

Student Name : [Bryan Lu We Zhern](#)

Group : [A52](#)

Date : [30/3/2023](#)

LAB 4: ANALYZING 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	RIPE Network Coordination Centre (RIPE)
2	155.69.160.32	2975	Asia Pacific Network Information Centre (APNIC)
3	130.14.250.11	2604	National Library of Medicine (NLM)
4	14.139.196.58	2452	Asia Pacific Network Information Centre (APNIC)
5	140.112.8.139	2056	Asia Pacific Network Information Centre (APNIC)

TOP 5 LISTENERS

Rank	IP address	# of packets	Organisation
1	103.37.198.100	3841	Asia Pacific Network Information Centre (APNIC)
2	137.132.228.15	3715	Asia Pacific Network Information Centre (APNIC)
3	202.21.159.244	2446	Asia Pacific Network Information Centre (APNIC)
4	192.101.107.153	2368	Battelle Memorial Institute, Pacific Northwest Division (PNNL-Z)
5	103.21.126.2	2056	Asia Pacific Network Information Centre (APNIC)

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	56064	82.37%
2	17	UDP	9462	13.90%
3	50	ESP	1698	2.49%
4	47	GREs	657	0.97%

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	Unassigned 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 2048)

Total Traffic(MB)	61.7769
--------------------	---------

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).

Top 5 Communication Pair

```
In [9]: # Top 5 Communication Pairs
top_5_comm_pairs = sflow_data.groupby([SRC_IP, DST_IP]).size().sort_values(ascending=False)[:5]
print(f"{'Source':<18}{'Destination':<18}Number of Communication Pairs")
for (a, b), y in top_5_comm_pairs.items():
    print(f"{a:<18}{b:<18}{y}")
```

Source	Destination	Number of Communication Pairs
193.62.192.8	137.132.228.15	3041
130.14.250.11	103.37.198.100	2599
14.139.196.58	192.101.107.153	2368
140.112.8.139	103.21.126.2	2056
137.132.228.15	193.62.192.8	1910

Visualisation

```
In [10]: comm_dataset = sflow_data.groupby([SRC_IP, DST_IP]).size().sort_values(ascending=False)

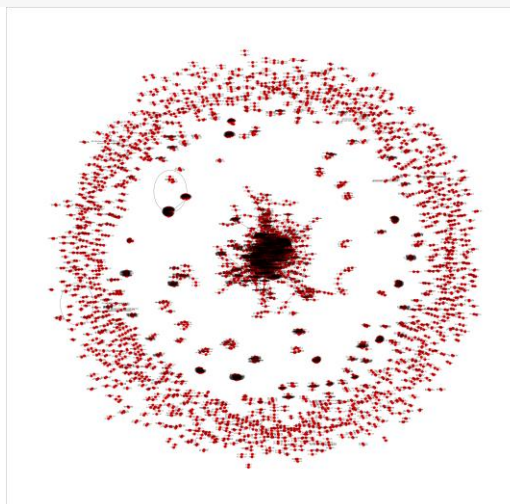
comm_dataframe = pd.DataFrame()
froms = []
tos = []
for (a, b), y in comm_dataset.items():
    froms.append(a)
    tos.append(b)

comm_dataframe["from"] = froms
comm_dataframe["to"] = tos

G = nx.from_pandas_edgelist(comm_dataframe, "from", "to")

plt.figure(figsize=(60, 60))
gp = nx.spring_layout(G)
nx.draw_networkx_nodes(G, gp, node_color="red")
nx.draw_networkx_edges(G, gp)
nx.draw_networkx_labels(G, gp, font_size=8)

plt.show()
```



EXERCISE 4F: SOFTWARE CODE

Please also submit your code to the NTU Learn lab site.