EasyChair Preprint

№ 3201

Distributed Denial of Service Attack Classification Using Artificial Neural Networks.

Bhargavi Goparaju and Srinivasa Rao Bandla

April 20, 2020

Distributed Denial of Service Attack classification using Artificial Neural Networks

1.Bhargavi Goparaju,
Research scholar,
Acharya Nagarjuna university.
2.Dr.Bandla Srinivas Rao,
Research guide,
Acharya Nagarjuna university.

Abstract

For all sizes of organizations and ISPs, most devastating attacks of all time are emerged by the DDoS Attacks (Distributed Denial of Service). The contribution of production of unsecured botnets and IoT devices in the range of billions number is lead to the increment of DDoS attacks due to the improved availability of services of DDoS-for-hire. Continuously, these DDoS attacks are growing in frequency, magnitude, and sophistication. Owing to the smarter growing of these attacks day by day and evasion of IDS, the legacy methods are challenged that include scrubbing and signature-based detection. As the scale of attacks mostly concentrating on the organizations, the security technologies of next-generation can't keep in the pace. Due to the higher demand of human intervention, various limitations are included the anomaly-based detection with false positives and accuracy. By using machine learning (ML) model, DDoS anomaly detection based on the dataset of open CICIDS2017 is presented in this paper. However, maximum accuracy is reached with the use of this ML model and tuning hyper parameters meticulously.

Keywords: DDoS Attacks, anomaly detection, machine learning, Intrusion detection system, accuracy.

I. Introduction

By exhausting the network services and resources, DDoS attacks are causing the denial of service to generate valid requests. From distributed sources, the attack will be initiated to improve the impact i.e. known as distributed denial of service attack. Based on the botnets, these attacks are started to launch in most of the cases. According to the reports exposed by git hub, the largest attack of DDoS on the recent records is occurred on Feb 2018. From different systems of autonomous with thousands of number, the attack DDoS is initiated across unique end points ranging from tens to thousands of number. By using the Memchached-based which is at peak level point of 1.35Tbps, it was considered as an amplification attack. In a DDoS attack with high-volumetric feature which is around at peak of 1.1 Tbps, another different DDoS attack known as Mirai [1] botnet was used that involved an internet with major part in October 2016. By utilizing home routers, poor security of cameras, printers, and DVRs with default credentials, Mirai is commanded nearly 100,000 bots successfully based on the usage of telnet ports. However, it's not an infection but it's just included the policy with lowest security and attention is not presented by vendors.

DDoS attacks are included three categories such as protocol attacks, application-level attacks, and volumetric attacks. In this literature paper, some of the details like work and dataset under each category will discuss.

1.1 Application-level Attacks:

In layer 7 protocols like HTTP, the vulnerability can be used in the application level attacks which usually come under the low-volumetric attacks [2] category. In general, application-level attacks are considered as the toughest among all available DDoS attacks since they are difficult to detect and restrict. The characteristics of the attacks [3] have been included cautious and sophisticated as they can be effective using a generated traffic based on a single machine at a low rate and the web server is crashed by them. By using the monitoring attacks

with traditional flow-based like IDS (Intrusion Detection System) [4], these attacks can't be monitored easily that means its difficult process for detecting these attacks proactively.

Under this category of application-level attacks, the most common attacks are Slow read attack, Slowloris, and Slow HTTP POST [5]. By raising CPU usage and considerable memory on the server, GET or POST floods, concurrent connections pool, and Apache Range Header attacks are exhausted.

Slowloris Attack:

HTTP Protocol design is targeted by Slowloris and Slow HTTP POST DoS that have the expectation of completed the requests received by the server before processing them. The resources are held in the state of busy waiting for the remaining data that reserved for incomplete requests if in case of incomplete requests of an HTTP protocol or slow transmission of packets. It's not ready to process or handle new requests in this process when allocating all available resources to that kind of requests and it leads to the denial of service. By slower transmission of HTTP headers to server within the allowable time by the server at maximum point, HTTP server is targeted by this tool without completion of a request.

HTTP GET or POST Flood Attack:

For various kinds of content such as images, files, or other web resources from a server, coordination of distributed clients is occurred in this form of an attack for sending multiple requests of HTTP GET. The denial of service is caused based on the denial of legitimate requests if the target is flooded with responses and requests.

Based on HTTP POST requests, the server can be flooded in a similar way. When compared to HTTP GET, the impact of attack will be more in POST requests which include in database write operation. Whereas HTTP GET requests involve operations with more intensive which requires high amount of bandwidth and processing power. To cause denial of service, the target server's capacity can be saturated easily with the transmission of flood of POST requests.

1.2 Protocol Attacks:

In Layer 3 and Layer 4 implementation of systems, the vulnerabilities are targeted by these attacks. The most commonly type of attacks is included Ping of Death, SYN Flood Attacks, Smurf Attacks [6], and fragmented packet attacks. Bandwidth of a network and resources of server are consumed in this kind of attack.

SYN Flood Attacks:

In the TCP Connection Sequence, the weakness is exploited by this attack. By exchanging sequence of TCP Packet with SYN, ACK flags, SYN-ACK initiating from client end, the establishment of TCP connection is done based on a 3-way handshake process between server and client in a scenario of normal connection. Without getting to the response of SYN-ACK, multiple SYN attacks are sent by the attacker when the SYN flood [7]. From IP addresses with source spoofed, TCP SYN can also be transmitted. By searching out resources for new connection and reserving the resources for each of request, the server is in wait state for SYN-ACK in either way that ultimately lead to the denial of service.

Ping of death:

The malicious pings or multiple malformed are sent to the target by an attacker in this case. On datalink layer, the limitations of MTU are posed by the IP packet's maximum length with 65,535 bytes and the datalink layer is divided the packet into the fragments of 1500 bytes. For benign packets, buffer over flow is caused by reassembling the target host that can cause the denial of service finally.

Smurf Attack:

Based on smurf malware, it is executed the operations by establishing a spoofed bucket through the setting of source IP to the target victim's IP address. With the sending of IP packets to an intermediate network's IP broadcast address, the attack is amplified. Using

ICMP echo, the intermediate network is responded that is connected by each host and send replies of packets to target. Potentially, the target is resulted in the denial of service for legitimating the traffic.

1.3 Volumetric Attacks:

For denying the legitimate services, the resources and bandwidth of a network are starved out in these kinds of attacks with the flooding of UDP traffic. In order to multiply the traffic bandwidth of traffic through the services such as NTP, DNS [8], and Memcached, reflection is utilized to hide the identity of a source through the amplification and spoofing source IP address. Sending of a request with spoofs to UDP services is the primary key feature of any amplification attack and the large amount of data is elicited as a response. Through amplification and reflection attacks and ICMP floods are covered in the major part of attacks under this category. In the below section, the popular reflection and amplification attacks are discussed:

Memcached Amplification:

The amplification factor of 51000 is reached by recent attacks which are one of the most evolved attacks. To make the data processing faster, Memcached [9] is used as a tool to cache the data. Since there is no mechanism of authentication, the main intention is to use the systems which are not connected to the internet. As per Akamai, over 50,000 known vulnerable systems are existed currently. A vulnerable server of UDP Memcached is included a request from the attacker spoofs. Additionally, a large response is flooded in a targeted victim and the resources of a victim are overwhelmed potentially. However, new requests can't be processed and the internet resource can't be accessed by regular traffic while overloading the internet infrastructure of a target that result in the denial of service.

DNS Amplification:

The vulnerability of DNS systems are made use in this attack where DNS systems are accessed publicly and support the open recursive relay. Using the victim's target queries and IP, the source IP is spoofed by an attacker. The information of zone is returned in a single request when querying for "ANY" resource record by DNS server and this is reflected to a target.

Based on EDNS (Extension Mechanisms for DNS) or DNSSEC (The Doman Name System Security Extention), the amplification is done by this attack. A response message with 4000 bytes to aim victim is converted from a request message of 60 bytes through these methods and ultimately the amplification factor of 1:70 is achieved. The response data is sent to the victim through the multiple DNS servers and thousands of Bots query if the attack is initiated by Botnets. The resources of target server will be depleted by accelerating the rate and increasing the volume of traffic. Based on a massive DDoS Attack, the Spamhaus project was targeted. The attackers are exploited the DNS Amplification [10] which was the primary approach.

NTP Amplification:

For synchronization of a clock, NTP was designed primarily between systems with internet connection. From NTP server, an attacker requests "get monlist" repeatedly in the most basic kind of NTP amplification attack with the IP address based on spoofing. Using the queried server which connected to a list of the recent 600 hosts, the NTP server is responded. The ratio of query-to-response is ranging between 20:1 to 200:1 or more is reached in an attack of NTP amplification. By using a tool like data or Metasploit from the open NTP project, these attacks are more ubiquitous as easier obtaining of list of open NTP servers. However, high volume and high bandwidth DDoS attack is generated easily by an attacker.

1.4 Challenges in traditional detection methods:

Today, most of the remediation activity of DDoS bots included a manual process. At the firewall or proxy, the required steps are considered by identifying the bots based on certain

domains or IP addresses. The traditional approaches to security that becomes less effective since the malicious bots sophistication level and other attacks are increased. The network bandwidth and web applications are targeted if in case of detecting the DDoS attacks [11]. The limitations are included in the traditional approaches like flow-based network parameter, frequency-based detection, poll-based monitoring, and deep packet inspection which depend on the attacks' signature.

The implementation of signature of an attack can't be done on its own. To be modelled each attack, human intervention is required. To improve the signature, the considerable time and effort will take. For stopping and catching a known attack, apply the signature. Different variants are created slightly by attackers to defeat the approaches of signature-based detection to bypass IDs. Here, this is the main reason for proliferating the variants of DDoS botnets [12].

1.5 Detection using ML:

To detect the attacks, a nonlinear way is provided by ML for appearing the marked things to be anomalies, finding beyond simple signatures, and detecting the similarities between this method and before happened things.

Based on threat intelligence about behaviours of DDoS bot externally, defense capabilities and detection are improved greatly by ML through collaborating with collected data about samples of traffic in order to study the patterns of new bot. However, the data is fed into the solution of ML. The various data points are consumed by the ML solution to run multiple models. In an approach of feedback loop, the human provides training for input. To get an understanding of new machines' of what kind of bot traffic, automated processes are launched by ML [13] solutions for blocking bot traffic. To detect unusual behaviour patterns, ML is needed to bring them to the attention of analysts. To block the traffic automatically, organizations can use ML solutions for repeated and common suspicious behaviours for alerting an analyst that helps to resolve the problem.

In case of quantifying, observing, and classifying inbound requests, ML and AI techniques are used exceptionally by including the degree of maliciousness.

1.6 Motivation for Work:

In the current market, security products are available already with the adoption of supervised ML. To detect the anomalous behaviours, providing of manual input to the ML engines is processed for most of the engines for examining thelog entries with massive numbers. For improving the identification of "significant events" in the logs, the manual input is augmented by the supervised ML system to bring out the events to the attention immediately.

Today, still, analyst interventions are demanded although practicing the devices of ML capable for analyzing the detection engine's response prior to the feeding of false positives with segregation into blocking engine. To identify DDoS attacks, the SGB algorithm is engineered in this paper for making the system with automation without including any misclassifications.

In the section II, the relevant works of DDoS detection methods and open dataset generation are discussed. The proposed methodology, dataset description, and developments are presented in the section III and IV. The conclusion is presented in the section V.

II. Related Work

In both academia and industry, machine learning is an active area of research when considering the domain of network security. Some of the research works will cover in this literature survey.

In [14], the details of one of the datasets CICIDS2017 [14] are explained that will be helpful to extract the DDoS traffic. By using seven ML classifiers, dataset is generated based on the evaluation. Accordingly, it will provide the best results based on Random Forests through the

execution time and precision. By using nonparametric CUSUM algorithm, a novel detection approach is proposed by authors in [15] for DoS attacks' application layer. By including various kinds of sampling attacks, the application layer of DoS attacks are detected in this paper. In [16], with the generation of different types of DDoS attacks of application layer, the dataset evaluation is created and mixed with benign traffic. In order to make the final dataset, the same dataset is utilized as one of the three datasets.

At the network of victim, employed the DDoS detection solutions generally. To identify DDoS attack at source end in the cloud environment, authors in [17] have proposed a model based on the statistical data from both the virtual machines and the cloud server's hypervisor. Nine ML algorithms are evaluated to restrict the packets of a network from being sent them to outside of the network. Finally, best prediction outcome is resulted using Random Forest model. At server or network end, employed the mechanism of strong defense against attacks of network security. But, dedicated security controls are not available on the internet at the devices of end-user. To identify the attacks from devices of end-user IOT with high accuracy, a model is proposed in [18] based on network behaviours with IoT-Specific. By including neural networks, variety of ML algorithms are developed.

The identification of DDoS attacks in an intelligent way is addressed in these studies. The shortcomings are reported by using different methods. Through the evaluation of DDoS datasets, no models are verified that the output in terms of accuracy based on SGB algorithm is not improved when compared to the Random Forest model. By comparing with the dataset, evaluated all models with datasets which including samples with fewer numbers. With the use of a DDoS flows' dataset based on over flows with 10 million bidirectional and meticulous tuning of hyper parameters, the performance is improved to reach 100 percent.

III. Proposed Method

For learning the supervised classification, Artificial Neural Network (ANN) is utilized which is nothing but a computational model. A number of neurons with the features of simple and highly interconnected have been included in ANN. As shown in figure 1, ANN based IDS have involved the step.

A. Dataset Selection

From the dataset CICIDS2017 taken from https://www.kaggle.com/cicdataset/cicids2017. The true real-world data (PCAPs) is resembled by the most common attacks which are up-to-date and the dataset of CICIDS2017 which contains benign. By using source and destination IPs, protocols and attacks (CSV files), and source and destination ports, and time stamp, the network traffic analysis results are also included based on the CICFlowMeter with labelled flows. The definition of extracted features is also available.

B. Feature Reduction

One class attribute and 78 attribute have included in the data set of CICIDS2017. Out of those 41 attribute, some have minimum role to detect the attack and some have no role. Almost all zero values of dataset have included in the features which is shown by the observation of dataset of CICIDS2017. From training and testing dataset, entire least usable features are removed, dataset's size is reduced and this is passed for testing and training. If in case of performing of testing and training with 78 features, the reduction of feature is not carry out on the dataset.

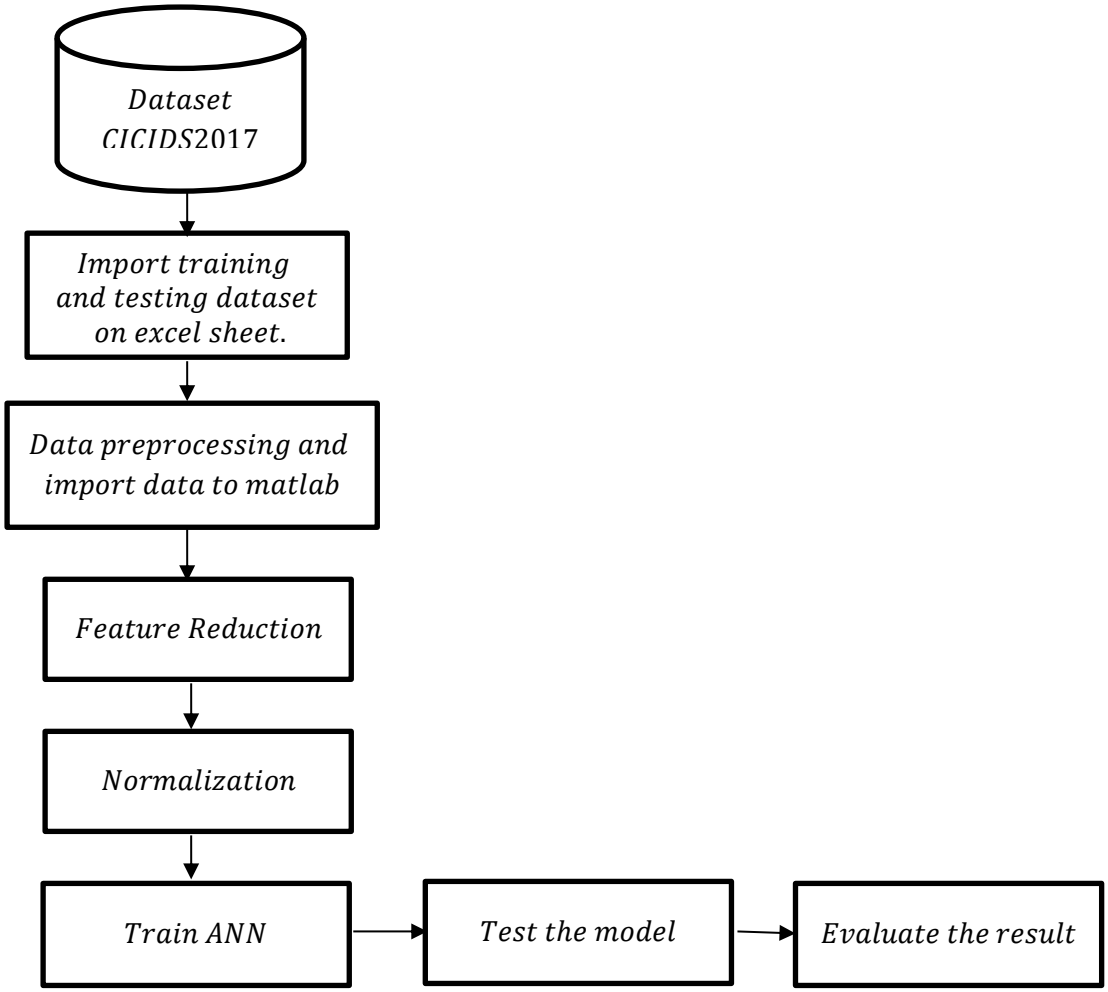

Figure 1: Flow chart of proposed IDS using ANN

C. Normalization

For normalization of attribute values, z-score normalization is utilized. After completion of normalization, the attribute value is normalized such that the standard deviation and mean have become one and zero respectively. This kind of normalization of z-score is also termed as zero mean normalization. The equation of normalization is mentioned below mathematically:

$$a(i) = \frac{a(i) - \ mean\ (A)}{std\ (A)}$$

Based on the above equation, the values are updated. Here, a(i) is the ith value of A and A is the attribute.

D. Training Neural Network

15000 records have included in the train dataset of CICIDS2017 and unequal distribution of patterns. The selection of 15000 patterns is done for maintaining the equality and for speedup training.

E. Testing Neural Network

Some unknown attacks have included in the dataset of test set of CICIDS2017 that is not available in the training set. For accurate classification of those attacks, it is main task to accomplish. With or without the reduction of feature, the testing of neural network is done against full test dataset with 8000 records.

F. Result Evaluation

Based on various parameters, neural network's performance is evaluated. However, the standard parameters are involved false positive rate, classification accuracy, and detection rate. By using False Negative (FN), True Positive (TP), True Negative (TN), and False Positive (FP), the given parameter is estimated. As shown in the table III, confusion matrix is utilized for evaluation of these parameters.

Table I Confusion Matrix

Attack		**Predicted Class**	
		Yes	No
Actual class	Yes	TP	FN
	No	FP	TN

Detection rate and high accuracy should have in the good IDS but low value should contain in the false positive rate. The misclassification rate is directly proportional to the false alarm rate.

Detection Rate (DR) = $\frac{TP}{TP+FN}$

False Positive Rate (FPR) = $\frac{FP}{FP+TN}$

Accuracy (ACC) = $\frac{TP+TN}{TP+TN+FP+FN}$

IV. Experimental Results

The current section shows the experimental results of the proposed system. Based on Matlab tool using Intel ® Core i3 CPU @ 2.20GHz processor with 8GB RAM, the experiment is performed. For training purpose of selected patterns with the number 15000 and testing the dataset of CICIDS2017 with 8000 patterns, neural network is utilized with the proposed algorithm and different hidden layer. By using different values of neural network architecture, training and testing results are processed based on 78 selected features CICIDS2017 dataset. Table II shows the training and testing samples taken fron the dataset for the experiments. In table III and IV, the results are shown.

Table II. Training and testing samples

	DDOS	BENIGN
Training	7500	7500
Testing	4000	4000

Table III. Training Confusion Matrix

	DDOS	BENIGN	Accuracy (%)	Error (%)
DDOS	7225	275	96.33	3.67
BENIGN	421	7079	94.38	5.62
			Over all accuracy	95.36

Table IV. Testing Confusion Matrix

	DDOS	BENIGN	Accuracy (%)	Error (%)
DDOS	3644	356	91.1	8.9
BENIGN	575	3425	85.62	14.38
			Over all accuracy	88.36

The confusion matrices are shown in tables III and IV. The training accuracy is 95.36 and the testing accuracy is 88.36.

Conclusion

For detection of DDoS attacks based on ANN classifier is proposed in this paper. The trainingis performed on CICIDS2017 dataset and compared the results have been demonstrated. The training and testing accuracies are 95.36 and 88.36 percentage respectively. To identify the DDoS attacks which are generated from customer-end devices such as home routers which are not in the protection range, this paper or research work is the prototype of ongoing work in the organization. From collected net flow data, the extraction of real-time flows is done at the internet gateways. According to the threat intelligence from the data points of security controls, the labelling of processed net flow data is accomplished. Here, the deploying of security controls is performed at different levels which include web application firewalls, firewalls, intrusion detection systems, intrusion protection systems, sandboxes, and solutions of privileged access management.

References:

[1].M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the mirai botnet," in Proc. of USENIX Security Symposium, 2017.

[2] A. Lohachab and B. Karambir, "Critical Analysis of DDoS—An Emerging Security Threat over IoT Networks," in Journal of Communications and Information Networks, vol. 3, no. 3, pp. 57-78, Sept. 2018. doi: 10.1007/s41650-018-0022-5

[3] W. Zhijun, L. Wenjing, L. Liang and Y. Meng, "Low-Rate DoS Attacks, Detection, Defense, and Challenges: A Survey," in IEEE Access, vol. 8, pp. 43920-43943, 2020. doi: 10.1109/ACCESS.2020.2976609

[4] Sperotto, Anna & Schaffrath, Gregor & Sadre, Ramin & Morariu, Cristian & Pras, Aiko & Stiller, Burkhard. (2010). An Overview of IP Flow-Based Intrusion Detection. IEEE Communications Surveys and Tutorials. 12. 343-356. 10.1109/SURV.2010.032210.00054.

[5] Indraneel Sreeram, Venkata Praveen Kumar Vuppala, “HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm”, Applied Computing and Informatics, Volume 15, Issue 1, January 2019, Pages 59-66.

[6] Kumar, Sanjeev. (2007). Smurf-based Distributed Denial of Service (DDoS) Attack Amplification in Internet. 25 - 25. 10.1109/ICIMP.2007.42.

[7] R. Oncioiu and E. Simion, "Approach to Prevent SYN Flood DoS Attacks in Cloud," 2018 International Conference on Communications (COMM), Bucharest, 2018, pp. 447-452. doi: 10.1109/ICComm.2018.8484802

[8] Guo, Fanglu & Chen, Jiawu & Chiueh, tzi-cker. (2006). Spoof Detection for Preventing DoS Attacks against DNS Servers. Proceedings - International Conference on Distributed Computing Systems. 37. 10.1109/ICDCS.2006.78.

[9] Petrovic, Jure. (2008). Using Memcached for Data Distribution in Industrial Environment. 368 - 372. 10.1109/ICONS.2008.51.
[10] Alieyan, Kamal & Kadhum, Mohammad & Anbar, Mohammed & Rehman, Shafiq & Alajmi, Naser. (2016). An overview of DDoS attacks based on DNS. 10.1109/ICTC.2016.7763485.
[11] S. Dong, K. Abbas and R. Jain, "A Survey on Distributed Denial of Service (DDoS) Attacks in SDN and Cloud Computing Environments," in IEEE Access, vol. 7, pp. 80813-80828, 2019. doi: 10.1109/ACCESS.2019.2922196
[12] Alomari, Esraa & Manickam, Selvakumar & Gupta, B B & Anbar, Mohammed & Alnakhalny, Redhwan & Alsaleem, Samer. (2016). A Survey of Botnet-Based DDoS Flooding Attacks of Application Layer: Detection and Mitigation Approaches. 10.4018/978-1-5225-0105-3.ch003.
[13] Livadas, Carl & Walsh, Robert & Lapsley, David & Strayer, Tim. (2006). Usilng Machine Learning Technliques to Identify Botnet Traffic. 2nd IEEE LCN Workshop on Network Security. 967 - 974. 10.1109/LCN.2006.322210.
[14] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018
[15].Hossein Hadian Jazi, Hugo Gonzalez, Natalia Stakhanova, and Ali A. Ghorbani. "Detecting HTTP-based Application Layer DoS attacks on Web Servers in the presence of sampling." Computer Networks, 2017
[16]. A. Shiravi, H. Shiravi, M. Tavallaee, A.A. Ghorbani, Toward developing a systematic approach to generate benchmark datasets for intrusion detection, Comput.
Security 31 (3) (2012) 357–374.
[17].Z. He, T. Zhang, and R. B. Lee, “Machine Learning Based DDoS Attack Detection from Source Side in Cloud,” in Proceedings of the 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud), pp. 114–120, New York, NY, USA, June 2017
[18].R. Doshi, N. Apthorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, 2018, pp. 29-35.