

Subject 4: Latency and capacity estimation for a network connection from asymmetric measurements

Quality of Presentation

The document is very well organized, the author follows the different steps asked in a logical order without much to add. The data parsing, regression and interpretation are all present. The figures are very easy to understand and are placed at the right moment in the analysis to follow a flow of reading throughout the entire report.

However, at times some explanations could be simplified or vulgarised to target a wider range of people that may be interested in reading or learning more about network connections and asymmetric measurements.

Technical Quality

The raw ping data is correctly parsed and fetched onto a graph of ping over time clearly showing the outliers and taking care of them as expected.

The time evolution plots, while useful, may be showing some biasness as the author gets rid of ping times above 100 ms very early in the analysis which could compromise the observation of cases with real network instability.

Linear regressions are performed correctly and superposed on scatter plots for better visualisation. However, I see that the regression is done in milliseconds and bytes but the final values of latency and capacity are not converted to their correct units. This problem could lead to numerical irregularities and biasness later on. Nonetheless, Quantile regression is a good idea because of how variable the data is but suffers from the same conversion of units issues as the linear regression.

Reproducibility

The work is made using a computational document, which is very important in any analysis to invoke reproducibility. The code, figures and explanations are all combined into a single document making the workflow much smoother.

However, the possibility to download the data and code to use it on its own would have been appreciated.

Relevance to the Assignment

The required steps of the assignment are covered, from data parsing, to visualization, regression and comparison between the two network connections as expected from the MOOC.

The analysis is not extended beyond the required tasks, but it does show a good understanding of the methodological issues in hand.

Conclusion

Overall, the work demonstrates a solid understanding of the problem and follows the structure of the assignment very well. The main issue only boils down to the technical mistakes of unit handling which could impact the estimation of latency and capacity.