



Assignment 1 of CPSC 641

Student name: Xue Kai

Student ID: 30160299



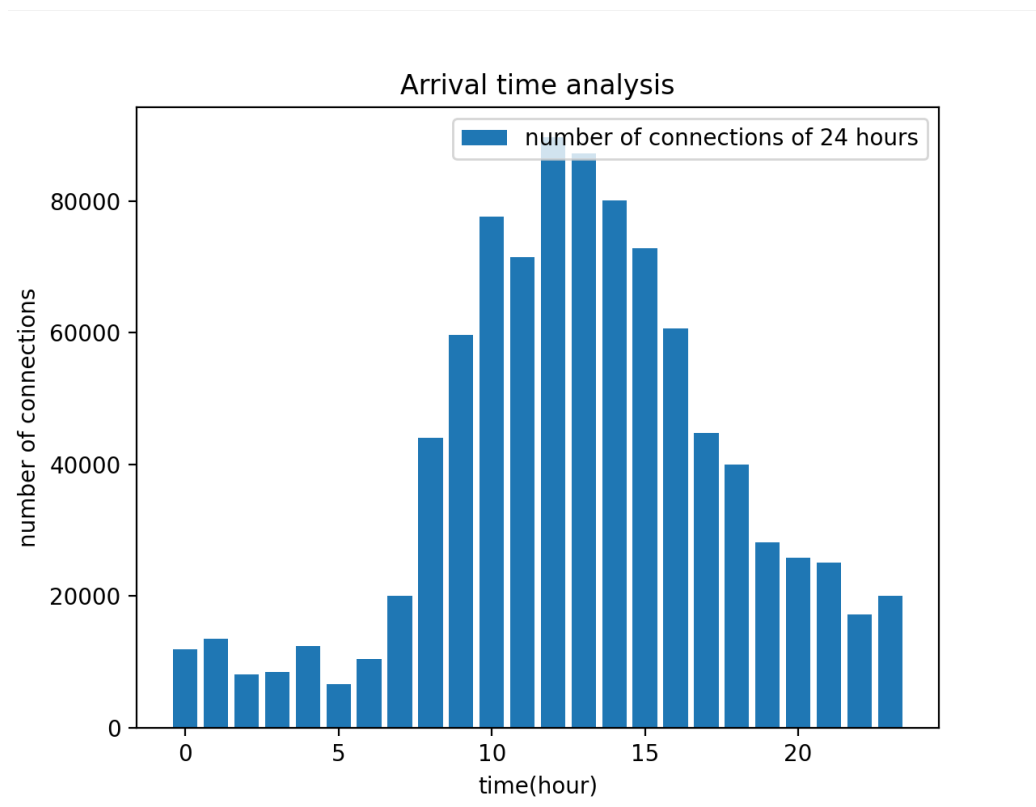
UNIVERSITY OF
CALGARY

Background

Spotify is a popular music streaming service that is heavily used by students, faculty, and staff at the University of Calgary. Users with accounts on Spotify can browse a catalog of over 50 million tracks, including songs, audio books, and podcasts. Media content is transferred using well-known Internet protocols, such as secure HTTP (HTTPS) and TCP/IP. Some of this content comes from Spotify's own servers, while the rest comes from Content Delivery Network (CDN) nodes, such as Akamai and Fastly.

Data Analysis Tasks

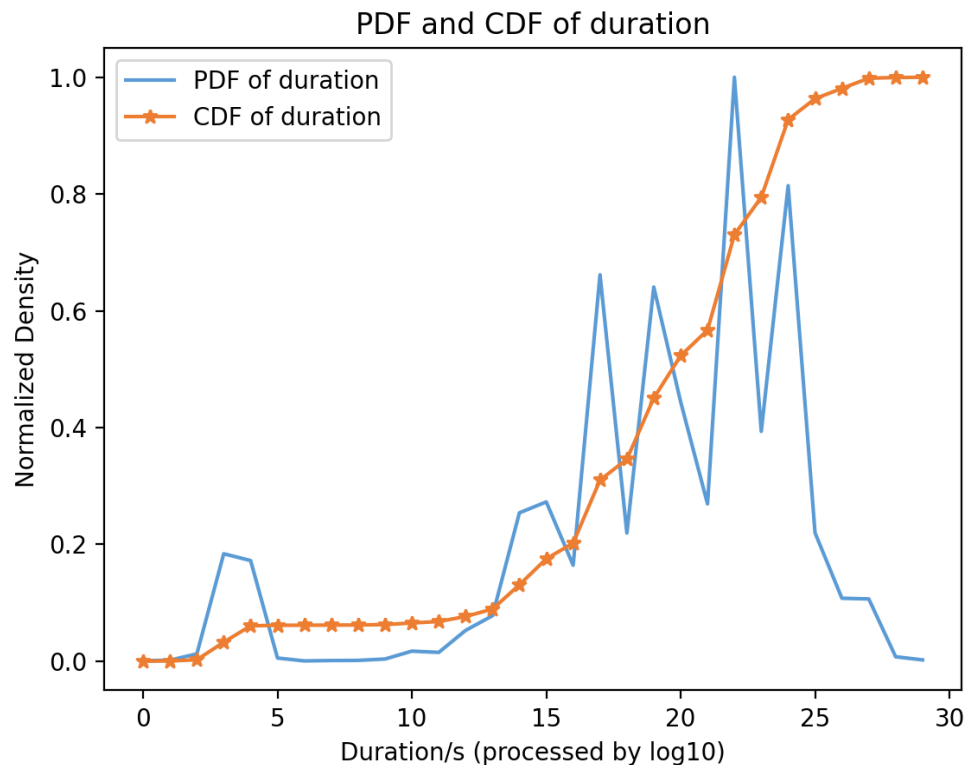
1. Arrival Time Analysis



This graph shows the information about the number of connections in each one-hour interval of a single day. The data is flat at first and then increases rapidly, hits the peak around 12 am. After that, the data gradually decreases until the end of the day. From the graph we can see that the majority of the connections are concentrated in the noon time, which means users prefer to use Spotify during this period.

2. Connection Duration Analysis (5 marks)

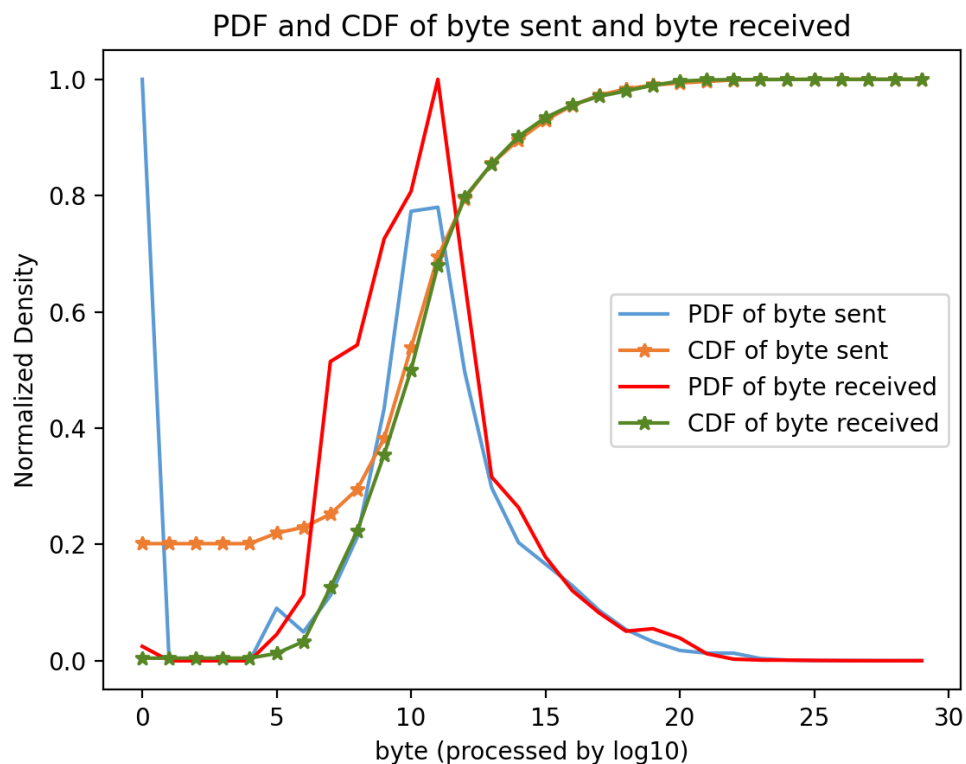
```
The maximum value of duration is: 21504.610639
The minimum value of duration is: 0.0
The mean value of duration is: 175.6282663659776
The median value of duration is: 10.2278
The standard deviation of duration is: 571.4785641884687
```



This graph shows the information about the PDF and CDF of duration. Both curves are flat at first and then increases rapidly. From the graph we can see that most of connections have durations in the thousands or hundreds, and very few have longer durations (as data is processed by log10 and the bin of the graph is 30 to show more fluctuations). We can also get this conclusion from the basic statistics of the data. The standard deviation of duration is high and there is a gap between the mean and the median of the data.

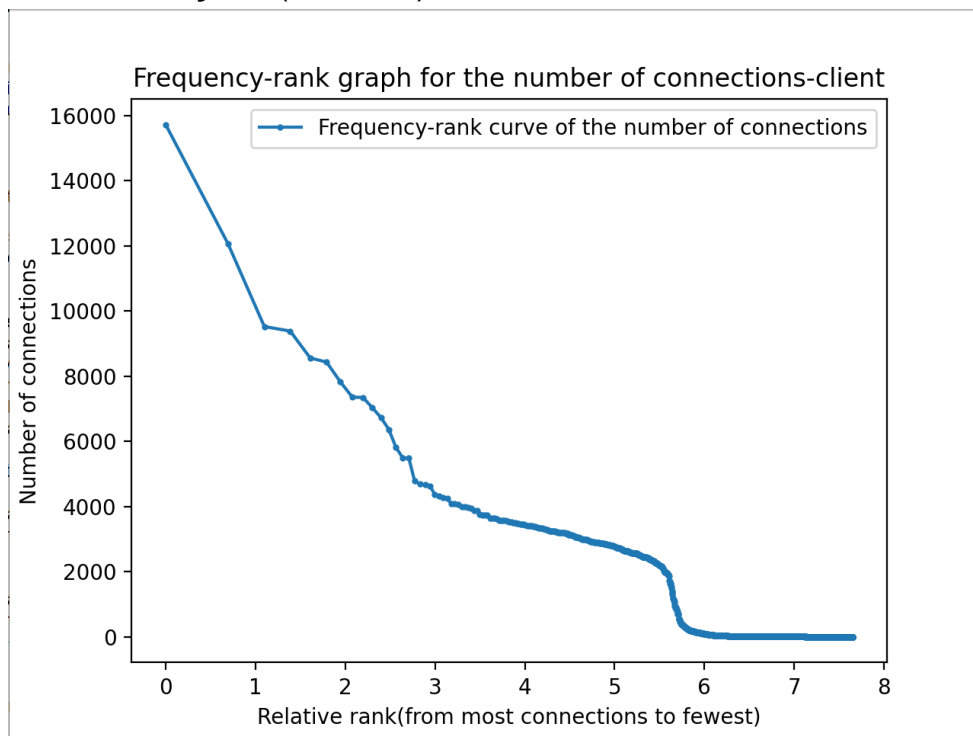
3. Transfer Size Analysis (5 marks)

```
The maximum value of byte sent is: 429825851
The minimum value of byte sent is: 0
The mean value of byte sent is: 43291.65363204911
The median value of byte sent is: 1267.0
The standard deviation of byte sent is: 1343060.0617704024
The maximum value of byte received is: 3131880934
The minimum value of byte received is: 0
The mean value of byte received is: 118294.65922014578
The median value of byte received is: 3034.0
The standard deviation of byte received is: 3980214.9131139056
```

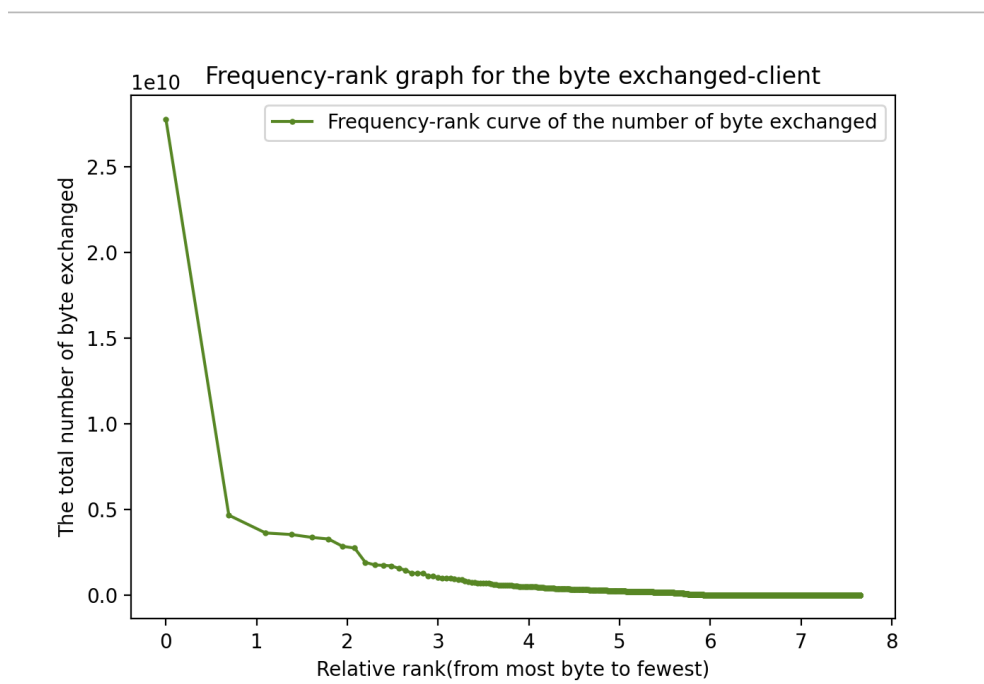


This graph shows the information about the PDF and CDF of byte sent and received. From the graph we can see that the PDF of byte sent and received have similar increase trend. The only difference is that the data of byte sent has much more small values (such as 0), which makes the curve of byte sent has a high density at the beginning of the horizontal axis. At the same time, this causes the CDF of byte sent has a higher starting point than the CDF of byte received. In addition, the data of byte sent and received both only have a few extreme large values.

4. Client Analysis (5 marks)

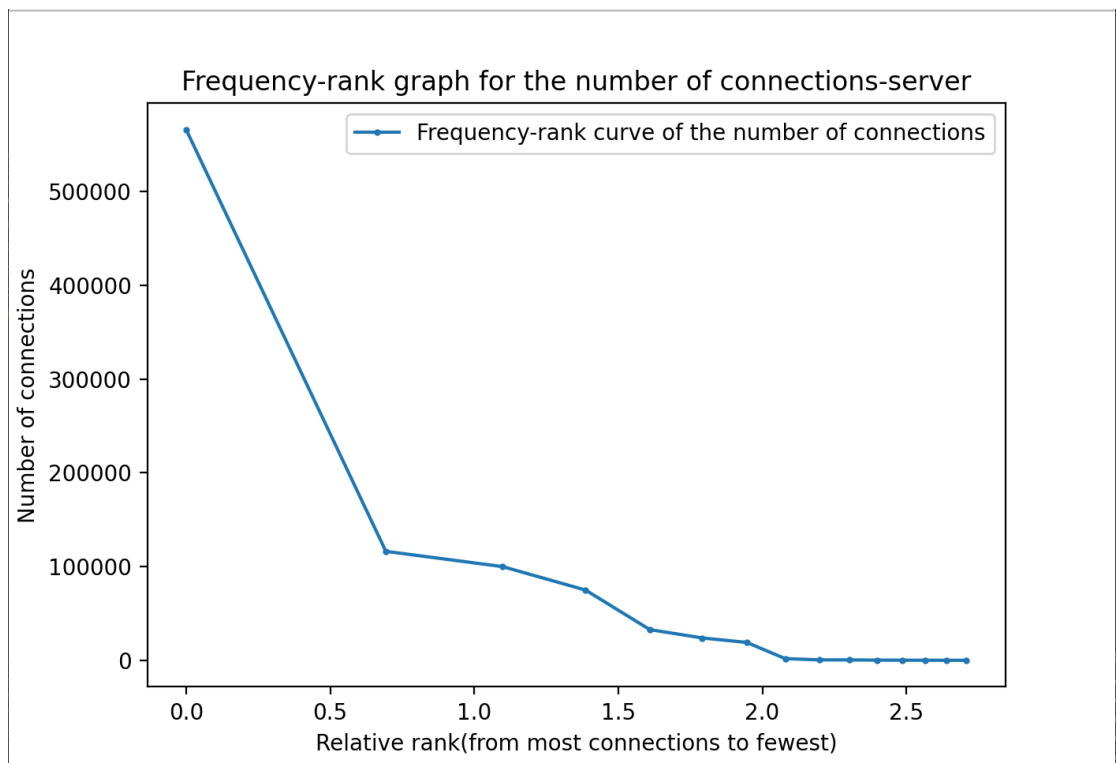


This graph shows the information about the number of connections initiated by each client IP address. From the graph we can see that only a few clients initiated a large number of connections, which means they intensively used Spotify. While more clients have a lower usage frequency. It is obvious that as the number of connections decreases, the density of ranks gradually increases. From my perspective, it means most of clients are light user.

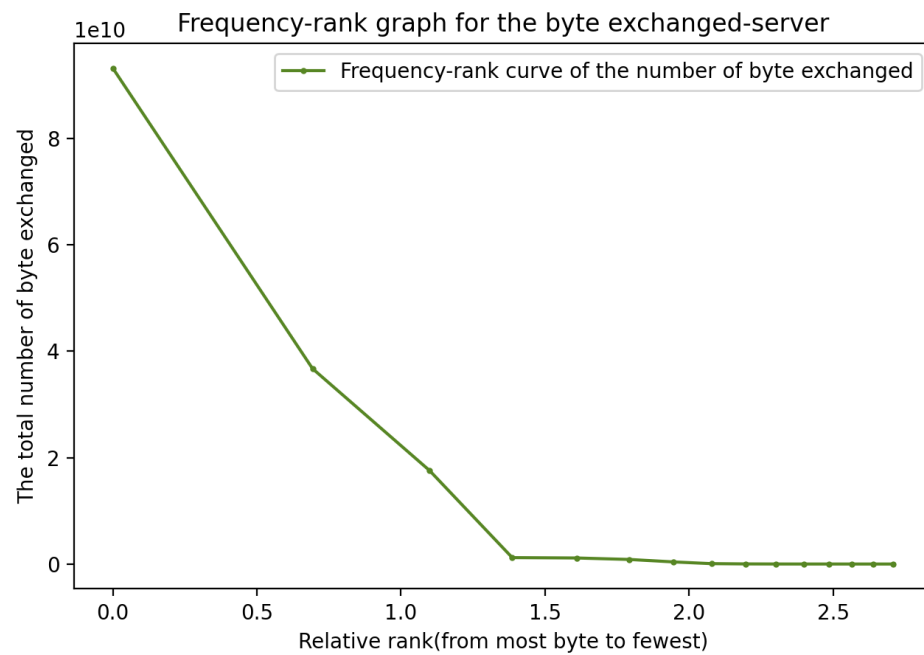


This graph shows the information about the number of bytes exchanged by each client IP address. It has similar trend with the number of connections initiated by each client IP address. Only a few clients exchanged massive data, which means they intensively used Spotify. While more clients exchanged data in a lower and ordinary amount. As the amount of exchanged bytes decreases, the density of ranks gradually increases. From my perspective, it means same conclusion with previous graph that most of clients are light user.

5. Server Analysis (5 marks)

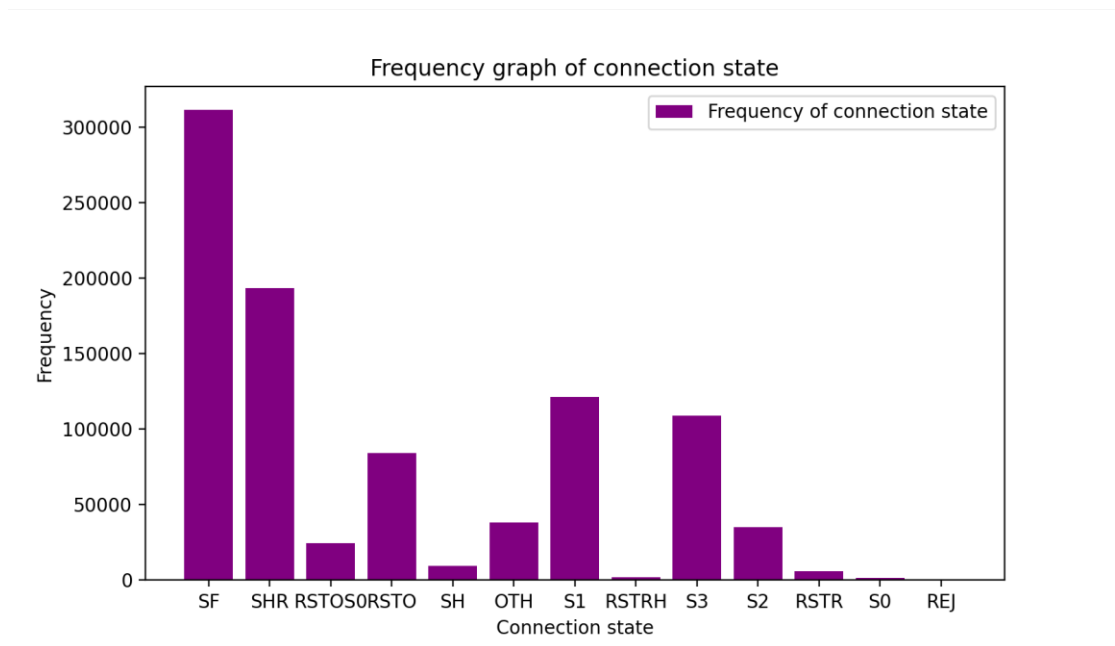


This graph shows the information about the number of connections involved by each server IP address. From the graph we can see that only a few servers involved in a large number of connections, which means they have heavier task volume. While more servers involved in less connections. It is obvious that as the number of connections decreases, the density of ranks gradually increases. From my perspective, it means most of server received minor and less requests from clients, while one server received main and more requests.



This graph shows the information about the number of byte exchanged by each server IP address. It has similar trend with previous graph. It is obvious that one server exchanged higher order of magnitude bytes, while most of server exchanged limited bytes. This confirms my inference from another aspects that most of server received minor and less requests from clients, while one server received main and more requests.

6. Your Own Analysis (5 marks)



For my own analysis, I choose to do some statistic about the connection state. This graph shows the information about the frequency of different connection states. From the graph we can see that the most common connection is SF. This means normal establish and termination. The second state is SHR. This symbol means responder sent a SYN ACK followed by a FIN, said we never saw a SYN from the originator. The least state is REJ, seems like most connection attempts were approved.