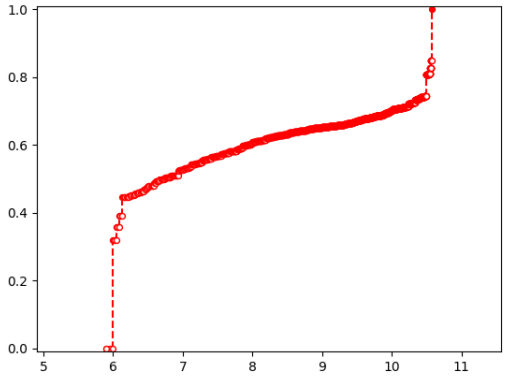
Course Project Report
CSCD58
Name: Minqi Wang
Student#: 1002105900
Utorid: wangmi56

Tools: Wireshark, Python
Wireshark download link: https://www.wireshark.org/download.html
Python libraries: numpy, matplotlib

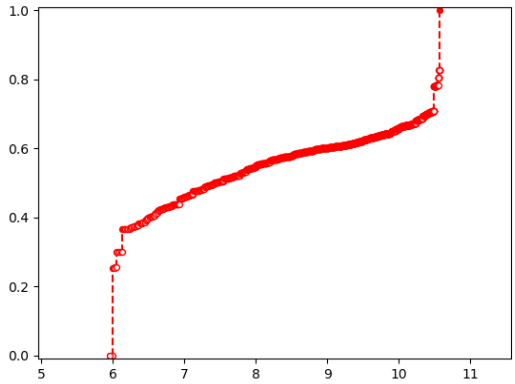
1. Per-packet statistics
    a. Packet type, size and count

```
Ethernet packets -- count: 1021724, percentage: 1.0, size: 554535994, size percentage: 1
IP packets -- count: 887647, percentage: 0.8687737588624717, size: 542800710, size percentage: 0.9788376514293498
ICMP packets -- count: 38775, percentage: 0.03795056199130098, size: 2491540, size percentage: 0.004493017634487402
TCP packets -- count: 666096, percentage: 0.6519333988435233, size: 474753364, size percentage: 0.8561272291370865
UDP packets -- count: 128726, percentage: 0.1259890146458339, size: 28847810, size percentage: 0.05202152847088227
```

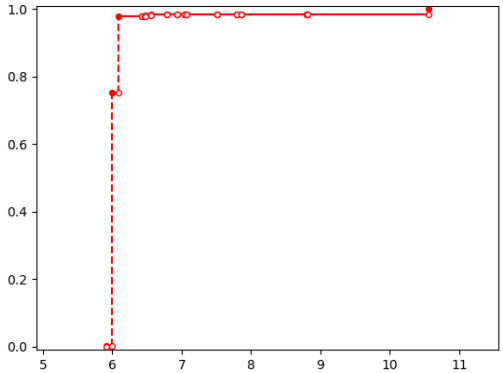
    b. CDF of size of packets (Note: Applied log_2 on the x axis)



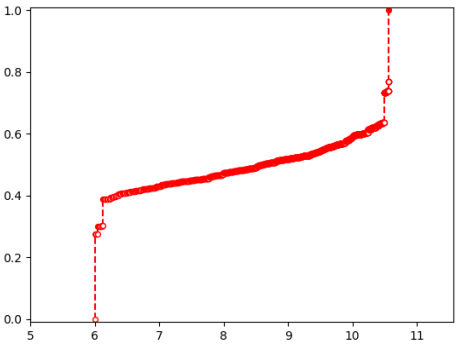
Size of all packets



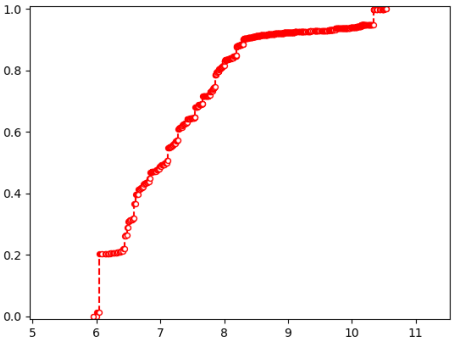
Size of all IP packets



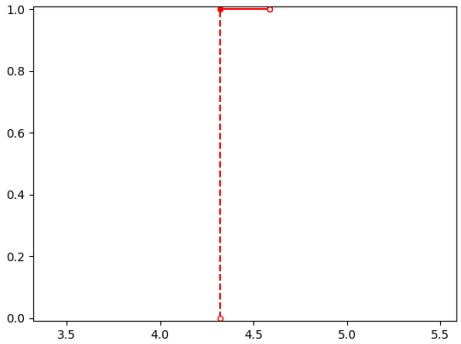
Size of non-IP packets



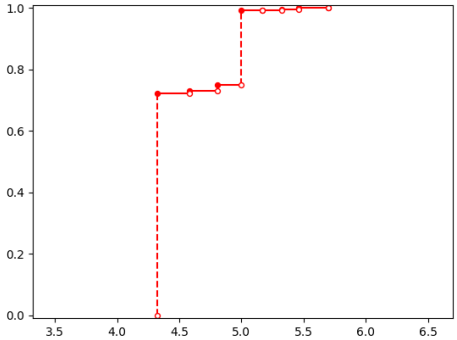
Size of all TCP packets



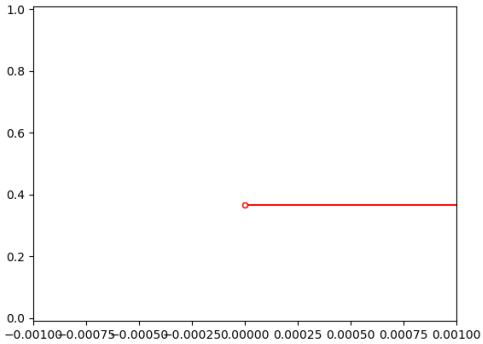
Size of all UDP packets



IP header size



TCP header size



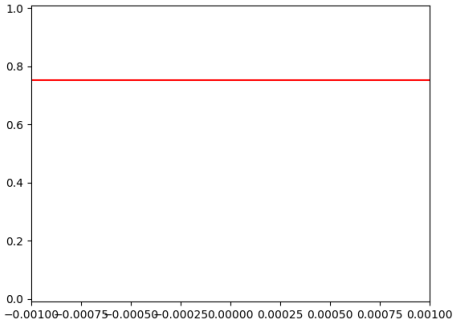
UDP header size

Analysis: UDP vs TCP: The size of TCP packets is more stable, which is shown in the graph that the curve is smooth. The size of UDP packets varies more frequently, which is shown as a steep curve. Both the TCP header size and UDP header size tend to be fixed, while from the graph it seems there are more variations in TCP header sizes.

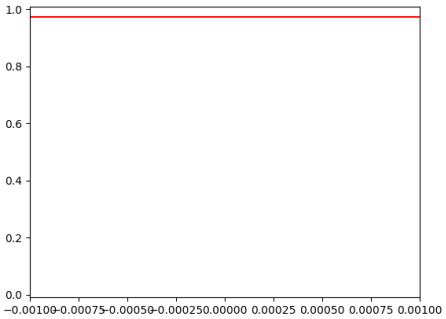
2. Flow statistics
    a. Flow Type

```
Total number of flows: 32303
Number of TCP flows: 13618, percentage: 0.421570751942544
Number of UDP flows: 18685, percentage: 0.578429248057456
TCP connection state -- Request count: 132, Ongoing count: 7047, Failed count: 1180, Reset count: 5174, Finished count: 85
```

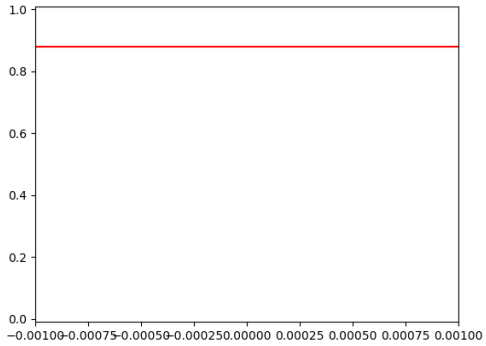
    b. Flow duration (Note: Applied log_2 on the x axis)



TCP flows duration CDF



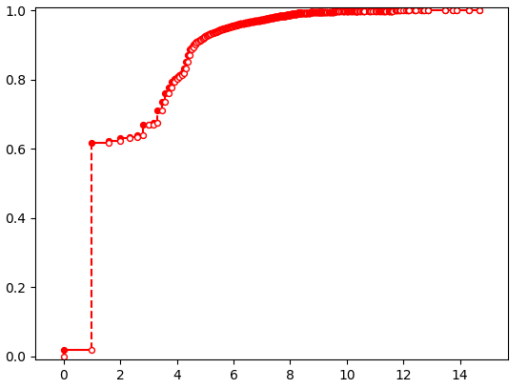
UDP flows duration CDF



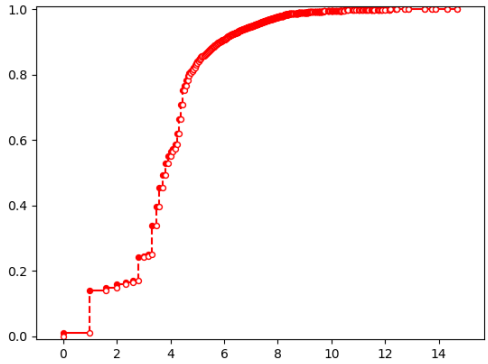
All flows duration CDF

In terms of duration, there are not much difference between UDP and TCP flows.

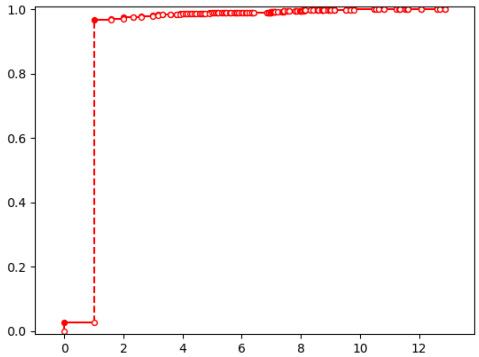
c. Flow Size (Note: Applied log_2 on the x axis)



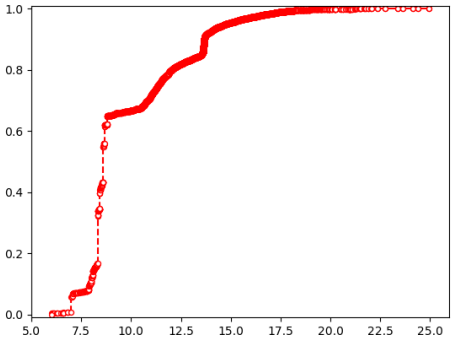
Packet count of all flows



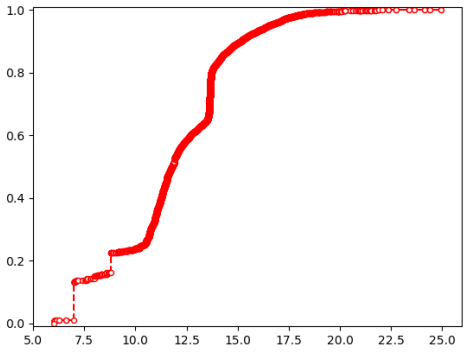
Packet count of TCP flows



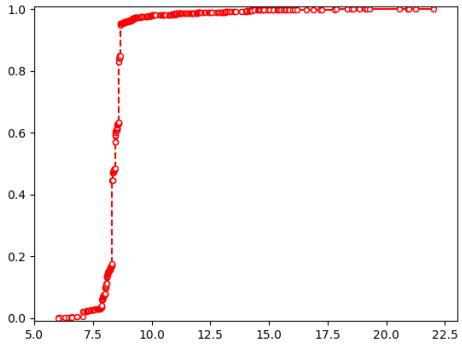
Packet count of UDP flows



Size of all flows

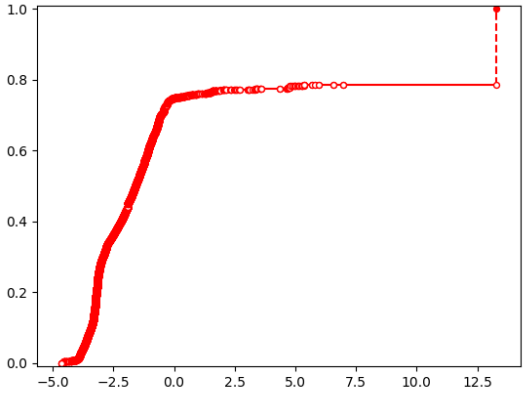


Size of TCP flows



Size of UDP flows

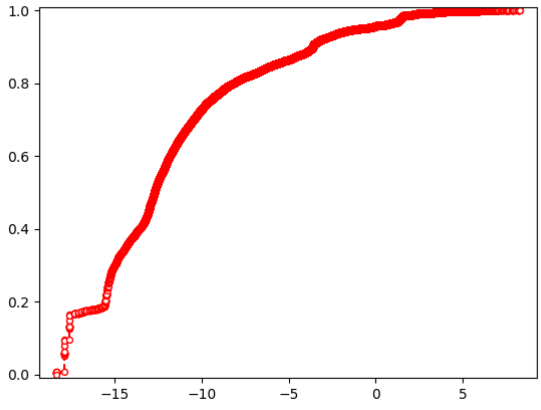
TCP's CDF curve is smooth, meaning the distribution of flow size is more "continuous". UDP's CDF tends to suddenly blow up when x (i.e. size) is still some small value, meaning the size of packets concentrates near that value.



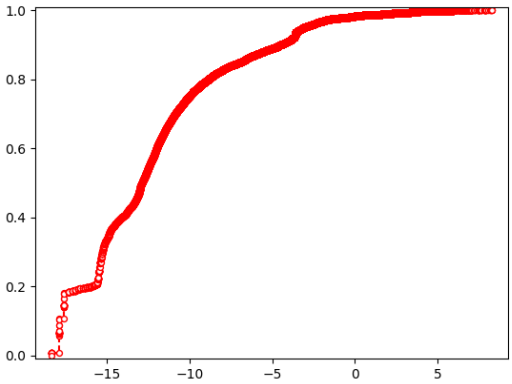
Header overhead of TCP flows

There are approximately 25% of the TCP flows having header overhead > 1, which means they have larger header than the actual data they carry.

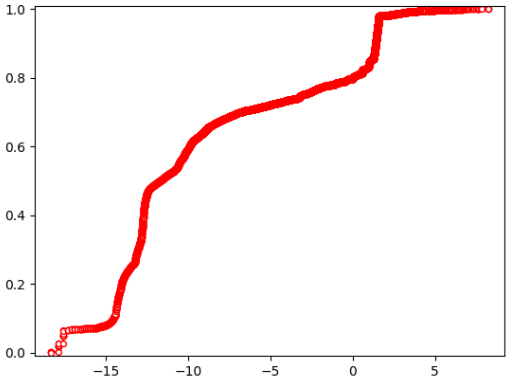
d. Inter-packet arrival time



All flows inter-packet arrival time



TCP flows inter-packet arrival time



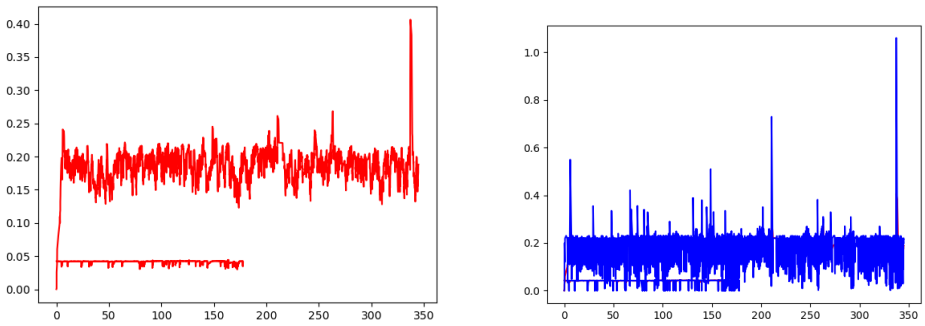
UDP flows inter-packet arrival time

The inter-packet arrival time distributes evenly so no outstanding common values.

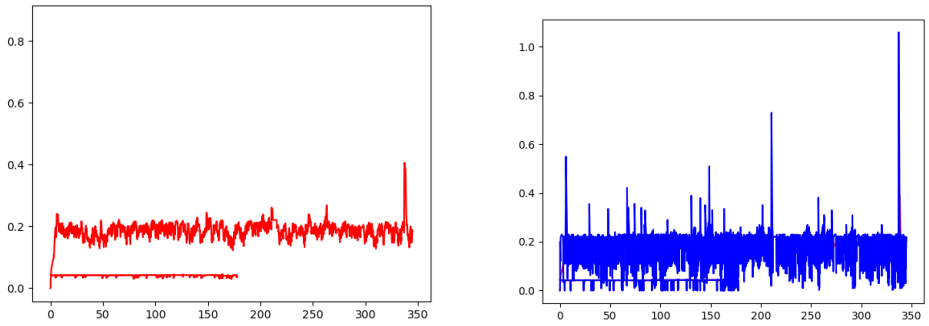
e. TCP State

```
Total number of flows: 32303
Number of TCP flows: 13618, percentage: 0.421570751942544
Number of UDP flows: 18685, percentage: 0.578429248057456
TCP connection state -- Request count: 132, Ongoing count: 7047, Failed count: 1180, Reset count: 5174, Finished count: 85
```

3. RTT Estimation

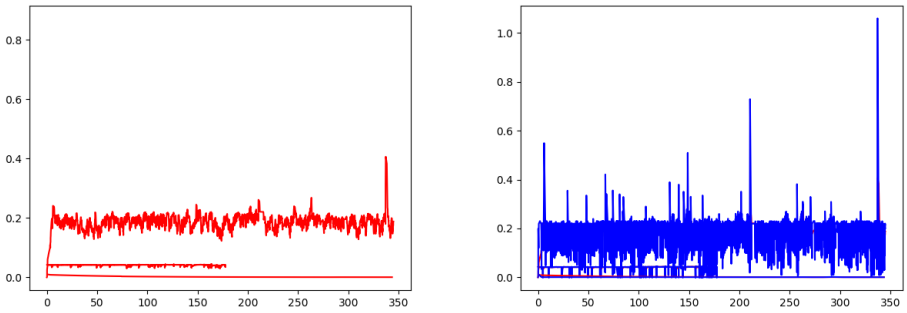


Top1 duration flow estimation RTT (left) vs Sample RTT (right)



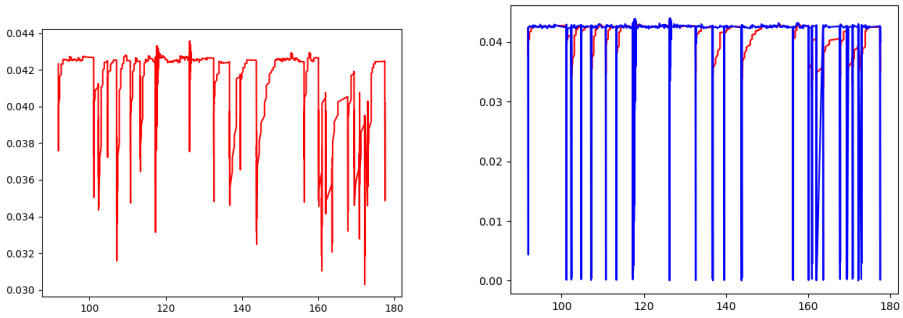
Top2

duration flow estimation RTT (left) vs Sample RTT (right)



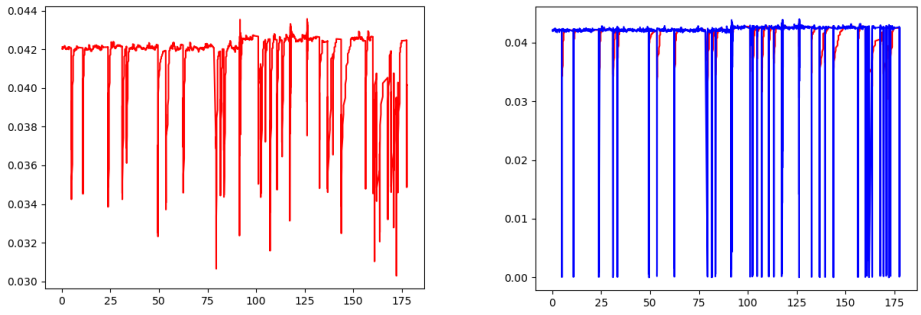
Top3

duration flow estimation RTT (left) vs Sample RTT (right)



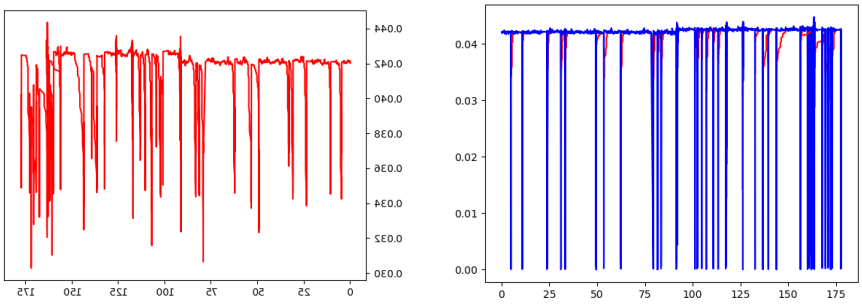
Top1

number of packet flow estimation RTT (left) vs Sample RTT (right)



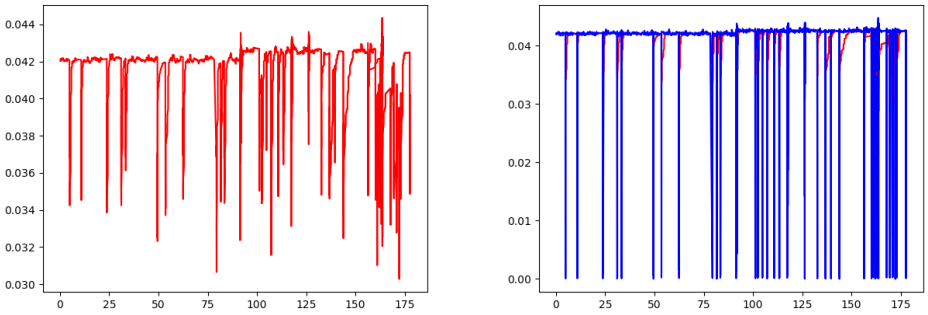
Top2

number of packet flow estimation RTT (left) vs Sample RTT (right)



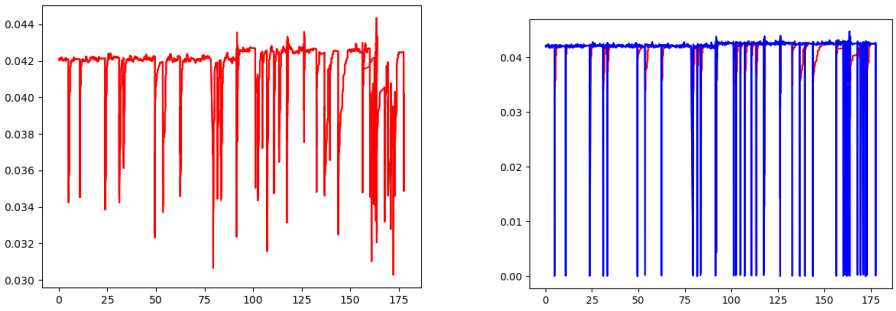
Top3

number of packet flow estimation RTT (left) vs Sample RTT (right)



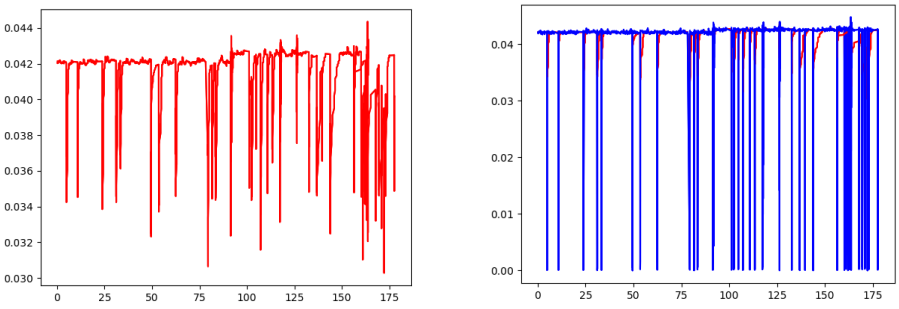
Top1

size flow estimation RTT (left) vs Sample RTT (right)



Top2 size

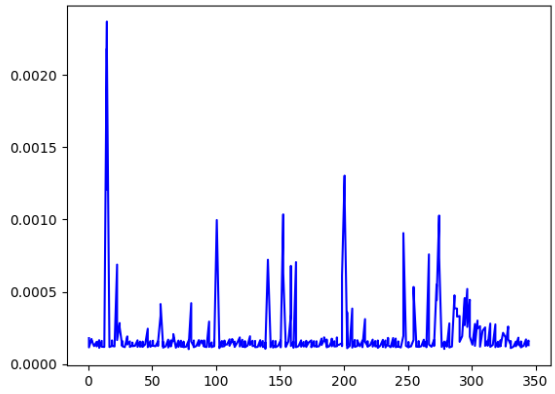
flow estimation RTT (left) vs Sample RTT (right)



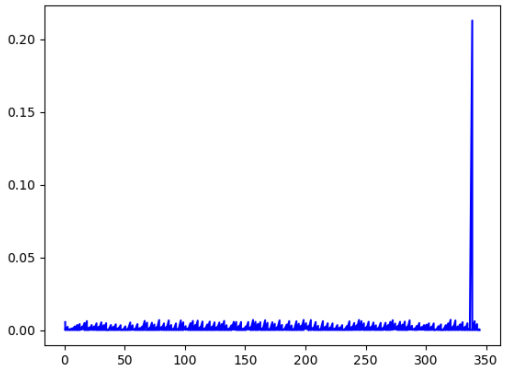
Top3 size

flow estimation RTT (left) vs Sample RTT (right)

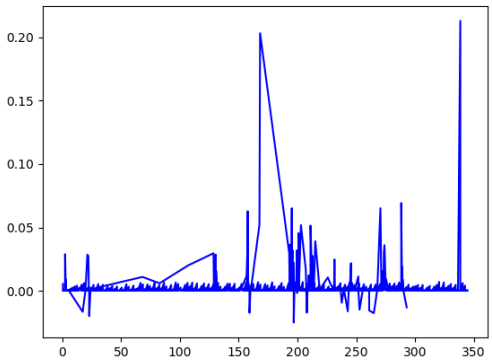
The estimated RTT is relatively more stable than the sample RTT. The congestion window size is changing during the lifetime of a connection, which means the flow rate is changing and thus the RTT gets changed.



Top1 number of connection host pair representative RTT per flow



Top2 number of connection host pair representative RTT per flow



Top3 number of connection host pair representative RTT per flow

The changes in representative RTTs seem to be random. This might be caused by the instabilities in the network such as congestion, host (endpoint) failure etc.