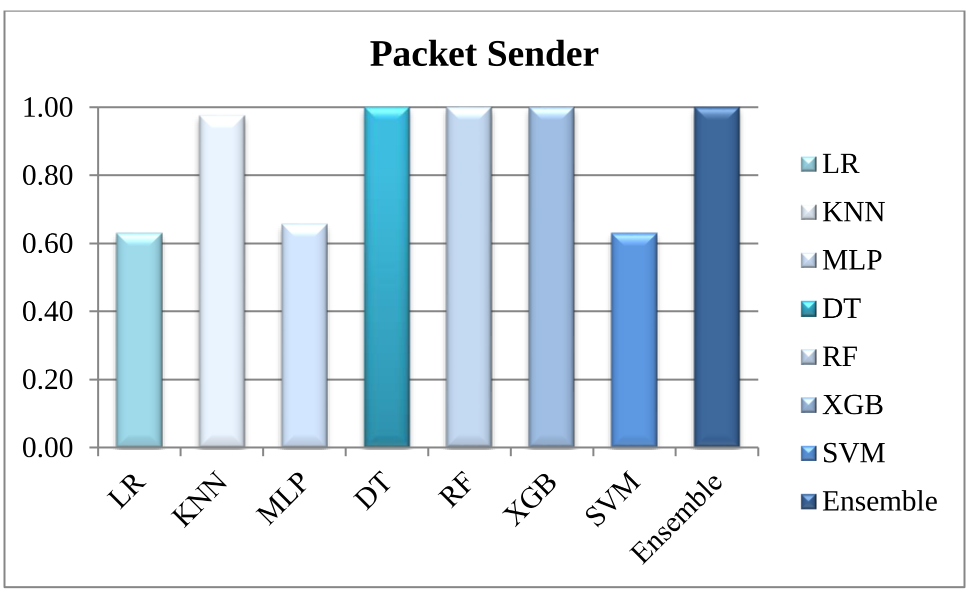


Figure 4.4.8 Confusion Matrix of Ensemble Method of PS



Accuracy Bar Graph 4.1.4 of Packet Sender

Chapter 5 Discussion

As many researchers worked on publically available datasets [22, 29] in detection of DDoS attacks whereas few of them generated their own simulated datasets based on UDP attack and concluded that simulated DDoS datasets are better because publically available datasets are not updated regularly [16] without comparing results on both types of datasets, also they concluded few algorithms are performing well.

Considering the problem statement, we have generated simulated DDoS datasets for Traditional Networks in Packet Sender tool whereas for Software Defined Networks in Mininet using Scapy tool. Dataset for traditional networks is based on UDP traffic whereas SDN DDoS dataset is based on TCP, UDP and ICMP traffic. In addition, publically available two datasets were also used for comparative analysis; ToN-IoT (real-time) and Mendeley DDoS (simulated). Seven different algorithms based upon classification, regression and neural network are investigated for detection of DDoS attacks & ensemble method is also applied on top three algorithms for better results.

Evaluation of Datasets by Algorithms:

As it can be seen from the results of both types of datasets; either it is real time based datasets or simulated datasets, in order to achieve maximum performance of any algorithm, feature selection and hyper-parameter tuning matters a lot. As we have selected best features from the datasets individually by evaluating feature importance and then we provided the Grid Search CV our desired algorithms and possible hyper-parameters which returned us the best parameters for each specific algorithm, then all the algorithms were trained based on those hyper-parameters. On investigation of algorithms we found out that Decision Tree, Random Forest, XG-Boost is giving highest accuracy rate in detection of malicious traffic whereas Logistic Regression has less accuracy rate overall as shown in Table 4.1. Other algorithms behave differently in each dataset e.g MLP gave 65.54% accuracy in Packet Sender, in Mendeley its accuracy is 78.26% whereas in SDN and in ToN-IoT the accuracy is in 90s because of the different nature of each dataset generated in particular environment either real-time or simulated and the change in qualitative features of each dataset.

Evaluation of algorithms on the basis of time shows that Support Vector Machine algorithm is not an efficient algorithm in our research because it took many days in training and testing. Although K-Nearest Neighbor is slow learner it took many hours as compare to other algorithms, Logistic Regression and Random Forest took shorter time for training and testing. Decision Tree, Multi-Layer Perception and XG-Boost took an average time for training and testing.

Following are the Receiver Operating Characteristic curves of both types of datasets:

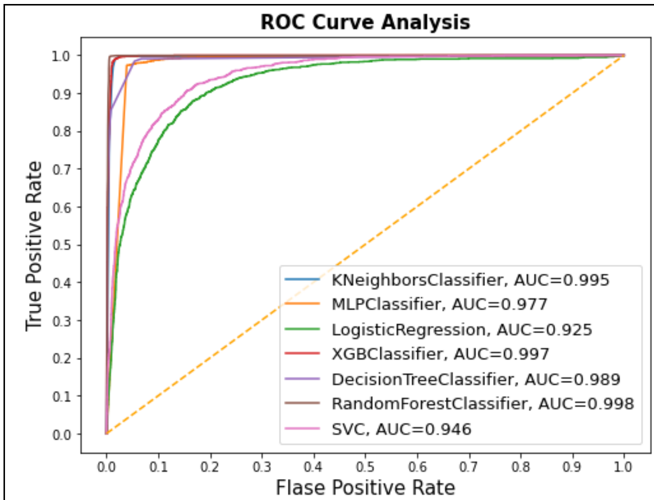


Figure 5.1.1 ROC of ToN-IoT

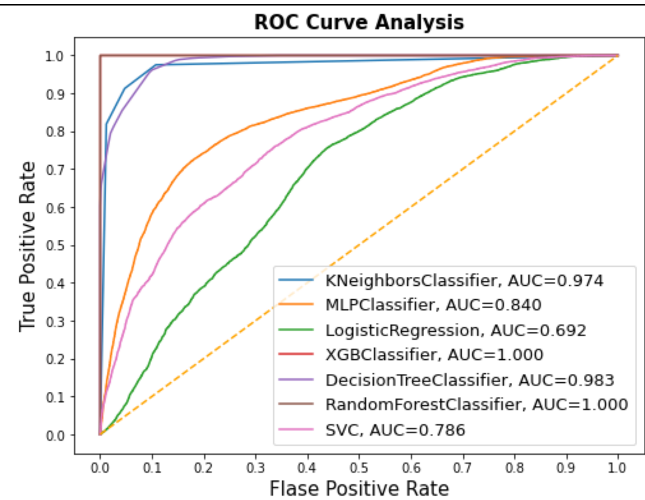


Figure 5.1.3 ROC of Mendeley DDoS

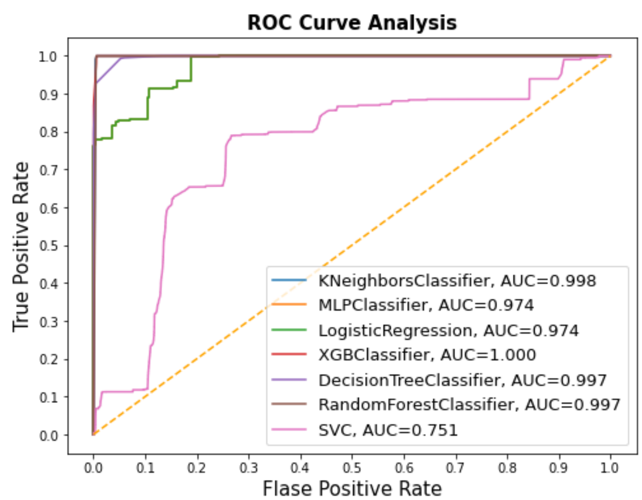


Figure 5.1.2 ROC of SDN

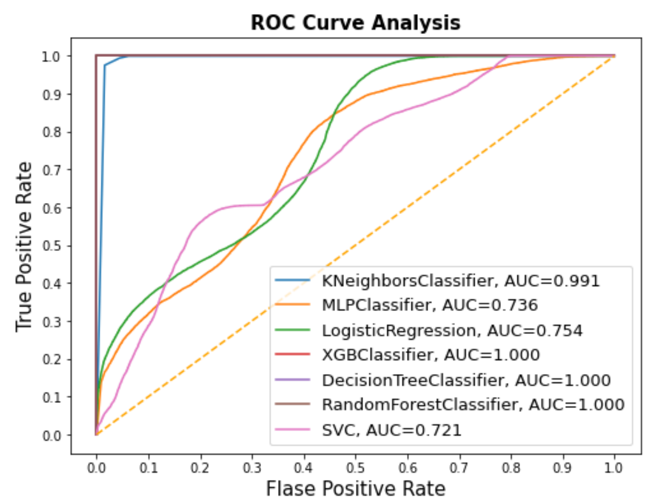


Figure 5.1.4 ROC of Packet Sender

If we look at the ROCs; Random Forest, Decision Tree and XG-Boost have outperformed in most of the datasets and got higher average accuracy than other algorithms. It is mainly due to the reason that tree splits on the basis of entropy.

Comparative Analysis of Real-Time & Simulated Datasets:

To support our research analysis first of all, in general consider strength and weaknesses of real-time based and simulated DDoS datasets that is:

- Real-time based scenarios are more complex than simulated.
- Real-time based datasets has more features than simulated datasets.
- Real-time based datasets generation is expensive than simulated datasets.
- Real-time based datasets took a lot of time in generating and detecting whereas simulated datasets took less time because of fewer features.

As author [16] considered that simulated DDoS datasets are better because publically available real time datasets are not updated regularly. However, in our research we found out that publically available datasets are also good in detection of DDoS attack because ToN-IoT performed very well just like other simulated datasets. Both types of datasets are better in their own ways such as: if we look at the nature of datasets and the criteria of generating attacks in particular environment they directly influence the performance of algorithms. It also depends on the qualitative number of features in each of the dataset that have been taken in the thesis have huge impact on the machine learning algorithms. As the total number of features of ToN-IoT is 127 whereas SDN has 8, Mendeley DDoS 24 and Packet Sender DDoS dataset has 9 features. Therefore, features and feature selection matters a lot in evaluation of algorithms results'. According to our research, simulated datasets took less time for each algorithm for training and testing because of fewer features whereas ToN-IoT took more time however, the accuracy rate of algorithms shows that both datasets are better in detection of DDoS attacks.

Chapter 6 Conclusion & Future Work

In our thesis, we have generated two datasets of DDoS attacks, in addition, we have also selected two well-known real-time and synthetic datasets [ToN-IoT & Mendeley] from internet for accuracy comparison and then applied machine learning techniques for the detection of attacks. In our generated datasets; one is based on traditional networks, generated on Packet Sender Tool whereas the other one is based on SDN generated in Mininet using Scapy tool and captured by Wireshark. We have selected seven different algorithms to investigate the accuracy of datasets as well as to evaluate the algorithms whether which of these are performing well in detection of DDoS attacks. On investigation we found out that every dataset, either it is real time dataset or simulated; Decision Tree, Random Forest and XG-Boost performed well and have highest accuracy rate in detecting DDoS attack whereas the performance of Logistic Regression was not good in most of datasets as compare to other algorithms. Furthermore, Ensemble Method was applied on DT, RF and XGB and after that datasets such as ToN-IoT, Mendeley, SDN and Packet Sender achieved the accuracy rate of 99.27%, 99.90%, 99.23% & 100% respectively. The motive of our thesis was to identify the performance of Machine Learning algorithms on both types of datasets and we found out that three algorithms have got highest accuracy rate among other algorithms and both types of datasets are better in detection process..

6.1 Future work:

In future, if any researcher will take these datasets, the results may differ because of selection of less or more number of features and hyper-parameters. Important Future research works includes prevention of DDoS attacks using machine learning, generation of massive traffic of other attacks such as; TCP SYN flood, Ping of Death attacks & HTTP flood. Researchers may use different software to generate simulated traffic such as Kali Linux, SolarWinds Event Manager (SEM), HULK, LOIC and XOIC. In mitigation process SEM, HULK and XOIC will be beneficial because these all block IPs which do the bombardment of packets and slow the system.

References

[1] Kargl, F., Maier, J., & Weber, M. (2001, April). Protecting web servers from distributed denial of service attacks. In *Proceedings of the 10th international conference on World Wide Web* (pp. 514-524).

[2] Sekar, V., Duffield, N. G., Spatscheck, O., van der Merwe, J. E., & Zhang, H. (2006, June). LADS: Large-scale Automated DDoS Detection System. In *USENIX Annual Technical Conference, General Track* (pp. 171-184).

[3] Ankali, S. B., & Ashoka, D. V. (2011). Detection architecture of application layer DDoS attack for internet. *International Journal of Advanced Networking and Applications*, *3*(1), 984.

[4] Swami, R., Dave, M., & Ranga, V. (2019). Software-defined networking-based DDoS defense mechanisms. *ACM Computing Surveys (CSUR)*, *52*(2), 1-36.

[5] Yihunie, F., Abdelfattah, E., & Odeh, A. (2018, May). Analysis of ping of death DoS and DDoS attacks. In *2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT)* (pp. 1-4). IEEE.

[6] Harshita, H. (2017). Detection and prevention of ICMP flood DDOS attack. *International Journal of New Technology and Research*, *3*(3), 263333.

[7] Hong, K., Kim, Y., Choi, H., & Park, J. (2017). SDN-assisted slow HTTP DDoS attack defense method. *IEEE Communications Letters*, *22*(4), 688-691.

[8] Swami, R., Dave, M., & Ranga, V. (2021). Detection and analysis of TCP-SYN DDoS attack in software-defined networking. *Wireless Personal Communications*, *118*(4), 2295-2317.

[9] Kumar, V., & Sinha, D. (2021). A robust intelligent zero-day cyber-attack detection technique. *Complex & Intelligent Systems*, *7*(5), 2211-2234.

[10] Dobrin, D., & Dimiter, A. (2021, November). DDoS attack identification based on SDN. In *2021 IEEE 20th International Symposium on Network Computing and Applications (NCA)* (pp. 1-8). IEEE.

[11] Santos, R., Souza, D., Santo, W., Ribeiro, A., & Moreno, E. (2020). Machine learning algorithms to detect DDoS attacks in SDN. *Concurrency and Computation: Practice and Experience*, *32*(16), e5402.

[12] Li, H., Wei, F., & Hu, H. (2019, March). Enabling dynamic network access control with anomaly-based IDS and SDN. In *Proceedings of the ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization* (pp. 13-16).

[13] Dina, A. S., & Manivannan, D. (2021). Intrusion detection based on machine learning techniques in computer networks. *Internet of Things*, *16*, 100462.

[14] Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., Wang, C., & Liu, Y. (2018). A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges. *IEEE Communications Surveys & Tutorials*, *21*(1), 393-430.

[15] Song, C., Park, Y., Golani, K., Kim, Y., Bhatt, K., & Goswami, K. (2017, July). Machine-learning based threat-aware system in software defined networks. In *2017 26th international conference on computer communication and networks (ICCCN)* (pp. 1-9). IEEE.

[16] Ahmad, A., Harjula, E., Ylianttila, M., & Ahmad, I. (2020, December). Evaluation of machine learning techniques for security in SDN. In *2020 IEEE Globecom Workshops (GC Wkshps* (pp. 1-6). IEEE.

[17] Ahuja, N., Singal, G., Mukhopadhyay, D., & Kumar, N. (2021). Automated DDOS attack detection in software defined networking. *Journal of Network and Computer Applications*, *187*, 103108.

[18] Erhan, D., & Anarım, E. (2020). Boğaziçi University distributed denial of service dataset. *Data in brief*, *32*, 106187.

[19] Abubakar, A., & Pranggono, B. (2017, September). Machine learning based intrusion detection system for software defined networks. In *2017 seventh international conference on emerging security technologies (EST)* (pp. 138-143). IEEE.

[20] Song, C., Park, Y., Golani, K., Kim, Y., Bhatt, K., & Goswami, K. (2017, July). Machine-learning based threat-aware system in software defined networks. In *2017 26th international conference on computer communication and networks (ICCCN)* (pp. 1-9). IEEE.

[21] Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., Wang, C., & Liu, Y. (2018). A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges. *IEEE Communications Surveys & Tutorials*, *21*(1), 393-430.

[22] Pérez-Díaz, J. A., Valdovinos, I. A., Choo, K. K. R., & Zhu, D. (2020). A flexible SDN-based architecture for identifying and mitigating low-rate DDoS attacks using machine learning. *IEEE Access*, *8*, 155859-155872.

[23] Mehr, S. Y., & Ramamurthy, B. (2019, December). An SVM based DDoS attack detection method for Ryu SDN controller. In *Proceedings of the 15th international conference on emerging networking experiments and technologies* (pp. 72-73).

[24] Al-Nashif, Y., Kumar, A. A., Hariri, S., Luo, Y., Szidarovsky, F., & Qu, G. (2008, June). Multi-level intrusion detection system (ML-IDS). In *2008 International Conference on Autonomic Computing* (pp. 131-140). IEEE.

[25] Shukla, P. (2017, September). ML-IDS: A machine learning approach to detect wormhole attacks in Internet of Things. In *2017 Intelligent Systems Conference (IntelliSys)* (pp. 234-240). IEEE.

[26] Yang, L., & Zhao, H. (2018, October). DDoS attack identification and defense using SDN based on machine learning method. In *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)* (pp. 174-178). IEEE.

[27] Sultana, N., Chilamkurti, N., Peng, W., & Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications*, *12*(2), 493-501.

[28] Suresh, M., & Anitha, R. (2011, July). Evaluating machine learning algorithms for detecting DDoS attacks. In *International Conference on Network Security and Applications* (pp. 441-452). Springer, Berlin, Heidelberg.

[29] Perez-Diaz, J. A., Valdovinos, I. A., Choo, K. K. R., & Zhu, D. (2020). A flexible SDN-based architecture for identifying and mitigating low-rate DDoS attacks using machine learning. *IEEE Access*, *8*, 155859-155872.

[30] Sarica, A. K., & Angin, P. (2020, November). A Novel SDN Dataset for Intrusion Detection in IoT Networks. In *2020 16th International Conference on Network and Service Management (CNSM)* (pp. 1-5). IEEE.

[31] Halimaa, A., & Sundarakantham, K. (2019, April). Machine learning based intrusion detection system. In *2019 3rd International conference on trends in electronics and informatics (ICOEI)* (pp. 916-920). IEEE.

[32] Lawal, B. H., & Nuray, A. T. (2018, May). Real-time detection and mitigation of distributed denial of service (DDoS) attacks in software defined networking (SDN). In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

[33] Nguyen, T., & Zakhor, A. (2004). Multiple sender distributed video streaming. *IEEE transactions on multimedia*, *6*(2), 315-326.

[34] Canadian Institute for Cybersecurity. (2019). DDoS Evaluation Dataset (CIC-DDoS2019)35

[35] Moustafa, N., Koroniotis, N., The BoT-IoT Dataset |UNSW Research (2019).

[36] Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., ... & Zhang, H. (2021). Machine learning: new ideas and tools in environmental science and engineering. *Environmental Science & Technology*, *55*(19), 12741-12754.

[37] Alam, S., & Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. *Computational and Mathematical Organization Theory*, *25*(3), 319-335.

[38] Booij, T. M., Chiscop, I., Meeuwissen, E., Moustafa, N., & den Hartog, F. T. (2021). ToN_IoT: The Role of Heterogeneity and the Need for Standardization of Features and Attack Types in IoT Network Intrusion Datasets. *IEEE Internet of Things Journal*.

[39] Ahuja, N., Singal, G., Mukhopadhyay, D., & Kumar, N. (2021). Automated DDOS attack detection in software defined networking. *Journal of Network and Computer Applications*, *187*, 103108.

[40] Erhan, D., & Anarım, E. (2020). Boğaziçi University distributed denial of service dataset. *Data in brief*, *32*, 106187.

[41] Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.

[42] Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), 660-674.

[43] Lin, W., Wu, Z., Lin, L., Wen, A., & Li, J. (2017). An ensemble random forest algorithm for insurance big data analysis. *Ieee access*, *5*, 16568-16575.

[44] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November). KNN model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 986-996). Springer, Berlin, Heidelberg.

[45] Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, *32*(14-15), 2627-2636.

[46] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1-4.

[47] Evgeniou, T., & Pontil, M. (1999, July). Support vector machines: Theory and applications. In *Advanced Course on Artificial Intelligence* (pp. 249-257). Springer, Berlin, Heidelberg.

[48] Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B., & Zhou, Y. (2015). A novel ensemble method for classifying imbalanced data. *Pattern Recognition*, *48*(5), 1623-1637.

[49] Adhikari, A., Tax, D. M., Satta, R., & Faeth, M. (2019, June). LEAFAGE: Example-based and Feature importance-based Explanations for Black-box ML models. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-7). IEEE.

[50] Agrawal, P. K., Gupta, B. B., Jain, S., & Pattanshetti, M. K. (2011, August). Estimating strength of a DDoS attack in real time using ANN based scheme. In *International Conference on Information Processing* (pp. 301-310). Springer, Berlin, Heidelberg.

[51] Bawany, N. Z., Shamsi, J. A., & Salah, K. (2017). DDoS attack detection and mitigation using SDN: methods, practices, and solutions. *Arabian Journal for Science and Engineering*, *42*(2), 425-441.