A Study of DDOS Attack Classification Using Machine Learning Classifiers

Soe Kalayar Naing
Cyber Security Research Lab
University of Computer Studies, Yangon
Yangon, Myanmar
soekalayarnaing@ucsy.edu.mm

Tin Thein Thwel
Cyber Security Research Lab
University of Computer Studies, Yangon
Yangon, Myanmar
tintheinthwel@ucsy.edu.mm

***Abstract*— Malware threats, security attacks and intrusion are security risks and among them, Distributed Denial of Service (DDOS) attack continues to smash the constructing of integrity and confidentiality of computer systems and network systems. Regardless of the development of security protection techniques, DDOS remains as a severe and challenging issues and hence still need to do more efficient and effective protective methods to expose these DDOS attacks. Specifically, this experiment aims on comparing the classification performance of machine learning algorithms using the open DDOS attack dataset. Some of the outstanding machine learning algorithms, namely, Stochastic Gradient Classifier, Support Vector Machine (SVM), k Nearest Neighbor (kNN), Naïve Bayes and Logistic Regression models are experimented on the open DDOS dataset. According to this experiment, the most outstanding classifier for the classification of DDOS attacks is Logistic Regression classifier. A clear and better understanding of the DDOS dataset for the network traffic is obtained as an additional advantage.**

Keywords—machine learning; DDoS attack

I. INTRODUCTION

DDOS attacks are one of the most threatening security threats in the future and they are growing day by day. This kind of attacks disallow the legitimate network traffic from connecting to its destination by bombarding it with malicious network traffic by blocking the network traffic. The attackers order and control the bots to flood with packets to cause the network traffic overflow.

DDoS attacks become an important threat and intrusion in the privacy and security of network security. Figure 1 illustrates the scenario of the DDOS attack. Distributed DoS attacks can be performed by utilizing the devices such as computers and IoT devices which are connected to the internet and infected with malicious software. Then, an attacker takes control of these devices remotely. These kinds of infected devices are called bots and the collection of those bots are known as botnet.

The effect of DDoS attack is to make a web site or web services unavailable to an authenticated user. Some of the symptoms of DDoS attack include a large number of network traffic coming from the same source internet address and a flood of abnormal network connection requests to a single web site or destination internet address.

Although many researches on machine learning have been done on detection of those attacks, there is still limitation on good quality DDOS datasets. This study experiments on the state-of-art of open dataset.

It is crucial to discriminate whether it is attack or normal behavior just by looking through the types of internet protocols and network services. Detecting the DDOS attack is hard to determine and becomes a challenging issue for organization operating and integrating their operation and technologies in public network.

It is important to discriminate the fraudulent attacks of bombarding requests so that the unavailability of legitimate service and economic losses can be avoided.

Developing new technologies and experimenting existing technologies have been done in the area of protecting information security and network security. Among these technologies, machine learning techniques gains the great success in that field. DDOS dataset is input to standard machine learning models and the dataset is split as training and testing data for efficient detection and classification of DDOS attacks. The performance of DDOS attack classification is computed and compared with various evaluation scores such as accuracy, precision, recall and F1 score. Hence, experimenting and selecting the optimal machine learning model with open DDOS dataset is essential to determine and detect DDOS attacks in advance before encountering the major security problems in the network.

This paper devotes the experimental results of machine learning algorithms to precisely detect and classify the DDOS attacks. Machine learning classifiers are built using the open DDOS dataset. Then, the classifier with the highest performance measure is selected for detecting the malicious network traffic.

The reference and related work are presented in second section of the paper. The third section presents the system design and the machine learning algorithms experimented in this paper. Section 4 describes the deploying machine learning models and experimental results of DDOS attack classification. Finally, section 5 summarizes the process of DDOS attack classification.

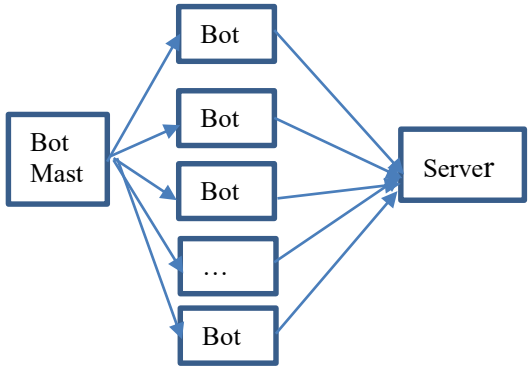

Fig. 1: A DDoS Attack Scenario

978-1-6654-3599-4/23/$31.00 ©2023 IEEE

2023 IEEE Conference on Computer Applications (ICCA) | 978-1-6654-3599-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCA51723.2023.10182146

II. RELATED WORK

DDOS attacks are still a security issue in business organization and networks. These security challenges may lead to security threats to determine and detect those DDOS attacks. While trying to detect and classify the DDOS attacks, the performance of the network of the organization can be slow. Under the classification problems of DDOS attacks, machine learning is one of the best efficient methodologies in detecting the known and unknown DDOS attacks with the aid of various machine learning models.

Few studies have been made related to the DDOS attacks classification using different kinds of techniques. Ashutosh [1] proposed a DDoS attack classification in a smarter way using machine learning techniques. The author presented SVM to be the better classifier than the Naïve Bayes in his experiment. Swathi[11] proposed an approach for DDOS attacks classification applying multiple liner regression. His proposed solution detects the attack depending on the types of attacks.

Zekri [6] presented an approach of machine learning models in cloud architecture which is built on flooding-based attack emphasizing layer 3 and 4 of the 7-layer model. In [11], authors proposed on unsupervised anomaly detection, under NLP literature concentrating on the flow of packets as a 'language' between machines and used Recurrent Neural network.

This kind of DDOS attacks detection is done in the work [2]. This author implemented using a Convolutional Neural Network (CNN) and compared its performance with a Recurrent Neural Network (RNN).

Therefore, it is essential to avoid the major security problems in advance and deploy the optimal machine learning classification model for detecting the DDOS attacks early.

III. SYSTEM DESIGN FOR DDOS ATTACK CLASSIFICATION

This study experiments the DDOS attack detection using APA-DDOS dataset [4]. Some of the machine learning models are testing with this DDOS data. The machine learning methodology implemented in this study for the classification of DDOS attack is provided in this section. SVM, kNN, Stochastic Gradient classifier and Logistic regression machine learning models are presented. Figure 3 demostrates the diagram of the experimented process for the detection of DDOS attacks. The input DDOS data is cleaned and pre-processed removing the null values. Then the cleaned data is input to model building process, in this step, the input data is split as training and testing dataset. After building the machine learning models, the classification performance of each model is computed and evaluated using the precision, recall, F-measure and accuracy scores.

The machine learning model with the highest evaluation score is selected as an optimal DDOS attack classifier.

A. Datasets

DDOS attack detection is experimented using the APA-DDOS dataset collected from Kaggle[4]. This dataset contains the fresh attacks related to DDOS-ACK and DDOS-PUSH_ACK flooding DDOS attack in the network. Both normal network traffic and malicious traffic are involved in this dataset. The dataset has a total of 151201 records. The number of data columns is twenty two and data columns are described in Table 1.

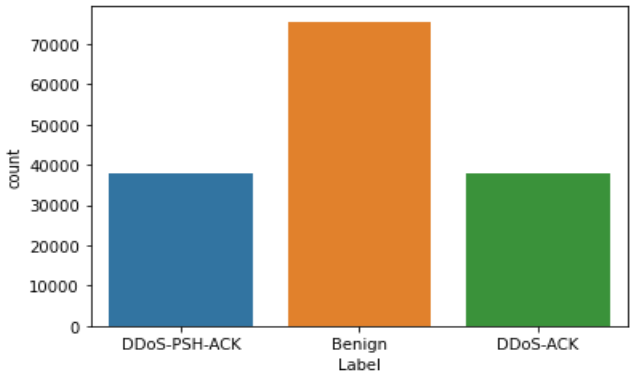

Fig. 2: DDOS Attacks Types in the APA_DDOS Dataset

The APA-DDOS dataset is constructed based on the source and destination IP, source and destination Port, types of protocols and timeframes to detect whether it is DDOS attack or not. In this dataset, DDOS-PSH-ACK, DDoS-ACK and normal network traffic or BENIGN are included.

TABLE 1: (A) ATTRIBUTES OF THE APA-DDOS DATASET

No.	Data columns
1.	ip.src
2.	ip.dst
3.	tcp.srcport
4.	tcp.dstport
5.	ip.proto
6.	frame.len
7.	tcp.flags.syn
8.	tcp.flags.reset
9.	tcp.flags.push
10.	tcp.flags.ack
11.	ip.flags.mf
12.	ip.flags.df
13.	ip.flags.rb
14.	tcp.seq
15.	tcp.ack
16.	frame.time
17.	Packets
18.	Bytes
19.	Tx Bytes
20.	Rx Packets
21.	Rx Bytes
22.	Label

(B) SOME INSTANCES OF THE APA-DDOS DATASET

	ip.src	ip.dst	tcp.srcpc	tcp.dstpc	ip.proto	frame.len	tcp.flags.	tcp.flags.	tcp.flags.	tcp.flags.	ip.flags.m	ip.flags.d	ip.flags.rt	tcp.seq	tcp.ack	frame.tim	Packets	Bytes	Tx Packe	Tx Bytes	Rx Packe	Rx Bytes	Label
2	192.168.1	192.168.2	2412	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	8	432	4	216	4	216	DDoS-PSH-ACK
3	192.168.1	192.168.2	2413	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	10	540	5	270	5	270	DDoS-PSH-ACK
4	192.168.1	192.168.2	2414	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	12	648	6	324	6	324	DDoS-PSH-ACK
5	192.168.1	192.168.2	2415	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	10	540	5	270	5	270	DDoS-PSH-ACK
6	192.168.1	192.168.2	2416	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	6	324	3	162	3	162	DDoS-PSH-ACK
7	192.168.1	192.168.2	2417	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	12	648	6	324	6	324	DDoS-PSH-ACK
8	192.168.1	192.168.2	2418	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	8	432	4	216	4	216	DDoS-PSH-ACK
9	192.168.1	192.168.2	2419	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	12	648	6	324	6	324	DDoS-PSH-ACK
10	192.168.1	192.168.2	2420	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	12	648	6	324	6	324	DDoS-PSH-ACK
11	192.168.1	192.168.2	2421	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	12	648	6	324	6	324	DDoS-PSH-ACK
12	192.168.1	192.168.2	2422	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	10	540	5	270	5	270	DDoS-PSH-ACK
13	192.168.1	192.168.2	2423	8000	6	54	0	0	1	1	0	0	0	1	1	16-Jun-2	12	648	6	324	6	324	DDoS-PSH-ACK

B. Machine Learning Algorithms

(1) Logistic Regression: Logistic Regression is a type of supervised learning algorithm which can classify the DDOS attacks whether DDOS or not based on the probability theorem.

(2) Gaussian Naive Bayes: In this Bayes theorem, the classification of the class value is done by computing the conditional attributes for the executed features.

(3) K-Nearest Neighbors (kNN): It is a kind of supervised learning method. It works well for both regression and classification process. The distance between *k* number of neighbors and a new unknown data is calculated for classification or prediction. It classifies the data point depending on the similarity measures computed using the distance between neighbor data points.

(4) Support Vector Machine (SVM): It is supervised machine learning model and suited for both classification and regression problems. Linear SVM and non-linear SVM. Linear SVM models are the two kinds of SVM classifier. The former is used for linearly separable data points using the single straight line. The latter is manipulated for non-linearly separable data and the straight line cannot be used to classify those data points. The main advantage of SVM is faster processing speed and higher performance compared with the neural network algorithms.

(5) Stochastic Gradient Descent: It is an optimization algorithm which finds an optimal solution to be the best fit between the actual and predicted output for a wide range of problems. It is an iteratively finding the parameters that minimize the cost function.

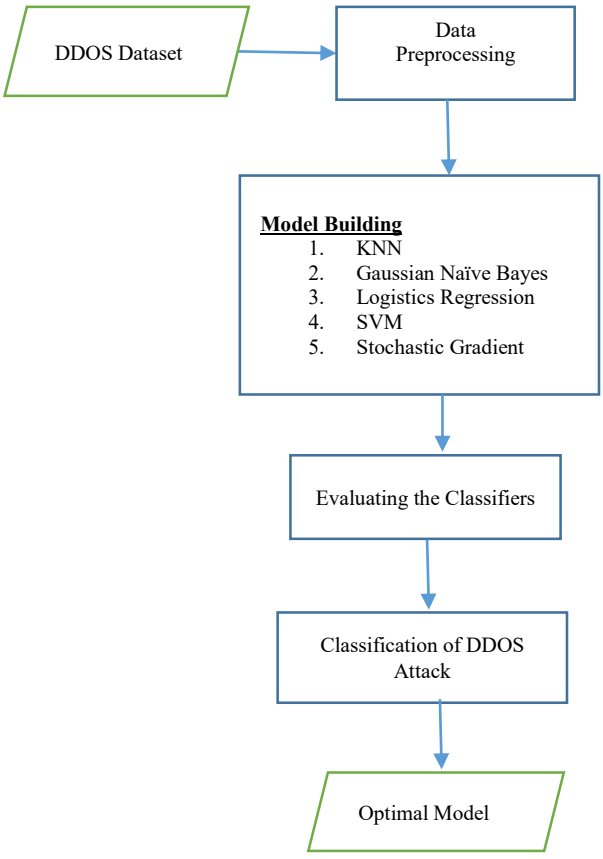

Fig. 3: System Design for DDOS Attack Detection

C. *Performance Measures*

The attack classification performance is measures using the confusion matrix presented in Table 2. Evaluation measures are identified as in Table. 2. The evaluation terms are:

- TP (True Positive) : Number of malicious requests classified as DDOS attack.
- TN (True Negative) : Number of legitimate requests classified as normal traffic.
- FP (False Positive) : Number of legitimate requests classified as DDOS attack.
- FN (False Negative) : Number of malicious requests classified as normal traffic.

TABLE 2. CONFUSION MATRIX

		Detected	
		Positive	Negative
Actual	Positive	True Positive (TP)	True Negative (TN)
	Negative	False Positive (FP)	False Negative (FN)

Accuracy, precision and recall will be calculated for the experimented classifiers using all the similarity measures and distance metrics to compare and assess the performance of the DDOS attack classifiers.

IV. EXPERIMENTAL RESULT AND DISCUSSION

DDOS network traffic attacks are still as severe problem in computerized system and network system. It is crucial to develop an effective and efficient network traffic classification system to reduce the number of DDOS security attacks. This testing uses the fresh DDOS security dataset and examined the optimal performer for security analysis.

DDoS dataset is split into 80-20 ratio as training and testing dataset and eighty percent of the data is used for building the learning models and the other rest of the twenty percent data applied to experiment the performance of the attack detection model to estimate whether it is normal or attack.

In order to detect DDoS attack, malicious or abnormal network traffic are identified depending on the packer header information and data contents. The receiving address examines the incoming data packet information such as source internet address, port number, destination internet protocol address and port number. Normally, establishing network communication between the web client and web server needs to make a three-way handshaking. But when a flood of TCP or UDP requests are submitted to the targeted server, the corresponding syn-ack replies cannot continue as computer resources are consumed by these malicious requests.

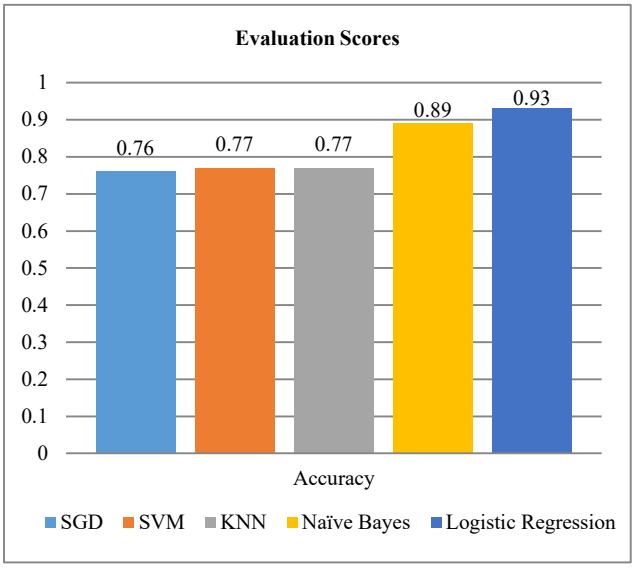

Fig. 4: Accuracy Measurement of ML Models

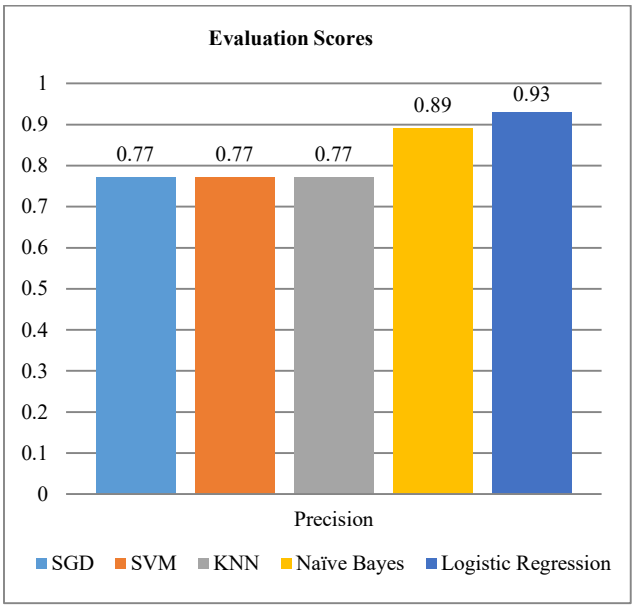

Fig. 5: Precision Measurement of ML Models

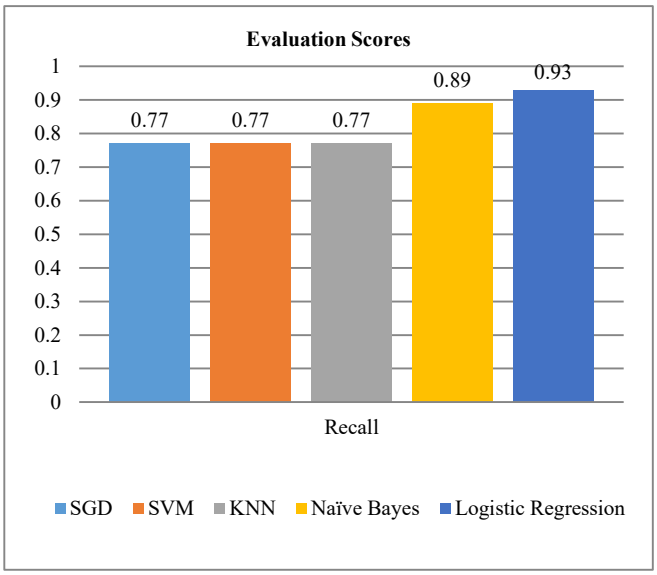

Fig. 6: Recall Measurement of ML Models

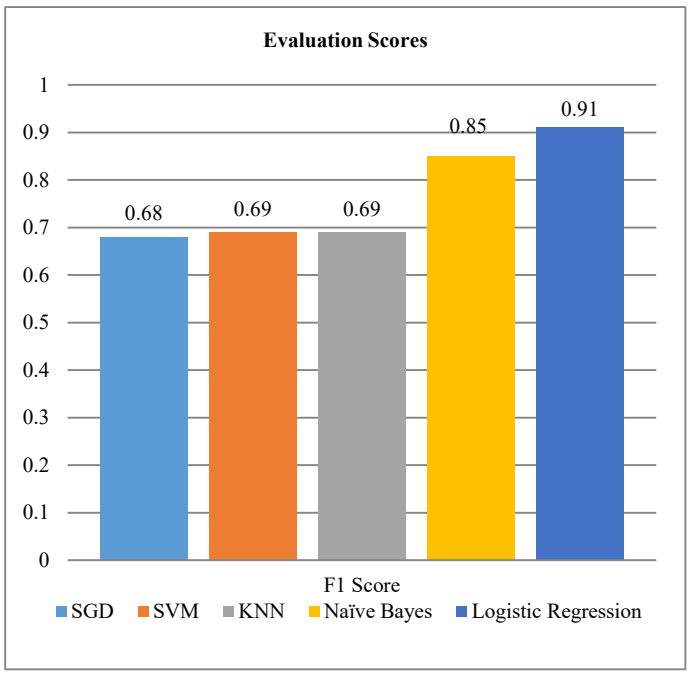

Fig. 7: F1-Score Measurement of ML Models

TABLE 3: EVALUATION SCORES OF THE MACHINE LEARNING CLASSIFIERS

ML Model	Precision	Recall	F1 Score	Accuracy
SGD	0.77	0.77	0.68	0.76
SVM	0.77	0.77	0.69	0.77
KNN	0.77	0.77	0.69	0.77
NAÏVE BAYES	0.89	0.89	0.85	0.89
LOGISTIC REGRESSION	0.93	0.93	0.91	0.93

Accuracy, precision, recall and F-measure are computed for the classifiers' performance analysis. Table 3 shows the classification performance's measures for DDOS attack classification applying machine learning classifiers. The machine learning classifiers and their respective scores for accuracy measurement are described comparatively in Figure 4, and scores for precision measurement are described comparatively in Figure 5. In figure 6 and 7, recall and f1-score measurement of machine learning models are described respectively.

APA-DDOS dataset [4] is used for the purpose of performance analysis of ML algorithms used in the classification of DDOS attacks. There are 24 features in this dataset. So, this measures are the result implemented without removing features from APA-DDoS dataset.

According to the classification results, it shows that machine learning classifiers are advantageous in detecting and discriminating the types of DDOS attacks. The results shows that Logistic regression classifier enhances the attack detection process increasing the accuracy compared with other machine learning classifiers.

The classification results show that the optimal accuracy is achieved with the Logistic Regression model with 93% accuracy score. SVM and kNN have obtained the equal performance scores of 77%. Naïve bayes classifier is a little

bit lower accuracy score than the optimal performer, Logistic regressor.

It is superior to build with logistic regression for classification problems whenever the relation among the data is non-linear. Logistic regression classifier is able to achieve in designing the complicated relations among variables. Hence, logistic regression model becomes the best performer in this classification problem and linear classifier estimate inferior classification results compared with this logistic regressor because of non-linearity nature of the APA-DDOS data tested in this experiment.

V. LIMITATIONS

This experiment has tested DDoS attack using machine learning methods based on the offline analysis. Actually, DDoS attack types are occurred in real time and not in offline situation. The main limitation of this experiment is detecting DDoS attack using real time network data flow.

VI. CONCLUSION

According to the experimental result based on machine learning classifiers, Logistic regressor gives the most outstanding performance measure among the four machine learning classifiers. Hence, logistic regression model becomes the well-performed DDOS attack detection classifier for the open DDOS attack dataset. The classification performance is evaluated and computed using accuracy, precision, recall and F-measure. Logistic regression classifier performs very well for this open dataset and SVM and kNN are the weak learners for this DDOS attack classification experiment.

REFERENCES

[1] Ashutosh Nath Rimal and Dr. Raja raveen; DDOS Attack Detection Using Machine Learning – 2020.

[2] Dong, S., Abbas, K., & Jain, R. (2019). A survey on distributed denial of service (DDoS) attacks in SDN and cloud computing environments. IEEE Access, 7, 80813-80828.

[3] Jiangtao Pei, Yunli Chen1 and Wei Ji1;A DDoS Attack Detection Method Based on Machine Learning; IOP Conf. Series: Journal of Physics: Conf. Series 1237 (2019) 032040.

[4] https://www.kaggle.com/datasets/yashwa nthkumbam/apaddos-dataset

[5] https://cybersecurity.att.com/blogs/security-essentials/types-of-ddos-attacks-explained

[6] Marwane Zekri, Said El Kafhali, Noureddine Aboutabit1 and Youssef Saadi DDoS Attack Detection using Machine Learning Techniques in Cloud Computing Environments

[7] Sharafaldin, I.; Lashkari, A.H.; Hakak, S.; Ghorbani, A.A. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In Proceedings of the International Carnahan Conference on Security Technology, Chennai, India,1–3 October 2019.

[8] Parvinder Singh Saini, Sunny Behal, Sajal Bhatia; Detection of DDoS Attacks using Machine Learning Algorithms 12-14 March 2020.

[9] Muthamil Sudar, K., & Deepalakshmi, P. (2020). A two-level security mechanism to detect a DDoS flooding attack in software-defined networks using entropy-based and C4.5 techniques. Journal of High-Speed Networks, (Preprint), 1-22.

[10] 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. Authorized licensed use is limited to Tencent.

[11] Swathi Sambangi and Lakshmeeswari Gondi; A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression 2020.