Background

The following consists of conclusions derived from an open analysis of the data collected during March by SamKnows and Measurement Lab as a part of the FCC's Measuring Broadband America program. The intent of this work is to explicate and describe the impact of server anomalies on the data collected during this period, and instruct researchers accessing the data on its proper use, given these anomalies. The analysis scripts used to achieve these results are included.

Acknowledgements and authors

The analysis and conclusions detailed in this report were conducted by:

- Steven Bauer, MIT
- Stephen Soltesz, Open Technology Institute
- William Hawkins, Open Technology Institute

Thanks to Samknows for providing their analysis of the March anomaly, for providing access to the raw data from March, and to the FCC for supporting this process.

Introduction to M-lab

Measurement Lab (M-Lab) is an open measurement platform that collects globally comparable, scientifically verifiable data on broadband performance. M-Lab has operated continuously since January 2009, growing steadily since inception. Currently, M-Lab hosts 97 servers at 32 sites around the world. This number is subject to change at any moment, as new sites are brought online. In this time, M-Lab has collected >600 TB of data, supported numerous academic papers, and informed policy within the United States, Greece, New Zealand, Guam and other global locations. M-Lab continues to grow, gaining partners, servers, tests, and research collaborators.

Incident Timeline

In over three years of operating M-Lab, the anomaly discovered in March, 2012 was the first of its kind. Upon discovery, Thomas Gideon of OTI immediately solicited the feedback of operating partners at PlanetLab and Google. In under eight days, the source of the anomaly was identified and the servers restored. Within one more week, new monitors were in place to detect and prevent the same event from recurring.

As a result of this event, the new, active monitors now run continuously across the platform. The new monitors augment the already extensive set of availability, connectivity, responsiveness, and other health checks run on M-Lab servers in order to detect anomalous events before they impact measurements.

Data Analysis

At M-Lab site LGA02, servers mlab1 and mlab3 were identified as affected. And, at site LAX01, server mlab3 was identified as affected. However, not all tests to these servers were impacted equally. The following outlines a method for discriminating which tests were affected on the identified servers and which were not.

Data

The data that inform the method and conclusions of this report come from these sources:

- Historical Monitoring by OTI (smokeping database and figures)
- Historical Monitoring by PlanetLab (restart history)
- SamKnows database of March measurements

Historical Monitoring by OTI: Smokeping

One of the tools OTI uses to monitor Measurement Lab is Smokeping. Smokeping periodically pings every server in the platform and records latency and loss statistics over time. This service was running during the incident covered in this report, and we have detailed measurements for the servers covered in this report.

Historical Monitoring by PlanetLab: Restart History

M-Lab can be conceptualized as a “private subset” of PlanetLab. PlanetLab Central at Princeton University helps with M-Lab operations and maintenance. The restart history of servers at sites LGA02 and LAX01 is provided by the PlanetLab monitoring service. This service runs across the entire PlanetLab platform. This service confirms that mlab1.lga02, mlab3.lga02 and mlab3.lax01 were restarted on March 31, 2012.

SamKnows Database of March Measurements

The dataset provided by SamKnows consists of measurements performed from 2012-03-01 to 2012-03-31. Our method of analysis is described in more detail below.

Method

The following describes the method used to filter individual tests in the SamKnows database. In general, individual tests that pass the filter criteria are regarded as unaffected, while those that do not pass the filter criteria are regarded as affected.

Each M-Lab site is equipped with three servers. During the period of March, all three servers in LGA02 were operational, and two were affected, and in LAX01 two servers were offline due to maintenance and the third was affected.

Based on the Smokeping measurements, hosts mlab1.lga02, mlab3.lga02, and mlab3.lax01 exhibit dramatic increases in latency through March. Hosts mlab2.lga02 and mlab2.nuq01 do not exhibit increased latency at any point in March.

The absence of latency increases to mlab2.lga02 and mlab2.nuq01 is an indication that these servers were not affected by the same factors as mlab1.lga02, mlab3.lga02, and mlab3.lax01. In their analysis of server behavior, SamKnows identified mlab2.lga02 and mlab2.nuq01 as unaffected. Therefore, as a result of the signal from Smokeping latency and the SamKnows analysis we too take mlab2.lga02 and mlab2.nuq01 to be unaffected and a valid a point of comparison for tests run against 1) mlab3.lga02 and mlab1.lga02, and 2) mlab3.lax01,

respectively.

LGA02

At LGA02, mlab2 was unaffected in March. All three servers at LGA02 are physically close to one another. Therefore, measurements to mlab2 characterize the expected performance from the affected servers, mlab1 and mlab3. The analysis here uses the average performance to mlab2 as a benchmark for performance to mlab1 and mlab3. Specifically, if a throughput measurement to mlab1 or mlab3, is greater than the *average-stddev* of measurements to mlab2 for the day of the test, then the test passes. Otherwise, the value is less than average-stddev, and the test fails. Tests run against mlab1 and mlab3 without corresponding package tests against mlab2 are not considered, so by default fail. For tests where smaller is better, such as latency and loss, measurements that are less than the *average+stddev* of measurements to mlab2 pass and otherwise fail.

LAX01

At LAX01, mlab3 was affected in March. The other servers physically close to mlab3 were offline. Therefore, the exact method describe for LGA02 is not applicable. Instead, we use a comparison server at a nearby site, mlab2 at NUQ01. In all other respects, the method is the same.

The table below outlines the data sets evaluated using the method described above along with the database table and column names used.

Test	Database Table Name	Value Compared
Download speed	curr_httpgetmt	bytes_sec
Upload speed	curr_httppostmt	bytes_sec
UDP latency	curr_udplatency	rtt_avg
UDP packet loss	curr_udplatency	failures/(successes+failures)
Video streaming	curr_videostream	latency
Voice over IP	curr_udpjitter	latency
ICMP latency	curr_ping	rtt_avg

Limitations of Method

Network measurements are inherently influenced by multiple dimensions. In this analysis, we use a single dimension to filter measurements. While a single dimension is appropriate for tests such as aggregate transfer rates, using a single metric to evaluate the quality of a Voice over IP connection may not be ideal. As well, we acknowledge that the geographical separation between LAX01 and NUQ01 may introduce error unrelated to the platform or measurements to the platform.

Results

Test	Primary measure(s)	Percentage of tests to server affected	Percentage of all tests affected
Download speed	Throughput in Megabits per second (Mbps) utilizing three concurrent TCP connections	mlab1.lga02: 28.43% mlab3.lga02: 29.37% mlab3.lax01: 41.25%	mlab1.lga02: 1.46% mlab3.lga02: 1.48% mlab3.lax01: 3.55%
Upload speed	Throughput in Mbps utilizing three concurrent TCP connections	mlab1.lga02: 25.39% mlab3.lga02: 26.69% mlab3.lax01: 34.35%	mlab1.lga02: 1.29% mlab3.lga02: 1.33% mlab3.lax01: 3.04%
Web browsing	Total time to fetch a page and all of its resources from a popular website	N/A	
UDP latency	Average round trip time of a series of randomly transmitted UDP packets distributed over a long time frame	mlab1.lga02: 79.34% mlab3.lga02: 81.81% mlab3.lax01: 49.37%	mlab1.lga02: 4.18% mlab3.lga02: 4.11% mlab3.lax01: 4.26%
UDP packet loss	Fraction of UDP packets lost from UDP latency test	mlab1.lga02: 74.63% mlab3.lga02: 75.29% mlab3.lax01: 82.10%	mlab1.lga02: 3.93% mlab3.lga02: 3.79% mlab3.lax01: 7.08%
Video streaming	Initial time to buffer, number of buffer under-runs and total time for buffer delays	mlab1.lga02: 30.31% mlab3.lga02: 30.40% mlab3.lax01: 35.70%	mlab1.lga02: 1.59% mlab3.lga02: 1.58% mlab3.lax01: 3.02%
Voice over IP	Upstream packet loss, downstream packet loss, upstream jitter, downstream jitter, round trip latency	mlab1.lga02: 40.16% mlab3.lga02: 41.03% mlab3.lax01: 41.89%	mlab1.lga02: 2.07% mlab3.lga02: 2.05% mlab3.lax01: 3.58%

DNS resolution	Time taken for the ISP's recursive DNS resolver to return an A record for a popular website domain name	N/A	
DNS failures	Percentage of DNS requests performed in the DNS resolution test that failed	N/A	
ICMP latency	Round trip time of five regularly spaced ICMP packets	mlab1.lga02: 33.15% mlab3.lga02: 33.78% mlab3.lax01: 35.52%	mlab1.lga02: 1.67% mlab3.lga02: 1.64% mlab3.lax01: 3.10%
ICMP packet loss**	Percentage of packets lost in the ICMP latency test	N/A	
Latency under load	Average round trip time for a series of regularly spaced UDP packets sent during downstream/ upstream sustained tests	Not analyzed.	
Availability	Total time the connection was deemed unavailable for any purpose, which could include a network fault or unavailability of a measurement point	0% (unaffected)	0% (unaffected)

** The method of analysis only chose tests where all pings were successful.

Conclusions

An anomaly affected the Measurement Lab platform during March 2012. This anomaly impacted 3 of 73 servers during March, 2012. Once Measurement Lab operations was aware of the

problem, we were able to identify the root cause and restore the affected servers within eight days. Within one more week, active monitors were in place to detect and prevent the same event from recurring.

Since the time of this anomaly, Measurement Lab has continued to expand the platform, as of September 2012 with 97 servers online or in deployment. As well, every server is monitored continuously for availability, connectivity, responsiveness, and many other health checks in order to detect anomalous events before they impact measurements.

For more information visit http://www.measurementlab.net

Scripts used in analysis:

[code repository to be linked from live page]