Student Name : Cholakov Kristiyan Kamenov

Group : FDDP1

Date : 18.10.2022

# 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 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.

TOP 5 TALKERS

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | IP address | # of packets | Organisation |
| 1 | 193.62.192.8 | 3041 | JISC Services Limited |
| 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 | NKN EDGE Network |
| 5 | 140.112.8.139 | 2056 | 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 | NUS |
| 3 | 202.21.159.244 | 2446 | Republic Polytechnic |
| 4 | 192.101.107.153 | 2368 | Battelle Memorial Institute |
| 5 | 103.21.126.2 | 2056 | Powai |

# 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 | 56064 | 80.818798 |
| **2** | 17 | UDP | 9462 | 13.639902 |
| **3** | 50 | ESP | 1698 | 2.447744 |
| **4** | 0 | HOPOPT | 1261 | 1.817789 |
| **5** | 47 | GRE | 657 | 0.947095 |
| **6** | 41 | IPv6 | 104 | 0.149921 |
| **7** | 1 | ICMP | 74 | 0.106674 |
| **8** | 381 | Header value should be in 0-255 range | 45 | 0.064870 |
| **9** | 58 | IPv6-ICMP | 4 | 0.005766 |
| **10** | 103 | PIM | 1 | 0.001442 |

# EXERCISE 4C: APPLICATIONS PROTOCOL

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

Page 3-1

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Destination IP port number | # of packets | Service |
| 1 | 443 | 13423 | HTTPS |
| 2 | 80 | 2647 | HTTP |
| 3 | 52866 | 2068 | Dynamic |
| 4 | 45512 | 1356 | Unassigned |
| 5 | 56152 | 1341 | Dynamic |

# 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. (The sampling rate is as given in the data provided. Possibly 1 in 2048)

Total calculated sampled traffic (MB): 61.77694511413574 MB (calculated with MB = 220 bytes)

126519.18359375 MB (calculated with MB = 220 bytes)

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

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

# EXERCISE 4E: ADDITIONAL ANALYSIS

Top 5 Connection Pairs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | Address 1 | Organization 1 | Address 2 | Organization 2 | # of Packets |
| 1 | 137.132.228.15 | NUS | 193.62.192.8 | European Bioinformatics | 4951 |
| 2 | 103.37.198.100 | A\*STAR | 130.14.250.11 | National Library of Medicine | 2842 |
| 3 | 14.139.196.58 | Indian Institute of Technology | 192.101.107.153 | Battelle Memorial Institute | 2368 |
| 4 | 140.112.8.139 | Taiwan Academic Network | 103.21.126.2 | Powai | 2056 |
| 5 | 140.90.101.61 | National Oceanic and Atmospheric Administration | 167.205.52.8 | Institut Teknologi Bandung | 1752 |

Icon

Description automatically generated

Based on the results from the top 5 connection pars, we can consider that the sampled data may be from a router from Singapore because the NUS and A\*STAR figure in the top 2 connection. Also, most of the organizations from the top 5 connection pairs are educational and research institutions. Thus, we may assume that the sampled network is used for educational/research purposes.

IP Geolocation

Talkers' Map
This map represents the number of talkers from the sampled data plotted on the world map. The bubble shapes are used to give an idea of the volume of talkers in each country. We can see that the countries with most talkers are Singapore, USA, China, India, etc. Following this and the data from the top connection pairs, we can say that the singaporean educational and research institutions communicate with many foreighn ones .

Talkers

Page 3-2

# 

Listeners

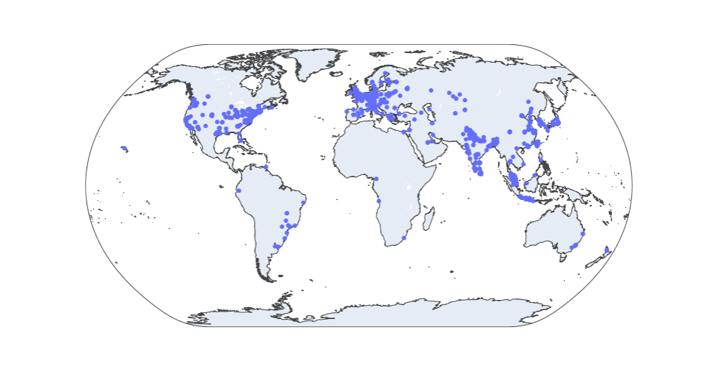
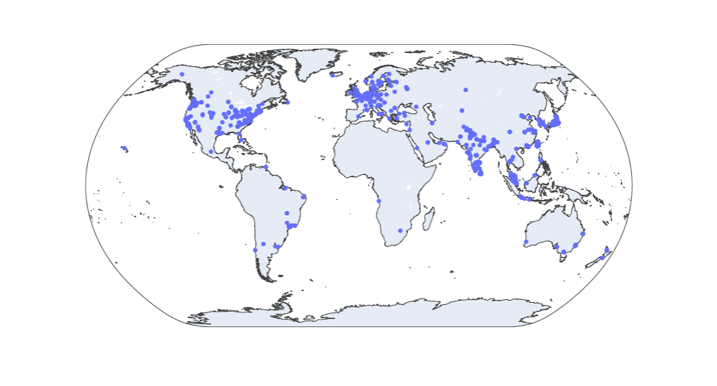
This map represents the same data but for the number of listeners in the countries. The difference is maybe that the number of listeners outside of Singapore is closer to the number of listeners in Singapore. Meaning the difference between Singaporean and foreign listeners is not so significant. Overall, the conclusion is again the same.

Chart

Description automatically generated with medium confidenceA picture containing graphical user interface

Description automatically generated

The above bar charts show that in fact the difference between the number of listeners in Singapore and the foreign ones is less than the one for the talkers



Listeners

Talkers

These maps show the actual locations of the talkers and listeners from the sampled data across the globe.

Network Geo Map

Diagram

Description automatically generated

The world map shows the interactions between the nodes in the top 10 connection pairs (bidirectional) with most packets sent between them. The red paths represent the path where each end is one of the two nodes from the connection pair. The graph further proves that the main traffic in the sampled network is between Singapore and some other foreign countries like UK, USA and China.

Page 3-3

Network Graph

Now lets explore the relations between the single-direction pairs with most packets sent between them.

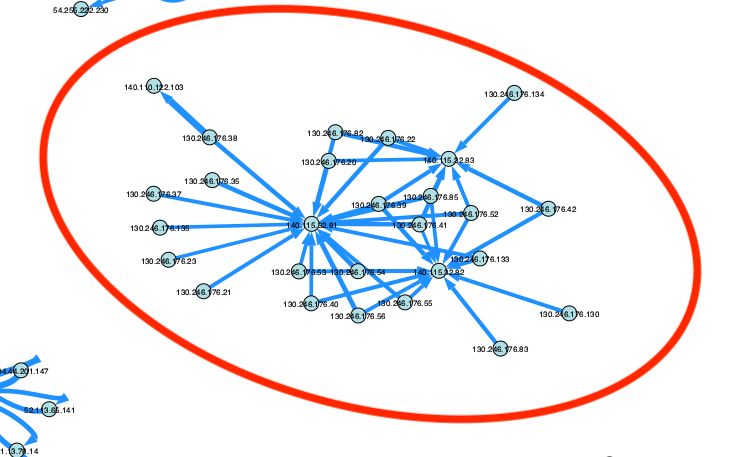
A picture containing chart

Description automatically generatedFirst, we will do the plotting based on the top 5% of the pairs.

**Diagram, radar chart

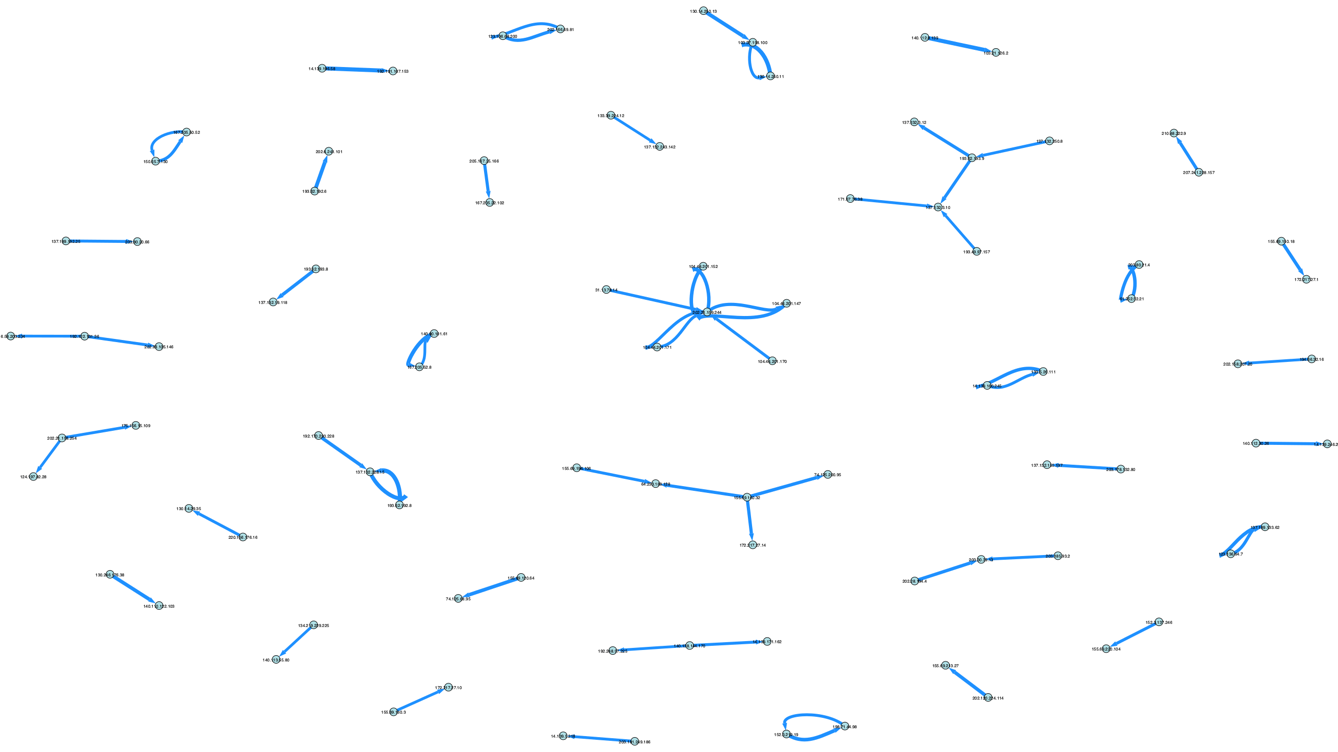
Description automatically generated**

This part of the graph represents the connections with most packets sent between NTU and other nodes. We can say that because the IP address (155.69.160.32) of the node that combines most of the other nodes in one graph is an NTU’s one.

****

As for this one, we can say this segment shows the interaction between UK’s Science Technology Facilities and Taiwan Academic Network as most of the IP addresses in the graph (130.246.176.41 and 140.115.32.82) are registered to these organizations.

Page 3-4

We can also, show the top 1% of the pairs with most packets sent between them. However, the nodes here are significantly less than in the graph about the top 5%. Thus, not many interactions between the different nodes can be monitored.

Port Scanning

Table

Description automatically generated

If we group all entries by their source and destination IP addresses, we will have all the single-directional pairs in the sampled network. Then by counting the unique destination ports we can check which listener’s ports a talker tries to access.

From the table on the left, we can see that talker node with IP address: 193.62.192.8 sends packets ot 1447 different ports of listener node with IP address: 137.132.228.15. Also, vice-versa packets are transmitted to 1009 unique ports.

Graphical user interface, application

Description automatically generatedIf we check the organization to which the IP addresses belong, we will see that 193.62.192.8 is registered under “European Bioinformatics”, while 137.132.228.15 belongs to “NUS”.

Further, if we plot the different listener’s ports that are accessed on a scatter plot, we can see that NUS’s node sends packets to EB’s ports in the range from 0 to ~ 65 000. With 65,535 being the maximum port number, it seems like NUS’s node is doing some kind of port scanned to EB. We can see the same thing vice-versa but EB sends to ports in the interval 35 000 – 60 000. Both connections access ports in the range 49 151 – 65 536 which are dynamic or private and can be accessed by basically everyone.

These two connections are the only conncections where these phenomenon can be spotted. Also, the scanning is between only two nodes, none of the nodes does the same thing with other nodes from the sampled data.

Page 3-5

# EXERCISE 4F: SOFTWARE CODE

The next pages represent the Jupyter Notebook file and some of the graphs in form of images with higher resolution.

Page 3-6