Machine Learning for Cyber-Security & Artificial Intelligence

CIC DDoS 2019 Dataset

Hermes Senger

Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization

Iman Sharafaldin, Arash Habibi Lashkari and Ali A. Ghorbani
Canadian Institute for Cybersecurity (CIC), University of New Brunswick (UNB), Canada

Keywords: Intrusion Detection, IDS Dataset, DoS, Web Attack, Infiltration, Brute Force.

Abstract: With exponential growth in the size of computer networks and developed applications, the significant increasing of the potential damage that can be caused by launching attacks is becoming obvious. Meanwhile, Intrusion Detection Systems (IDSs) and Intrusion Prevention Systems (IPSs) are one of the most important defense tools against the sophisticated and ever-growing network attacks. Due to the lack of adequate dataset, anomaly-based approaches in intrusion detection systems are suffering from accurate deployment, analysis and evaluation. There exist a number of such datasets such as DARPA98, KDD99, ISC2012, and ADFA13 that have been used by the researchers to evaluate the performance of their proposed intrusion detection and intrusion prevention approaches. Based on our study over eleven available datasets since 1998, many such datasets are out of date and unreliable to use. Some of these datasets suffer from lack of traffic diversity and volumes, some of them do not cover the variety of attacks, while others anonymized packet information and payload which cannot reflect the current trends, or they lack feature set and metadata. This paper produces a reliable dataset that contains benign and seven common attack network flows, which meets real world criteria and is publicly avaliable. Consequently, the paper evaluates the performance of a comprehensive set of network traffic features and machine learning algorithms to indicate the best set of features for detecting the certain attack categories.

1 INTRODUCTION

datasets a perfect dataset is yet to be realized (Ne-

CIC IDS 2017

- Contains benign and the most up-to-date common attacks, which resembles the true/real-world data (PCAPs).
- Generates (normal) realistic background traffic and having the abstract behaviour of 25 users based on the HTTP, HTTPS, FTP, SSH, and email protocols.
- Period:
  - From 9a.m. Monday, July3,2017 to 5p.m. on Friday July7,2017
  - Monday: normal day and only includes the benign traffic.
- Attacks:
  - Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS.
  - Attacks executed both morning and afternoon on Tuesday, Wednesday, Thursday andFriday.

Testbed

	Machine	OS	IPs
Victim-Network	Servers	Win Server 2016 (DC and DNS) Ubuntu 16 (Web Server) Ubuntu 12	192.168.10.3 192.168.10.50-205.174.165.68 192.168.10.51-205.174.165.66
	PCs	Ubuntu 14.4 (32, 64) Ubuntu 16.4 (32-64) Win 7Pro Win 8.1-64 Win Vista Win 10 (Pro 32-64) Mac	192.168.10.19-192168.10.17 192.168.10.16-192.168.10.12 192.168.10.9 192.168.10.5 192.168.10.8 192.168.10.14-192.168.10.15 192.168.10.25
	Firewall	Fortinet	
Attackers	PCs	Kali win 8.1 Win 8.1 Win 8.1	205.174.165.73 205.174.165.69 205.174.165.70 205.174.165.71

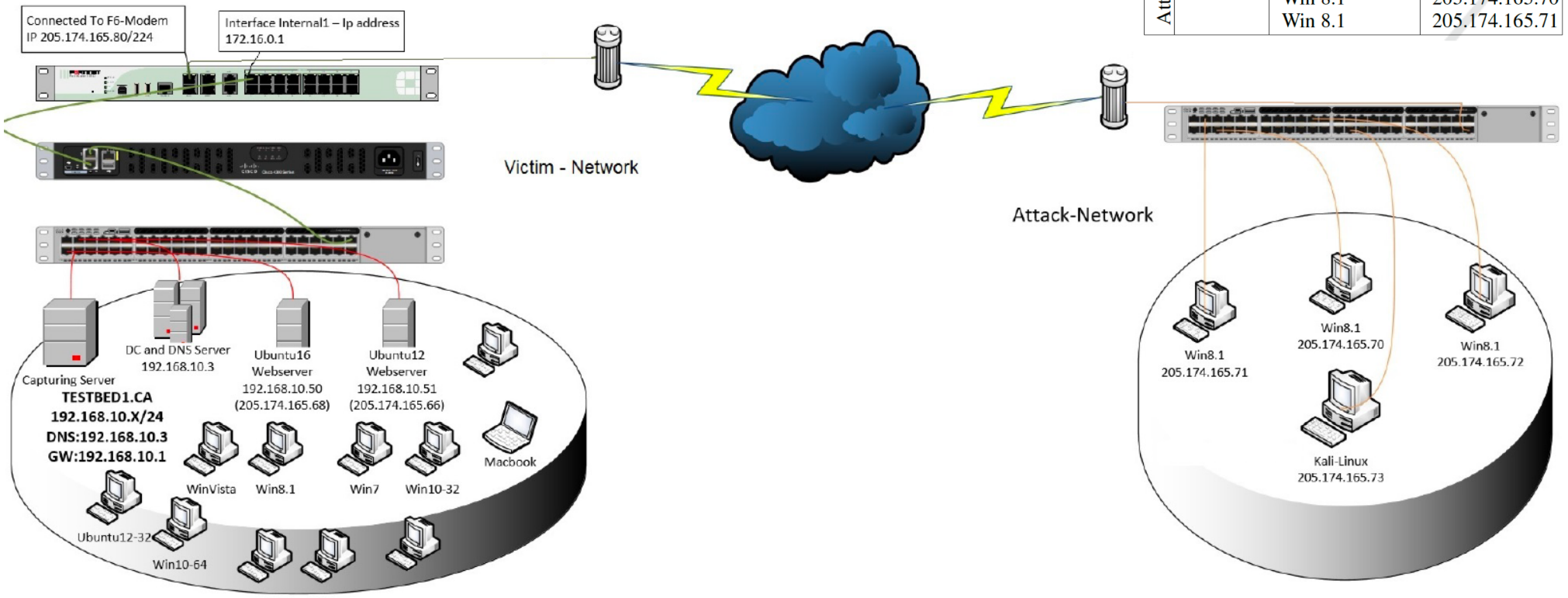

Figure 1: Testbed Architecture.

Attack Types

Brute Force Attack: This is one of the most popular attacks that only cannot be used for password cracking, but also to discover hidden pages and content in a web application. It is basically a hit and try attack, then the victim succeeds.

Heartbleed Attack: It comes from a bug in the OpenSSL cryptography library, which is a widely used implementation of the Transport Layer Security (TLS) protocol. It is normally exploited by sending a malformed heartbeat request with a small payload and large length field to the vulnerable party (usually a server) in order to elicit the victim's response.

Botnet: A number of Internet-connected devices used by a botnet owner to perform various tasks. It can be used to steal data, send spam, and allow the attacker access to the device and its connection.

DoS Attack: The attacker seeks to make a machine or network resource unavailable temporarily. It typically accomplished by flooding the targeted machine or resource with superfluous requests in an attempt to overload systems and prevent some or all legitimate requests from being fulfilled.

DDoS Attack: It typically occurs when multiple systems, flood the bandwidth or resources of a victim. Such an attack is often the result of multiple compromised systems (for example, a botnet) flooding the targeted system with generating the huge network traffic.

Web Attack: This attack types are coming out every day, because individuals and organizations take security seriously now. We use the SQL Injection, which an attacker can create a string of SQL commands, and then use it to force the database to reply the information, Cross-Site Scripting (XSS) which is happening when developers dont test their code properly to find the possibility of script injection, and Brute Force over HTTP which can tries a list of passwords to find the administrator's password.

Infiltration Attack: The infiltration of the network from inside is normally exploiting a vulnerable software such as Adobe Acrobat Reader. After successful exploitation, a backdoor will be executed on the victim's computer and can conduct different attacks on the victim's network such as IP sweep, full port scan and service enumerations using Nmap.

Attack Types

- Bruteforce attack; They used FTP Patator and SSH Patator tools to collect these type of attacks.
- DoS attack; They used Hulk, GoldenEye, Slowloris and Slowhttptest tools to collect these type of attacks.
- Web attack; They used Damn Vulnerable Web App (DVWA) and In-house selenium framework (XSS and Brute-force) tools to collect these type of attacks.
- Infiltration attack; They used Nmap and portscan tools to collect these type of attacks.
- Botnet attack; They used screenshots and keylogging
- DDoS attack; They used Low Orbit Ion Canon (LOIC) for UDP, TCP, or HTTP requests.
- Heartleech; It is a DoS attack.

- The CIC team recorded the raw data each day including the
- network traffic and event logs. In features extraction process
- from the raw data, they used the CICFlowMeter-V3 and
- extracted more than 80 network traffic features. Finally, they
- saved them as a CSV file per machine

Table 3: Feature Selection.

Label	Feature	Weight
Benign	B.Packet Len Min	0.0479
	Subflow F.Bytes	0.0007
	Total Len F.Packets	0.0004
	F.Packet Len Mean	0.0002
DoS GoldenEye	B.Packet Len Std	0.1585
	Flow IAT Min	0.0317
	Fwd IAT Min	0.0257
	Flow IAT Mean	0.0214
Heartbleed	B.Packet Len Std	0.2028
	Subflow F.Bytes	0.1367
	Flow Duration	0.0991
	Total Len F.Packets	0.0903
DoS Hulk	B.Packet Len Std	0.2028
	B.Packet Len Std	0.1277
	Flow Duration	0.0437
	Flow IAT Std	0.0227
DoS Slowhttp	Flow Duration	0.0443
	Active Min	0.0228
	Active Mean	0.0219
	Flow IAT Std	0.0200
DoS slowloris	Flow Duration	0.0431
	F.IAT Min	0.0378
	B.IAT Mean	0.0300
	F.IAT Mean	0.0265
SSH-Patator	Init Win F.Bytes	0.0079
	Subflow F.Bytes	0.0052
	Total Len F.Packets	0.0034
	ACK Flag Count	0.0007
FTP-Patator	Init Win F.Bytes	0.0077
	F.PSH Flags	0.0062
	SYN Flag Count	0.0061
	F.Packets/s	0.0014
Web Attack	Init Win F.Bytes	0.0200
	Subflow F.Bytes	0.0145
	Init Win B.Bytes	0.0129
	Total Len F.Packets	0.0096
Infiltration	Subflow F.Bytes	4.3012
	Total Len F.Packets	2.8427
	Flow Duration	0.0657
	Active Mean	0.0227
Bot	Subflow F.Bytes	0.0239
	Total Len F.Packets	0.0158
	F.Packet Len Mean	0.0025
	B.Packets/s	0.0021
PortScan	Init Win F.Bytes	0.0083
	B.Packets/s	0.0032
	PSH Flag Count	0.0009
DDoS	B.Packet Len Std	0.1728
	Avg Packet Size	0.0162
	Flow Duration	0.0137
	Flow IAT Std	0.0086

References

[1] Sharafaldin I, Habibi Lashkari A and Ghorbani A 2018 Toward generating a new intrusion detection dataset and intrusion traffic characterization *4th International Conference on Information Systems Security and Privacy (ICISSP)* pp 108–116