Examine and Visualising Packet Capture Files

S. Leelalakshmi, K. Tamilselvi, K. Rameshkumar

Abstract*: The field of security Visualisation is an interesting and tough field of research. Enormous amount of (big) data is involved in the networking of devices. In order to analyse and get data for solving the problem, visualisation can be very helpful. Combination of security world as well as the network world is discussed in this paper. Identifying various visualisation techniques for security log data and executing workflow based composition of multiple analytic components will be identified. Interactive modes of the techniques will be discussed. Making the security files to be readable and the format for analysing are identified. More network visualisation tool allows the security analysts to quickly examine the large amount of information by rendering a millions of events and log entries in a single graphical view. Extracting files from full packet captures can save security analyst a great deal of time. There are tools available for capturing PCAP(Packet Capture) files. This PCAP files will be analysed for further details. In the proposed solution, the PCAP files will be generated with the help of Wireshark and it will be processed with the help of Apache drill for converting it into a readable format and the Visualisation can be done with R Studio. Various Visualisation tools in R will be used to visualise the PCAP files. This in order will thoroughly give some insight on the log files for any detection and prediction of malicious data.

Keywords: PCAP, Security Visualisation, Analytic components, Visualisation tools

I. INTRODUCTION

The field of Security Visualisation is an interesting and tough field for research. Huge amount of data is involved in the networking of the computers. In order to analyse and get data for solving the problem Visualisation can be very helpful.

This paper focuses on why there is a need for the combination of security world as well as the Visualisation world. Both these fields have enough research on their own fields. This paper gives us an overview of security visualisation.[1]

In analysis of Cyber security the following fields constitutes: Mathematics, Statistics, Information Science and high performance computing. For cyber security the expertise required in the subject matter is cyber security and intelligence. Data collected is extensive on cyber security. And also expertise is required in Modelling , Simulation, statistical methodology. The following analyses from the cyber data helps in exploring the data, analysis, prediction and uncertainity quantification

Diverse interdisciplinary capabilities are needed to address the challenges of Cyber Security including mathematics, statistics, information science, computer science, and high performance computing, as well as subject matter expertise in cyber security, homeland security, and intelligence. Expertise is needed in several areas including, but not limited to the following: modeling and simulation; statistical methodology for exploration, analysis, prediction, and uncertainty quantification; network analysis and graph theory, machine learning and anomaly detection, streaming data, data intensive computing, and visualisation. Cyber Security provides exciting opportunities to pursue innovative research motivated need to ensure the security and privacy of networks, systems, and data.

Applications
Network Services
Operating System
Traffic Flows
Packet Captures
Proxies
Intrusion Detection Systems
Firewalls
Passive Network

This is probably the network stack.

Most security visualisation tools are written by security people who do not know much about visualisation theory and human-computer interaction; the rest are written by visualisation people who do not know much about computer security and adjacent technical fields, such as operating systems or networking. Therefore, tools lack one of two important aspects: either the security domain knowledge and accuracy or the visual efficiency. Complete security visualisation expertise requires knowledge of two worlds: the security world and the visualisation world .The security world consists of bits and bytes, of exploits and security policies ,of risk and compliance mandates. It is absolutely necessary to know these concepts to build a tool that is easy to use and effective for security .Information visualisation is the use of interactive, sensory representations, typically visual, of abstract data to reinforce cognition.

The knowledge of the visualisation world encompasses visual perception and human-interface design. These two aspects are necessary to build a usable tool. There are many tools but whether they are useful for security data is a question. There can be nice program also but that will not be very helpful because it was developed for one specific use-case that has nothing to do with real-world applications and problems that security professionals are facing. There should not be a gap or a dichotomy between these two disciplines.

Revised Version Manuscript Received on 16 September, 2019.

S.Leelalakshmi, Research Scholar, Bharathiar University sleein@rediffmail.com
K.Tamilselvi, Research Scholar, Bharathiar University ktselviravi@hotmail.com
K.Rameshkumar, Research Guide, Bharathiar University rameshkumark.dr@gmail.com

Cyber Security poses a number of interdisciplinary research challenges.

i. Intrusion tolerance and resilience; containing, removing, and surviving intrusions

ii. Mining peta-scale network data to detect changes and anomalies and predict consequences

iii. Creating trust (confidentiality, integrity, availability, and privacy) in systems that contain untrusted components

The idea of information visualisation is overview first, zoom and filter and details on demand-Ben Schneider man[2]

The tools used can help to analyse anything related to visualisation components (size, color, sequence etc)

Information visualisation can help in identifying extremes, comparison, trends etc.[6]

Visualisation can be helpful in hot research areas like visualising vulnerabilities,

Visualising worms, forensic visualisation, feature selection and construction ,

Incremental, online learning forensic analysis.

There can be approaches where data can be looked at TCP/IP data and some interesting visualisations can be found out.

II. METHODOLOGY FOR ANALYSING THE PCAP FILES [4]

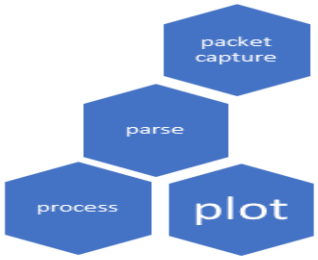


Diagram for the Process

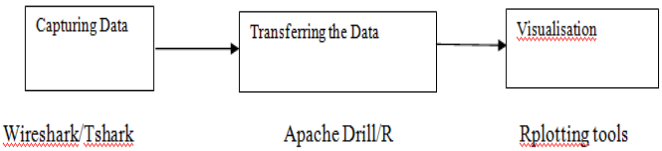
Packet Capture (can be done with tshark,wireshark etc)

Parsing can be done with R,RStudio

Process can be done with Apache drill as well as R

Plotting can be done with R

System Architecture



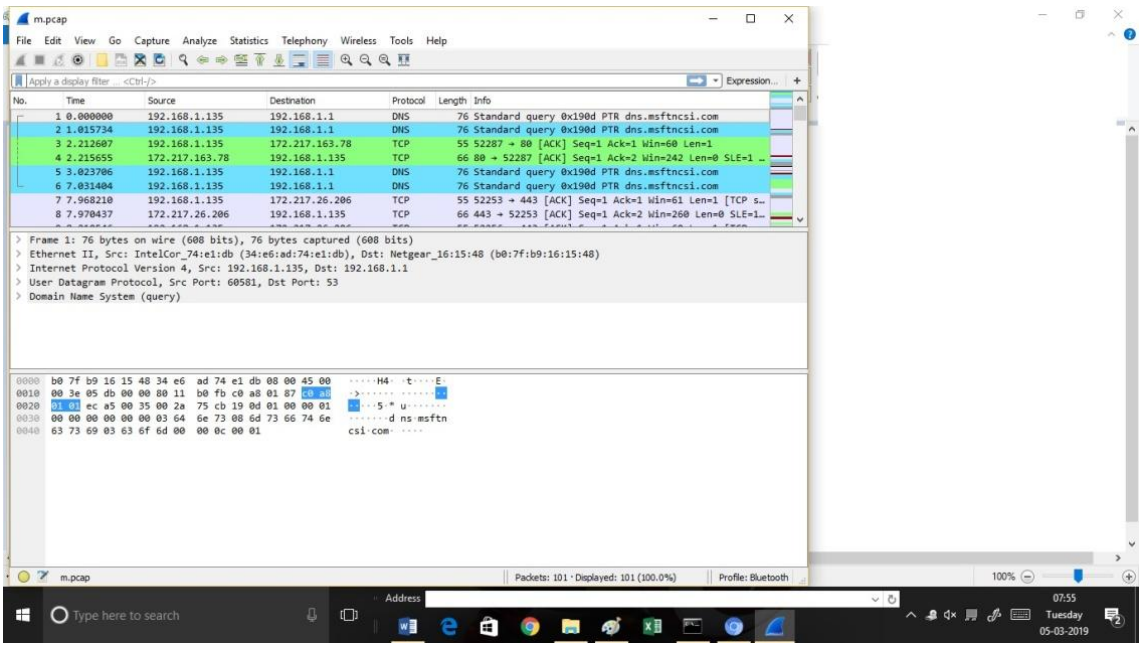
The following methodology can be used for capturing and analysing PCAP files

A. Log files has to be captured from Wireshark and Tshark

B. After that captured files are transferred into apache drill

C. From apache drill the following data can be used for visualisation

A. PCAP FILE has to be captured in wireshark



B. Before handling this file Apache Drill has to be connected

In the RStudio the following codes are required and some of the library
has to be installed[3]

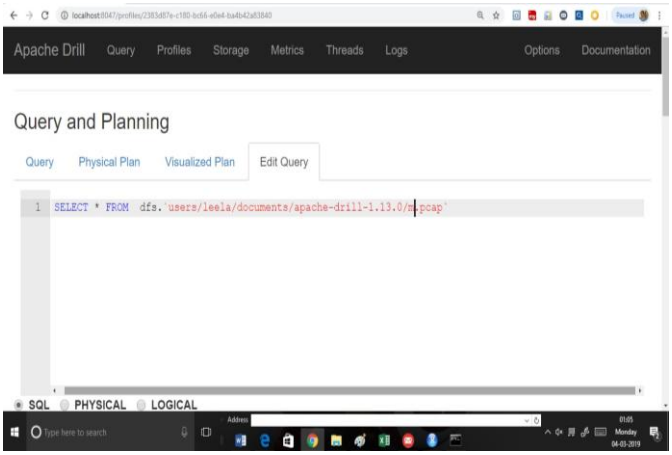
library(sergeant)
library(iptools)
library(tidyverse)
library(cymruservices)

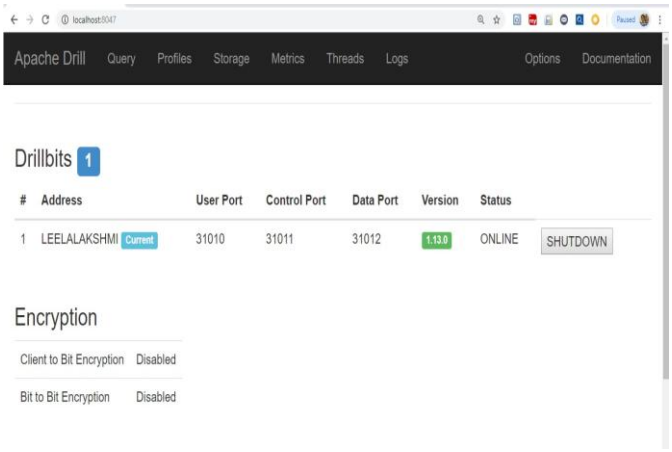
This particular codes gets connected Apache-drill with R-Studio [5]

```
db <- src_drill("localhost")

my_pcaps <- tbl(db,
"c:\users\leela\documents\apache-drill-1.13.0\m.pcap`")
```

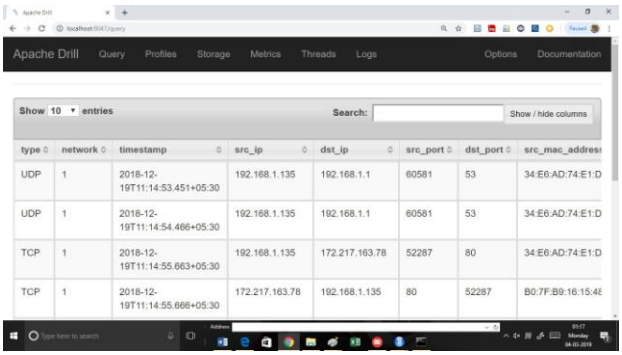
The following codes in R create corresponding Query result in apache
And it also stores the pcap fields in the variable my_pcaps





The following codes in R create corresponding Query result in apache

The resultant of the query will be like this.



Whatever been discussed is on the Apache drill side. Since Apache and R are connected it is very visible that querying on Apache side can be done easily. Querying in the apache drill provides more flexibility, fragmentation, Query timing can be found out very easily.

But the objective of the paper is to analyse the PCAP files with the help of R and Apache Drill.

After PCAP files are converted into a much readable format if individual field has to be identified it can be done. Simultaneous plots can be drawn with the field.
The following commands in R can be used

```
    > count(my_pcaps,data)
# Source:   lazy query [?? x 2]
# Database: DrillConnection
   data                                    n
   <chr>                                <dbl>
 1 "dns.msftncsi.com.....(.\\....L...L"     1
 2 "dns.msftncsi.com...../(.\\. ..7...7"    1
 3 []                                      31
 4 "q...r0(.\\.>..L...L"                    1
 5 "dns.msftncsi.com.....4(.\\.\\..L...L"   1
 6 "dns.msftncsi.com.....5(.\\.e..7...7"    1
 7 "\\.n..B...B"                            1
 8 ".....5(.\\....7...7"                    1
 9 "\\A%..B...B"                            1
10 "..8...8.?(.\\m*..O...O"                 1
# ... with more rows
```

The following code from R generates the following output

```
filter(my_pcaps, type=="TCP") %>%
+     count(dst_port, sort=TRUE)
# Source:     lazy query [?? x 2]
# Database:   DrillConnection
# Ordered by: desc(n)
   dst_port     n
      <dbl> <dbl>
 1      443    37
 2       80     8
 3    52302     8
 4    52301     8
 5    52304     7
 6    52303     4
 7     2048     2
 8    52292     2
 9    52253     1
10    52246     1
# ... with more rows
```

And the resultant in the Apache Drill generates the following query and output

```
SELECT `dst_ip`, COUNT(*) AS `n`
FROM  dfs.`c:/Users/leela/documents/apache-drill-1.13.0/m.pcap`
WHERE (`type` = 'TCP')
GROUP BY `dst_ip`
ORDER BY `n` DESC
```

C. Visualisation

The change in the format of PCAP file like database format will definitely help in the anal ysis and Visualisation.The Vis ualisations provided by R are more than any language, offer

s particularly for the analysis of PCAP files. The packages like ggplot2,dplyr,heatmap,zoo,knitr,heatmap can do the visualisations in a better manner[7]. Thorough understanding and applications of the tools may provide good analysis of the pcap files as these files are generally difficult to read.

III. CONCLUSION

Analysing the PCAP files can provide more valuable information in the intrusion detection.
But handling the files which is huge will be possible with Apache Drill and R. Visualisation in deep will provide more answers to cyber security analysis.

REFERENCES

1. Raffel Marty , Applied Security Data Visualisation,1st Edition
2. Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualisations." *The Craft of Information Visualisation*. Morgan Kaufmann, 2003. 364-371.
3. https://rud.is/b/2017/07/27/reading-pcap-files-with-apache-drill-and-the-sergeant-r-package/
4. http: // cds. iiitb. ac. in/ gvcl
5. https://**github.com/apache/drill**
6. Conti, Greg. *Security data visualisation: graphical techniques for network analysis*. No Starch Press, 2007.
7. Tomar, Amit, et al. "A Survey of Visualisation Techniques for Network Security Analytics." (2015)
8. Jay Jacobs, Bob Rudis, Data–Driven Security: Analysis, Visualization and Dashboards April 2014