Student Name : ___________GONG ZERUI______

Group : ___________A57______________

Date : ___________2023/03/24_______

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. 192

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	JANET Jisc Service Limited, GB
2	155.69.160.32	2975	NU-AS-AP Nanyang Technological University, SG
3	130.14.250.11	2604	NLM-GW, US
4	14.139.196.58	2452	NKN-EDGE-NW NKN EDGE Network,IN
5	140.112.8.139	2056	NTU-TW Nantional Taiwan University, TW

TOP 5 LISTENERS

Rank	IP address	# of packets	Organisation
1	107.37.198.100	3841	A-STAR-AS-AP A-STAR, SG
2	137.132.228.15	3715	NUS-AS-AP NUS Information Technology, SG
3	202.21.159.244	2446	Republicpolytechnic-AS Republic Polytechnic, Multihoming AS Singapore, SG
4	192.101.107.153	2368	ESNET-AS,US
5	103.21.126.2	2056	HTB-IN Powai, IN

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	56063(80.82%)
2	17	UDP	9462(13.64%)

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 and/or Private Ports by IANA, Xsan,by Apple Inc
4	45512	1356	Unassigned
5	56152	1341	Dynamic and/or Private Ports by IANA, Xsan,by Apple Inc

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)	126516.254 MB

EXERCISE 4E: ADDITIONAL ANALYSIS

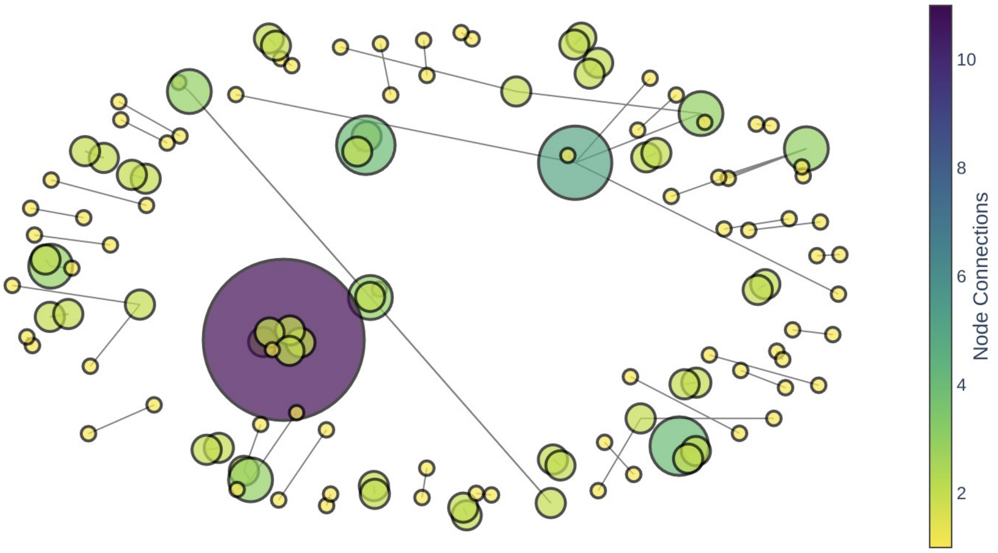
Please append ONE page to provide additional analysis of the data and the insight it provides.
Examples include:
Top 5 communication pairs;
Visualization of communications between different IP hosts;
etc.

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

I have created a visualization of the communication network between IP hosts using the Python package Plotly. The network is represented as a graph with IP addresses as nodes and connections between them as edges. The size and color of each node corresponds to the number of connections it has, while the edges represent the communication flow between them. To enhance the visualization, I have added additional information to each node, such as the country and organization associated with the IP address. Hovering over a node reveals this information as well as the number of connections it has. The resulting plot provides a clear and informative view of the communication network, making it easy to identify the most active nodes and connections.

This is a screenshot of how it looks like, the code is appended in my submission, be mind it takes around 3 minutes to run:

This is just a screenshot of my work, the actual demo is interactive and is available in my code.



EXERCISE 4F: SOFTWARE CODE

Please also submit your code to the NTULearn lab site.