Student Name : Lee Ci Hui

Group : A58

Date : 31/3/2023

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 Institute of Technology
5	140.112.8.139	2056	NIL (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	National University of Singapore
3	202.21.159.244	2446	Rpnet
4	192.101.107.153	2368	Battelle Memorial Institute, Pacific Northwest Division
5	103.21.126.2	2056	Indian Institute of Technology Bombay

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	Percentage (%)
1	6	TCP	56063	80.818521
2	17	UDP	9462	13.640099
3	(Total)	(All)	69369	100.000000

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	433	13423	https
2	80	2647	http
3	52866	2068	Dynamic and/or Private Ports
4	45512	1356	Unassigned
5	56153	1341	Dynamic and/or 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 2048)

Total Traffic (MB)	7.722MB

EXERCISE 4E: ADDITIONAL ANALYSIS

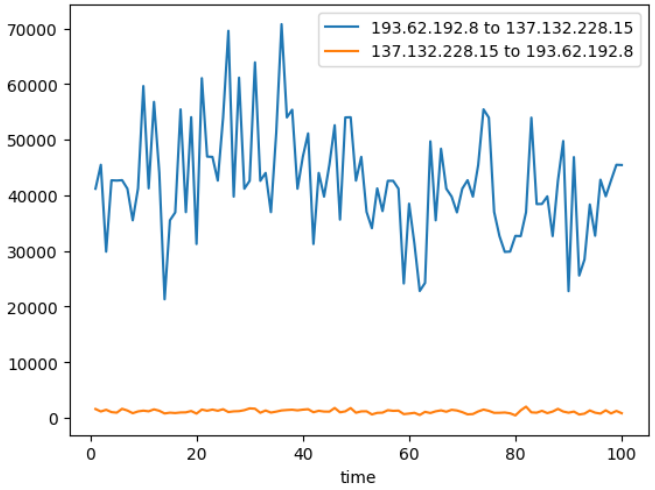
Please append ONE page to provide additional analysis of the data and the insight it provides.
Examples include:
Top 5 communication pairs;
Visualization of communications between different IP hosts;
etc.

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

1. Top 5 Communication Pairs

	src_IP	dst_IP	Number of times paired	Source Organisation	Destination Organisation
1	193.62.192.8	137.132.228.15	3041	European Bioinformatics Institute	National University of Singapore
2	130.14.250.11	103.37.198.100	2599	National Library of Medicine	A*STAR
3	14.139.196.58	192.101.107.153	2368	Indian Institute of Technology	Battelle Memorial Institute, Pacific Northwest...
4	140.112.8.139	103.21.126.2	2056	Taiwan Academic Network	Indian Institute of Technology Bombay
5	137.132.228.15	193.62.192.8	1910	National University of Singapore	European Bioinformatics Institute

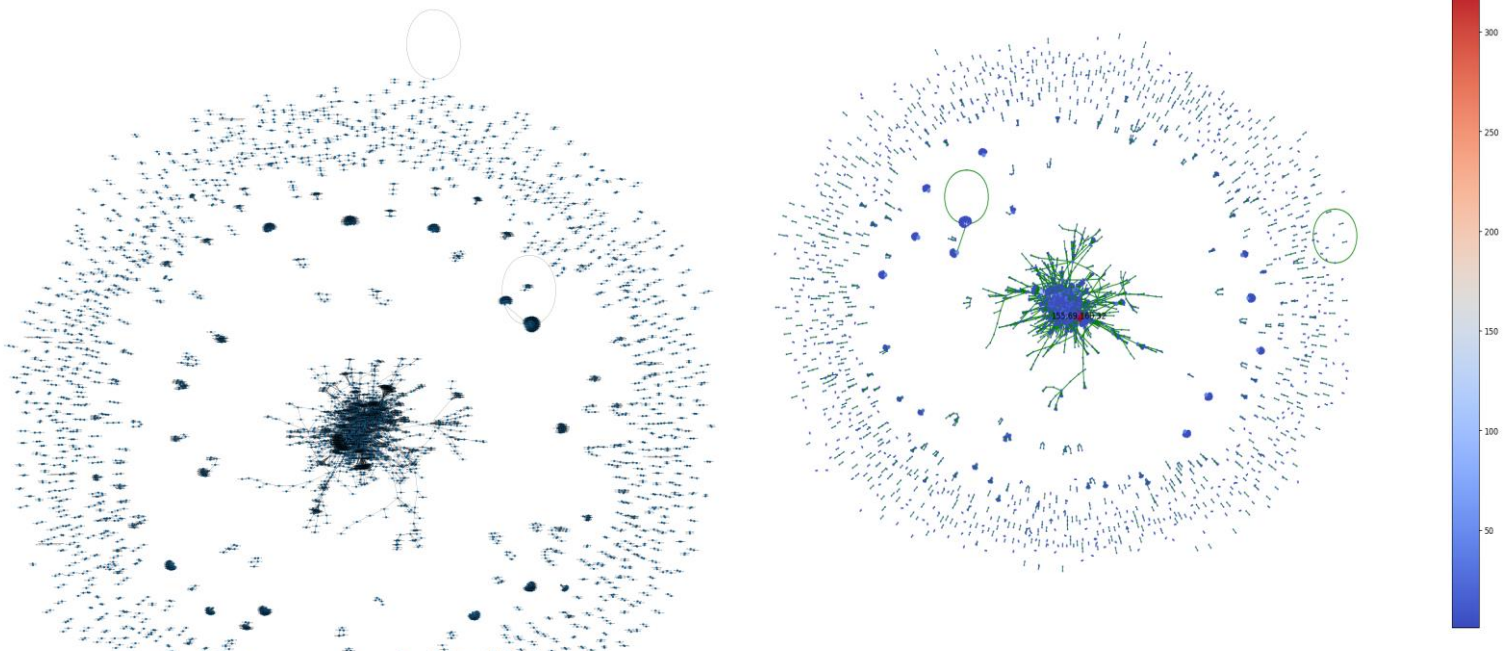
Figure 1: The table of the top 5 most communication pairs.



The traffic flow from 193.62.192.8 to 137.132.228.15 is higher than the opposite direction of traffic flow

Figure 2: The graph shows the traffic between the top 1 communication pair

2. Visualizing of communications between different IP hosts



The network graphs above show the communications between the hosts, ie. the source IP and the destination IP.

EXERCISE 4F: SOFTWARE CODE

Please also submit your code to the NTULearn lab site.