A NOVEL SDN INTRUSION DATASET report final.docx plag

by Sudesh Kumar

Submission date: 29-Dec-2021 11:08AM (UTC+0530)
Submission ID: 1216036364
File name: A_NOVEL_SDN_INTRUSION_DATASET_report_final.docx (2.33M)
Word count: 3419
Character count: 18075

A NOVEL SDN INTRUSION DATASET

MINI PROJECT REPORT

Submitted by

SIDDHANT KUMAR(18BCS077)

Submitted in partial fulfilment of the requirements for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

SHRI MATA VAISHNO DEVI UNIVERSITY,KATRA

(School of Computer Science &Engineering)

JAMMU &KASHMIR – 182 320

2021-2022

SHRI MATA VAISHNO DEVI UNIVERSITY

CERTIFICATE

Certified that this project report **"A NOVEL SDN INTRUSION DATASET"** is the work of **" SIDDHANT KUMAR(18BCS077) "** who carried out the mini project work under my guidance/supervision. And this whole project work is done in **INTERNET LAB** on Linux Desktop system.

Mr.Sudesh Kumar
Assistant professor
School of computer science & engineering

Submitted to the Viva voce Examination held on _____

INTERNAL EXAMINER

ACKNOWLEDGEMENT

I would like to thank my project guide **"Mr.Sudesh Kumar"** under whose guidance the successful completion of this project has been possible. The project could not have been implemented without constant inputs and encouragement from him.. Also , I would like to thank my almighty and parents for their blessings.At last I would like to thank the **"INTERNET LAB"** incharge for providing me a LINUX desktop system to complete my project successfully.

ABSTRACT

Software defined network (SDN) is an emerging technology of network that is manageable,cost effective,and adaptable to use in the application.It is developed to reduce the network complexity as the traditional network system has a dedicated way but it does not have .SDN work on the centralized method .Apart from the introducing the SDN in abstract what i have in this project is that i had created a virtual topology and performed the cyber security attacks and captured the data packets to generate the dataset.The motto of generation of dataset is make available publicly and to analysis IDS .In short ,there is no publicly available dataset for the anomaly detection in SDN . If available then ,it is outdated and inappropriate which might not be helpful for the performance analysis.

INDEX

INTRODUCTION

1.Introduction:

The researcher and the engineer had developed the 8th wonder of the world which is “Network”.Network is considered to be a wonder because of how easily we transfer the information and data from one place to another destination.

Network had divided into two parts based on their function and architecture are as follows:

- ➔ Conventional network
  - ◆ Decision formation process known as control plane.
  - ◆ Network administrators had to implement the network policies on each device independently.
- ➔ SDN

Now , coming to the point SDN first comes to mind what SDN is and its advantage over the traditional network system. The answer is simple, SDN comes with an advantage over the traditional network distribution system as the conventional network device had to be configured particularly, but sdn had resolved this by providing the centralised control data plane . the following are the points for the deployment of SDN in the network are as follow:-

- Separated the control plane from the data plane which becomes easier for the administrator to make any update or changes in the network configuration .
- Reduces the human mistakes while configuration of the devices.
- Implementation on the network devices easily without any constraints or restrictions .
- Reduces cost.
- The devices which come under the application do not need any programming language to configure or to utilize the network resources.

Although SDN has many benefits over the conventional network distribution, it has security threats of being exploited by the attackers.If unfortunately ,attackers are able to take control over the SDN control plane then automatically the hacker is able to control the related whole system without any much effort.The research is still going to make and perform better the SDN.

BACKGROUND

2.1 Literature review:

Before starting the project I read about the dataset that are available for the analysis of the intrusion detection system. Some of the dataset are outdated but important to learn why it is not an appropriate dataset to analyse as well as why I was required to generate my dataset.

A. KDD '99:- It is one of the most well known dataset which are widely used for the intrusion system ,which originated from the **DARPA** having 41 traffic features which are sub grouped as:

- **i.** basic feature
- **ii.** traffic features
- **iii.** content features.

a. Problems:-

- Redundancy 78 percent in training sets and 75 percent in test sets.

B. NSL-KDD:- This dataset is the updated version of the KDD '99 dataset that solves the duplicacy problem .

It contains the two parts:

(1) Training set

(2) Testing set.

a. Problems:-

i. Data is not realistics to current network
ii. Low detection rate
iii. High false alarm
iv. Applicable to DOS attack only.

C. KYOTO:- It originated from the honeypot server in Kyoto university.It comprises 24 statistical features out of which 14 are shared with KDD.Basically the dataset that was generated was based on the malicious attack.

a. **Problems:-**
   i. Performed on the real traffic of the network but not having considered information regarding the type of attack.

D. ISCX 2012:-The administrator used two profiles to generate the traffic on a simulated network environment.

   i. ***ALPHA*** used to generate attack traffic
   ii. ***BETA*** for normal traffic generation.

➔ The dataset mainly includes **DOS** and **BRUTE FORCE** attacks with 20 packet features.

➔ **Problems:-** The major issue with this dataset is that it includes only **HTTP** traffic but nowadays traces are based on the **HTTPS** traffic.

E. CICIDS 2017:-As it is the newest dataset ,which attracts the researcher to analyze it as it overcomes the problem of the previous dataset namely **"ISCX2012"** and has a large number of attack cases.

a. **Problems:-**
   i. The dataset is so large that it is irrelevant for intrusion detection training.

F. CSE-CIC-IDS 2018:- It is due to the collaboration of a project between the Canadian Institute for Cyber Security(**CIC**) and communications Security Establishment(**CSE**) .The dataset generation implementation took place on the **"AWS"** platform.

It comprises two classes are as:

(1) B-profile for the generation of the normal traffic

(2) M-profiles for the attack.

a. **Problems:-**
   i. Although the dataset is used to analyse the IDS research on SDN network .But the dataset is not generated from SDN network platform and that's why the dataset generated earlier is of no use except for learning and understanding purposes.
   ii. The drawbacks occurred because SDN and conventional networks work on different theories

So, from reading the above mentioned theory about the generation of dataset and evolution of the network ,I understood the necessity of generation of dataset from SDN network . Till now there is no publicly available dataset for SDN network that accalaims the genuinity.

Data-set	Year	Types of network	Realistic Traffic	Total no.of attributes	Balanced	Network Environment
KDD' 99	1998	Small network	NO	41	No	conventional network
NSL-KDD	2009	Small network	No	41	No	conventional network
KYO-TO	2006-2009	Honeypots	Yes	24	No	conventional network
ISCX 2012	2012	small network	Yes	unknown	No	conventional network
CIC-IDS 2017	2017	Small network	Yes	83	No	conventional network
CSE-CIC-IDS 2018	2018	Small network	Yes	83	No	AWS platform
InSDN	2020	Small network	Yes	83	No	SDN network

TABLE 1. Comparison between the different attack scenarios

2.2 Attack Phases:-

Attacker's goal is to control the network system by gaining unauthorized access to network resources.If an attacker is able to attack the control plane then it would become much easier to control over the rest of devices easily and steal valuable data from it.

According to the theory of ethical hacking it has five phase which are as follow:-

1) **Reconnaissance:-**The first step of the ethical hacking states that before attacking any network we should have proper knowledge about that like what version of the network device is using,which operating system is it,Ip addresses, ….etc.

2) **Scanning:-** After reconnaissance the utmost important step is to prepare for the attacks and to discover the vulnerabilities of the system.

3) **Gaining Access:-** After finding the proper vulnerability in the device the attacker can exploit the vulnerability to gain the system control.The exploitation can be done in different methods such as bruteforce attack,dos attack,ddos attack …etc.

4) **Maintaining Access:-** To keep the system in control for the long term access the attacker maintains the acces through backdoor in the system, which provides the remote shell to the attacker.

5) **Clearing Tracks:-**After performing the operation, attackers clear the path of entering into the system to show them as they were not in the system. This is important because leaving any proof can lead an attacker to get caught.

ARCHITECTURE

3.1 Description:-According to the proposed theory of SDN network we require four machines. It can be four different physical machines or one physical machine and another virtual machine.

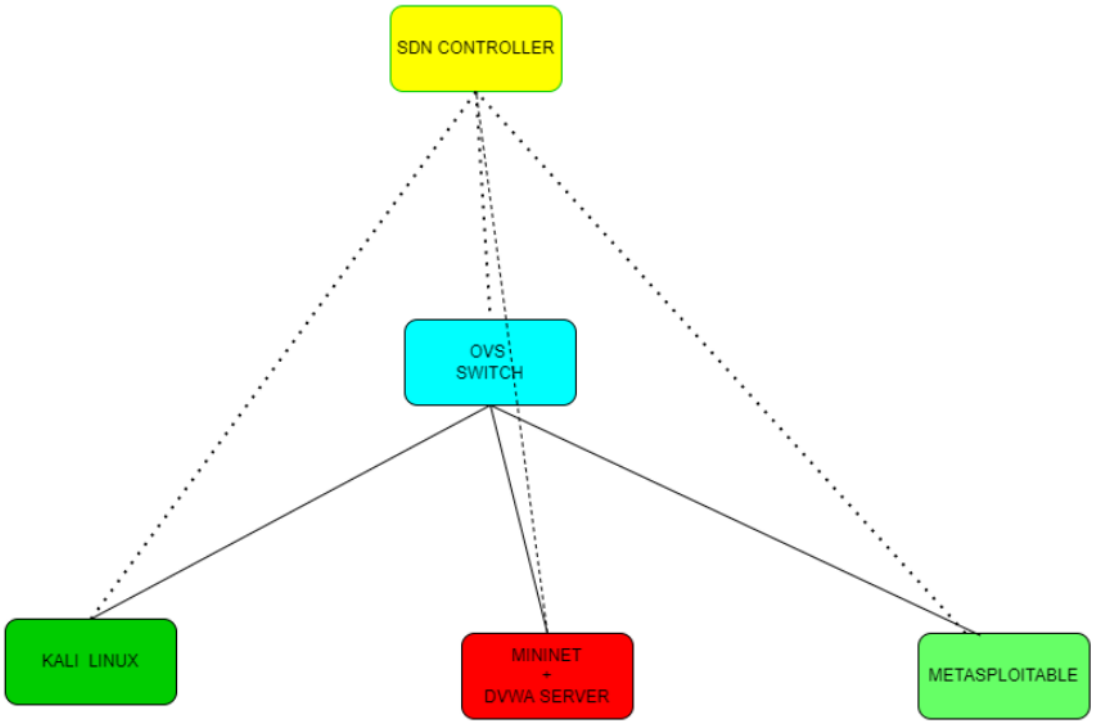
But in this project the requirement of the machine are as follow:-

- Ubuntu (as a base machine)
  - Latest version **20.04.3 LTS**
- Virtualbox
  - Latest version **6.1**
- Mininet
- Kali linux
  - Latest version **2021.4a**
- OWASPBWA
  - Act as a web server
- OVS SWITCH
- RYU controller
  - Version 2.3.0
- Metasploitable
- Wireshark
  - For data flow packet capturing.
- CIC FLOWMETER
  - To convert the .pcap file into a csv file.

NOTE:-

- **L3** switch connectivity is used for the communication between machines.
- As a function of **L3** switching the **OVS** switch is configured with OVS software and kernel routing.
- The below diagram depicts the working model of the SDN network

3.2 Diagram:-



3.3 Installation process:-

1. First install Ubuntu on the desktop as a base machine.
2. When ubuntu desktop is installed then install mininet on the base machine
3. Install the RYU controller on the base machine.
4. Create the mininet topology
5. Install the virtualbox.
6. In virtualbox, install the OWASPBWA as a new machine.
7. Install kali linux on the other machine or on the virtualbox as a machine.
8. Metasploitable is pre-installed in kali linux . No need to install separately.
9. Connect all the different machines with the RYU controller.
10. Ping all the machines ,if successful then connection is established .

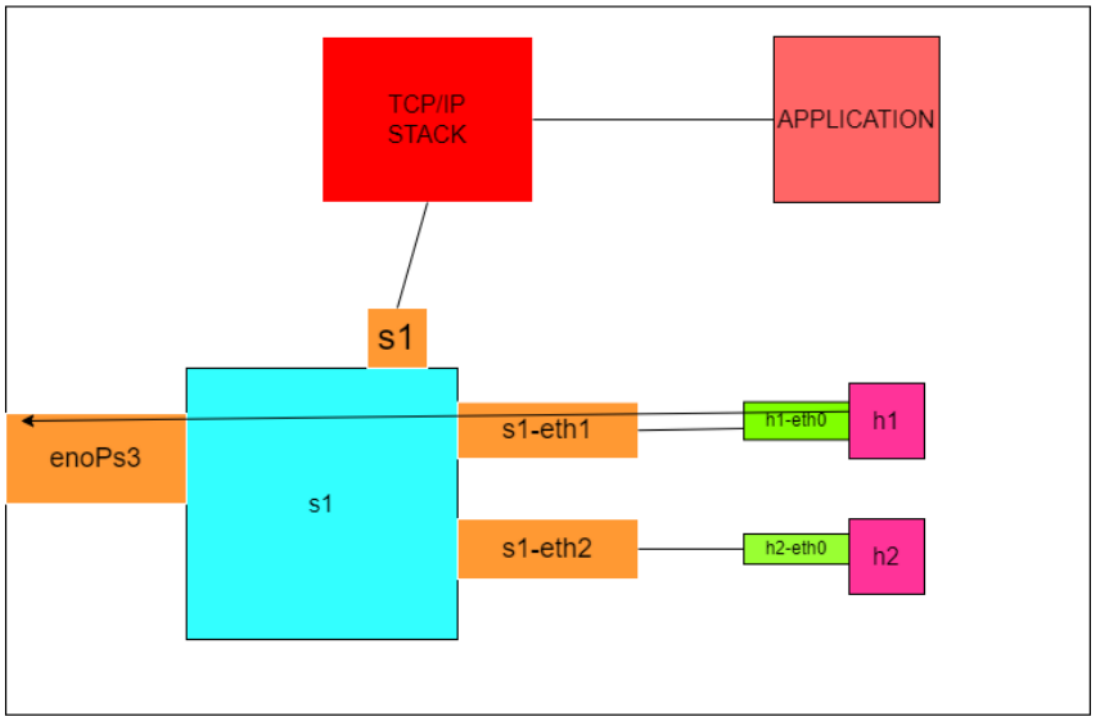
3.4 InstallationCommands:-

- **MININET:-**
  - git clone git://github.com/mininet/mininet
  - cd mininet
  - git tag
  - git checkout -b 2.3.1
  - mkdir my_mininet
- **RYU controller:-**
  - apt install python3-pip
  - pip3 install ryu
  - ryu-manager --version
- Use “sudo” for the administration purpose in linux.

3.5 Test command:-

- **RYU:-**
  - ryu-manager ryu.app.simple_switch_13
- **MININET Topology:-**
  - Creating a virtual network with default topology which connects to the RYU controller.
    - sudo mn - - mac - - switch ovs,protocols=OpenFlow13 --controller remote
    - sh ovs-vsctl add-port s1 enopS3
    - sh ovs-vsctl show
    - sh ovs-ofctl -O openflow13 dump-flows s1 | grep NORMAL - - color
    - sh ifconfig enopS3 0
    - sh dhclient s1
    - sh route
    - h1 ifconfig h1-eth0 0
    - h1 dhclient h1-eth0
    - h2 ifconfig h2-eth0 0
    - h2 dhclient h2-eth0

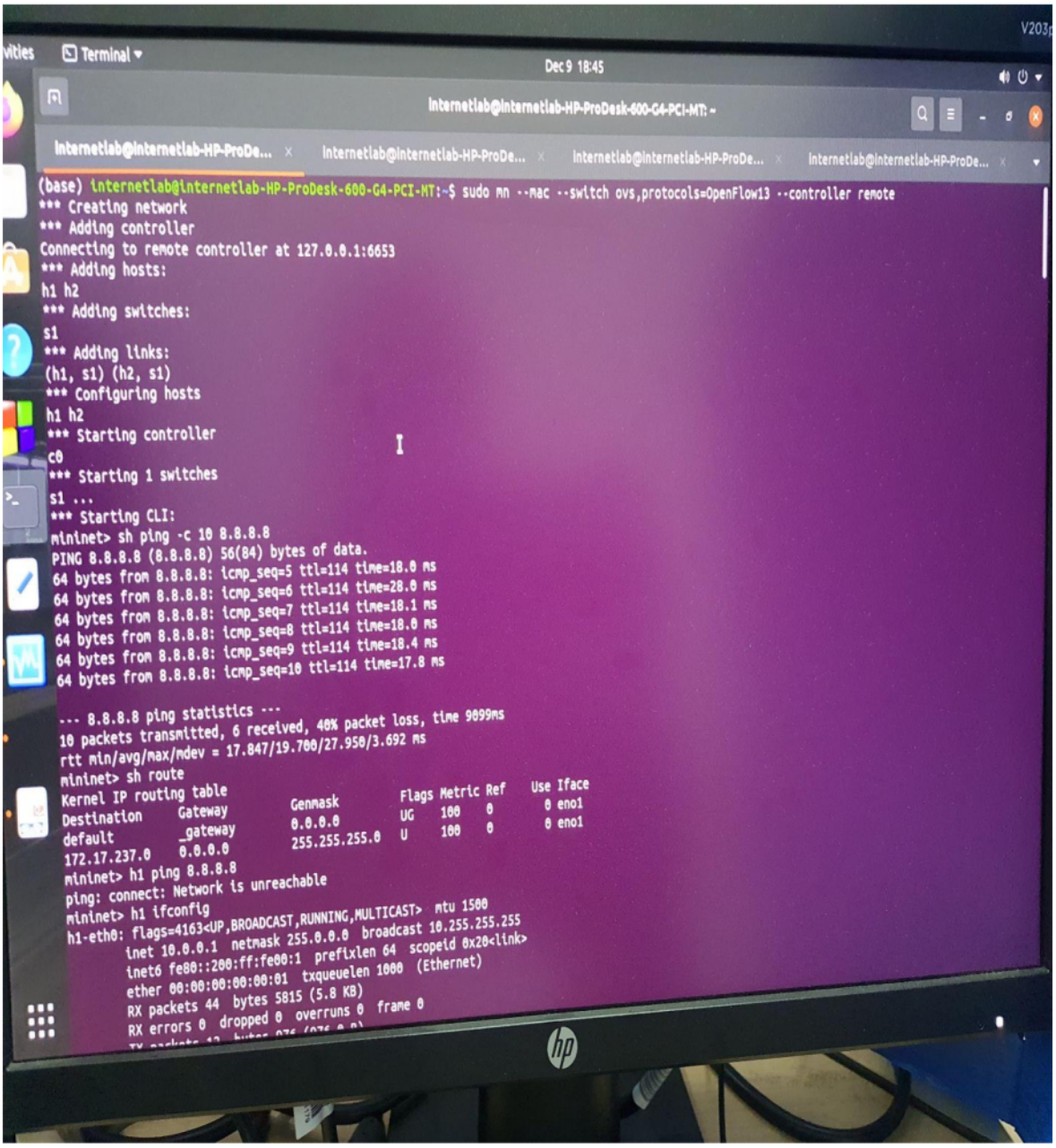
❖ After performing the above simulation test command the connection of the internet would be in accordance with the depicted below picture.



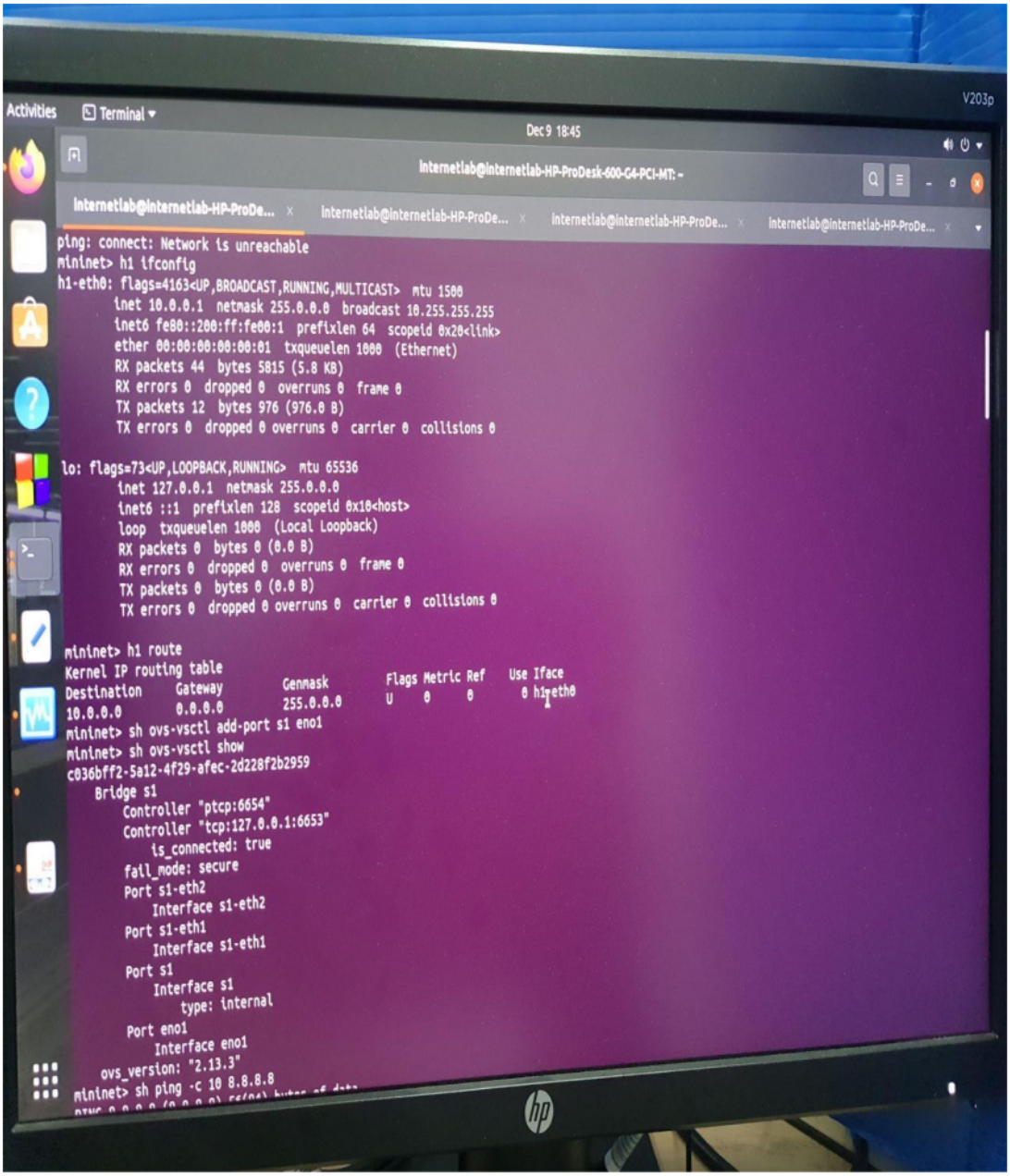
NOTE:-

- To run the google chrome in mininet node then first download it on the base machine and the open the xterm of the mininet node and type following command:-
    - google-chrome - -no-sandbox - -user-data-dir

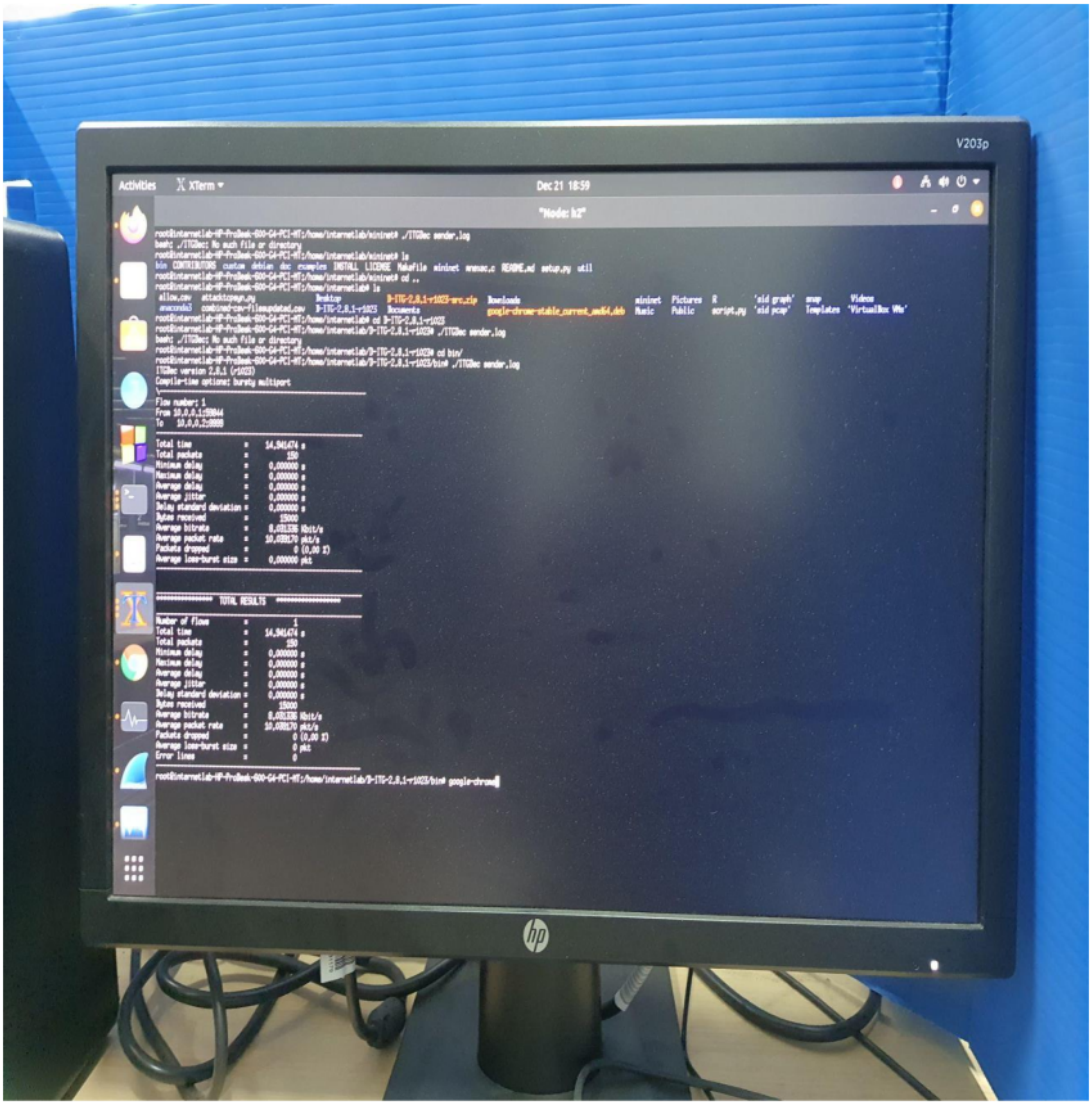
3.6 Topology Diagram:-



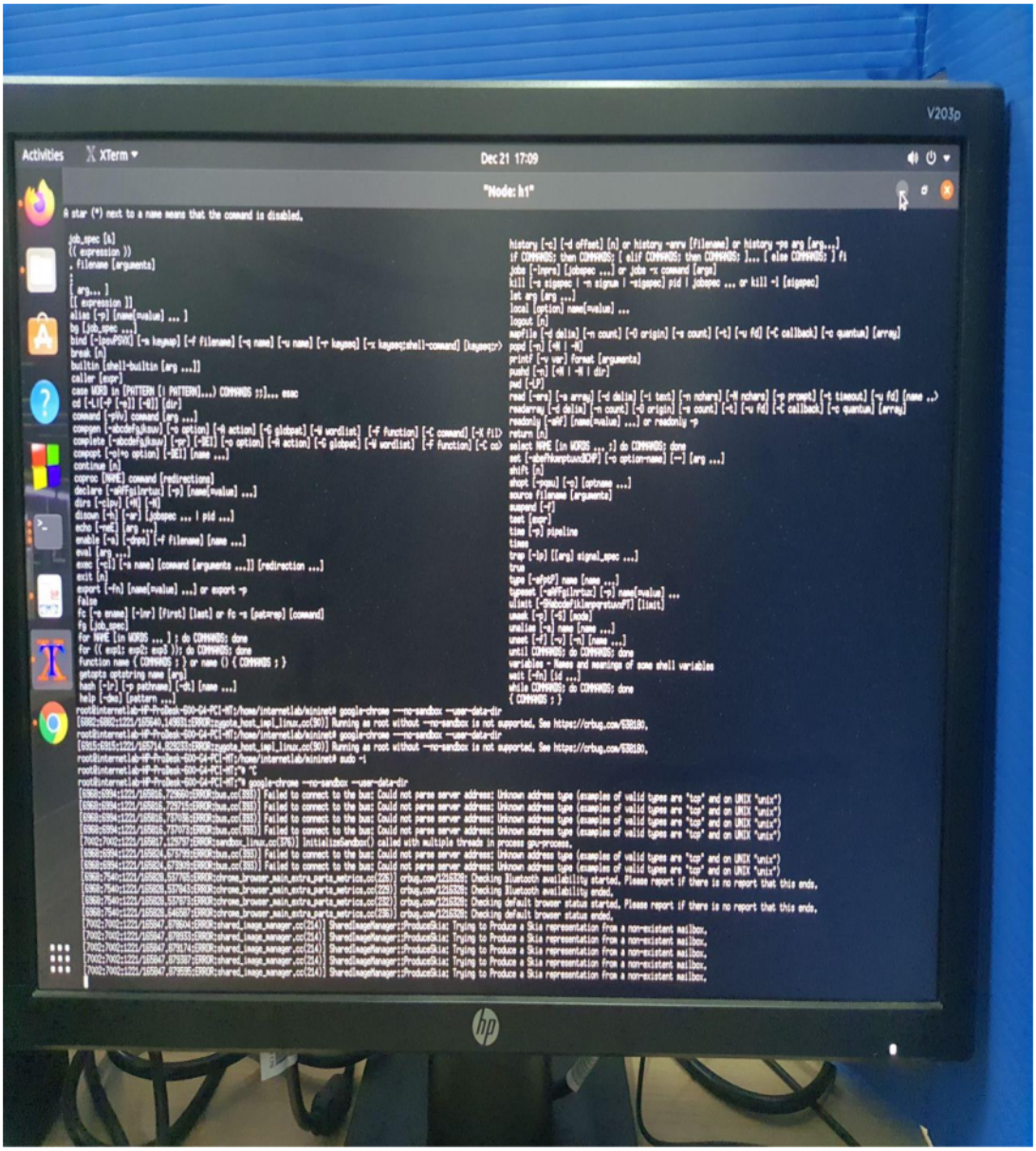
Mininet topology


Activities
Terminal
Dec 9 18:45
internetlab@internetlab-HP-ProDesk-600-G4-PCI-MT: ~
ping: connect: Network is unreachable
mininet> h1 ifconfig
h1-eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
inet 10.0.0.1 netmask 255.0.0.0 broadcast 10.255.255.255
inet6 fe80::200:ff:fe00:1 prefixlen 64 scopeid 0x20<link>
ether 00:00:00:00:00:01 txqueuelen 1000 (Ethernet)
RX packets 44 bytes 5815 (5.8 KB)
RX errors 0 dropped 0 overruns 0 frame 0
TX packets 12 bytes 976 (976.0 B)
TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
inet 127.0.0.1 netmask 255.0.0.0
inet6 ::1 prefixlen 128 scopeid 0x10<host>
loop txqueuelen 1000 (Local Loopback)
RX packets 0 bytes 0 (0.0 B)
RX errors 0 dropped 0 overruns 0 frame 0
TX packets 0 bytes 0 (0.0 B)
TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
mininet> h1 route
Kernel IP routing table
Destination Gateway Genmask Flags Metric Ref Use Iface
10.0.0.0 0.0.0.0 255.0.0.0 U 0 0 0 h1-eth0
mininet> sh ovs-vsctl add-port s1 eno1
mininet> sh ovs-vsctl show
c036bff2-5a12-4f29-afec-2d228f2b2959
Bridge s1
Controller "ptcp:6654"
Controller "tcp:127.0.0.1:6653"
is_connected: true
fail_mode: secure
Port s1-eth2
Interface s1-eth2
Port s1-eth1
Interface s1-eth1
Port s1
Interface s1
type: internal
Port eno1
Interface eno1
ovs_version: "2.13.3"
mininet> sh ping -c 10 8.8.8.8

Adding port



Sender log file picture on xterm node h1



Executing google from h1 mininet node

ATTACKS FOR DATA GENERATION

4.1 Dataset Attack scenarios:-

The goal is to generate the SDN network traffic data by using different attack scenarios .

The given table below shows the detailed information about the types of attack we are going to perform .

Types of Attack	Activities	Tools used
Malware	1. Botnet attack	1. ARES
DoS	1. UDP flood 2. HTTP flood 3. TCP-ACK flood, 4. Slowloris	1. Slowhttptest 2. LOIC 3. Torshammer 4. HULK
	1. Slowloris 2. TCP flood	1. Nping 2. Metasploit
Web attacks	1. sql InjectXSS 2. sql Inject	1. Metasploit 2. sqlmap
DDoS	1. UDP Flood 2. TCP-SYN Flood 3. ICMP Flood	1. Hping3
U2R	1. IRCd 2. Vsftpd 3. Samba and distcc	1. Metasploit
Probes	1. port scan 2. Version scan 3. discover services	1. Nmap
	1. Port scan,vulnerability scan(WMAP)	1. Metasploit
R2L	1. Password-Guessing	1. Hydra 2. BurpSuite 2. Metasploit

4.2 Description about attacks:-

A. **DoS:-** It is the most common attack that is usually performed in every network,which not only damages the victim machine but also utilizes the resources in a short period of time. When a DoS attack is performed then it acts as "*a body with no brain*". Hence, a legitimate user request cannot be fulfilled.

➔ Types of Dos attack:-

i. **Network Dos attack:-** Aim of these attacks is to flood the victim machine by large amounts of spoofed packets. The general protocols used are ICMP,UDP,TCP.

ii. **Application DoS attack:-** Although this attack does not require the high bandwidth but damages it by consuming the resources in a very short amount of time.

B. DDoS attack:- In dataset this attacke scenarios also include such as ICMP Flood attacks,TCP-SYN Flood and UDP Flood.

C. Password-Guessing attack:- It is performed when an attacker requires the username and password for the login credentials .

D. Web applications attacks:- In web attacks we frequently perform:

i. Cross-site scripting(XSS)
ii. SQL Injection

.

- **XSS attacks:-**This type of attack is performed by writing or injecting the malicious code on the website . If a user accesses the infected website then the malicious script will be executed which results in revealing of sensitive information such as "login credentials,sensitive tokens,cookies".
- **SQL injection attacks:-**
  - Manipulation by using malicious queries

E. Probe attacks:-

a. It is included in the first phase of the ethical hacking which is used to detect remote vulnerabilities like open ports ,versions of operating systems, etc...

F. Botnet attacks:- Although almost all devices or things are based on the internet access ,with some security features.but this security features does not guarantee that you are always secure. The attacker can control several infected devices ,known as botnets, to execute the malicious activities,such as stealing information ,fraud attacks,launching DDos against the victim server.

G. U2R(Exploitation) attack:- when an attacker attacks any victim then the attacker always has to keep backdoor ,so that the attacker can utilize the benefit when needed anymore.

H. Normal Traffic:-

Protocol are:-

a. HTTP,HTTPS,SSH,mail,DNS, etc

4.3 <u>Performed activity</u>:-

➔ Dos attack
➔ DDoS attack
➔ Probes attack

DATASET

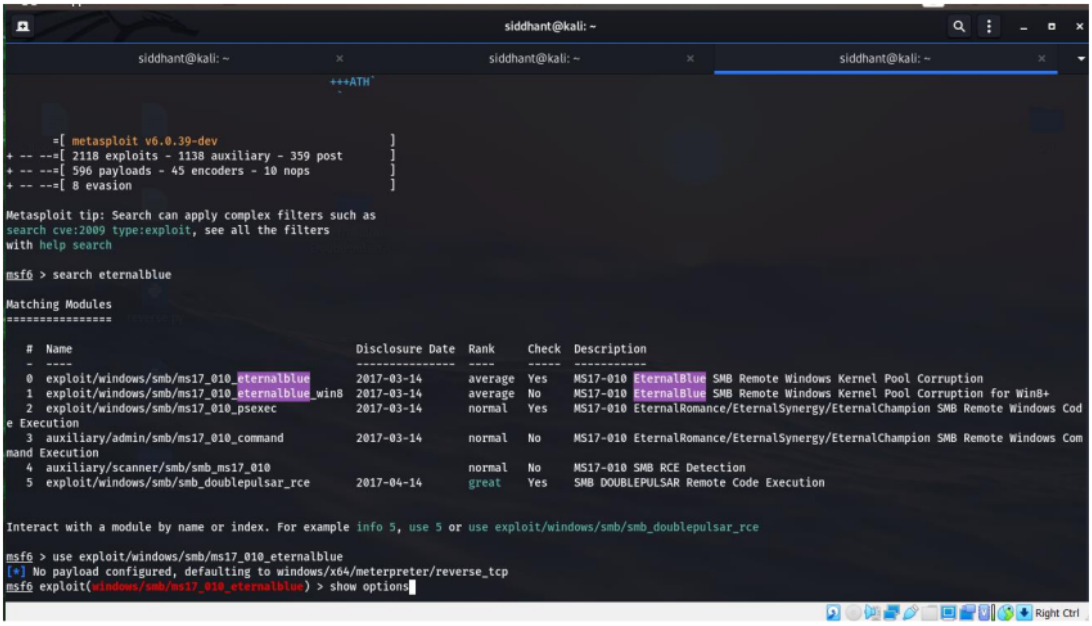
5.1 Dataset Description:-

According to our attack preparation on sdn network the dataset is generated for the analysis purpose to detect the intrusion in the network as well as to establish the better security features in it.

This dataset comprises the **84** column and **237478** instances of having file size 76,345KB when all other dataset files merge together.

Attacks Types	File size
DDos	25,499KB
Normal	6KB
Web attack	35,352KB
Dos	185KB
Probe	267KB

Some of the attacks are repeated but performed with different tools to DoS attacks are repeated with other tools like HULK,Metasploit framework etc… Not only restricted to DoS attacks there are many more attacks that have been repeated with different tools.Many other attacks that are majorly performed by using the "Metasploit framework ".Metasploit have a handful amount of attacks libraries which is called as payloads , auxiliary etc. These payloads or auxiliary contains the malicious code file as a script to execute when needed to perform the activity.

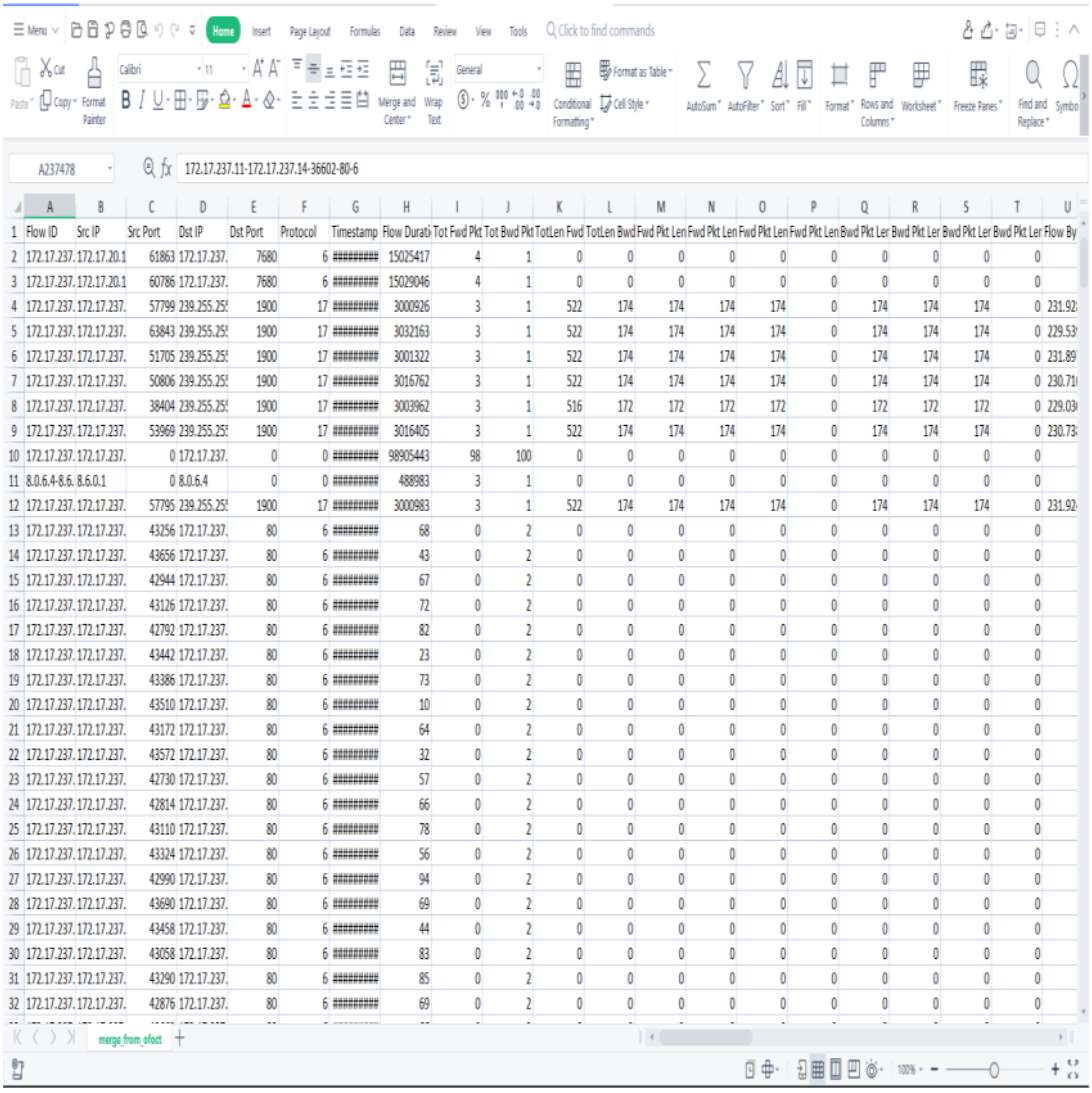


Metasploit framework

The instance of the dataset comprises of are as following:-

- Flow id
- Source ip
- Source port
- Destination port
- Protocol
- Timestamp
- Flow Duration
- Packet length
- Byte length
- Flag
- etc

5.2 Dataset Diagram:-



Flow ID	Src IP	Src Port	Dst IP	Dst Port	Protocol	Timestamp	Flow Durati	Tot Fwd Pkt	Tot Bwd Pkt	TotLen Fwd	TotLen Bwd	Fwd Pkt Len	Fwd Pkt Len	Fwd Pkt Len	Bwd Pkt Len	Bwd Pkt Ler	Bwd Pkt Ler	Bwd Pkt Ler	Flow By
172.17.237.	172.17.20.1	61863	172.17.237.	7680	6	########	15025417	4	1	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.20.1	60786	172.17.237.	7680	6	########	15029046	4	1	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	57799	239.255.25	1900	17	########	3000926	3	1	522	174	174	174	174	0	174	174	174	0 231.92
172.17.237.	172.17.237.	63843	239.255.25	1900	17	########	3032163	3	1	522	174	174	174	174	0	174	174	174	0 229.53
172.17.237.	172.17.237.	51705	239.255.25	1900	17	########	3001322	3	1	522	174	174	174	174	0	174	174	174	0 231.89
172.17.237.	172.17.237.	50806	239.255.25	1900	17	########	3016762	3	1	522	174	174	174	174	0	174	174	174	0 230.71
172.17.237.	172.17.237.	38404	239.255.25	1900	17	########	3003962	3	1	516	172	172	172	172	0	172	172	172	0 229.03
172.17.237.	172.17.237.	53969	239.255.25	1900	17	########	3016405	3	1	522	174	174	174	174	0	174	174	174	0 230.73
172.17.237.	172.17.237.	0	172.17.237.	0	0	########	98905443	98	100	0	0	0	0	0	0	0	0	0	
8.0.6.4-8.6.	8.6.0.1	0	8.0.6.4	0	0	########	488983	3	1	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	57795	239.255.25	1900	17	########	3000983	3	1	522	174	174	174	174	0	174	174	174	0 231.92
172.17.237.	172.17.237.	43256	172.17.237.	80	6	########	68	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43656	172.17.237.	80	6	########	43	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	42944	172.17.237.	80	6	########	67	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43126	172.17.237.	80	6	########	72	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	42792	172.17.237.	80	6	########	82	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43442	172.17.237.	80	6	########	23	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43386	172.17.237.	80	6	########	73	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43510	172.17.237.	80	6	########	10	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43172	172.17.237.	80	6	########	64	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43572	172.17.237.	80	6	########	32	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	42730	172.17.237.	80	6	########	57	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	42814	172.17.237.	80	6	########	66	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43110	172.17.237.	80	6	########	78	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43324	172.17.237.	80	6	########	56	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	42990	172.17.237.	80	6	########	94	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43690	172.17.237.	80	6	########	69	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43458	172.17.237.	80	6	########	44	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43058	172.17.237.	80	6	########	83	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	43290	172.17.237.	80	6	########	85	0	2	0	0	0	0	0	0	0	0	0	
172.17.237.	172.17.237.	42876	172.17.237.	80	6	########	69	0	2	0	0	0	0	0	0	0	0	0	

ANALYSIS

For the analysis of the dataset for our attack scenarios that I described in the section proposed architecture and dataset attack generation, I deployed some machine learning technique for the analysis of it.Te deployed algorithm is tree-based algorithm are Random Forest.The learning classifier are trained with the cross-validation technique "k=5".

6.1 Results:-

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 116.62 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	237466	99.9954 %
Incorrectly Classified Instances	11	0.0046 %
Kappa statistic	0.1538	
Mean absolute error	0.0001	
Root mean squared error	0.0064	
Relative absolute error	73.4233 %	
Root relative squared error	94.4002 %	
Total Number of Instances	237477	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.909	1.000	1.000	1.000	0.213	0.818	1.000	DDOS
	0.091	0.000	0.500	0.091	0.154	0.213	0.818	0.247	NORMAL
Weighted Avg.	1.000	0.909	1.000	1.000	1.000	0.213	0.818	1.000	

=== Confusion Matrix ===

a	b	<-- classified as
237465	1	a = DDOS
10	1	b = NORMAL

LIMITATIONS

1) As we know that SDN network environment deployment technology is still under development.

2) We used a RYU controller in this project but there are many other different controllers too and it has been ignored. But in a different research journal paper the author claimed that different controllers have different security models which is also a great issue.

3) Although the SDN requires the one controller, in some cases it may need another sub controller to control the sub-devices.

4) To generate the more accurate data ,SDN should be implemented on the physical machine rather than on virtual machines.

5) In some cases the flow tables of the network devices may be ignored.

6) If the dataset is unbalanced in large amounts then there is a risk of providing the low detection rate and accuracy.

CONCLUSION

The project when started looks somewhat easier but as long as I worked upon the project,the level of difficulty also increased.At the starting there were many errors in connections established between the resources .But with continuous effort ,finally effort got something and succeeded . The conclusion came in my mind that the SDN network is the most promising network that can reduce not only establishment cost as well as the cost of the network resources. The number of resources used in the conventional resources also reduces in this SDN network.Although ,my project work depicts the attacks on the SDN network,but it defines the ways we can attack it and generate dataset to analyze. It comes out as a possible way to reduce the possibility of attack upon it .According to me, dataset generation and analysis of IDS detection by using the machine learning technique will help the network administrator to secure it more securely. Yes,SDN requires some more improvement and it is developing as it is under development. In the upcoming years ,soon SDN will come with a new network revolution.

REFERENCES

- H. Z. Jahromi and D. T. Delaney, ‘‘An application awareness framework based on SDN and machine learning: Defining the roadmap and challenges,’’ in Proc. 10th Int. Conf. Commun. Softw. Netw. (ICCSN), Jul. 2018,pp. 411–416

- M. S. Elsayed, N. -A. Le-Khac and A. D. Jurcut, "InSDN: A Novel SDN Intrusion Dataset," in IEEE Access, vol. 8, pp. 165263-165284, 2020, doi: 10.1109/ACCESS.2020.3022633.

- D. Firdaus, R. Munadi and Y. Purwanto, "DDoS Attack Detection in Software Defined Network using Ensemble K-means++ and Random Forest," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2020, pp. 164-169, doi: 10.1109/ISRITI51436.2020.9315521.

- T. Zoppi, A. Ceccarelli and A. Bondavalli, "Unsupervised Algorithms to Detect Zero-Day Attacks: Strategy and Application," in IEEE Access, vol. 9, pp. 90603-90615, 2021, doi: 10.1109/ACCESS.2021.3090957.

- H. A. Hassan, E. E. Hemdan, W. El-Shafai, M. Shokair and F. E. A. El-Samie, "An Efficient Intrusion Detection System for SDN using Convolutional Neural Network," 2021 International Conference on Electronic Engineering (ICEEM), 2021, pp. 1-5, doi: 10.1109/ICEEM52022.2021.9480383.

- Q. -V. Dang, "Studying the Fuzzy clustering algorithm for intrusion detection on the attacks to the Domain Name System," 2021 Fifth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4), 2021, pp. 271-274, doi: 10.1109/WorldS451998.2021.9514038.

- https://shantoroy.com/sdn/sdn-mininet-ryu/
- https://groups.geni.net/geni/wiki/HowTo/ConfigureOVSWithLayer3Routing

- https://stackoverflow.com/questions/38845033/connecting-open-vswitch-with-two-virtual-machines
- https://www.youtube.com/results?search_query=sdn+network+project

A NOVEL SDN INTRUSION DATASET report final.docx plag

ORIGINALITY REPORT

9%	5%	5%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Mahmoud Said Elsayed, Nhien-An Le-Khac, Anca D. Jurcut. "InSDN: A Novel SDN Intrusion Dataset", IEEE Access, 2020 Publication	3%
2	Submitted to University of Wolverhampton Student Paper	3%
3	docplayer.net Internet Source	1%
4	www.coursehero.com Internet Source	1%
5	Submitted to Liverpool John Moores University Student Paper	1%
6	repository.library.teimes.gr Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		