Student Name : Adrian Goh Jun Wei

Group : TS5

Date : 1 April 2020

LAB 4: ANALZING NETWORK DATA LOG

You are provided with the data file, in .csv format, in the working directory. Write the program to extract the following informations.

EXERCISE 4A: TOP TALKERS AND LISTENERS

One of the most commonly used function in analyzing data log is finding out the IP address of the hosts that send out large amount of packet and hosts that receive large number of packets, usually know as TOP TALKERS and LISTENERS. Based on the IP address we can obtained the organization who owns the IP address.

List the TOP 5 TALKERS

Rank	IP address	# of packets	Organisation
1	193.62.192.8	3041	European Bioinformatics Institute
2	155.69.160.32	2975	Nanyang Technological University
3	130.14.250.11	2604	National Library of Medicine
4	14.139.196.58	2452	Indian Institue of Technology (IIT) Guwahati
5	140.112.8.139	2056	Taiwan Academic Network

TOP 5 LISTENERS

Rank	IP address	# of packets	Organisation
1	103.37.198.100	3841	A*STAR
2	137.132.228.15	3715	Nationaly University of Singapore
3	202.21.159.244	2446	Republic Polytechnic
4	192.101.107.153	2368	Pacific Northwest National Laboratory
5	103.21.126.2	2056	Powai

EXERCISE 4B: TRANSPORT PROTOCOL

Using the IP protocol type attribute, determine the percentage of TCP and UDP protocol

	Header value	Transport layer protocol	# of packets
1	6	TCP	56064 (80.82%)
2	17	UDP	9462 (13.64%)
3	50	ESP	1698 (2.48%)
4	0	HOPOPT	1261 (1.82%)
5	47	GRE	657 (0.95%)

EXERCISE 4C: APPLICATIONS PROTOCOL

Using the Destination IP port number determine the most frequently used application protocol.
(For finding the service given the port number https://www.adminsub.net/tcp-udp-port-finder/)

Rank	Destination IP port number	# of packets	Service
1	443	13423	HTTPS
2	80	2647	HTTP
3	52866	2068	Dynamic / private ports
4	45512	1356	Dynamic / private ports
5	56152	1341	Dynamic / private ports

EXERCISE 4D: TRAFFIC

The traffic intensity is an important parameter that a network engineer needs to monitor closely to determine if there is congestion. You would use the IP packet size to calculate the estimated total traffic over the monitored period of 15 seconds. (Assume the sampling rate is 1 in 1000)

Total Traffic (MB)	64,777.822

EXERCISE 4E: ADDITIONAL ANALYSIS

Please append ONE page to provide additional analysis of the data and the insight it provides.
Examples include:
Top 5 communication pairs;
Visualization of communications between different IP hosts;
etc.

Please limit your results within one page (and any additional results that fall beyond one page limit will not be assessed).

EXERCISE 4F: SOFTWARE CODE

Please attach a copy of your code in an appendix.

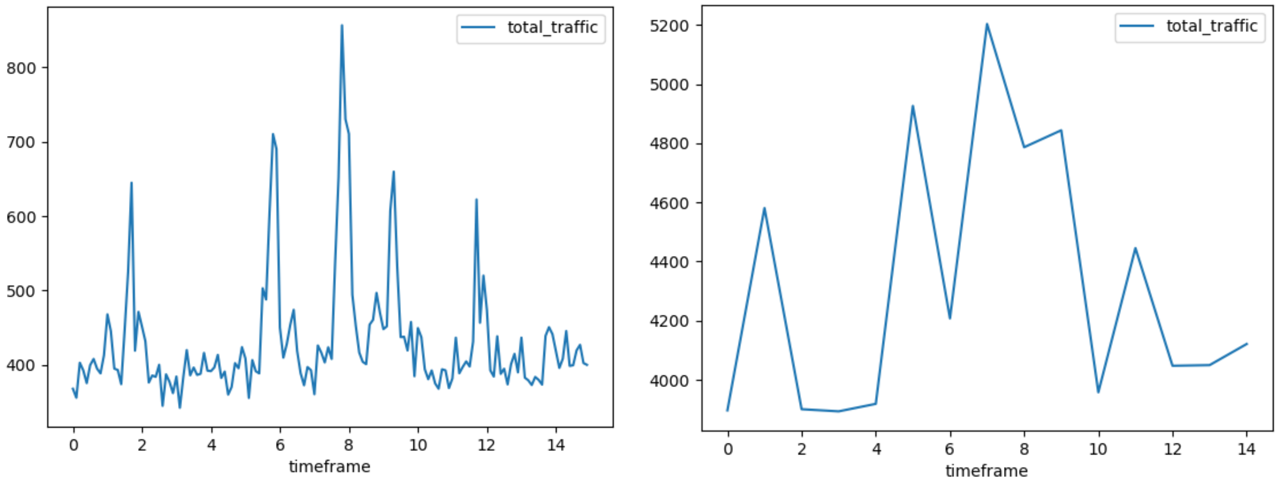
Peak traffic of the network

(Code can be found in /Analysis - Peak Traffic)

As network is an essential part of the infrastructure, it is crucial to understand how much traffic a network is experiencing on average and during it's peak, in order to ensure that the infrastructure can support the load. Thus, further analysis was carried out to determine the peak traffic of the network.

Assumptions:

- Sampling rate of 1 in 1000
- Estimated total traffic over a monitored period of 15 seconds
- Packet transmission is spread and distributed equally over the 15 seconds



Left image shows the network traffic (in MB) per 100 milliseconds
Right image shows the network traffic (in MB) per second.

From the analysis, it is understood that:

- The network traffic per 100 milliseconds averages around 403MB and peaks at 856MB.
- The network traffic per second averages around 4121MB and peaks at 5203MB.

Most communicated pairs of IP addresses

(Code can be found in /Analysis - Most communicated pairs)

Rank	IP Address (1)	Organisation (1)	IP Address (2)	Organisation (2)	Count of packets
1	137.132.228.15	National University of Singapore	193.62.192.8	European Bioinformatics Institute	4951
2	103.37.198.100	A*STAR	130.14.250.11	National Library of Medicine	2842
3	14.139.196.58	Indian Institue of Technology (IIT) Guwahati	192.101.107.153	Pacific Northwest National Library	2368
4	103.21.126.2	Powai	140.112.8.139	Taiwan Academic Network	2056
5	140.90.101.61	National Oceanic and Atmospheric Adminstration	167.205.52.8	Institut Teknologi Bandung	1752

From the interactions between the different IP addresses and their respective organisations, it can be concluded that this network is one used between educational and research institutes.