CSC458 Project Report

For this project, we have decided to use wireshark to read the pcap file. If interested, the following link is the link to our converted csv file for the packets of univ1\_pt11:

<https://www.dropbox.com/s/rza3wodi49y7cjb/11.csv?dl=0>

This file is required to be in the code folder if any of the codes are ran.

**Dataset Statistics:**

Q1: **For many of the CDF charts, it is better to use logarithmic scale for the x-axis. Explain why this is usually the case.**

There are two reasons to use logarithmic scale in this case and the first is to respond to skewness towards large values; i.e., cases in which one or a few points are much larger than the bulk of the data. The second is to show percent change or multiplicative factors.

Q2: **Type of packets:**Total number of packets: 1008977

Link Layer packets:

Ethernet type packets: 1008977

accounting for 100.0%

516857456bytes

Network Layer Packets:

ICMP type packets: 32209

accounting for 3.192243232501831%

2071991bytes

ARP type packets: 94270

accounting for 9.343126751154882%

6039440bytes

IPv4 type packets: 3271

accounting for 0.32418974862657923%

1357402bytes

other type packets: 879227

accounting for 87.14044026771671%

507388623bytes

Transport Layer Packets:

Network Layer Packets:

UDP type packets: 113364

accounting for 11.235538570254823%

45506229bytes

TCP type packets: 533074

accounting for 52.83311710772396%

363074981bytes

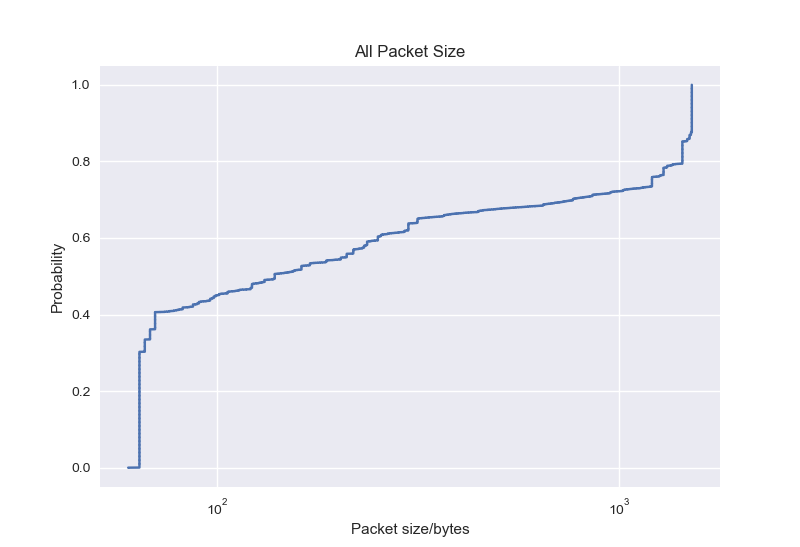
other type packets: 362539

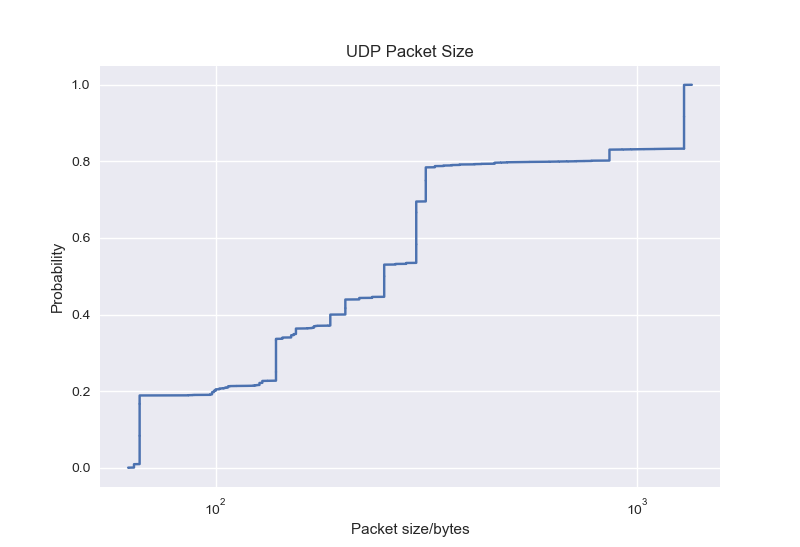
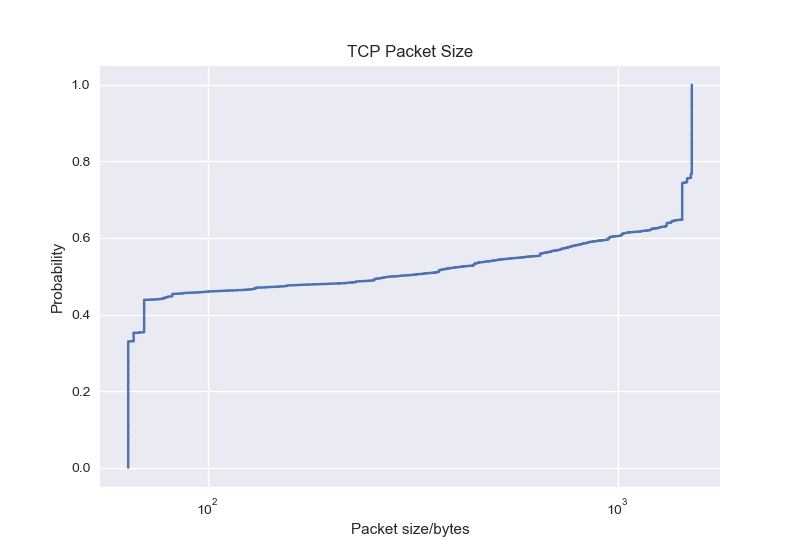
accounting for 35.93134432202121%

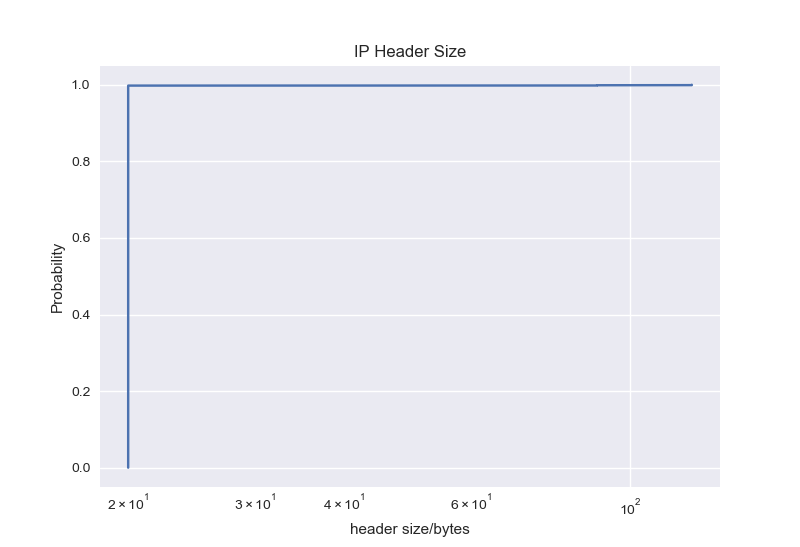
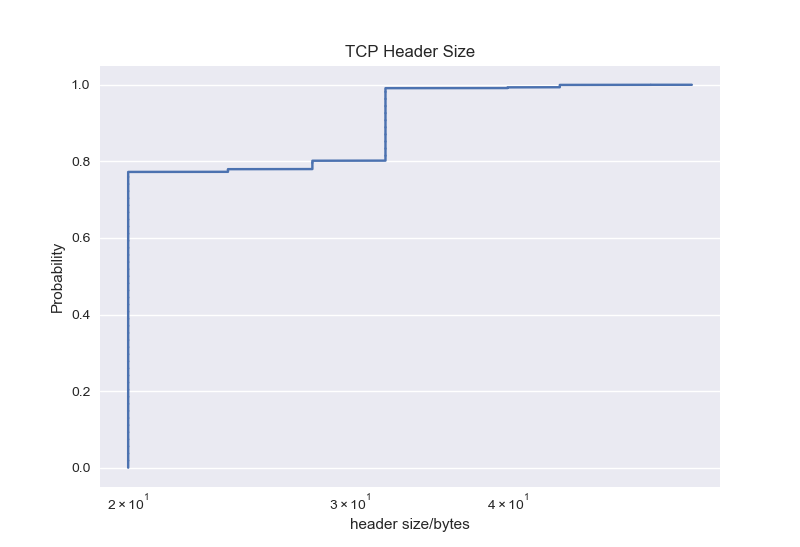
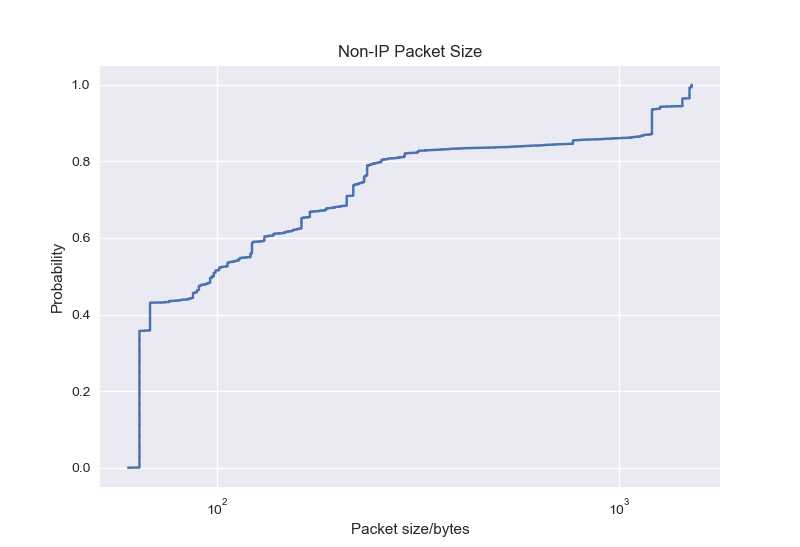
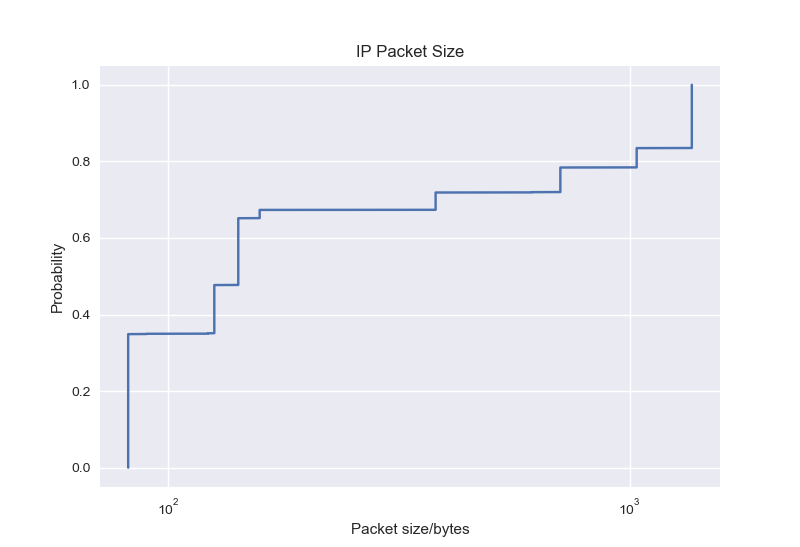
108276246bytes

Q3: **Analysis of the size of packets. For example, what difference you observe between TCP and UDP**

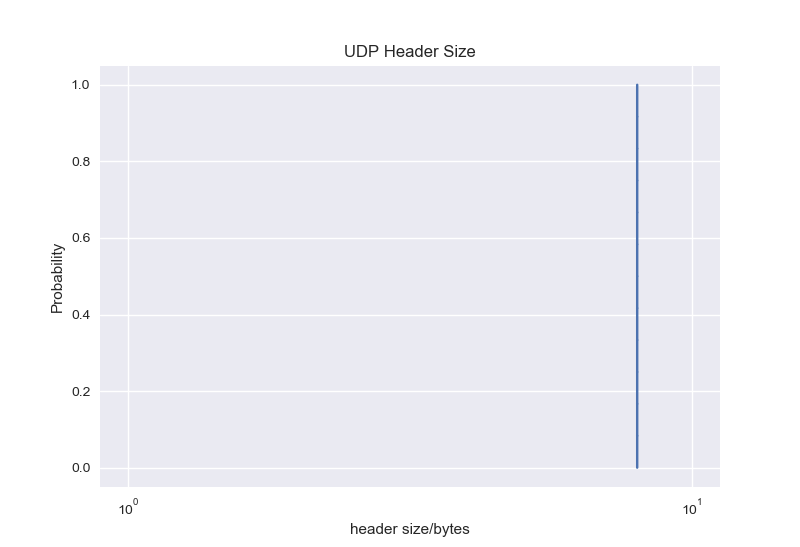
There is not much difference between TCP and UDP in terms of the size of packets and the only thing we can notice is that UDP has slightly more occurrences of higher packet size.

For IP and non-IP packets, the curves of the size of packets have a very similar shape apart from the curve Non-IP packet being smoother than the IP packets sizes in general. As for header sizes, the TCP header tends to be greater than 20 and mostly 30 to 40. The header size of UDP packets appears to be consistant (8 Bytes according to wireshark) and IP header size is mostly 20 with some exceptions





Since the header size for all UDP packets is all 8 bytes, the cdf of UDP header graph should not display anything other than a straight line due to all same values across the dataset.



Q4: **Flow type: Report the number of TCP and UDP packets relative to the total number of IP packets**

The total number of IP packets is 886944

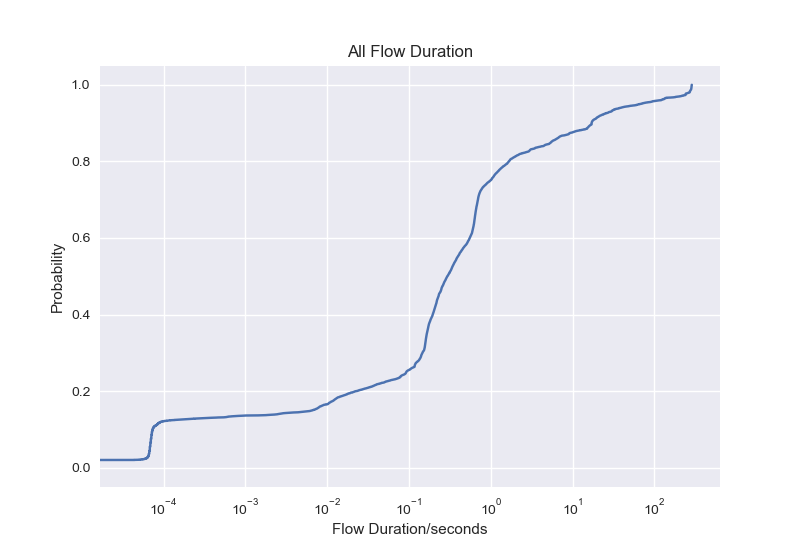
The total number of TCP packets is 533074

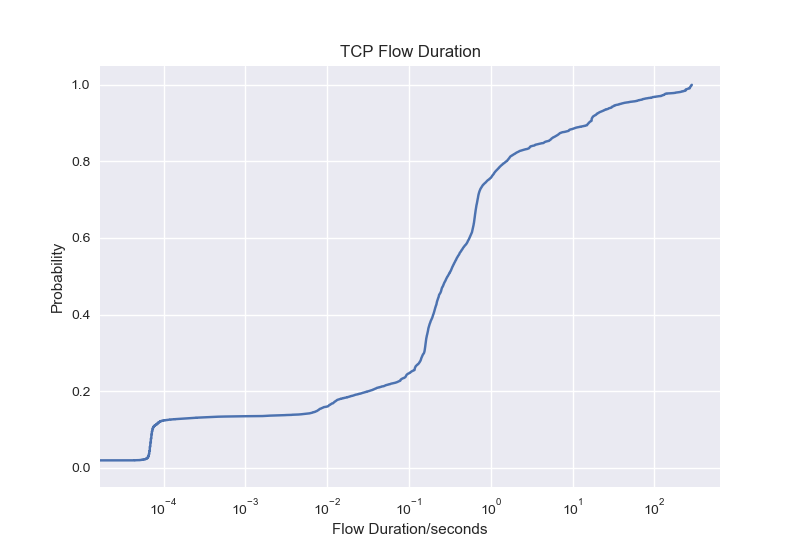
The total number of UDP packets is 113364

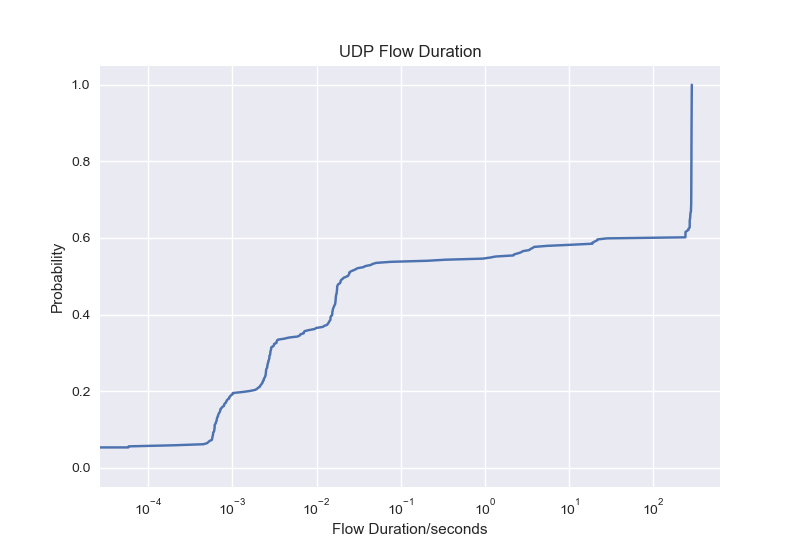
The remaining other types of packets is 240506

Q5: **Flow duration: Is there any difference between TCP and UDP flows?**

The TCP Flow duration graph indicates that there are a higher percentage of flows being shorter (such as the percentage of TCP flows under 100 seconds, compared to the percentage of UDP flows under 100 seconds). UDP flows generally last longer. Overall, since TCP flows account for most of the flows in this sample, this would dictate the shape of the All Flow Duration graph showing a similar story such as most flows being shorter than 100 seconds.

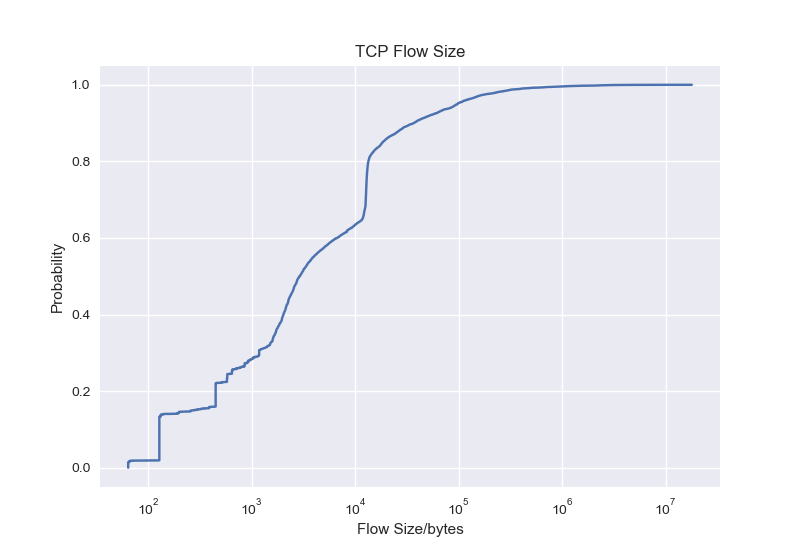


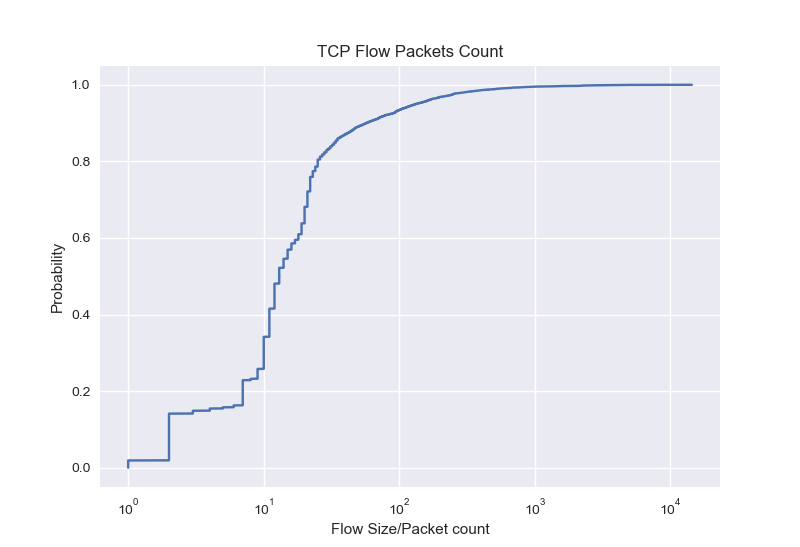
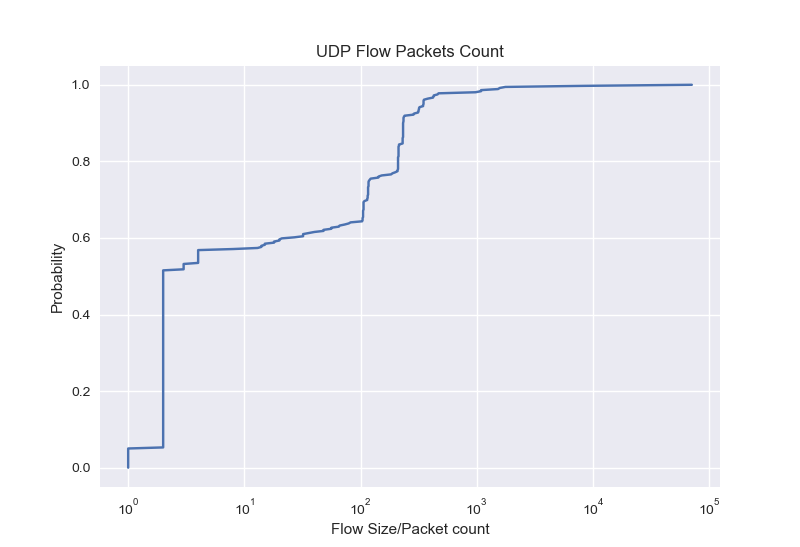
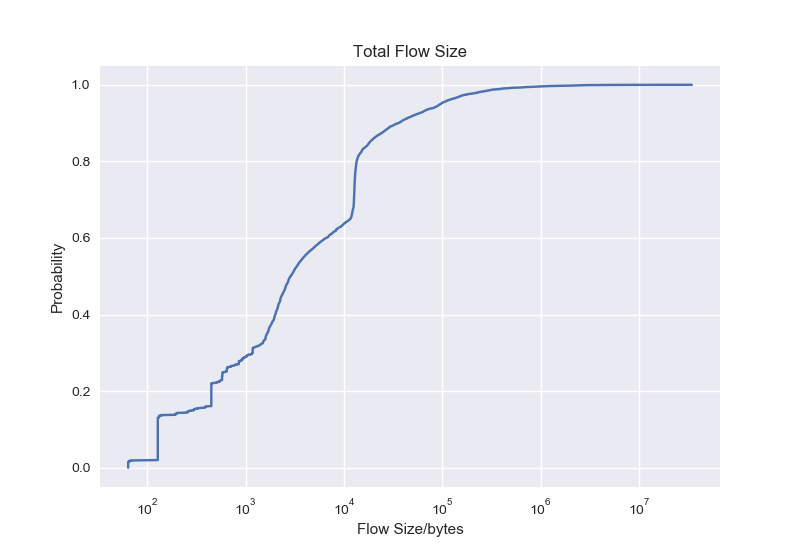
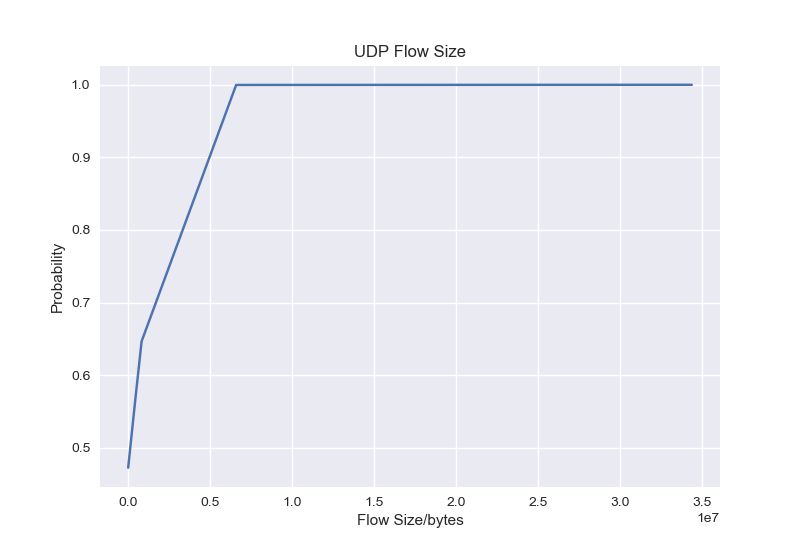


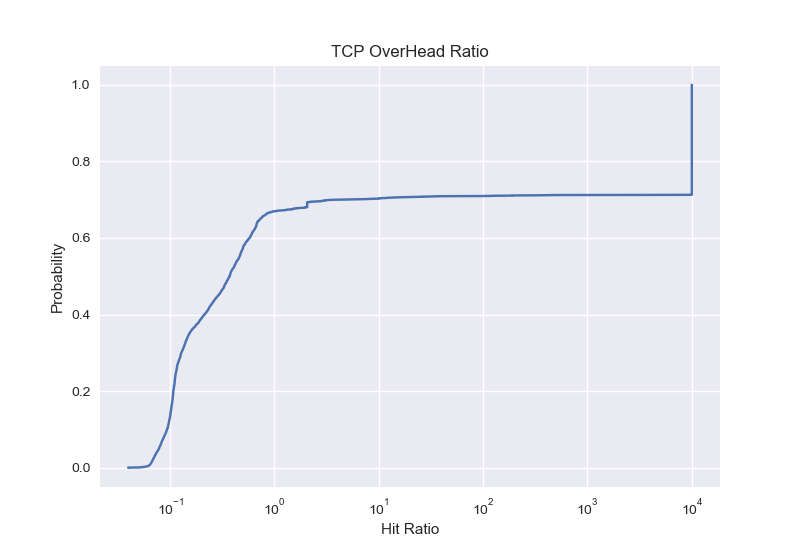
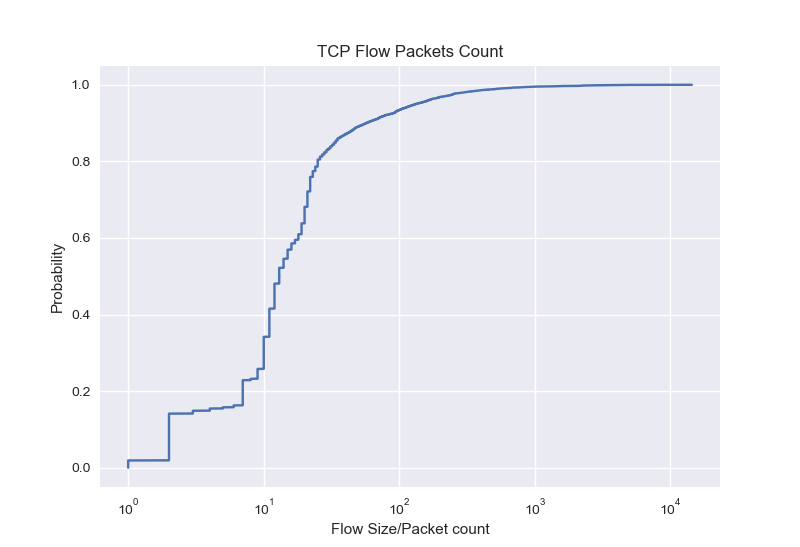


Q6: **Flow size: Do you see any difference between TCP and UDP? What about the case of using packet count vs byte sum? Draw the CDF of hit ratio. What can you say about TCP overhead base on this chart?**

Upon drawing the CDF graph of both UDP and TCP flow byte sum and flow size, there appears to be a correlation between byte sum and packet count where the flow size and packet count of TCP graphs as well as flow size and packet count UDP graphs have the same shape. Which shows that the more packet counts there are among a flow, the higher the byte sum would be. Generally, when comparing TCP and UDP, UDP would have a percentage of flows being larger both in terms of packet count and byte sum. Finally, the overhead ratio graph also shows that there is a large portion of data on the network that does not have any data showing that a lot of the data sent on the network is wasted on headers of packets.

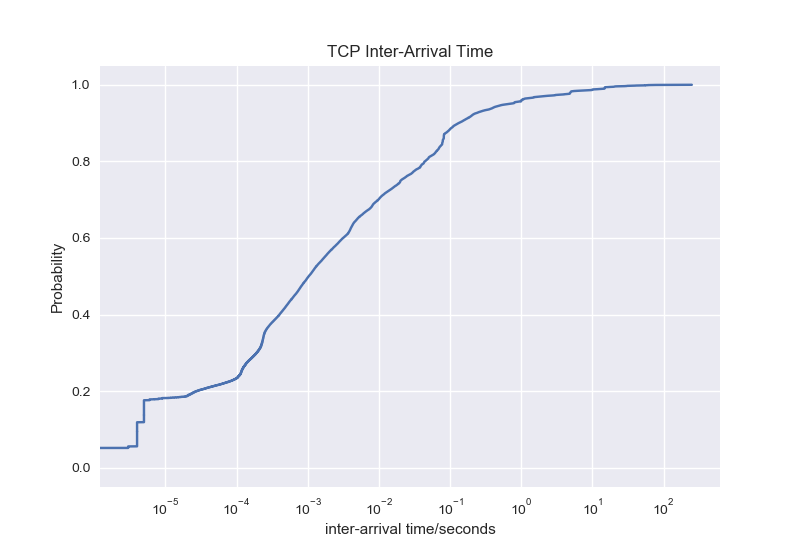
****

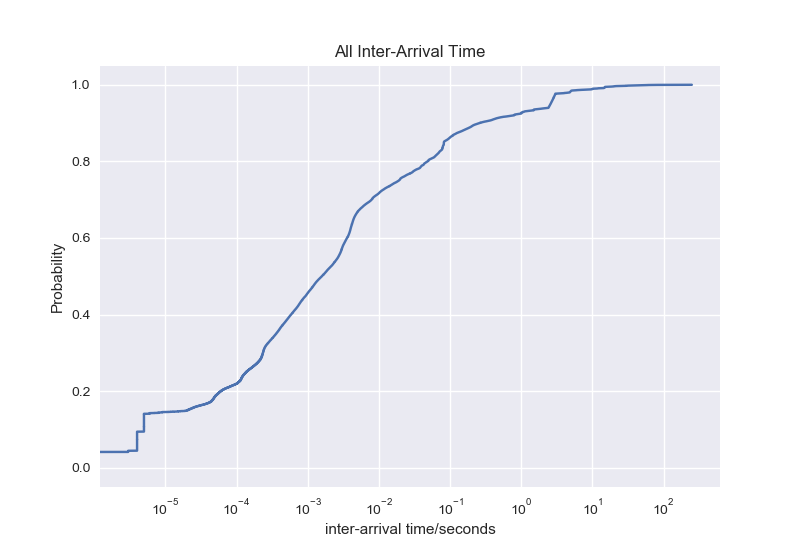
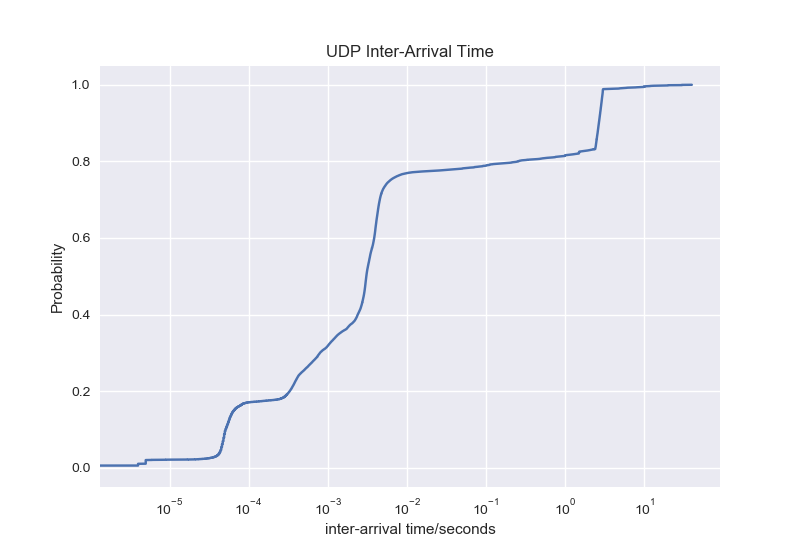
****

****

**Q7**:**Is  there any  specific interarrival  time  that appears more commonly? If yes, is it present in all flows, TCP flows, or UDP flows? Do you see any difference between TCP and UDP flows?**

For both TCP and UDP flows, inter-arrival time that is more than 100 seconds appears more commonly than others. It presents in all flows both TCP and UDP. There is no specific difference between TCP and UDP in general other than the increase from 10^(-4) to 10^(-1) and the shape of the CDF are the same and both stay within the same range.



****

**For TCP State: The following is the connection to each one of the classes.**

**The total number of flows: 12005**

**The total number of Request: 156**

**The total number of Reset: 4292**

**The total number of Finish: 25**

**The total number of Ongoing: 7532**

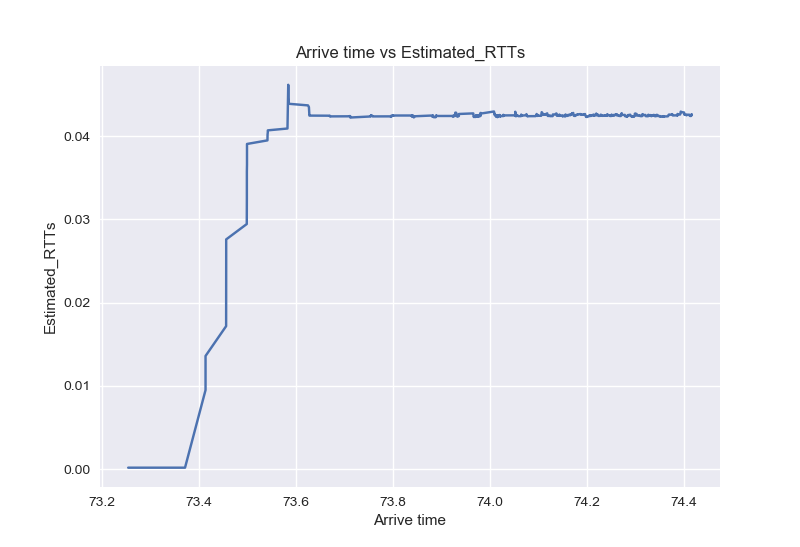
**The total number of Failed: 0**

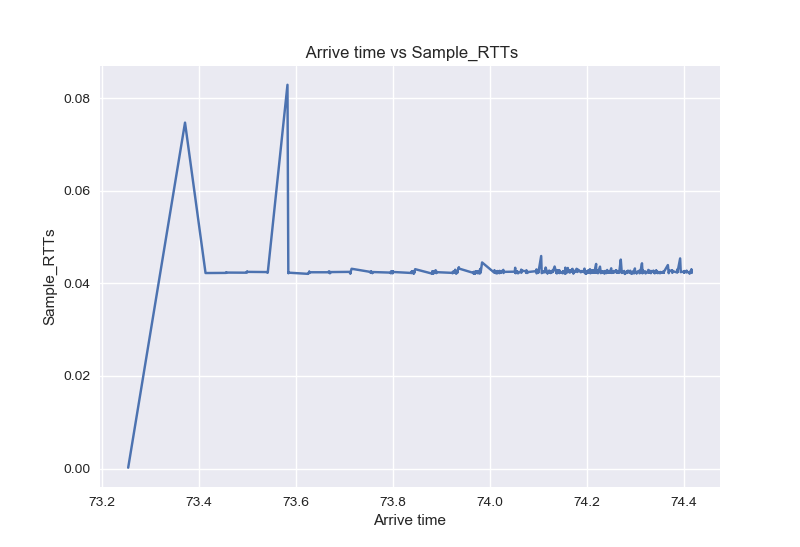
**RTT estimation:**

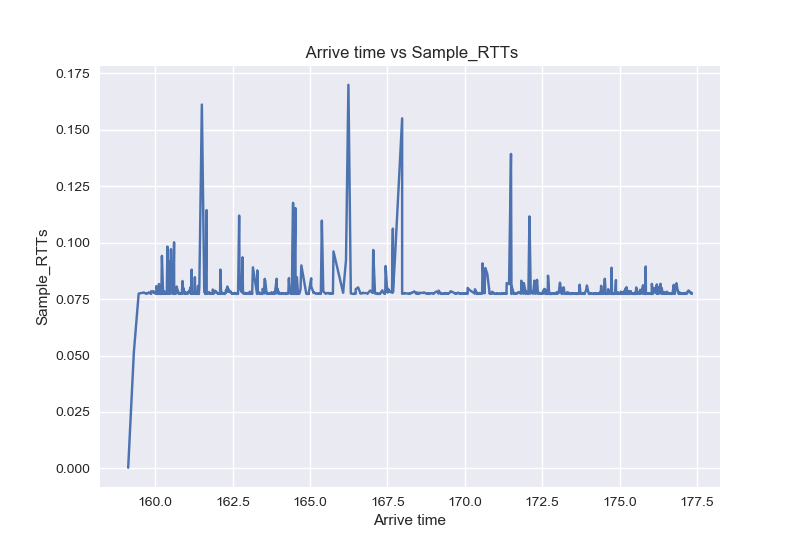
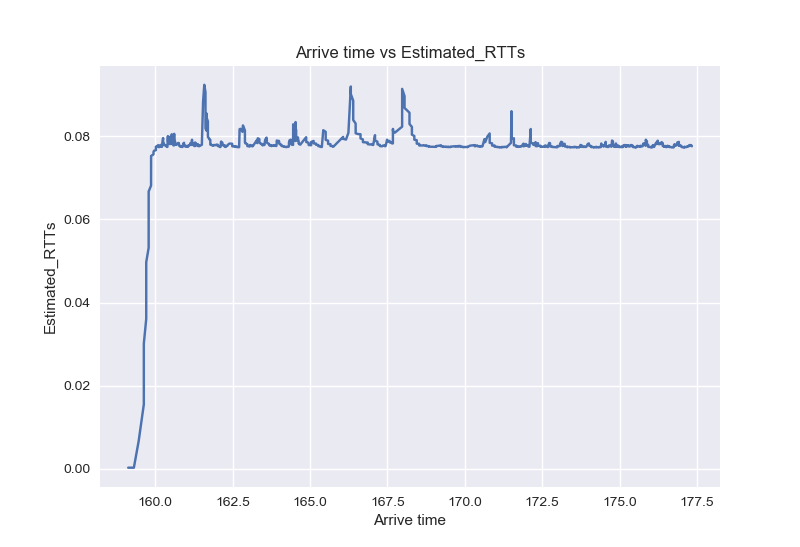
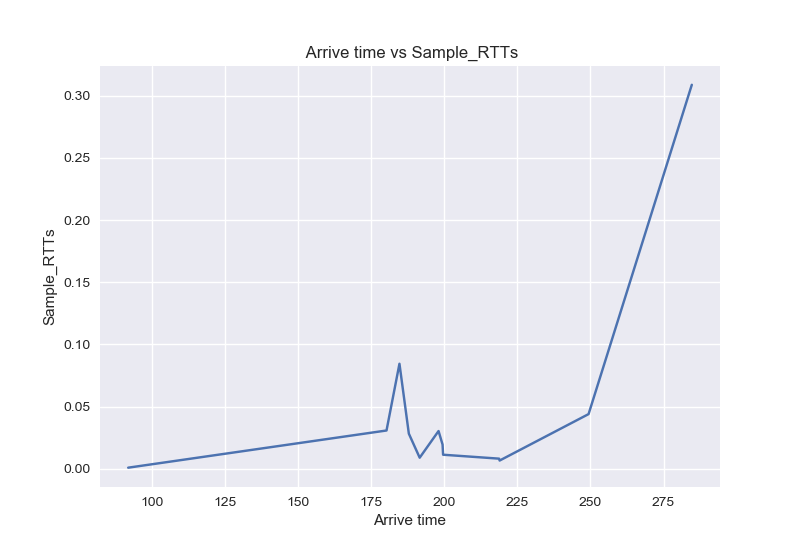
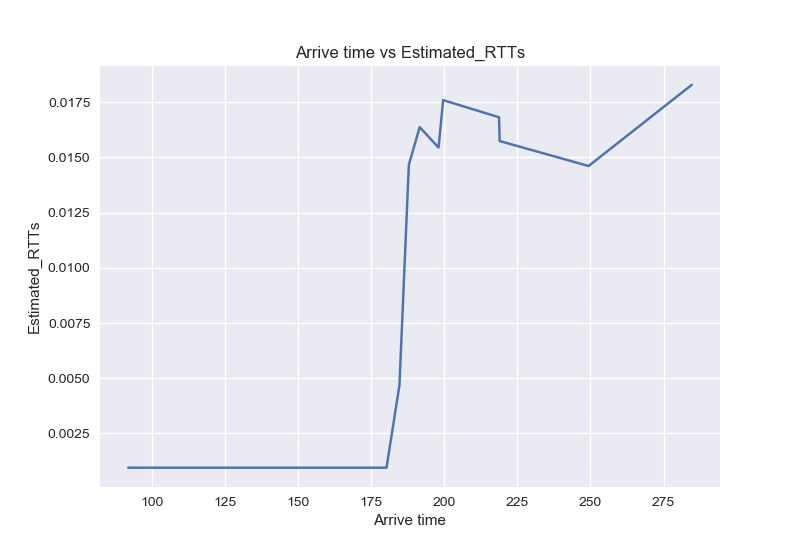
**Q8: For each of these flows, draw the sample RTT and estimated RTT as a function of time (as new packets arrive). Then analyze this chart. For example, is estimated RTT relatively stable? What about the sample RTTs? Do you see any increase or decrease in RTT? If yes, what can be the reason that the RTT is changing during the life of a connection?**

Across all top three flows in terms of packet number and total byte size for this analysis, the estimated RTT is relatively stable after 2-3 seconds range in general, whereas the sample RTTS tend to be stable after 1s but varies significantly within the 0.2-0.5 seconds. However, there are two exceptions for the descriptions for the above. The first exception for this would be the second largest packet number graph since it takes a while to increase and varies a lot in the process. The second exception is the longest TCP flows in terms of duration which includes significant large number of peaks. There is a huge increase of RTTs in the very first 0-3 seconds and an unstable change within that time range as well. Unlike, the previous ones, the top three longest TCP flows in terms of duration has shown ad extreme differences across all RTTs from the graph. However, it is relatively stable if you look closely to the change in RTTs and the difference stays within 0-2. Congestion control on the network is the reason that RTT is changing since it sets a limit on how many flow should flow.

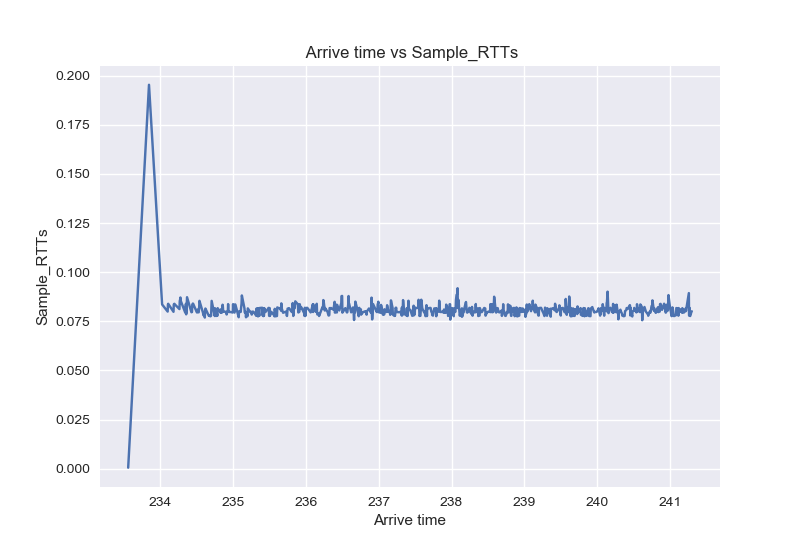
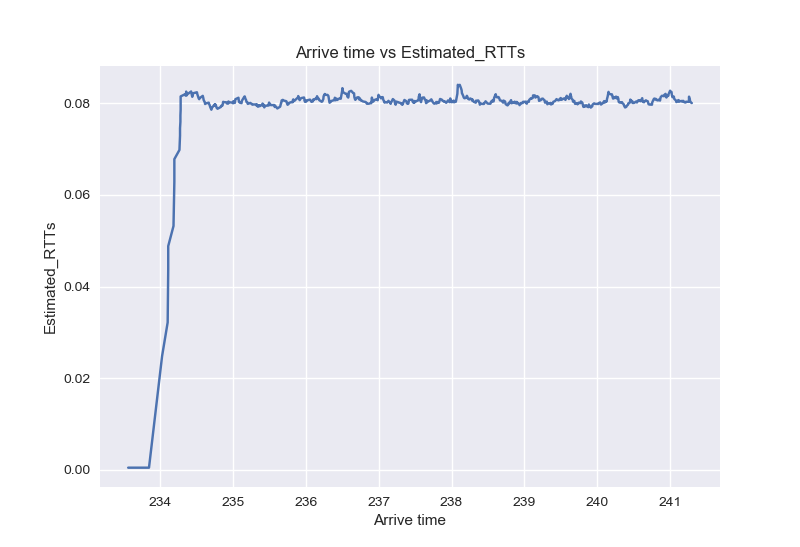
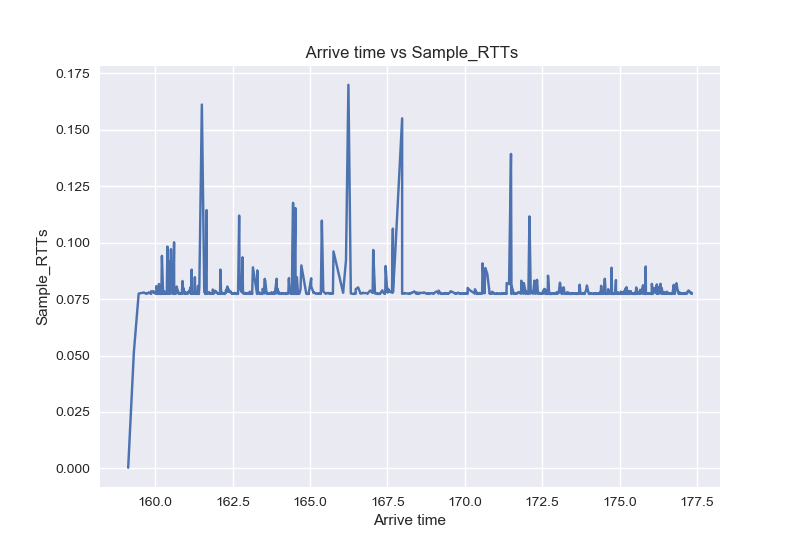
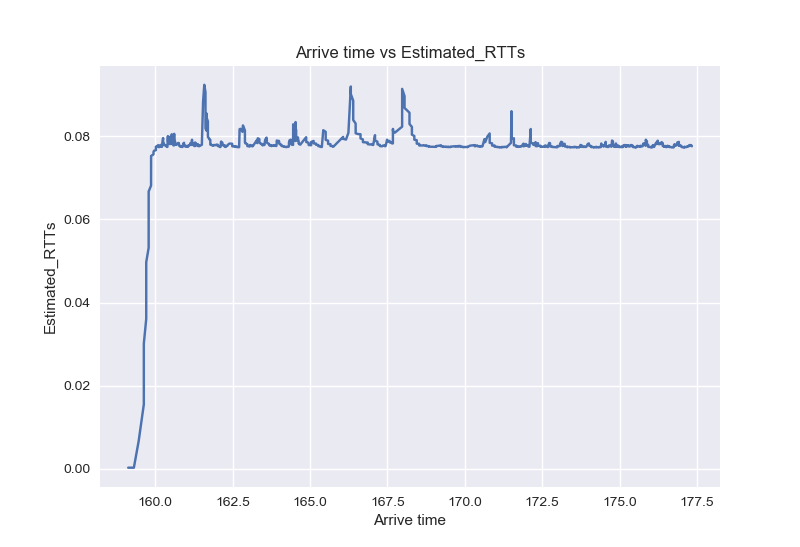
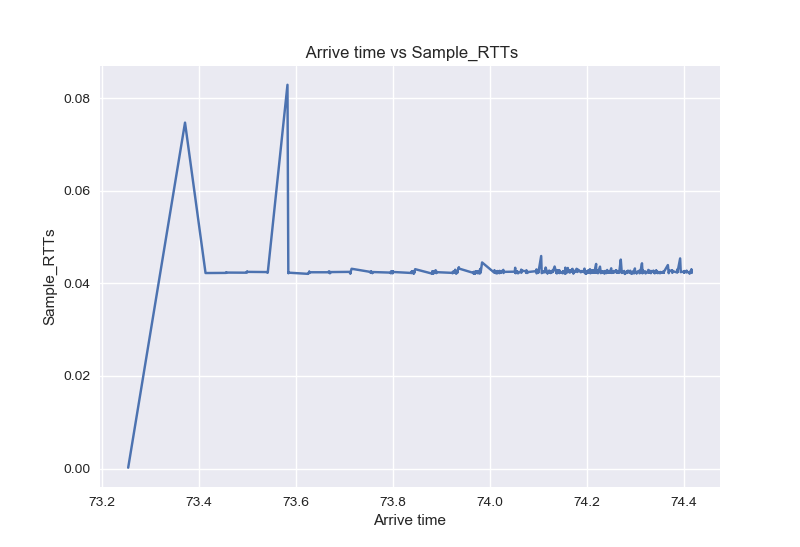
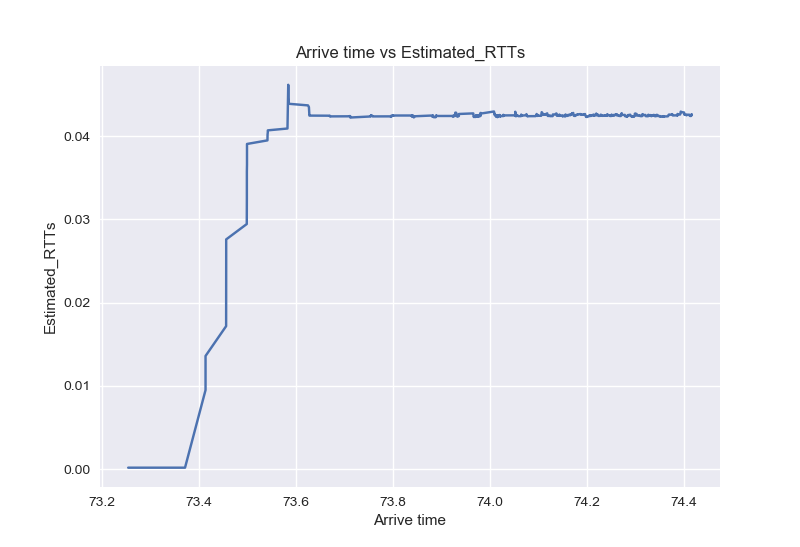
**Top 3 Longest TCP flows in terms of packet number:**

****

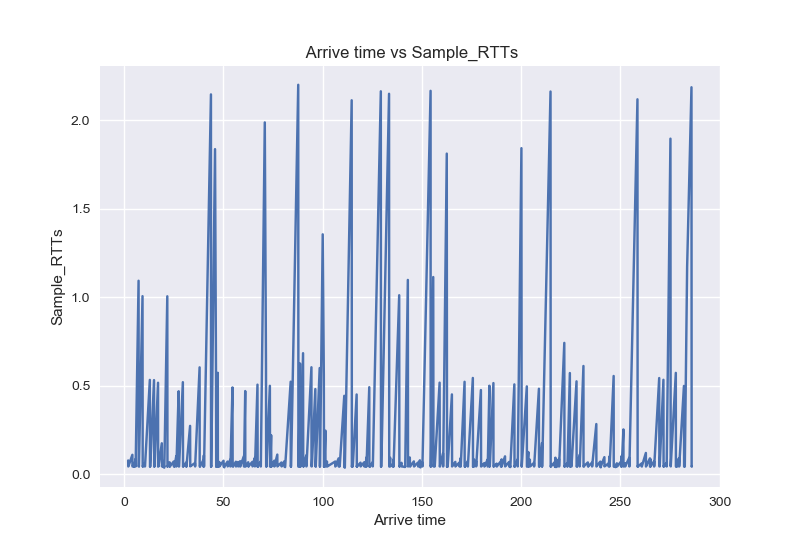
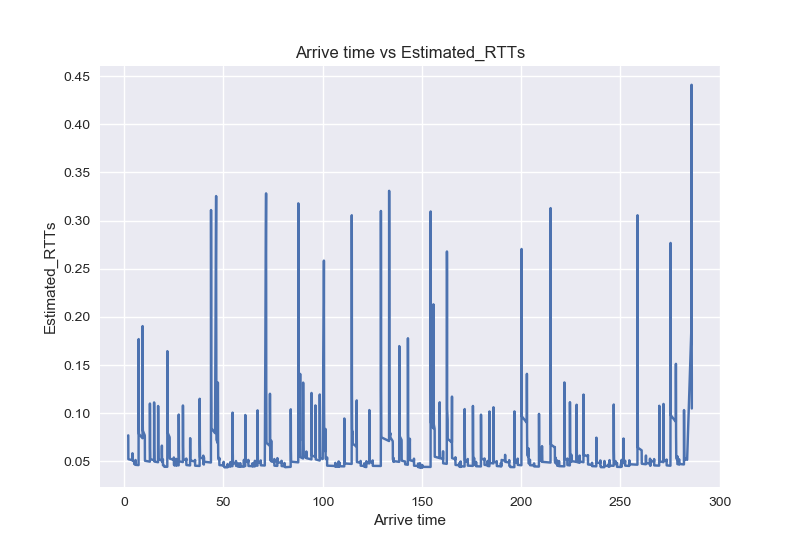
****

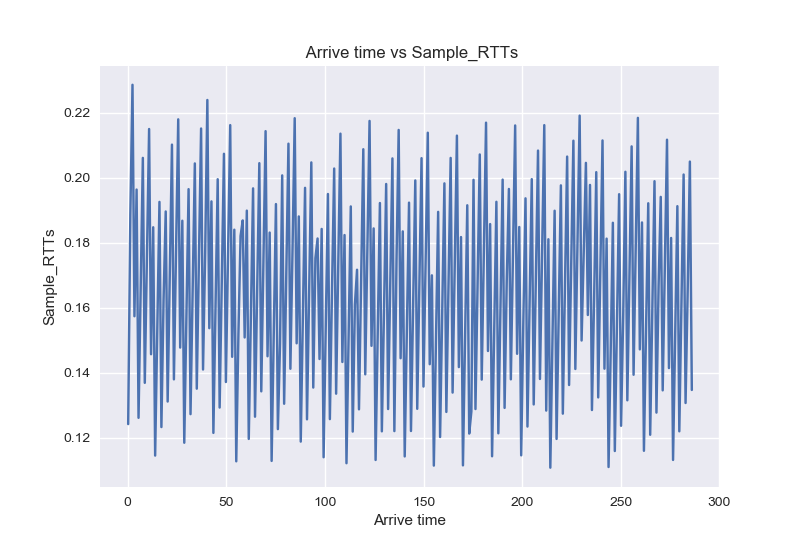
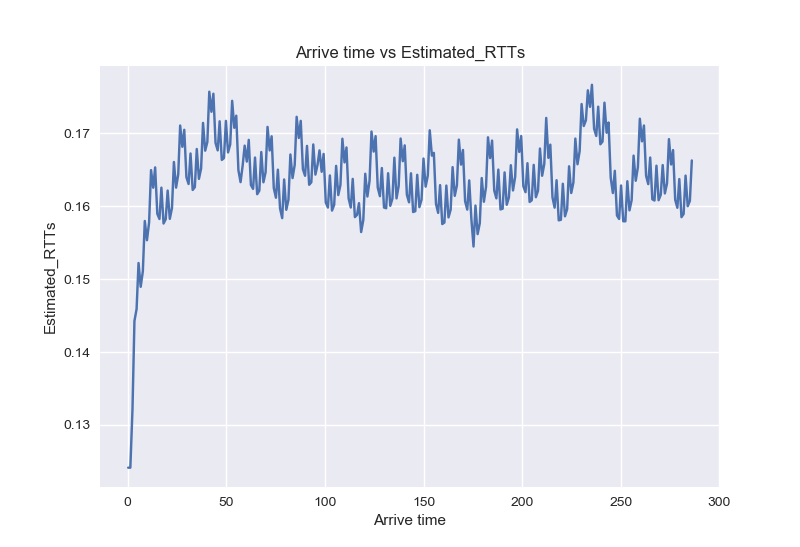
****

**Top 3 Longest TCP flows in terms of total byte size**

****

**Top 3 Longest TCP flows in terms of duration(Blank graph or no graph for the third):**

****

****

**Q9: For each of those flows, calculate the estimated RTT base on RFC 6298 and then use the median estimated RTT as the representative RTT to represent this flow. Now draw chart of representative RTTs of different flows as a function of their start time (to demonstrate how these RTTs change over time) for this pair of hosts. Do you see any specific pattern? Is the representative RTT changes over time? Is this change random or is it following a specific patter (e.g., first increase, and then decrease)? Explain what the reason of this change could be. Also compare the 3 different host pairs together. Is there any difference between them? If yes, what could be the reason.**

The pattern of the representative RTT vs start time graph is simply a linear function and the representative RTT changes over time. There is a specific pattern in our second and third graphs, but there is no other pattern than the huge increase and decrease within short amount of time. Based on our result, the second and the third graphs are similar in terms of peak, whereas the first one has a huge peak between 100 seconds and 150 seconds instead of 200 and 250 seconds. However, all these significant differences are relatively small and all stays within 0-0.00030. Distance between the host pairs is one of the main reasons why there is different in terms of the shape of representative RTT. In addition, the propagation delay is also a reason since it is the distance divided by the speed of light which in this case depends on the distance.

