Student Name : Ernest Ang Cheng Han

Group : FDDP1

Date : 23 October 2021

# LAB 4: ANALZING NETWORK DATA LOG

You will be provided with the data file, in .csv format, in the working directory. Write the program to extract the following information.

# 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 | 13.107.4.50 | 5960 | Microsoft Corporation |
| 2 | 130.14.250.7 | 4034 | National Library of Medicine |
| 3 | 155.69.160.38 | 3866 | Nanyang Technological University |
| 4 | 171.67.77.19 | 2656 | Stanford University |
| 5 | 155.69.199.255 | 2587 | Nanyang Technological University |

TOP 5 LISTENERS

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | IP address | # of packets | Organisation |
| 1 | 137.132.228.33 | 5908 | National University of Singapore |
| 2 | 192.122.131.36 | 4662 | A-STAR |
| 3 | 202.51.247.133 | 4288 | NUS Gigapop |
| 4 | 137.132.228.29 | 4022 | National University of Singapore |
| 5 | 103.37.198.100 | 3741 | A-STAR |

# 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 | 137707 | 77.69 |
| 2 | 17 | UDP | 36852 | 20.8 |
| 3 | 0, 50, 47, 1, 58, 41, 2 | Others | 2673 | 1.51 |

# EXERCISE 4C: APPLICATIONS PROTOCOL

Using the Destination IP port number determine the TOP 5 most frequently used application protocol.

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Destination IP port number | # of packets | Service |
| 1 | 443 | 43208 | HTTPS |
| 2 | 80 | 11018 | HTTP |
| 3 | 50930 | 2450 | Dynamic/unassigned |
| 4 | 15000 | 2103 | Dynamic/unassigned |
| 5 | 8160 | 1354 | patrol |

# EXERCISE 4D: TRAFFIC INTENSITY

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 calculated sampled traffic (MB):

169934750 bytes, 169.93475 MB

Estimated Total Traffic taking into account the sampling rate (MB):

169934750 \* 2048 = 348026368000 bytes, 348026.368 MB

# EXERCISE 4E: ADDITIONAL ANALYSIS

Please described additional analysis of the data and how it is useful. Please use a separate sheet to submit your new graphs and observations. Your report for this exercise is limited to 2 pages. The answer template and the two page additional analysis are to be submitted to your e-learning drive.

Examples

* Visulisation using scatter graph of port and IP address to determine if a specific node been port scanned by another node.
* Visualisation using network graph
* Other methods

You must analise and explain the graphs. Please do not be limited by the above examples.

# EXERCISE 4F: SOFTWARE CODE

Please attach a softcopy of your code to the e-learning drive.

ALL code related 4A – 4E can be found in the source code in python and/or Jupyter notebook that has been submitted together with this report on NTULearn:

CZ3006\_Lab\_2\_Source\_Code\_Jupyter\_Notebook.ipynb

CZ3006\_Lab\_2\_Source\_Code\_Python.py

**Top 5 Communication Pairs**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | IP Address 1 | Organization 1 | IP Address 2 | Organization 2 | Number of packets |
| 1 | 130.14.250.7 | A\*Star | 103.37.198.100 | National library of Medicine | 4201 |
| 2 | 192.122.131.36 | Stanford University | 171.67.77.19 | A\*STAR | 3628 |
| 3 | 129.99.230.54 | National Aeronautics and Space Administration | 137.132.22.74 | National University of Singapore | 2417 |
| 4 | 137.131.17.212 | The Scripps Research Institute | 137.132.228.42 | National University of Singapore | 2370 |
| 5 | 104.146.199.27 | Microsoft Corporation | 202.21.159.246 | Republic Polytechnic | 1794 |

From the above table, we can see that the top 5 communication pairs are usually between universities and research institutes, which is expected given the relationship between universities (as an academic and semi-research institute) and actual specialized research institutions and technology firms in the form of joint-research ventures, joint-education ventures

**Analyzing Traffic (assume dataset is time-series)**

Network is an important part of infrastructure. It affects the performance and response time of many applications we use such as WhatsApp and Instagram. It is important to understand trends that we observe in terms of traffic in our network infrastructure, which gives us the flexibility to plan and implement solutions such as load balancers and auto scaling groups

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

*Network Traffic in MB per second Network Traffic in MB per 100 millisecond*

As shown above, the peak traffic in the network is at 17314.935MB per second and 1844.855MB per 100 milliseconds (0.1 seconds). Averagely, the network that has been mined experience a traffic of 16993.475MB per second and 1699.348MB per millisecond

**Visualization of communicating pairs using Graph Visualization library**

![Chart, scatter chart

Description automatically generated]()

The graph illustrates all of the current active connections from all communicating pairs within the data file. The highest concentration of active connections is centered in the middle of the graph, where red dots where the source IP have more than 75 > active connections at present.

**Visualization of communicating pairs using Bar Chart**

Shape

Description automatically generated

The bar chart above visualizes the top 100 pairs of IPs connecting and communicating with one another, as well as the respective quantity of traffic that has been sent between the pairs.