

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	EUR-BIO-INST
2	155.69.160.32	2975	NTUNET1
3	130.14.250.11	2604	NLM-ETHER
4	14.139.196.58	2452	NKN-IIT-GUW
5	140.112.8.139	2056	T-NTU.EDU.TW.NET

TOP 5 LISTENERS

Rank	IP address	# of packets	Organisation
1	103.37.198.100	3814	A-STAR-AS-AP
2	137.132.228.15	3715	NUSNET
3	202.21.159.224	2446	RPNET
4	192.101.107.153	2368	PNNL
5	103.21.126.2	2056	IITB-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	TCP	Transmission Control Protocol	56064(80.82%)
2	UDP	User Datagram Protocol	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	52899	2068	Dynamic and/or Private Ports

4	4512	1356	Unassigned
5	56152	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)	7.722
--------------------	-------

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).

You may want to refer to the jpynb file

The screenshot shows a Jupyter Notebook interface with the following content:

Exercise 4E: Additional Analysis

We can identify the top 5 communication pairs (source IP to destination IP) to understand the most active network interactions. This will provide insights into the major communication flows between devices in the network.

```
# Calculate the top 5 communication pairs using the correct column names
top_communication_pairs = df.groupby(['src_ip', 'dst_ip']).size().reset_index(name='Count')
top_communication_pairs = top_communication_pairs.sort_values(by='Count', ascending=False).head(5)

# Display the top 5 communication pairs
print("Top 5 Communication Pairs:\n", top_communication_pairs)
```

```
[100]
```

```
Top 5 Communication Pairs:
      src_ip      dst_ip  Count
4224  193.42.132.8    137.132.228.15    3841
975   138.14.228.11   183.37.198.108    2539
1587  14.151.196.58   192.181.187.153    2588
1839  146.112.5.139   193.21.135.2    2056
1297  137.132.228.15  193.42.192.8    1938
```

Using the IP_protocol column, we can calculate the percentage of TCP and UDP traffic in the dataset. This helps identify the dominant transport protocols in the traffic.

```
# Calculate the count of each transport protocol
protocol_counts = df['IP_protocol'].value_counts(normalize=True) * 100

# Print the protocol distribution
print("Protocol Distribution (TCP vs UDP):")
print(protocol_counts)
```

```
[102]
```

```
Protocol Distribution (TCP vs UDP):
IP_protocol
0      88.818798
17     13.439980
38     2.447744
0      1.817789
47     0.947895
43     0.148921
1      0.186674
381    0.864878
38     0.887766
183    0.881442
Name: proportion, dtype: float64
```

EXERCISE 4F: SOFTWARE CODE

Please also submit your code to the NTULearn lab site.

