MLAB

Network Science at Scale

Meredith Whittaker
Tiziana Refice
Dominic Hamon

MIT, Nov 18 2012

What and who is M-Lab?

Measurement Lab is a collaborative, researcher-driven platform that empowers **Internet users**, **researchers**, and **regulators** with freely accessible open data about network performance.

At every level, data are necessary

For researchers & data analysts

- **Replicable science at scale**

For policy makers

- **Data based policy**

For Internet users

- **Scientifically-founded answers** to important questions

M-Lab's founding principle: Openness

Openness means making room for real science

Independent peer-reviews; Reproduction of existing results; Building on top of existing research, instead of reinventing the wheel over and over and over...; **Long-term validity** and **credibility**

How does M-Lab do this?

- **Open source**, publicly documented **server platform**
- **Open source** **experiments** built by researchers
- **Openly available**, freely accessible **data**

Open, globally-distributed platform

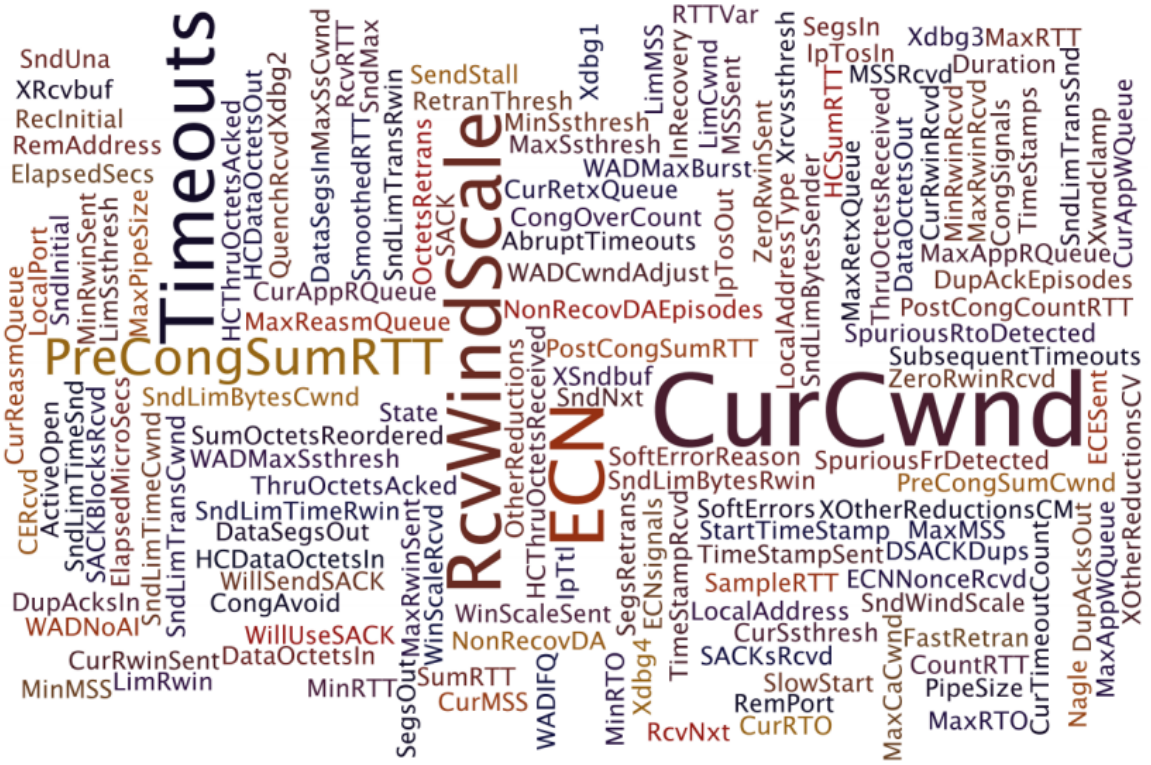
Open, globally-distributed platform

- **Globally-consistent**
- **IPv6 and IPv4**
- **PlanetLab based**

- **Dedicated resources to every experiment**
  - 1 Gb of dedicated upstream
  - One dedicated public IP addresses per experiment
- **Full access** to dedicated VM per experiment per server
- **Web100 instrumentation**
  - Provides rich information about measurements

Web100 - RFC4898

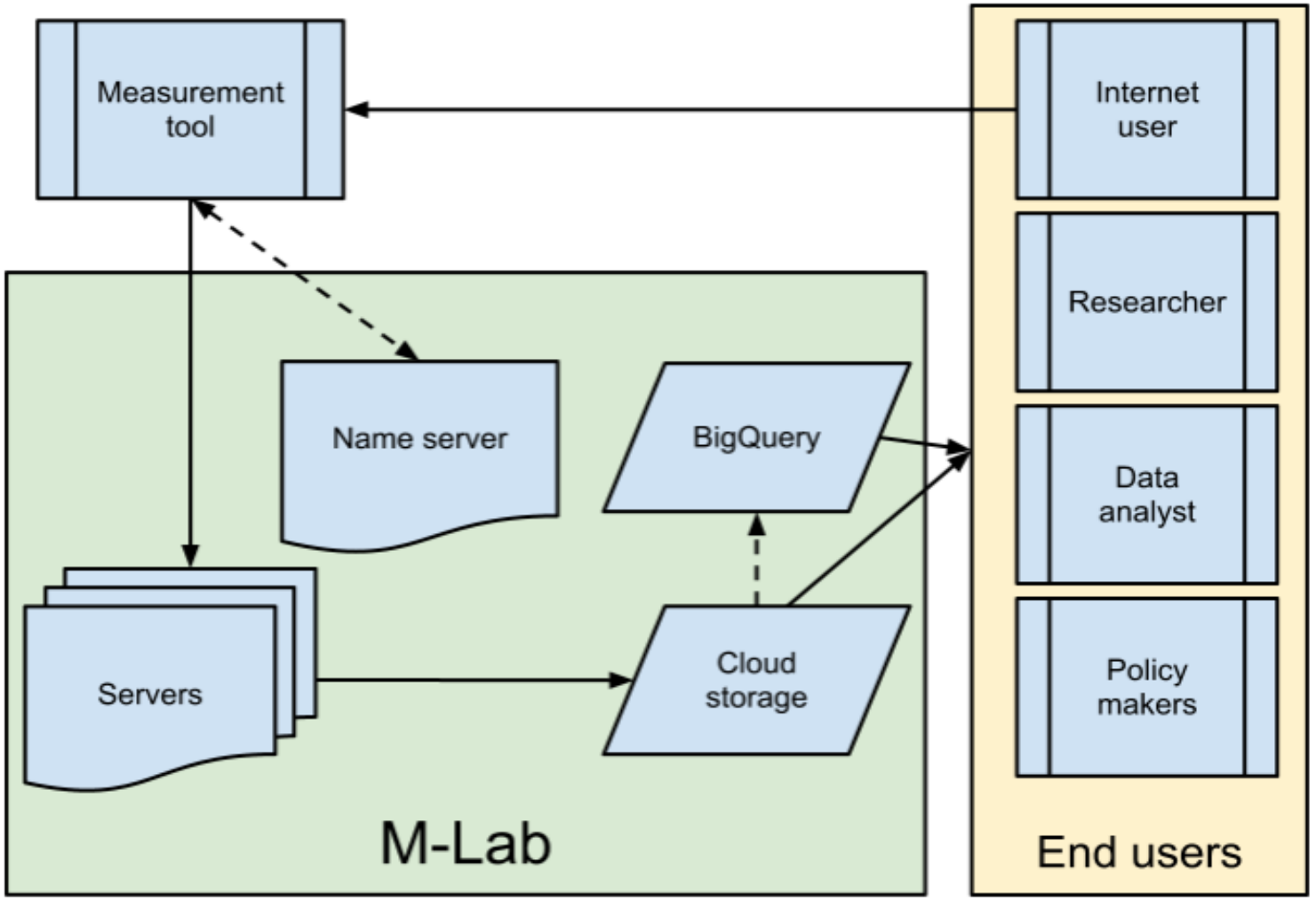
- Kernel-level instrumentation in the Linux TCP/IP stack
- User-level tools for accessing the kernel instrumentation
- 159 variables that fully describe a TCP state



Requirements for tools to run on M-Lab

- Open source
- User-facing
- Client-initiated
- Active measurement
- No personally identifiable information
- Open data

The M-Lab platform



LOTS of open data

- Public Domain (CC-Zero)

- Multiple ways to access the data, via Web or APIs
  - Raw format, as collected on the M-Lab servers
    - non aggregated, non anonymized
    - 630 TBytes since Jan 2010
    - 200k tests per day
  - SQL-line interface
    - 700B rows
    - MaxMind geolocation

Open data promotes research

- H. Asghari, M. van Eeten, M. Mueller. **Unraveling the Economic and Political Drivers of Deep Packet Inspection.** GigaNet 7th Annual Symposium. 2012
- E. Katz-Bassett, C. Scott, D. Choffnes, I. Cunha, V. Valancius, N. Feamster, H. Madhyastha, T. Anderson, A. Krishnamurthy. **LIFEGUARD: Practical Repair of Persistent Route Failures.** ACM SIGCOMM 2012.
- S. Basso, M. Meo, A. Servetti, J. C. De Martin. **Estimating Packet Loss Rate in the Access Through Application-Level Measurements**. ACM SIGCOMM W-MUST 2012.
- B. Lehr, S. Bauer, D. Clark. **Measuring Internet Performance when Broadband is the New PSTN**. MIT technical report. 2012.
- P. Bardowski, J. Klink, M. J. Podolska, T. Uhl. **Broadband Access to the Internet via Mobile Interfaces**. IEEE WMCNT 2012
- P. Kanuparthy, C. Dovrolis. **ShaperProbe: End-to-end Detection of ISP Traffic Shaping using Active Methods**. IMC 2011
- M. L. Mueller, H. Asghari. **Deep Packet Inspection and Bandwidth Management: Battles over BitTorrent in Canada and the United States**. TPRC 2011.
- S. Sundaresan, W. Donato, N. Feamster, R. Teixeira, S. Crawford, A. Pescape. **Broadband Internet Performance: A View From the Gateway**. SIGCOMM 2011.
- M. Dischinger, M. Marcon, S. Guha, K. P. Gummadi, R. Mahajan, S. Saroiu. **Glasnost: Enabling End Users to Detect Traffic Differentiation**. NSDI 2010.
- E. Katz-Bassett, H. V. Madhyastha, V. K. Adhikari, C. Scott, J. Sherry, P. van Wesep, T. Anderson, A. Krishnamurthy. **Reverse Traceroute.** NSDI 2010.
- S. Bauer, D. Clark, W. Lehr. **Understanding Broadband Speed Measurements**. MIT technical report 2010.
- C. Dovrolis, K. Gummadi, A. Kuzmanovic, S. D. Meinrath. **Measurement Lab: Overview and an Invitation to the Research Community**. SIGCOMM CCR 2010.
- M. Dischinger, A. Mislove, A. Haeberlen, K. P. Gummadi. **Detecting BitTorrent Blocking**. IMC 2008.
- M. Mathis, J. Heffner, P. O'Neil, P. Siemsen, **Pathdiag: Automated TCP Diagnosis**, PAM 2008.

Open source and data promotes regulator use

- **Greece's Telecom regulator**, EETT, built (and open-sourced) SPEBS
- **FCC**'s Measuring Broadband America 2011 report
  - New study in 2012.
- **European Commission** study
  - 30 countries
  - 10,000 users
  - 3 years, starting in 2012
- **Austria's Telecom regulator**, RTR, support an M-Lab node and have developed a mobile measurement tool
- **Cyprus' Telecom regulator** support an M-Lab node
- **Canada's CIRA** are deploying servers and utilising M-Lab and baseline for measurement

Why use M-Lab?

- You are doing all these things that are ***not* research**:
  - Deploy and manage servers
  - Collect and store data
  - Publish data in an easily accessible way

M-Lab does it all for you

- You can now do these things that ***are* research**:
  - Create new measurement methodologies
  - Build new measurement tools
  - Analyze and visualize data collected by your tools **or others'**

How M-Lab supports researchers

- Provide **developer resources**
- **Support deployment** of tools on the M-Lab platform
- Supply **name service** to help choose nearest server
- Provide a suite of **native libraries**
- Help in accessing and **processing the data**
- Advertise and **promote research tools**

M-Lab's limitations

- Active measurements only
- Client-server tests only
- Incomplete geographical coverage
- Biased user population
  - Tests are mostly run
    - When there is a problem
    - By "technical" users
  - µTorrent has a different user population

Data are better than no data!

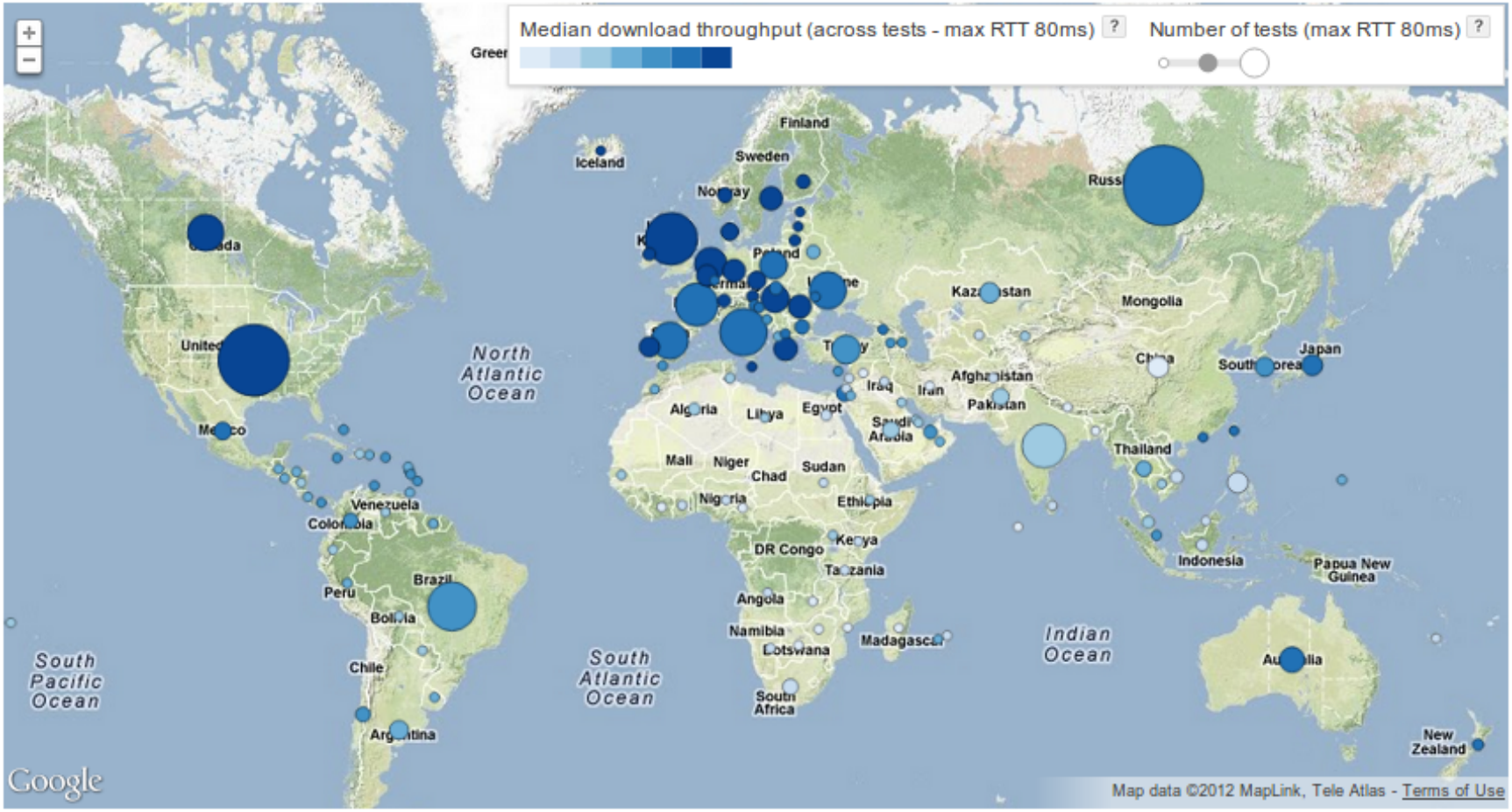
Future plans

- **Easier access to data**
  - Open data collection pipeline even easier to use
  - More structured data in BigQuery and Cloud Storage
  - Metrics server with API for pre-built queries
  - Standards for (mobile) data collection and tagging
- **Extend server platform**
  - More servers in more countries
  - Testing on 1G+ networks
  - Adding Lite option for developing areas that can't support our requirements
- **Make M-Lab the go-to platform for network performance measurement and analysis**

Demos and data visualizations

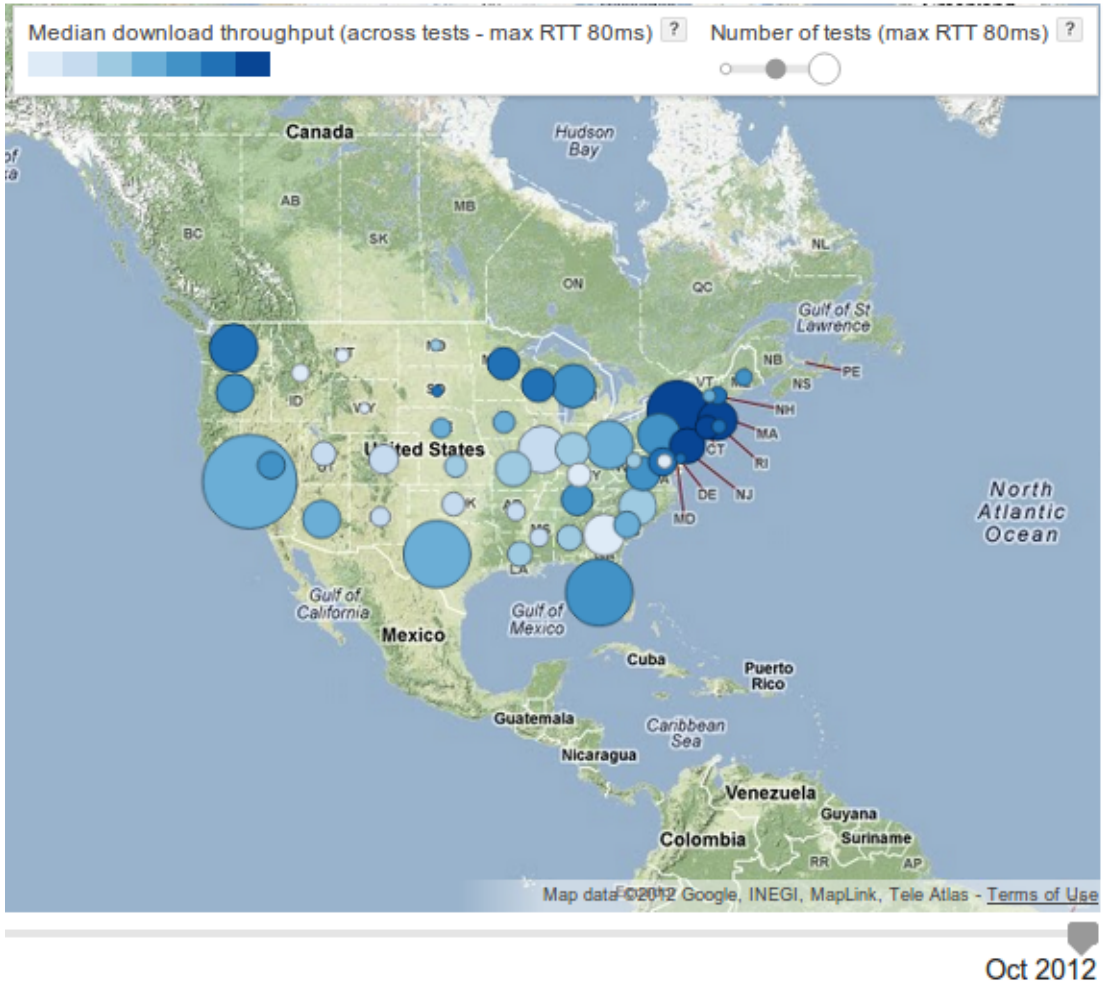
[Visualizing M-Lab data with BigQuery]

Download throughput worldwide



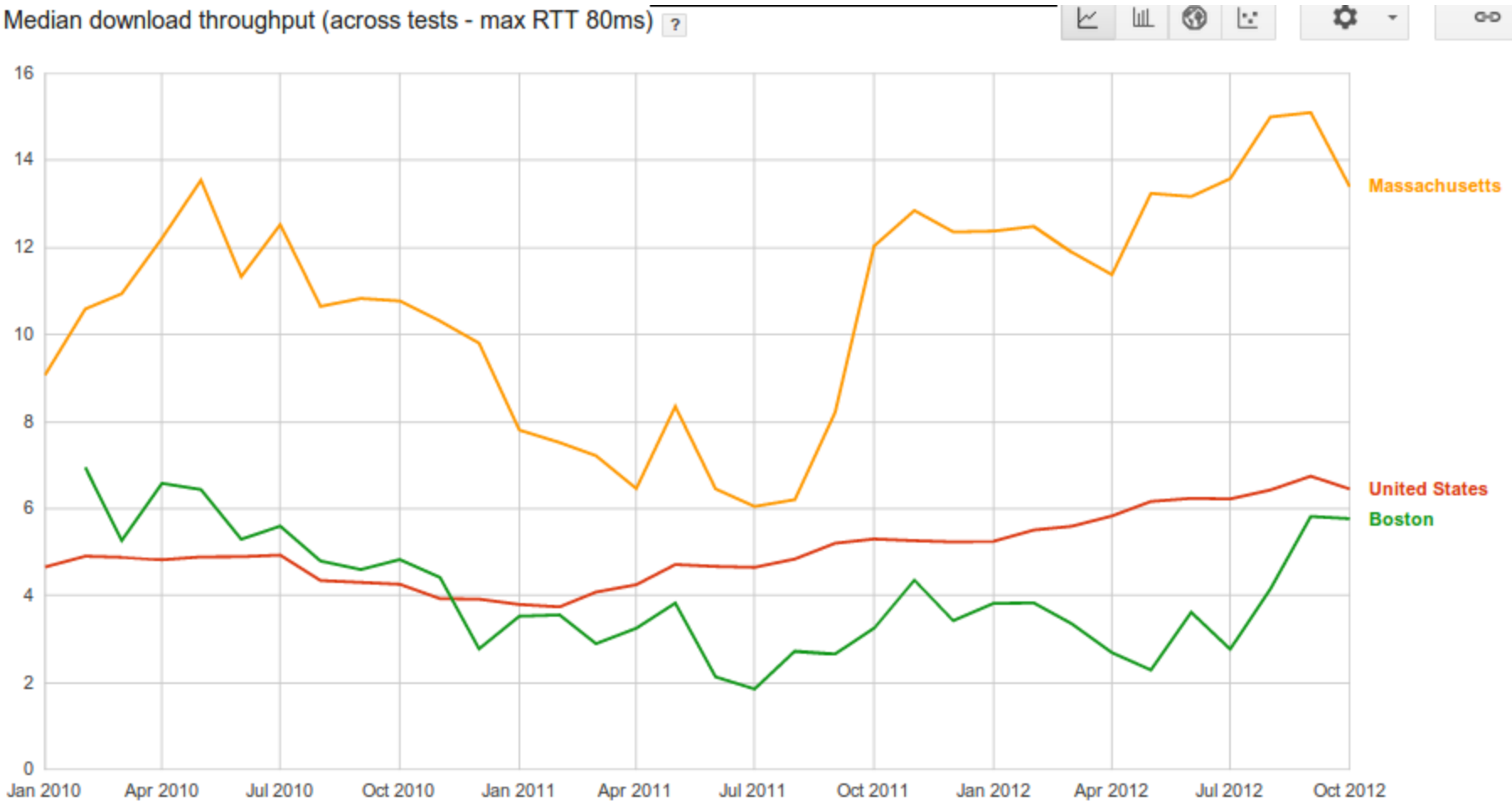
Link to Public Data Explorer chart

Download throughput in the US



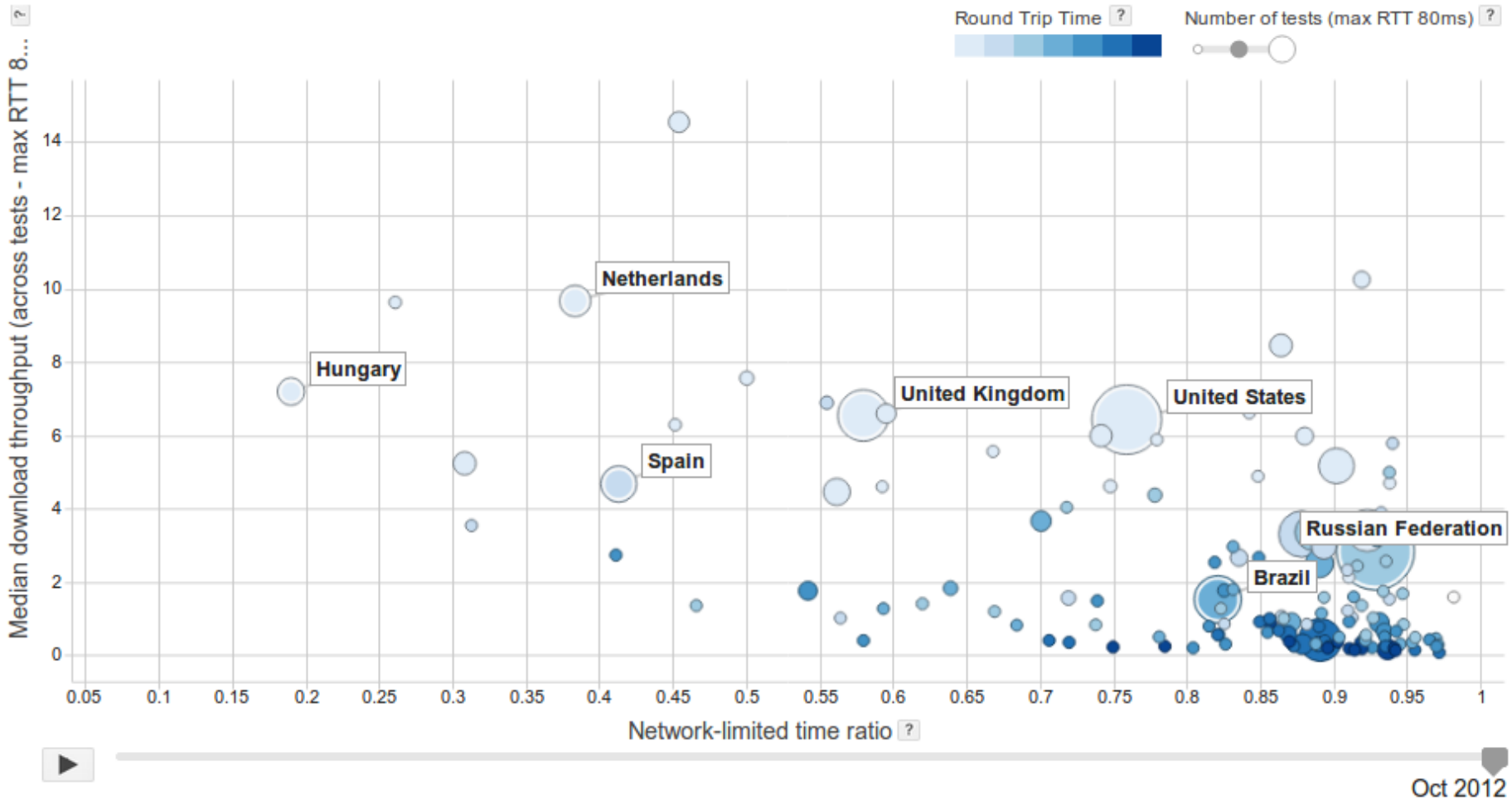
[Link to Public Data Explorer chart]

Download throughput in US, MA, Boston



[Link to Public Data Explorer chart]

Correlation between download and net-limited



[Link to Public Data Explorer chart]

Thanks!

More info at http://measurementlab.net