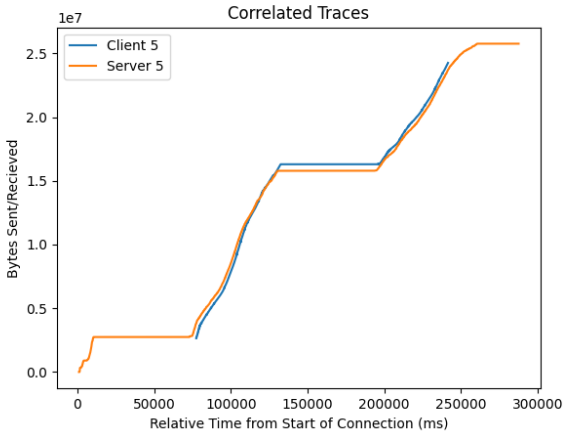

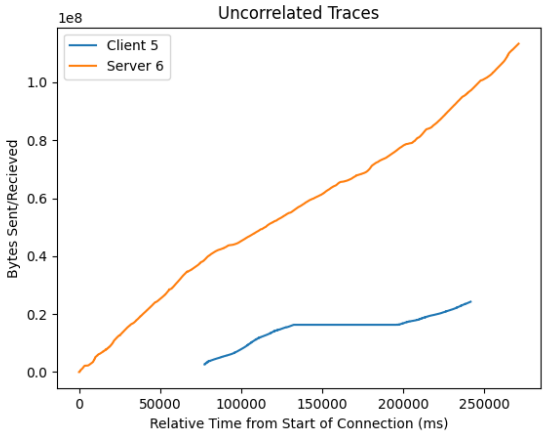

1. **If you use the above method to perform correlation, what parameters did you use? For example, start/end time, aggregation time interval, seq or ack number, etc.**

   In order to obtain the correlation, I first had to clean the data. First, I removed all of the network traffic that was not between the client/entry node and the server/exit node. Then I converted all the absolute timestamps to relative timestamps based on the start of the connection between the client/server and the respective node. Then I removed the bottom 5% and the top 5% of the data to remove some noise. I then aggregated the sequence and acknowledgement numbers based on the relative timestamp with 2 second buckets, and used the mean bytes sent/received in each bucket. Finally, I computed the Pearson coefficient, using the aggregated bytes sent/received and successfully correlated 19 of the 20 client/server pairs. The only incorrect correlation was Client 5 being correlated to Server 6 instead of Client 5. Note that the graphs mark the raw data and not the aggregates.

2. **If you did not use the above method, or you have tried other methods, please explain here what you did.**

   I used the method described above.

3. **Based on the two plots, what do you think contributed to the failure of correlating the client to the correct server?**

   Based on the two plots above I think that Client 5 was not correlated correctly because of the sustained pauses (the flat lines) on the network traffic and the extra data points on the Server 5 graph, caused the correlation to be skewed. Since I correlated the total bytes sent and received in each bucket using the Pearson correlation, which takes covariance of the two

variables and divides it by the product of their standard deviations, the sustained lag in traffic made Server 5 seem less correlated with Client 5. While, Server 6 had a lot more bytes sent than Client 5 had received, the trends of the data followed a similar pattern to Client 5, making it seem more correlated than it actually was.