CIC
About the CIC
Membership
Research
Datasets
Webinars
Global EPIC Program
Cybersecurity Workshop

DDoS Evaluation Dataset (CIC-DDoS2019)

Distributed Denial of Service (DDoS) attack is a menace to network security that aims at exhausting the target networks with malicious traffic. Although many statistical methods have been designed for DDoS attack detection, designing a real-time detector with low computational overhead is still one of the main concerns. On the other hand, the evaluation of new detection algorithms and techniques heavily relies on the existence of well-designed datasets.

In this paper, we first review the existing datasets comprehensively and propose a new taxonomy for DDoS attacks. Secondly, we generate a new dataset, namely CICDDoS2019, which remedies all current shortcomings. Thirdly, using the generated dataset, we propose a new detection and family classification approach based on a set of network flow features. Finally, we provide the most important feature sets to detect different types of DDoS attacks with their corresponding weights.

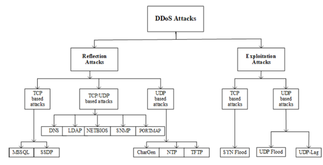
1. Introduction

There are a number of survey studies that have proposed taxonomies with respect to DDoS attacks. Although all have done a commendable job in proposing new taxonomies, the scope of attacks has so far been limited. There is a need to identify new attacks and come up with new taxonomies. Hence, we have analyzed new attacks that can be carried out using TCP/UDP based protocols at the application layer and proposed a new taxonomy. The rest of this sub-section has been explained the detailed taxonomy of DDoS attacks and illustrated in Figure 1, in terms of reflection-based and exploitation-based attacks.

Reflection-based DDoS: Are those kinds of attacks in which the identity of the attacker remains hidden by utilizing legitimate third-party component. The packets are sent to reflector servers by attackers with source IP address set to target victim’s IP address to overwhelm the victim with response packets. These attacks can be carried out through application layer protocols using transport layer protocols, i.e., Transmission control protocol (TCP), User datagram protocol (UDP) or through a combination of both. As Figure 1 shows, in this category, TCP based attacks include MSSQL, SSDP while as UDP based attacks include CharGen, NTP and TFTP. There are certain attacks that can be carried out using either TCP or UDP like DNS, LDAP, NETBIOS and SNMP.

Exploitation-based attacks: Are those kinds of attacks in which the identity of the attacker remains hidden by utilizing legitimate third-party component. The packets are sent to reflector servers by attackers with the source IP address set to the target victim’s IP address to overwhelm the victim with response packets. These attacks can also be carried out through application layer protocols using transport layer protocols i.e., TCP and UDP. TCP based exploitation attacks include SYN flood and UDP based attacks include UDP flood and UDP- Lag. UDP flood attack is initiated on the remote host by sending a large number of UDP packets.

These UDP packets are sent to random ports on the target machine at a very high rate. As a result, the available bandwidth of the network gets exhausted, system crashes and performance degrade. On the other hand, the SYN flood also consumes server resources by exploiting TCP-three-way handshake. This attack is initiated by sending repeated SYN packets to the target machine until server crashes/ malfunctions. The UDP-Lag attack is that kind of attack that disrupts the connection between the client and the server. This attack is mostly used in online gaming where the players want to slow down/ interrupt the movement of other players to outmaneuver them. This attack can be carried in two ways, i.e., using a hardware switch known as a lag switch or by a software program that runs on the network and hogs the bandwidth of other users.



2. Dataset

CICDDoS2019 contains benign and the most up-to-date common DDoS attacks, which resembles the true real-world data (PCAPs). It also includes the results of the network traffic analysis using CICFlowMeter-V3 with labeled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack (CSV files).

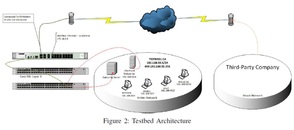


Figure 2: Testbed Architecture

Generating realistic background traffic was our top priority in building this dataset. We have used our proposed B-Profile system (Sharafaldin, et al. 2016) to profile the abstract behaviour of human interactions and generates naturalistic benign background traffic in the proposed testbed (Figure 2). For this dataset, we built the abstract behaviour of 25 users based on the HTTP, HTTPS, FTP, SSH and email protocols.

Machine	OS	IPs
Server	Ubuntu 16.04 (Web Server)	192.168.50.1 (first day) 192.168.50.4 (second day)
Firewall	Fortinet	205.174.165.81
PCs (first day)	Win 7	192.168.50.8
	Win Vista	192.168.50.5
	Win 8.1	192.168.50.6
	Win 10	192.168.50.7
PCs (second day)	Win 7	192.168.50.9
	Win Vista	192.168.50.6
	Win 8.1	192.168.50.7
	Win 10	192.168.50.8

In this dataset, we have different modern reflective DDoS attacks such as PortMap, NetBIOS, LDAP, MSSQL, UDP, UDP-Lag, SYN, NTP, DNS and SNMP. Attacks were subsequently executed during this period. As Table III shows, we executed 12 DDoS attacks includes NTP, DNS, LDAP, MSSQL, NetBIOS, SNMP, SSDP, UDP, UDP-Lag, WebDDoS, SYN and TFTP on the training day and 7 attacks including PortScan, NetBIOS, LDAP, MSSQL, UDP, UDP-Lag and SYN in the testing day. The traffic volume for WebDDoS was so low and PortScan just has been executed in the testing day and will be unknown for evaluating the proposed model.

Days	Attacks	Attack Time
First Day	PortMap	9:43 - 9:51
	NetBIOS	10:00 - 10:09
	LDAP	10:21 - 10:30
	MSSQL	10:33 - 10:42
	UDP	10:53 - 11:03
	UDP-Lag	11:14 - 11:24
	SYN	11:28 - 17:35
Second Day	NTP	10:35 - 10:45
	DNS	10:52 - 11:05
	LDAP	11:22 - 11:32
	MSSQL	11:36 - 11:45
	NetBIOS	11:50 - 12:00
	SNMP	12:12 - 12:23
	SSDP	12:27 - 12:37
	UDP	12:45 - 13:09
	UDP-Lag	13:11 - 13:15
	WebDDoS	13:18 - 13:29
	SYN	13:29 - 13:34
	TFTP	13:35 - 17:15

3. Using the dataset

The dataset has been organized per day. For each day, we recorded the raw data including the network traffic (Pcaps) and event logs (windows and Ubuntu event Logs) per machine. In features extraction process from the raw data, we used the CICFlowMeter-V3 and extracted more than 80 traffic features and saved them as a CSV file per machine.

If you want to use the AI techniques to analyze, you can download our generated data (CSV) files and analyze the network traffic.

If you want to use a new feature extractor, you can use the raw captured files (PCAP) to extract your features. And then, you can use the data mining techniques for analyzing the generated data.

4. License

You may redistribute, republish and mirror the CICDDoS2019 dataset in any form. However, any use or redistribution of the data must include a citation to the CICDDoS2019 dataset and related published paper. A research paper outlining the details of analyzing the similar IDS/ IPS dataset and related principles:

- Iman Sharafaldin, Arash Habibi Lashkari, Saqib Hakak, and Ali A. Ghorbani, **"Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy"**, IEEE 53rd International Carnahan Conference on Security Technology, Chennai, India, 2019.

Download this dataset