

September 21, 2012

Via Electronic Filing

Marlene H. Dortch
Secretary
Federal Communications Commission
445 Twelfth Street, SW
Washington, DC 20554

Re: Notice of Oral *Ex Parte* Communications, CG Docket No. 09-158, CC Docket No. 98-170, WC Docket No. 04-36

On September 19th, 2012, representatives from ISPs and other organizations and companies met with members of the Federal Communications Commission to discuss issues associated with the Commission's 2012 broadband measurement and performance program.

During this meeting representatives from M-Lab discussed recently completed analysis of data collected by SamKnows using the M-Lab platform during March 2012. In brief, anomalies were discovered in a limited amount of data collected over the course of a few days in March 2012 on 3 M-Lab servers. After executing an NDA to gain access to the March data, researchers from MIT and the Open Technology Institute conducted thorough analysis of the impacted data, and prepared a 6-page report¹ explicating the conclusions they reached via analysis, and including the analytic scripts. This rigor and openness ensures that anyone curious can retrace exactly how these researchers arrived at their conclusions, understand the character of the data precisely, and if needed ask clear, technical questions that can result in responsible research that benefits the field at large.

Based on this careful analysis, M-Lab drafted a verifiable, scientifically sound disclaimer statement, intended to enlighten researchers accessing the data:

Between March 23 and March 31, 2012 anomalies were detected on M-Lab servers mlab1.lga02, mlab3.lga02 and mlab3.lax01. These anomalies impacted 4.1% of the platform, as measured between March 1st and March 31, 2012. Of 20.5 million tests run in March, 1.6 million show signs of impact.

Open analysis of the data out of MIT and the Open Technology Institute shows that for the month of March, 7.8% of data in total were impacted. This is deemed to be a statistically significant portion of the total data. Looking specifically at data collected from the impacted servers, 41.7% was affected. The impact can be characterized as a progressively increasing delay

¹ This report, along with the analysis scripts, is included as an attachment to this filing.

in server process scheduling. The delay occasionally resulted in socket buffers filling before processes could empty them, leading to “lost packets” on UDP connections and lower throughput on TCP connections. For throughput based tests, speeds under 20Mbps were less likely to be affected than speeds above 40Mbps. Tests of upload, download, and video streaming show the least impact, as low as 25.39%, while tests of UDP latency and packet loss show impact as high as 82.10% of tests to a single server.

Researchers accessing this data should feel free to use data from unimpacted servers without concern. Use of data from impacted servers should proceed with caution, cognizant of the specific nature of the impact, and avoid characterizing performance based on impacted data. The longer analysis provides technical details on proper and improper uses of the released data.

Analysis can be found [here] (link to attached conclusions).

We submit this statement, along with the accompanying report, as a guide that can technically and specifically assist researchers accessing the raw March 2012 data. In addition, we offer it as a precedent that can guide subsequent attempts of this nature. It is important that any claim characterizing data be based on open, sound, and wholly replicable scientific analysis.

Respectfully submitted,
/s/ Thomas Gideon

Thomas Gideon
M-Lab/Open Technology Institute
New America Foundation
1899 L Street NW Suite 400
Washington, DC 20036

Meredith Whittaker
M-Lab/Google
Google New York
76 9th Ave
New York, NY 10011