Cyber Attack Evaluation Dataset for Deep Packet Inspection and Analysis

Chirag Ganguli[a], Shishir Kumar Shandilya[a], Ivan Izonin[b], Prof. Atulya Kumar Nagar[c]

[a] *Vellore Institute of Technology, VIT Bhopal University, Bhopal, India*

[b] *Lviv Polytechnic National University, Ukraine*

[c] *Liverpool Hope University, United Kingdom*

1. Introduction

To determine the effectiveness of any defense mechanism, there is a need for comprehensive real-time network data that solely references various attack scenarios based on older software versions or unprotected ports, and so on. This presented dataset has entire network data at the time of several cyber attacks to enable experimentation on challenges based on implementing defense mechanisms on a larger scale. For collecting the data, we captured the network traffic of configured virtual machines using Wireshark and tcpdump. To analyze the impact of several cyber attack scenarios, this dataset presents a set of ten computers connected to Router1 on VLAN1 in a Docker Bridge network, that try and exploit each other. It includes browsing the web and downloading foreign packages including malicious ones. Also, services like FTP and SSH were exploited using several attack mechanisms. The presented dataset shows the importance of updating and patching systems to protect themselves to a greater extent, by following attack tactics on older versions of packages as compared to the newer and updated ones. This dataset also includes an Apache Server hosted on the different subset on VLAN2 which is connected to the VLAN1 to demonstrate isolation and cross-VLAN communication. The services on this web server were also exploited by the previously stated ten computers. The attack types include: Distributed Denial of Service, SQL Injection, Account Takeover, Service Exploitation (SSH, FTP), DNS and ARP Spoofing, Scanning and Firewall Searching and

Indexing (using Nmap), Hammering the services to brute-force passwords and usernames, Malware attack, Spoofing and Man-in-the-Middle Attack. The attack scenarios also show various scanning mechanisms and the impact of Insider Threats on the entire network.

2. Dataset Files

1. L1_Cap_10PC_1S.pcapng
2. L1_Cap_10PC_1S_dissec.xlsx
3. L1_Cap_10PC_1S_dissec.csv
4. L1_Cap_10PC_1S_dissec_complete.csv
5. L1_Cap_202209051617.csv
6. L1_Cap_202209051620.sql

3. Files Description

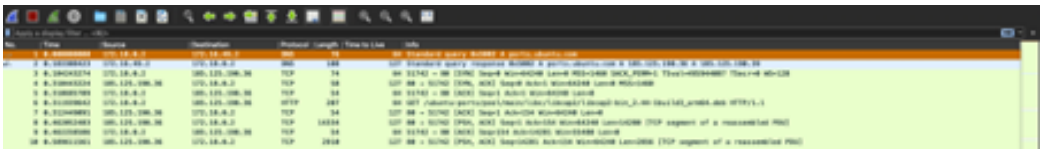
- L1_Cap_10PC_1S.pcapng : This is the Raw pcap file captured using Wireshark for a specific time period on 11 machines (10 Machines & 1 Apache Server). Total Packets: 3,962,784
- L1_Cap_10PC_1S_dissec.xlsx: This file is segregated first 1,048,576 rows of the Raw pcap file in modern Excel format.
- L1_Cap_10PC_1S_dissec.csv: This file is the segregated first 1,048,576 rows of the Raw pcap file for small compute analysis.
- L1_Cap_10PC_1S_dissec_complete.csv : This the complete exported Raw pcap file in .csv format. Total Packets: 3,962,784
- L1_Cap_202209051617.csv : This is the labeled dataset in which two columns added to the Raw Dataset – Source_Known and Destination_Known which flags known IP address in the Source and Destination fields to belong to the 10+1 machine range. Here, '1' represents 'True', and '0' represents 'False'.
- L1_Cap_202209051620.sql : This is the labeled dataset compiled in .sql format so that this can be easily imported into an SQL Database and analyzed based on the user requirement. This makes the analysis of huge datasets easier and more convenient.

4. Machine Information

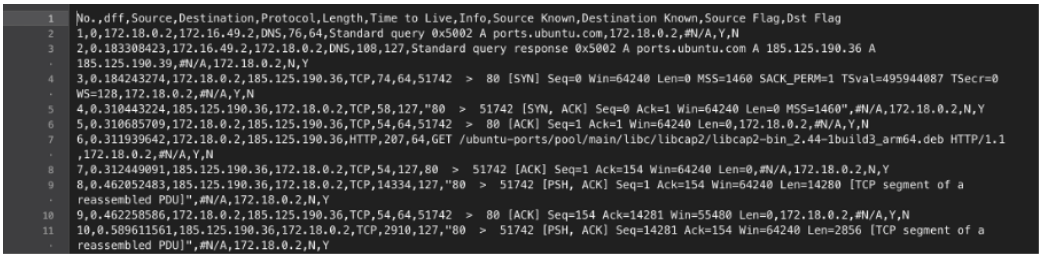
172.18.0.1	Router1
172.18.0.2	dev
172.18.0.3	devone
172.18.0.4	devtwo
172.18.0.5	devthree
172.18.0.6	devfour
172.18.0.7	devfive
172.18.0.8	devsix
172.18.0.9	devseven
172.18.0.10	deveight
172.18.0.11	devnine
172.17.0.1	Router2
172.17.0.2	server

5. How to open the Files

- File 1 (L1_Cap_10PC_1S.pcapng):
  - Install Wireshark from https://www.wireshark.org/#download
  - Double Click on the file to open it using Wireshark for packet level deep inspection and analysis

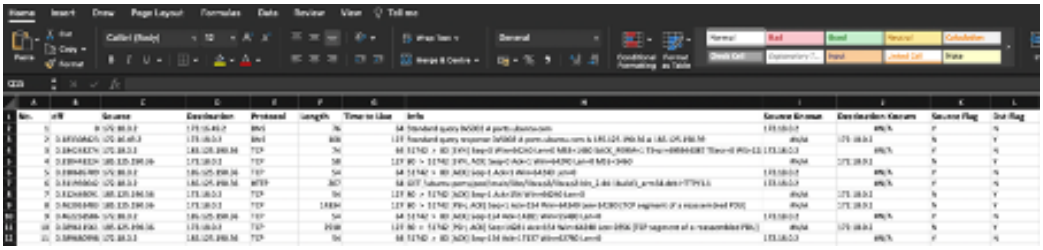


- File 2 and 3 (L1_Cap_10PC_1S_dissec.xlsx, L1_Cap_10PC_1S_dissec.csv)
  - Open any Text Editor to view these files – Notepad (Windows), TextEdit (Linux), TextMate (Mac)

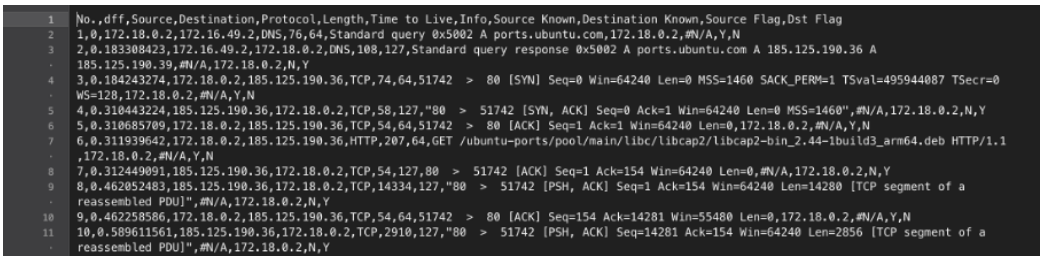


```
No.,dff,Source,Destination,Protocol,Length,Time to Live,Info,Source Known,Destination Known,Source Flag,Dst Flag
1,0,172.18.0.2,172.16.49.2,DNS,76,64,Standard query 0x5002 A ports.ubuntu.com,172.18.0.2,#N/A,Y,N
2,0.183308423,172.16.49.2,172.18.0.2,DNS,108,127,Standard query response 0x5002 A ports.ubuntu.com A 185.125.190.36 A
185.125.190.39,#N/A,172.18.0.2,N,Y
3,0.184243274,172.18.0.2,185.125.190.36,TCP,74,64,51742  >  80 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM=1 TSval=495944087 TSecr=0
WS=128,172.18.0.2,#N/A,Y,N
4,0.310443224,185.125.190.36,172.18.0.2,TCP,58,127,"80  >  51742 [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1460",#N/A,172.18.0.2,N,Y
5,0.310685709,172.18.0.2,185.125.190.36,TCP,54,64,51742  >  80 [ACK] Seq=1 Ack=1 Win=64240 Len=0,172.18.0.2,#N/A,Y,N
6,0.311939642,172.18.0.2,185.125.190.36,HTTP,207,64,GET /ubuntu-ports/pool/main/libc/libcap2/libcap2-bin_2.44-1build3_arm64.deb HTTP/1.1
,172.18.0.2,#N/A,Y,N
7,0.312449091,185.125.190.36,172.18.0.2,TCP,54,127,80  >  51742 [ACK] Seq=1 Ack=154 Win=64240 Len=0,#N/A,172.18.0.2,N,Y
8,0.462052483,185.125.190.36,172.18.0.2,TCP,14334,127,"80  >  51742 [PSH, ACK] Seq=1 Ack=154 Win=64240 Len=14280 [TCP segment of a
reassembled PDU]",#N/A,172.18.0.2,N,Y
9,0.462258586,172.18.0.2,185.125.190.36,TCP,54,64,51742  >  80 [ACK] Seq=154 Ack=14281 Win=55480 Len=0,172.18.0.2,#N/A,Y,N
10,0.589611561,185.125.190.36,172.18.0.2,TCP,2910,127,"80  >  51742 [PSH, ACK] Seq=14281 Ack=154 Win=64240 Len=2856 [TCP segment of a
reassembled PDU]",#N/A,172.18.0.2,N,Y
```

  - These files are limited to 1,048,576 rows which is the standard Excel row limit so these can also be opened and viewed using Excel.
    * Right Click on the File and Open With : Microsoft Excel



- File 4 and 5 (L1_Cap_10PC_1S_dissec_complete.csv, L1_Cap_202209051617.csv):
  - Open any Text Editor to view these file – Notepad (Windows), TextEdit (Linux), TextMate (Mac)



```
No.,dff,Source,Destination,Protocol,Length,Time to Live,Info,Source Known,Destination Known,Source Flag,Dst Flag
1,0,172.18.0.2,172.16.49.2,DNS,76,64,Standard query 0x5002 A ports.ubuntu.com,172.18.0.2,#N/A,Y,N
2,0.183308423,172.16.49.2,172.18.0.2,DNS,108,127,Standard query response 0x5002 A ports.ubuntu.com A 185.125.190.36 A
185.125.190.39,#N/A,172.18.0.2,N,Y
3,0.184243274,172.18.0.2,185.125.190.36,TCP,74,64,51742  >  80 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM=1 TSval=495944087 TSecr=0
WS=128,172.18.0.2,#N/A,Y,N
4,0.310443224,185.125.190.36,172.18.0.2,TCP,58,127,"80  >  51742 [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1460",#N/A,172.18.0.2,N,Y
5,0.310685709,172.18.0.2,185.125.190.36,TCP,54,64,51742  >  80 [ACK] Seq=1 Ack=1 Win=64240 Len=0,172.18.0.2,#N/A,Y,N
6,0.311939642,172.18.0.2,185.125.190.36,HTTP,207,64,GET /ubuntu-ports/pool/main/libc/libcap2/libcap2-bin_2.44-1build3_arm64.deb HTTP/1.1
,172.18.0.2,#N/A,Y,N
7,0.312449091,185.125.190.36,172.18.0.2,TCP,54,127,80  >  51742 [ACK] Seq=1 Ack=154 Win=64240 Len=0,#N/A,172.18.0.2,N,Y
8,0.462052483,185.125.190.36,172.18.0.2,TCP,14334,127,"80  >  51742 [PSH, ACK] Seq=1 Ack=154 Win=64240 Len=14280 [TCP segment of a
reassembled PDU]",#N/A,172.18.0.2,N,Y
9,0.462258586,172.18.0.2,185.125.190.36,TCP,54,64,51742  >  80 [ACK] Seq=154 Ack=14281 Win=55480 Len=0,172.18.0.2,#N/A,Y,N
10,0.589611561,185.125.190.36,172.18.0.2,TCP,2910,127,"80  >  51742 [PSH, ACK] Seq=14281 Ack=154 Win=64240 Len=2856 [TCP segment of a
reassembled PDU]",#N/A,172.18.0.2,N,Y
```

- File 6 (L1_Cap_202209051620.sql):
  - To import the MySQL file install MySQL from https://www.mysql.com/downloads/

– Open a terminal prompt and type ‘ mysql -u root -p ‘. If prompted for a password, enter the password used during installation of MySQL or keep blank if None was set. Then execute the below commands in MySQL prompt

  * mysql> create database pcdataset;

```
mysql> create database pcdataset;
Query OK, 1 row affected (0.00 sec)
```

  * mysql> use pcdataset;
  * mysql> CREATE TABLE L1_Cap(‘No.‘ int, ‘Time‘ int, Source varchar(50), Destination varchar(50), Protocol varchar(50), ‘Length‘ int, ‘Time to Live‘ int, Info varchar(1024), Source_Known int, Destination_Known int);

```
mysql> CREATE TABLE L1_Cap(`No.` int, `Time` int, Source varchar(50), Destination varchar(50), Protocol varchar(50), `Length` i
nt, `Time to Live` int, Info varchar(1024), Source_Known int, Destination_Known int);
Query OK, 0 rows affected (0.01 sec)

mysql> show tables;
+---------------------+
| Tables_in_pcdataset |
+---------------------+
| L1_Cap              |
+---------------------+
1 row in set (0.00 sec)
```

  * mysql> exit

– Now, Open a terminal prompt and type ‘ mysql -u root -p pcdataset < file-path/ L1_Cap_202209051620.sql ‘. The dataset will be imported in the SQL Database “pcdataset”.

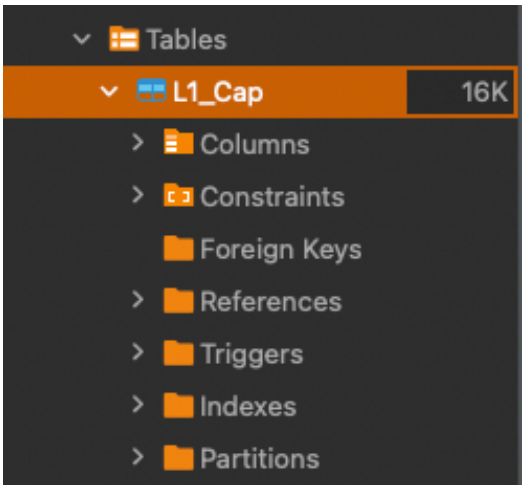
– To view the dataset, use:

  * prompt> mysql -u root -p
  * mysql> use pcdataset;
  * mysql> show tables;
  * mysql> select * from L1_Cap;

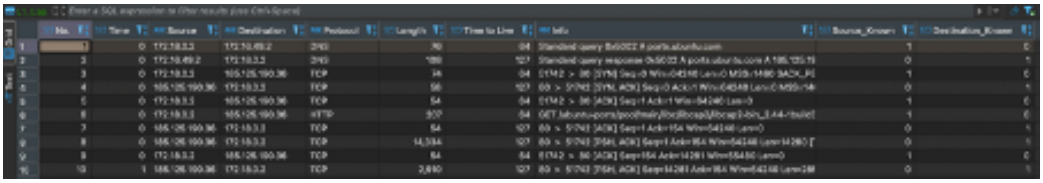
– Several other queries can now to executed to perform analysis on the data

– Alternatively, this data can be viewed using GUI Database Managers like DBeaver (https://dbeaver.io/download/)

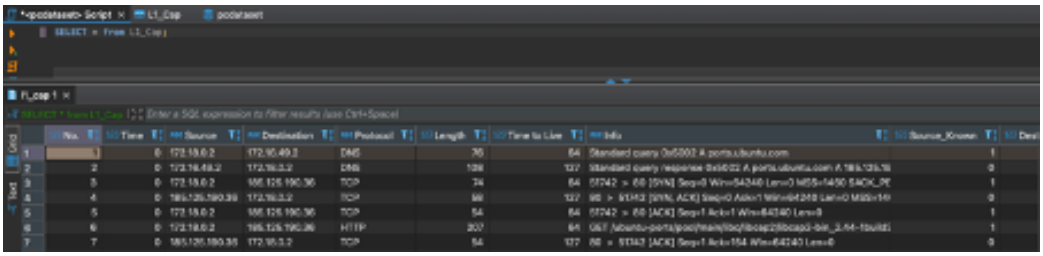
  * Open the DBeaver app and connect to local host to view the tables and databases



  * Click on Tables and Double on the Table Name to view the table



  * Also, SQL queries can be executed using the build-in SQL Query Editor



6. Deployment

To generate the dataset in its complete form in the local system, navigate to the 'L1_Cap_202209051620.sql' and import it into MySQL on the local system. Please note that the table with the defined properties needs to be created before the import process. Inside the database, L1_Cap file can be found which consists of the PCAP file database for the analysis of the packets that were captured to display both attack and defense mechanisms. The Info tab includes the Information about each packet which can be filtered from the generated database in MySQL and analyzed. For the packets to be analyzed graphically, the 'L1_Cap_10PC_1S.pcapng' can be directly imported into Wireshark. This can properly segregate the packets and can provide a graphical representation for analysis.