Senior Project Proposal

Department of Computer Science
Calvin University

Title: Netlog Mining for Speedtest-based Internet Performance Measurement
Author(s): Aishwarya Joshi, Anwesha Pradhananga
Date: 01/16/2024
Advisor: Prof. Rocky Chang

Background and Problem

Ensuring satisfactory and even guaranteed service-of-quality of the Internet is important for almost all business sectors nowadays and individual users. Governments also set up programs, such as Broadband America, to obtain the performance data from different ISPs and different regions, and make them available to the public. There have been various approaches to gauge the Internet performance in the past, broadly classified into passive measurement and active measurement. Most of the Internet performance platforms today are based on the active measurement approach which sends measurement probes from one or more vantage points to measure latency, throughput, packet loss, packet reordering, jitter, and many other derived metrics.

This project is about using various existing speedtest platforms, such as Ookla, Xfinity, Fast.com, Speedof.me, Cloudflare and M-Lab, to perform on-demand or continuous network measurement. The main advantage of this approach is bypassing the expensive task deploying and maintaining yet another measurement platform. These existing platforms, particularly Ookla, are growing rapidly, and they cover more geographical areas and ISPs. However, the main disadvantage is that they use different methodologies and configurations to obtain the measurement data. Moreover, most of the platforms, except for M-Lab, do not make their code available to the public. As a result, it is very difficult to compare the measurement results obtained from different platforms. It is also impossible to change their configurations according to different measurement objectives.

Brief Description of Solution Being Provided

To address these cross-speedtest-platform issues, a CAIDA (Center for Applied Internet Data Analysis) research team is recently developing a new web-based measurement toolkit for Reproducible Assessment of Broadband Internet Topology and Speed (RABBITS) which can utilize all the existing speedtest platforms to conduct configurable Internet measurement. The basic idea is to use a headless browser to mimic a manual execution of speedtest measurement. The toolkit also allows an experimenter to configure the measurement parameters, such as the number of measurement flows and packet size. The toolkit captures all the Netlog data between the toolkit and the remote speedtest server. A main challenge is to filter the relevant Netlog

events from a large amount of Netlog data for computing delay and throughput. This task is particularly challenging, because there is very little documentation available for understanding all the details in the data.

Research or Development Objective

The overall goal of this senior-year project is to develop a Netlog analytic engine that will accurately and efficiently glean from the vast amount of Netlog data obtained from the RABBITS toolkit for computing Internet performance metrics.

1. The engine is able to glean all the relevant network events from the six major speedtest platforms: Ookla, Xfinity, Fast.com, SpeedOf.Me, Cloudflare, and M-Lab. The first five platforms use the HTTP for measurement, whereas M-Lab deploys its own measurement protocol. The output is a small file consisting of those events in a standardized data format which can be used by the research community to analyze the data according to their objectives.
2. The network events are for computing three main performance metrics: round-trip latency, download throughput, and upload throughput. The round-trip latency is performed under two scenarios: loaded (when the measurement is ongoing) and unloaded (when there is no measurement). The raw data for latency can also be used to compute delay jitter which is important for video and audio streaming. The engine also supports different throughput computations, including those used by the speedtest platforms.
3. This engine will be seamlessly integrated with the RABBITS's tookit to perform large-scale measurement from different cloud platforms and network measurement platforms. Therefore, the engine must be able to produce the measurement metrics very efficiently in terms of the speed and use of memory. A console with data visualization will also be developed to monitor the continuous speedtest measurement from different vantage points and speedtest platforms.

Your Interest and Qualifications

We have taken some networking classes in Calvin and want to learn more about networking in general. This project gives us an opportunity to explore more about the layers of networking and also dive into the academic realm of things. We were also interested in doing a research project which introduces us to reading papers and working with the academics in the field. We have some experience with javascript which we will be using in this project to write netlog file analysis to get the roundtrip latency and download and upload throughput measurements. We started reading the papers and preparing ourselves a little bit before winter break so that we can get started this semester. We might also be putting in a few hours in the summer for this project and continue it back again in the Fall semester.

Collaboration with Advisor, Outside Experts and Users

Our weekly 1-2 hour meetings with Professor Rocky Chang, scheduled for Tuesdays are invaluable for refining our direction, troubleshooting challenges we've been facing to solve a particular problem or to understand more about what we are doing in the week. We have also been working closely with Dr. Ricky Mok, an assistant research scientist at CAIDA/UC San Diego. We have been meeting Dr. Mok every week for an hour to show him our progress and to learn about what the next steps could be. We also have the opportunity to hear what other grad students who are working on the same research have been working on. These interactions help us make steady progress toward meeting our objectives for the project.