Intrusion detection evaluation dataset (CIC-IDS2017)

Intrusion Detection Systems (IDSs) and Intrusion Prevention Systems (IPSs) are the most important defense tools against the sophisticated and ever-growing network attacks. Due to the lack of reliable test and validation datasets, anomaly-based intrusion detection approaches are suffering from consistent and accurate performance evolutions.

Our evaluations of the existing eleven datasets since 1998 show that most are out of date and unreliable. Some of these datasets suffer from the lack of traffic diversity and volumes, some do not cover the variety of known attacks, while others anonymize packet payload data, which cannot reflect the current trends. Some are also lacking feature set and metadata.

CICIDS2017 dataset contains benign and the most up-to-date common attacks, which resembles the true real-world data (PCAPs). It also includes the results of the network traffic analysis using CICFlowMeter with labeled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack (CSV files).

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 behavior of human interactions and generates naturalistic benign background traffic. For this dataset, we built the abstract behaviour of 25 users based on the HTTP, HTTPS, FTP, SSH, and email protocols.

The data capturing period started at 9 a.m., Monday, July 3, 2017 and ended at 5 p.m. on Friday July 7, 2017, for a total of 5 days. Monday is the normal day and only includes the benign traffic. The implemented attacks include Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack,

Infiltration, Botnet and DDoS. They have been executed both morning and afternoon on Tuesday, Wednesday, Thursday and Friday.

In our recent dataset evaluation framework (Gharib et al., 2016), we have identified eleven criteria that are necessary for building a reliable benchmark dataset. None of the previous IDS datasets could cover all of the 11 criteria. In the following, we briefly outline these criteria:

Complete Network configuration: A complete network topology includes Modem, Firewall, Switches, Routers, and presence of a variety of operating systems such as Windows, Ubuntu and Mac OS X.

Complete Traffic: By having a user profiling agent and 12 different machines in Victim-Network and real attacks from the Attack-Network.

Labelled Dataset: Section 4 and Table 2 show the benign and attack labels for each day. Also, the details of the attack timing will be published on the dataset document.

Complete Interaction: As Figure 1 shows, we covered both within and between internal LAN by having two different networks and Internet communication as well.

Complete Capture: Because we used the mirror port, such as tapping system, all traffics have been captured and recorded on the storage server.

Available Protocols: Provided the presence of all common available protocols, such as HTTP, HTTPS, FTP, SSH and email protocols.

Attack Diversity: Included the most common attacks based on the 2016 McAfee report, such as Web based, Brute force, DoS, DDoS, Infiltration, Heart-bleed, Bot and Scan covered in this dataset.

Heterogeneity: Captured the network traffic from the main Switch and memory dump and system calls from all victim machines, during the attacks execution.

Feature Set: Extracted more than 80 network flow features from the generated network traffic using CICFlowMeter and delivered the network flow dataset as a CSV file. See our PCAP analyzer and CSV generator.

MetaData: Completely explained the dataset which includes the time, attacks, flows and labels in the published paper.

The full research paper outlining the details of the dataset and its underlying principles:

- 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), Purtogal, January 2018

Day, Date, Description, Size (GB)

- Monday, Normal Activity, 11.0G
- Tuesday, attacks + Normal Activity, 11G
- Wednesday, attacks + Normal Activity, 13G
- Thursday, attacks + Normal Activity, 7.8G
- Friday, attacks + Normal Activity, 8.3G

Victim and attacker networks information

Firewall: 205.174.165.80, 172.16.0.1

DNS+ DC Server: 192.168.10.3

Outsiders (Attackers network)

- Kali: 205.174.165.73
- Win: 205.174.165.69, 70, 71

Insiders (Victim network)

- Web server 16 Public: 192.168.10.50, 205.174.165.68
- Ubuntu server 12 Public: 192.168.10.51, 205.174.165.66
- Ubuntu 14.4, 32B: 192.168.10.19
- Ubuntu 14.4, 64B: 192.168.10.17
- Ubuntu 16.4, 32B: 192.168.10.16
- Ubuntu 16.4, 64B: 192.168.10.12
- Win 7 Pro, 64B: 192.168.10.9
- Win 8.1, 64B: 192.168.10.5
- Win Vista, 64B: 192.168.10.8
- Win 10, pro 32B: 192.168.10.14
- Win 10, 64B: 192.168.10.15
- MAC: 192.168.10.25

Monday, July 3, 2017

Benign (Normal human activities)

Tuesday, July 4, 2017

Brute Force

FTP-Patator (9:20 – 10:20 a.m.)

SSH-Patator (14:00 – 15:00 p.m.)

Attacker: Kali, 205.174.165.73

Victim: WebServer Ubuntu, 205.174.165.68 (Local IP: 192.168.10.50)

NAT Process on Firewall:

Attack: 205.174.165.73 -> 205.174.165.80 (Valid IP of the Firewall) -> 172.16.0.1 -> 192.168.10.50

Reply: 192.168.10.50 -> 172.16.0.1 -> 205.174.165.80 -> 205.174.165.73

Wednesday, July 5, 2017

DoS / DDoS

DoS slowloris (9:47 – 10:10 a.m.)

DoS Slowhttptest (10:14 – 10:35 a.m.)

DoS Hulk (10:43 – 11 a.m.)

DoS GoldenEye (11:10 – 11:23 a.m.)

Attacker: Kali, 205.174.165.73

Victim: WebServer Ubuntu, 205.174.165.68 (Local IP192.168.10.50)

NAT Process on Firewall:

Attack: 205.174.165.73 -> 205.174.165.80 (Valid IP of the Firewall) -> 172.16.0.1 -> 192.168.10.50

Reply: 192.168.10.50 -> 172.16.0.1 -> 205.174.165.80 -> 205.174.165.73

Heartbleed Port 444 (15:12 - 15:32)

Attacker: Kali, 205.174.165.73

Victim: Ubuntu12, 205.174.165.66 (Local IP192.168.10.51)

NAT Process on Firewall:

Attack: 205.174.165.73 -> 205.174.165.80 (Valid IP of the Firewall) -> 172.16.0.11 -> 192.168.10.51

Reply: 192.168.10.51 -> 172.16.0.1 -> 205.174.165.80 -> 205.174.165.73

Thursday, July 6, 2017

Morning

Web Attack – Brute Force (9:20 – 10 a.m.)

Web Attack – XSS (10:15 – 10:35 a.m.)

Web Attack – Sql Injection (10:40 – 10:42 a.m.)

Attacker: Kali, 205.174.165.73

Victim: WebServer Ubuntu, 205.174.165.68 (Local IP192.168.10.50)

NAT Process on Firewall:

Attack: 205.174.165.73 -> 205.174.165.80 (Valid IP of the Firewall) -> 172.16.0.1 -> 192.168.10.50

Reply: 192.168.10.50 -> 172.16.0.1 -> 205.174.165.80 -> 205.174.165.73

Afternoon

Infiltration – Dropbox download

Meta exploit Win Vista (14:19 and 14:20-14:21 p.m.) and (14:33 -14:35)

Attacker: Kali, 205.174.165.73

Victim: Windows Vista, 192.168.10.8

Infiltration – Cool disk – MAC (14:53 p.m. – 15:00 p.m.)

Attacker: Kali, 205.174.165.73

Victim: MAC, 192.168.10.25

Infiltration – Dropbox download

Win Vista (15:04 – 15:45 p.m.)

First Step:

Attacker: Kali, 205.174.165.73

Victim: Windows Vista, 192.168.10.8

Second Step (Portscan + Nmap):

Attacker:Vista, 192.168.10.8

Victim: All other clients

Friday, July 7, 2017

Morning

Botnet ARES (10:02 a.m. – 11:02 a.m.)

Attacker: Kali, 205.174.165.73

Victims: Win 10, 192.168.10.15 + Win 7, 192.168.10.9 + Win 10, 192.168.10.14 + Win 8, 192.168.10.5 + Vista, 192.168.10.8

Afternoon

Port Scan:

Firewall Rule on (13:55 – 13:57, 13:58 – 14:00, 14:01 – 14:04, 14:05 – 14:07, 14:08 - 14:10, 14:11 – 14:13, 14:14 – 14:16, 14:17 – 14:19, 14:20 – 14:21, 14:22 – 14:24, 14:33 – 14:33, 14:35 - 14:35)

Firewall rules off (sS 14:51-14:53, sT 14:54-14:56, sF 14:57-14:59, sX 15:00-15:02, sN 15:03-15:05, sP 15:06-15:07, sV 15:08-15:10, sU 15:11-15:12, sO 15:13-15:15, sA 15:16-15:18, sW 15:19-15:21, sR 15:22-15:24, sL 15:25-15:25, sI 15:26-15:27, b 15:28-15:29)

Attacker: Kali, 205.174.165.73

Victim: Ubuntu16, 205.174.165.68 (Local IP: 192.168.10.50)

NAT Process on Firewall:

Attacker: 205.174.165.73 -> 205.174.165.80 (Valid IP of the Firewall) -> 172.16.0.1

Afternoon

DDoS LOIT (15:56 – 16:16)

Attackers: Three Win 8.1, 205.174.165.69 - 71

Victim: Ubuntu16, 205.174.165.68 (Local IP: 192.168.10.50)

NAT Process on Firewall:

Attackers: 205.174.165.69, 70, 71 -> 205.174.165.80 (Valid IP of the Firewall) -> 172.16.0.1

Webinar example of dataset use: "Enhancing Generalizability in DDoS Attack Detection Systems through Transfer Learning and Ensemble Learning Approaches" by Dr. Mahdi Rabbani, Postdoctoral Fellow, Canadian Institute for Cybersecurity and Q&A with Dr. Windhya Rankothge.

License

The **CICIDS2017** dataset consists of labeled network flows, including full packet payloads in pcap format, the corresponding profiles and the labeled flows (GeneratedLabelledFlows.zip) and CSV files for machine and deep learning purpose (MachineLearningCSV.zip) are publicly available for researchers. If you are using our dataset, you should cite our related paper which outlining the details of the dataset and its underlying principles:

- 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.

If you are interest in CIC-IDS2017, you may also be interested in the BCCC-CIC-IDS2017 dataset made available by our colleagues at the Behaviour-Centric Cybersecurity Center, York University.