INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 4 - V7I4-1809)

Available online at: https://www.ijariit.com

Detecting DDoS attack with AI/ML models

Arun Narayanan N.
arun.cs03@reva.edu.in
REVA University, Bengaluru, Karnataka

Dhruv Kalaan
dhruvkalaan@gmail.com
OLX Group, Bengaluru, Karnataka

ABSTRACT

The enormous evolution of web-based applications on the public Internet has accomplished universal communication among billions of Smart IoT (Internet of Things) devices and such smart devices communicate with each other through the network. Simultaneously, cyber-attackers have gained enough knowledge on attacking the Internet and information on Secured systems. Distributed Denial of Services evolved as a serious threat for organizations, these attacks must be mitigated immediately for preventing damages to the business. In Information and Cloud Security, DDoS attack is one of the main concerning successful attack procedures. Since DDoS attack can deny and disturb the service of one or more IoT nodes simultaneously, which causes the loss of data for any company. In other words, this will be a serious threat to the organization. ML models are implemented for resolving various Cybersecurity related problems such as Malware detections, spam mails, vulnerabilities in software, detection of an anomaly, biometric and facial recognition, and many other areas. It is essential to simulate the discrete event DDoS attack model to evaluate the efficiency of the detection model for in-house study on DDoS attacks and preventions. This paper proposes a model to analyze the impact of DDoS (Distributed Denial of Service). The implemented model was developed using Mininet and POX controller on open source. The resultant network traffic is generated as the dataset. The values of the dataset were compared and plotted with various Machine learning methods for the prediction of DDoS attacks.

Keywords: *DDoS attack, SDN, Mininet, AI/ML and DDoS attack, Cyber Security and AI/ML*

1. INTRODUCTION

Today, Internet-based application is crucial for business growth. The size of internet-connected systems in the network growing exponentially. Therefore, the requirement for Mobile Cloud Computing (MCC), Big Data Analytics, Information Cloud Security, Vehicular ad hoc networks (VANET), etc. are always growing in terms of scalability, elasticity, trustworthiness, and accessibility which are likely to add network security risks like cyber-attacks on systems and network. Today data traffic through the internet growing exponentially, Distributed Denial of Attacks is a common issue that affects any internet-connected business. The increasing trend of working from home increased the usage of the internet. Pandemic situations boosted the use of e-commerce for purchases, utility payments, and banking transactions. The number of cyber-crimes and related attacks increased very largely in number.

1.1 DDoS attack and SDN

The DDoS attacks in the conventional IoT Network architecture can deny the service for the allowed clients by overwhelming the excessive amount of malicious bandwidth on the IoT nodes and overloading the computing resources on the smart IoT devices. To overcome these issues, today the network world needs an open network architecture that can easily configure and deploy new network protocols for uninterpreted services. Software-Defined Network (SDN) emerged as a technology to resolve this issue.

The underlying theory of Software Defined Network (SDN) is to separate the control plane from the data plane in the networking stack and transmit each other using non-proprietary open flow protocol. It shows the exciting developments from the traditional network architecture. In the traditional network architecture, the elements of the control plane and data plane is implemented using proprietary software by the commercial networking providers.

1.2 DDoS attack and Machine Learning

Every organization to have a proper cybersecurity design and prevention infrastructure in place and these teams need to prepare for such attacks at any time. Another challenge is the capability of organizations on investigating the logs and traces generated during such kinds of attacks. There will be a huge volume of logs generated however the team may not be having the skills or they may not be capable of analyzing the contents of generated logs.

Artificial Intelligence /Machine Learning emerged as a learning solution where different supervised and unsupervised machine learning models which can detect DDoS attacks are used. These are models capable of analyzing the logs generated during cyber-attacks, these models can analyze the logs based on the available parameters for a specific period using different algorithms. The results of various machine learning models are analyzed at a later stage based on the input parameters or dataset attributes. These Machine Learning techniques can determine the legitimate and malicious network traffic based on the volume of traffic for a specific duration.

In Machine Learning, features like scaling and feature reduction bundled with K-nearest neighbor, Random Forest, Linear regression, Naïve Bayes, Decision tree, SVC when analyzing the dataset values these models detect the accuracy rate of various models based on the training and test data taken.

2. RELATED WORK

In the current IT infrastructure, all connected systems like Cloud, Storage, networks, etc. are closely interlinked and complimenting each other to deliver high performance. Most of these are specific to separate manufactures, DDoS attack emerges as a threat for the current Cyber Infrastructure since an attack can occur any time thus lead to business disruptions. DDoS can occur as a coordinated attack can affect entirely. [1]

A DDoS can simulate using the software-defined network. In a software-defined network, an SDN Application Layer deals with user data via an interface, Control Layer which manages the SDN software, and Infrastructure Layer which manages the API and networking devices like Routers, Hubs, and Switches [2]

The development and growth of Software Defined Network (SDN) and its promises in networking technology are enormous. However, it is believed that every technological development comes with its challenges of which the most prominent in this case is security. The researcher aimed to develop a DDoS testing platform created in a simulated isolated environment to generate a DDoS attack for testing. The implementation was done by emulating the network by SDN which runs on a Virtual Machine (VM) and they showed that the proposed method effectively detects and mitigates DDoS attacks. [3]

To empower most IoT based applications, IoT Bot security will repeatedly be a vital aspect, creating a complete detection technique that essentially protects against various types of DDoS attacks and can provide high accuracy on prediction for IoT BoT DDoS attacks in IoT environment is a principal objective for the forthcoming of IoT enhancement. The advancement of such a technique involves a great knowledge of the approaches that have been applied thus much in the prediction of IoT BoT DDoS attacks in the IoT devices in a smart home environment. In this research, the author tries to emphasize on most relevant Machine Learning (ML) algorithms established for the detection of various categories of DDoS attacks in IoT environment along with their benefit and drawbacks. [4]

Machine Learning got the ability to predict the results based on historical data. Machine Learning is used as a tool to predict patterns based on structured data and unstructured data. The majority of the Software-based networks are used for traffic pattern analysis, intrusion detection, DDoS attack detection, and predictions. Naive-Bayes algorithms and Bayesian network decision tables are used for predicting the host which can get attacked. TensorFlow emerged as another technique that can be used for predicting the pattern. Feature reduction is a method by which data to be selected for analysis can be reduced, Dimension reduction is another technique by which input variables in the dataset can be reduced and limit the variables considered for analysis. [5]

To detect an anomaly, Machine Learning and Artificial Intelligence are extensively used, which are used primarily to classify test data and train data, and algorithms such as, Decision Tree, Confusion Matrix, Support Vector Machine are mainly used. This algorithm helps when using historical data can predict the real-time network traffic and packet transfer which helps in classifying if there is DoS or DDoS attack or the traffic can consider as legitimate. [6]

Machine learning is an ideal way to detect DDoS attacks. K Nearest Neighbours (KNN) is one of the techniques used for traffic classification and correlation. Artificial Neural Network is an algorithm to classify an attack and normal traffic which are analyzed based on patterns. [7]

3. OBJECTIVE & METHODOLOGY

3.1. Objective

The objective of the study is that even IT departments of smaller organizations can develop their DDoS attack test environment the idea is to develop a small DDoS attack test environment with python scripts and simulated environments using SDN and Mininet. The resultant dataset generates from the simulated environment used as input for Machine Learning predictions. Analyze the performance results based on various machine learning algorithms, various available attributes or parameters from the dataset are given as key values, the results were analyzed to identify which parameters can give optimum results. Results are taken based on the yield shared by different data values and the data classification for test and train data. for a specific period. This needs to identify the accuracy based on the input data set based on the actual data or simulated data.

3.2. Methodology

Simulate the SDN network and execute the DDoS attack with the use of python scripts. The intended use of the system is to set up an environment that can generate DDoS attack equivalent traffic with help of software scripts.
The scripts will generate DDoS attack equivalent network traffic. The networking log will collect as data set. The below figure 3.1. explains the attack environment.

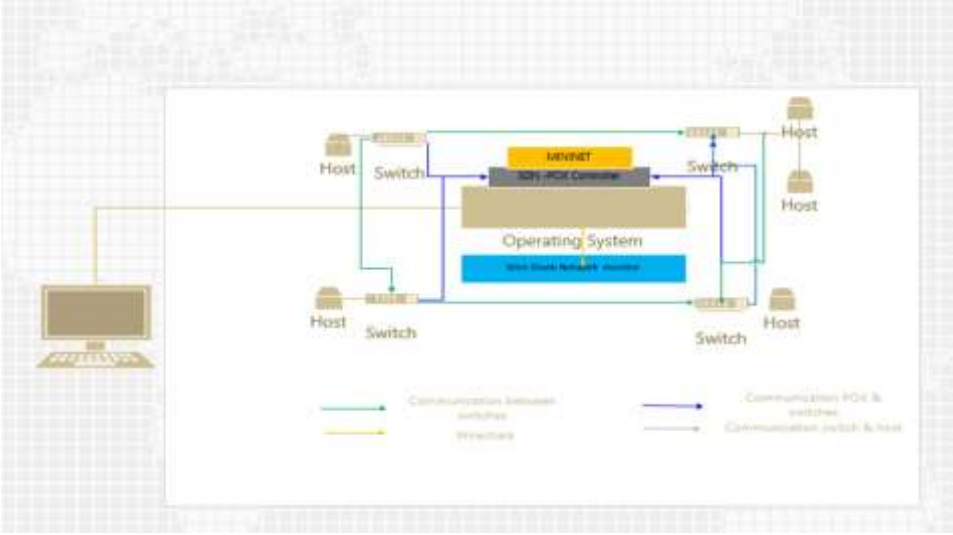


Figure 3.1. DDoS attack system setup

The dataset generated from the attack environment is used for Machine Learning predictions. Data generation, data understanding are the very first steps in an AI/ML model development, next comes the data preparation where unwanted attributes getting omitted for further processing, and later implement Machin algorithm to prepared fine-tuned data to generate analysis and findings. The very first step in Machine Learning is verifying the dataset, operations, and algorithm to use. In this project the identified dataset is manually validated, the data collected is then shared with the Machine Learning techniques as input with the identified header information those are suitable for sharing an optimal result. The header attributes of the selected dataset are to be carefully analyzed when selecting the headers identified for deriving the result. The data was fed into Machine Learning programs by selecting various fields using filtering with the help of the program. The below figure 3.2. details the typical process flow.

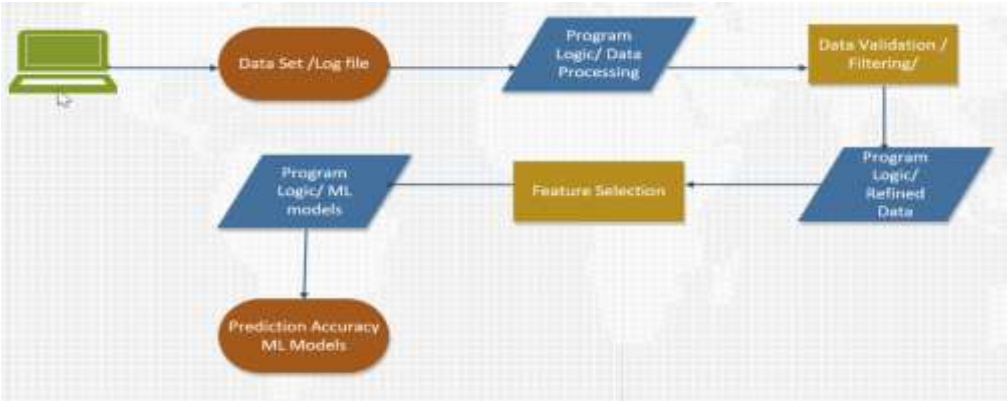


Figure 3.2. Machine Learning Process flow

4. IMPLEMENTATION

4.1. DDoS attack

The Project setup is on a physical system, required virtual computer hosts, switches and networks are being set up using software programs that are available as open-source. The software design starts with identifying the scope and finalizing the scope. Once the scope got finalized then identify the software and programs required for the DDoS environment. The Web tool required for building the environment is finalized at this stage.

The python programs were developed after going through quality checks, the aim is to validate if the program runs as expected, codes developed got any performance-related issues. If there are issues fix them at the quality gate approval stage. To work on emulated hosts, the desired host must call at the operating system level, the programs developed are executed at the host's level

i. Setup the Mininet emulation software.
ii. The SDN network for POX controller setup along with Mininet
iii. At the operating system level using a terminal window, the POX controller got started.
iv. Mininet creates the hosts with switches, controllers, and hosts when starting with input parameters.
v. The corresponding network, switches, and hosts created
vi. Check and verify if the basic connectivity from hosts are established
vii. To generate network traffic python scripts are used.
viii. The resultant network traffic can be analyzed from system logs or with help of tools like Wireshark. This can be saved as CSV files for analysis

The network traffic needs to be monitored and verified at the operating system level; the details of network traffic can capture using a network monitoring tool.

4.2. Machine Learning model

The AI/ML model developed at six stages such as Data Identification, Pre-Processing, Feature Selection & Feature Engineering, Vectorization, Pooling and Later adding weightage, the below brief the activities in each stage.

Data Identification: To develop the implementation of the AI/ML module start with data identification. The log details collected for network traffic are used as the input dataset for the AI/ML models, which will be the data input for the AI/ML model.

Pre-Processing: The dataset may contain a huge number of records that may be irrelevant. Pre-processing in which all the available fields are studied, their relevance is identified, once the details are identified the necessary features required are selected and the rest of the fields are eliminated from further processing since they don't have any relevance in deciding the results.

Feature Engineering: Feature engineering is an area that is getting the expertise to choose the features based on attributes, properties, and characteristics from unsalted data. There are different feature engineering methods available, some of them are Scaling, Log transform, Grouping operations. Add additional variables in data for getting better accuracy of the machine learning model.

Feature Selection: In Feature selection which has a significant advantage on model prediction, the attributes that are selected for the model have benefits on results. Irrelevant attributes harm the results, or they may give negative predictions and the entire decision may go wrong. In future selection either manually or automatically it selects the features which provide the desired outcome. By using future selection, that can have the following benefits.

Pooling: Used pooling for downsampling the dataset and for training parameters to get desired results.

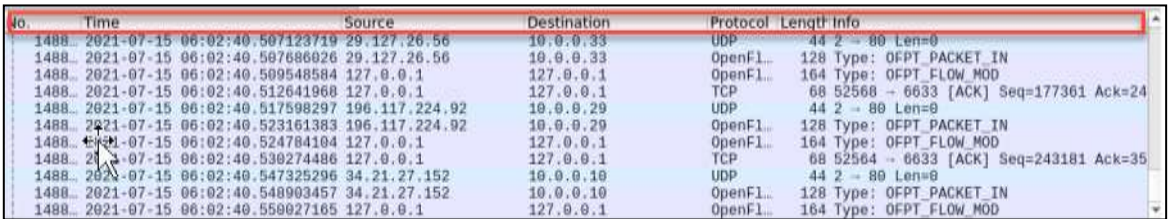
Weightage: Weightage decides the influence of input parameters in the results. Model weightage parameters are train and untrained data.

Prediction: Those are the outcomes receiving for various Machine Learning models after processing various algorithms, that will analyze the results based on the accuracy those are getting as output for various models, the models with more accuracy considered for future predictions.

5. RESULTS

5.1. DDoS attack results

To generate the network communication, the SDN controller (POX) needs to be up and running, below will be the resultant output when the POX controller is available on the host. The developed python program got executed for generating the network traffic. The network traffic when running the python program can be captured at the operating system level or with help of network monitoring tools such as Wireshark as below figure 5.1. The network traffic can capture as "CSV" files for analysis.



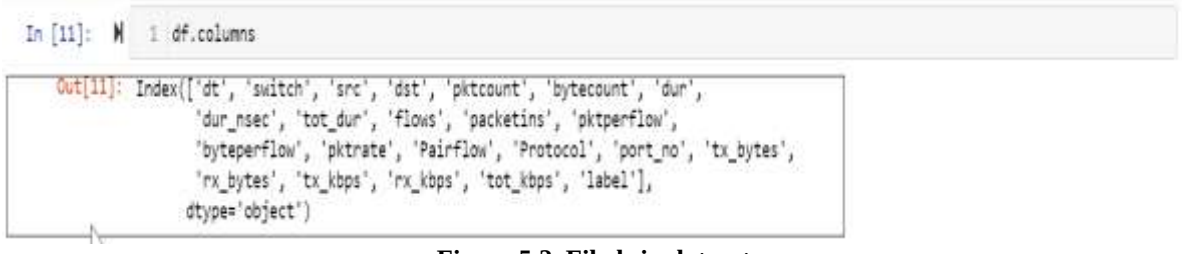
No.	Time	Source	Destination	Protocol	Length	Info
1488…	2021-07-15 06:02:40.507123719	29.127.26.56	10.0.0.33	UDP	44	2 → 80 Len=0
1488…	2021-07-15 06:02:40.507686026	29.127.26.56	10.0.0.33	OpenFl…	128	Type: OFPT_PACKET_IN
1488…	2021-07-15 06:02:40.509548584	127.0.0.1	127.0.0.1	OpenFl…	164	Type: OFPT_FLOW_MOD
1488…	2021-07-15 06:02:40.512641968	127.0.0.1	127.0.0.1	TCP	68	52568 → 6633 [ACK] Seq=177361 Ack=24
1488…	2021-07-15 06:02:40.517598297	196.117.224.92	10.0.0.29	UDP	44	2 → 80 Len=0
1488…	2021-07-15 06:02:40.523161383	196.117.224.92	10.0.0.29	OpenFl…	128	Type: OFPT_PACKET_IN
1488…	2021-07-15 06:02:40.524784104	127.0.0.1	127.0.0.1	OpenFl…	164	Type: OFPT_FLOW_MOD
1488…	2021-07-15 06:02:40.530274486	127.0.0.1	127.0.0.1	TCP	68	52564 → 6633 [ACK] Seq=243181 Ack=35
1488…	2021-07-15 06:02:40.547325296	34.21.27.152	10.0.0.10	UDP	44	2 → 80 Len=0
1488…	2021-07-15 06:02:40.548903457	34.21.27.152	10.0.0.10	OpenFl…	128	Type: OFPT_PACKET_IN
1488…	2021-07-15 06:02:40.559027165	127.0.0.1	127.0.0.1	OpenFl…	164	Type: OFPT_FLOW_MOD

Figure 5.1. network traffic

5.2. Machine Learning Analysis:

To make the Machine Learning analysis us the "CSV" files as a dataset.

- **i.** Verify the contents of Dataset (number of records and total available fields) to get an idea of the total number of records that are considered when developing the model.
- **ii.** The selected data set contains a total of 104345 records and 23 fields.
- **iii.** The below output presents the total number of fields, available in the dataset those used for generating the ML model (figure 5.2)



```
In [11]: M  1 df.columns

Out[11]: Index(['dt', 'switch', 'src', 'dst', 'pktcount', 'bytecount', 'dur',
                'dur_nsec', 'tot_dur', 'flows', 'packetins', 'pktperflow',
                'byteperflow', 'pktrate', 'Pairflow', 'Protocol', 'port_no', 'tx_bytes',
                'rx_bytes', 'tx_kbps', 'rx_kbps', 'tot_kbps', 'label'],
               dtype='object')
```

Figure 5.2. Fileds in dataset

- **iv.** When selecting the field with the field attribute 'Label' the contents are shared with resultant values '0' and '1' there are 63561 records for '0' and 40784 records with value '1' which is a significant amount for developing the models. (figure 5.3.)

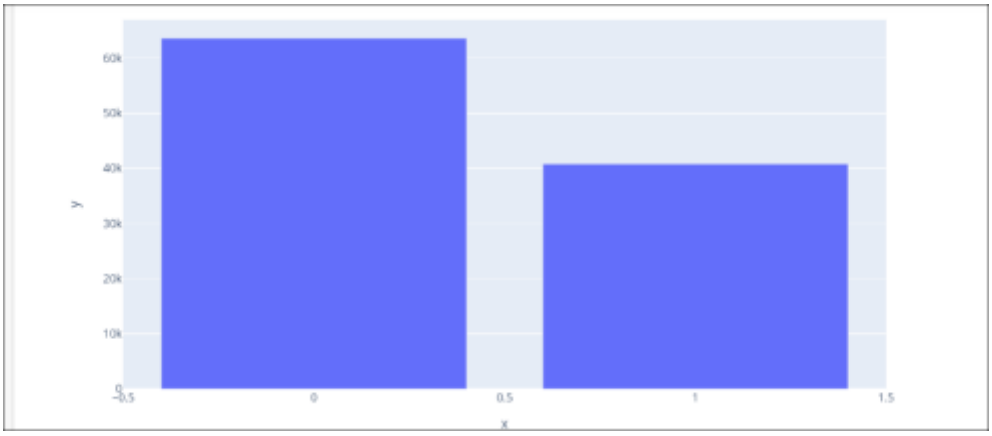


Figure 5.3. Number of records based on field 'Label'

v. Feature selection such as additional values got added to field "Label" to classify it as attack traffic and normal traffic for better prediction.

vi. Dataset whose fields recorded with value as zero got dropped from further processing

vii. Receiving Operating Character which is used for error classification implemented in the predictions.

viii. Confusion matrix for various ML models predicted

ix. The predictions for K-nearest neighbor as below in figure 5.4.

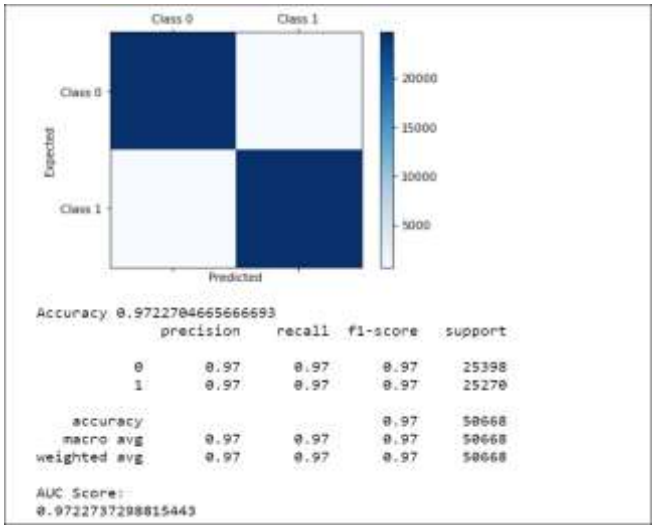


Figure 5.4. KNN accuracy of prediction

x.Consolidated findings of various models as below in Table 5.1.

Table 5.1. Consolidated results of various Machine Learning models

AI/ML model	Accuracy
Logistic Regression	.75
Decision Tree Classification	1
Random Forest Tree	1
K-nearest neighbors	.97
Support Vector Machine	.97
Gaussian NB	.69
TensorFlow	.97

6. CONCLUSION

This paper is about developing a DDoS attack in-house test environment. The network traffic generated at the environment when running the programs developed using python program generates network traffic like DDoS attack. The network traffic collected was stored in human-readable format for further analysis and studies. The analysis is about if such abnormal network traffic affects networks or systems also makes use of the log (dataset) generated for developing AI/ML-based model DDoS attack predictions.

The performance of various learning models was analyzed in this project, various AI/ML models had given different outputs on accuracy which explains the selection of attributes or parameters that affects the selected AI/ML model. Feature selection plays a vital role in the accuracy of outcome, the comparison of results was shared under 'Analysis and Results'.

The project was run on a dataset generated using a simulated dataset that has around twenty-three attributes during fine-tuning omitted seven attributes which don't have any impact on the predictions. As per the observations, it is evident that more the number of attributes resultant in more combination of predictions.

The prediction result may vary based on the combination of the dataset used for analysis. The predictions can vary for a larger number of records and datasets with a larger number of fields. Feature engineering helps in identifying the data input.

7. REFERENCES

[1] A. Sangodoyin, B. Modu, I. Awan, and J. P. Disso, "An Approach to Detecting Distributed Denial of Service Attacks in Software Defined Networks Defined Networks," IEEE-2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud), vol. DOI 10.1109/FiCloud.2018.00069, no. Date Added to IEEE Xplore: 10 September 2018, p. 8, 2018.

[2] O. J. J. E. Saravanan Krishnan, "Mitigating DDoS Attacks in Software Defined Networks," IEEE 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), vol. INSPEC Accession Number: 19896187, no. 10.1109/ICOEI.2019.8862589, p. 4, 2019

[3] N. A. Babatunde Hafis LAWAL, "Real-Time Detection and Mitigation of Distributed Denial of Service (DDoS) Attacks in Software-Defined Networking (SDN)," IEEE, vol. 10.1109/SIU.2018.8404674, no. 09 July 2018, p. 4 Fig 2.2, 2018.

[4] S. G. S. A. S. AHMED A. AWAD, "Collaborative Framework for Early Detection of RAT-Bots Attacks," IEEE-, no. date of publication May 29, 2019, p. 11, 2019.

[5] B. G. Assefa and O. Ozkasap, "MER-SDN: Machine Learning Framework for Traffic-Aware Energy Efficient Routing in SDN," IEEE, vol. IEEE 16th Internation Conference, no. IEEE Xplore: 29 October 2018, 2018.

[6] S. Gangadhar and J. P. Sterbenz, "Machine learning aided traffic tolerance to improve resilience for software-defined networks," IEEE -2017 9th International Workshop on Resilient Networks Design and Modeling (RNDM), vol. 2017 9th international workshop, no. 02 November 2017, p. 7, 2017.

[7] Z. Chen, F. Jiang, Y. Cheng, X. Gu, W. Liu and J. Peng, "XGBoost Classifier for DDoS Attack Detection and Analysis in SDN-Based Cloud," IEEE, vol. 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), no. IEEE Xplore: 28 May 2018, 2018

