

Measurement Lab

MLAB

@ **CS&S** Code for
Science &
Society

History of M-Lab

- M-Lab founded in 2009 to answer the question:

How do we measure Internet performance, at scale?

- Founders: New America's Open Technology Institute, the PlanetLab Consortium, Google and academic researchers

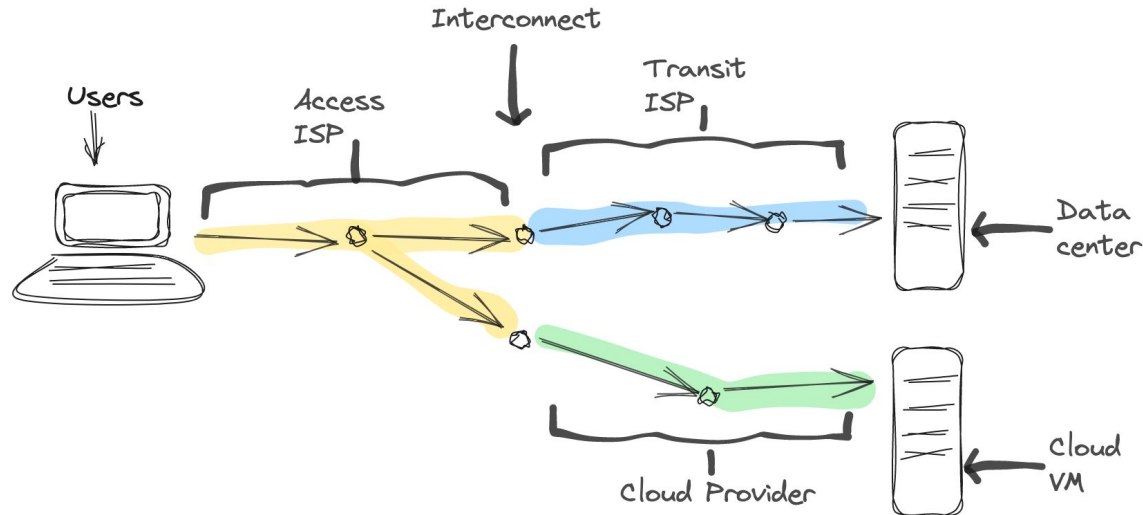
M-Lab Today

- Provides an open, verifiable measurement platform for global network performance
- Hosts one of the largest open Internet performance

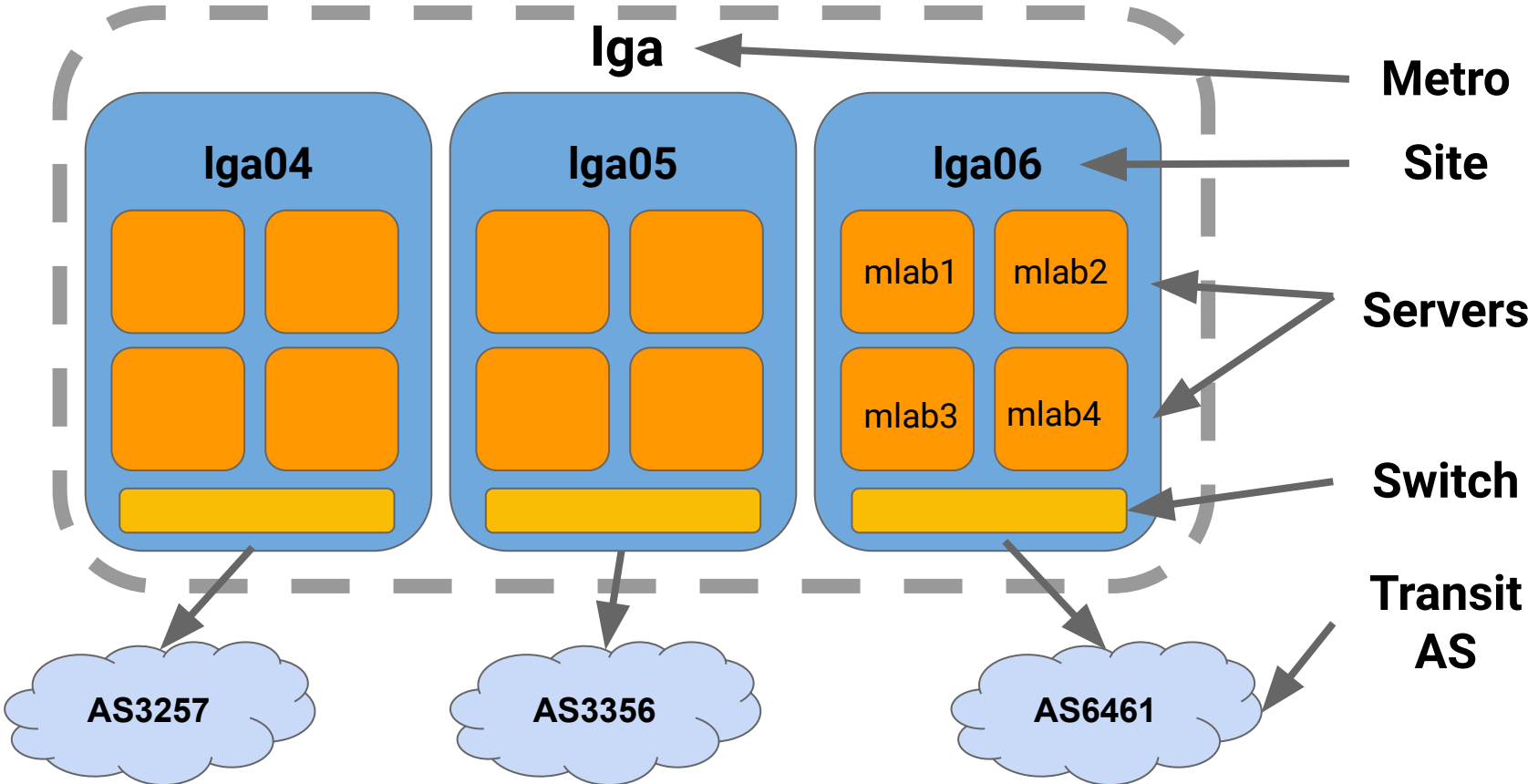
M-Lab's mission is to measure the Internet, save the data, and make it universally accessible and useful

How M-Lab Measures the Internet

- Servers deployed on transit networks & Cloud providers
- User measurements cross interdomain boundaries
- So that M-Lab can measure the “Inter-” part of the Internet



Metros and Sites

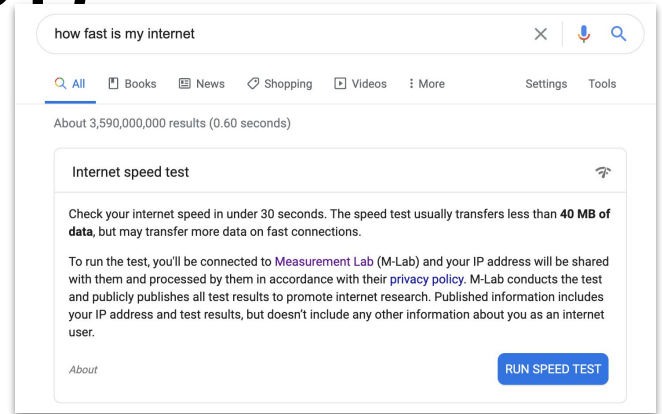


Datasets



Network Diagnostics Tool (NDT)

- NDT is our most frequently run test
 - 6 billion+ tests total
 - 4 million+ tests per day, on average
- NDT measures the single-stream performance of bulk-transport capacity, more commonly referred to as a “speed test”

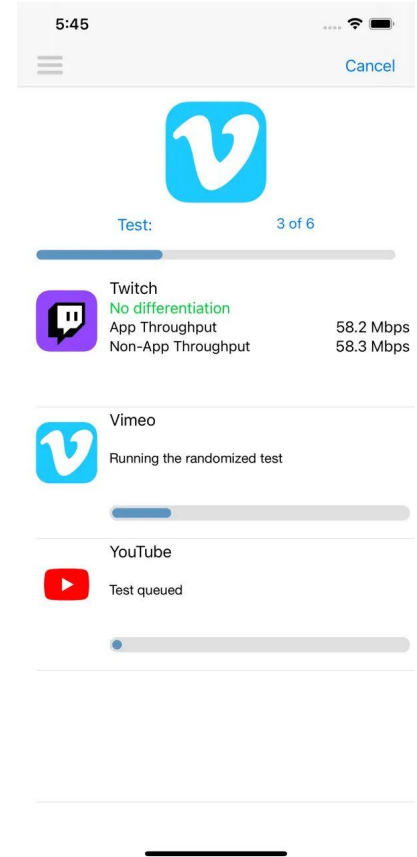
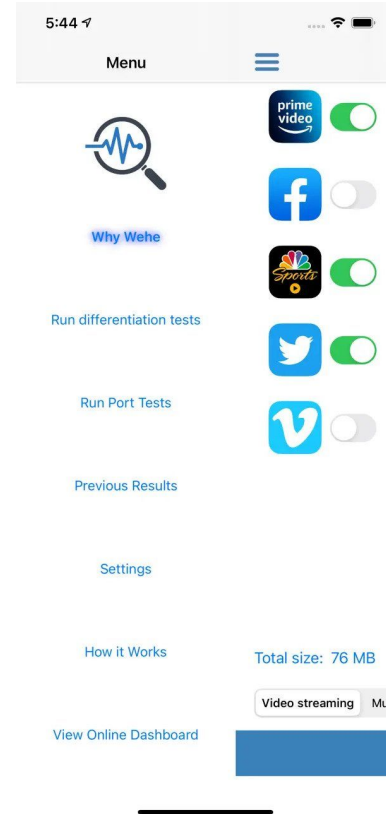


A screenshot of the "Results" page from the NDT tool. The page has a dark purple background with white text. The title "Results" is at the top left. Below it is a table with five rows of performance metrics. Each row includes a small icon representing the metric, the name of the metric, and the measured value.

	Test Server	New York, US
Download	Download	64.88 Mb/s
Upload	Upload	19.98 Mb/s
Latency	Latency	16 ms
Retransmission	Retransmission	0.26%

Wehe

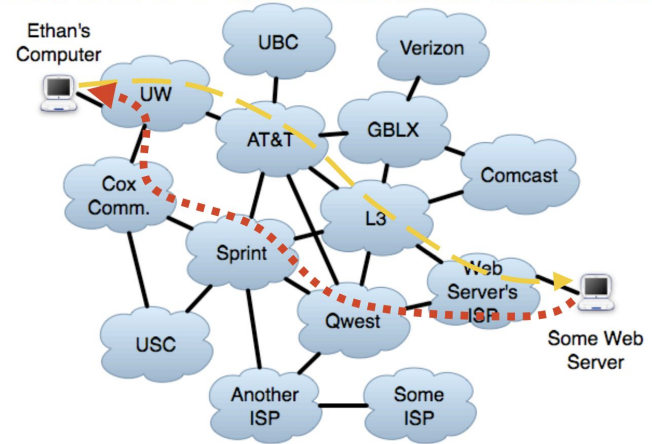
- The Wehe mobile application is available on iOS and Android and allows users to detect whether or not various applications are being throttled by their ISP.
- Led by Dave Choffnes at Northeastern University



Reverse Traceroute

- **Traceroute** collects network path information from our server back to the client
- **RevTr** measures the network path back to a user from selected network endpoints, and provides a rich source of information on network routing and topology.

Traceroute's Fundamental Limitation

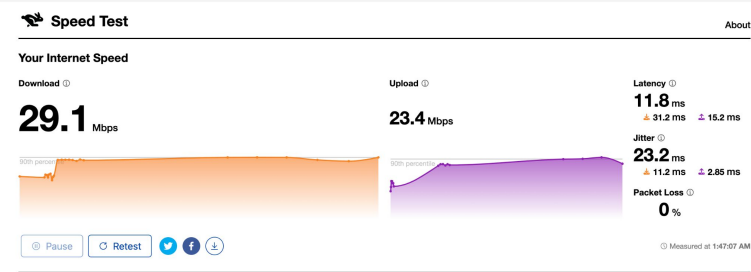


Traceroute: Tool to measure path FROM YOU to anywhere

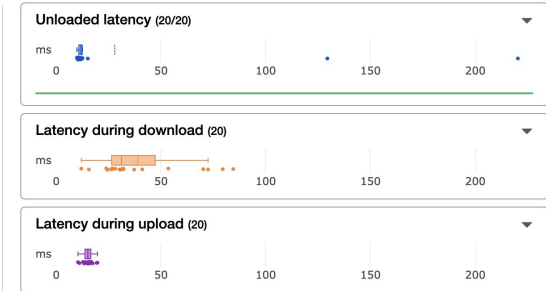
What about the path from anywhere back to you?

Cloudflare SpeedTest & AIM

- speed.cloudflare.com measures the speed and consistency of your connection to the Cloudflare's network
- M-Lab publishes the data produced by this tool
- We also publish Aggregated Internet Measurement (AIM data) which assigns point values based on speed tests to help you understand how your Internet quality will perform for streaming, gaming, and webchat/real-time communication (RTC).
- Noteworthy: Cloudflare's speed test includes latency under load



Latency Measurements



Sidecar Services

- **Traceroute** collects network path information from our server back to the client
- **PCAP** collects packet headers for all incoming TCP flows
- **TCP Info** collects statistics for every open TCP socket

Accessing the Data

Accessing M-Lab Data

MLAB

@ CS&S Code for Science & Society

MLAB

Measurement Lab is led by teams based at Code for Science & Society; Google, Inc; and supported by partners around the world.

Learn more about M-Lab. Get Involved.



Home / Data / BigQuery QuickStart

BigQuery QuickStart

M-Lab provides query access to our datasets in BigQuery at no charge to interested users. Following the steps below will allow you to use BigQuery to search M-Lab datasets without charge when the [measurement-lab](#) project is selected in your Google Cloud Platform console, or set as your project in the Google Cloud SDK. **Queries from projects you create, saving query results to BigQuery tables, etc. will incur costs to you.**

Please follow the steps below to configure free query access. If you have questions, please contact us at support@measurementlab.net

Subscribe your Google account to the [M-Lab Discuss group](#)

To gain access to M-Lab's open dataset, you are asked to sign up for the M-Lab discuss mailing list. The M-Lab team uses this mailing list to communicate updates to the project including changes to the data's format and updates to the platform as well information about upcoming events that M-Lab is hosting or participating in.

The mailing list is also a place for discussion for M-Lab users. We invite our community to ask questions, become familiar with each other's work, and directly communicate with our team. To facilitate constructive conversation, we ask all participants to adhere to [community guidelines](#). If you prefer to communicate with the M-Lab team directly, we are available at support@measurementlab.net

Members of this group are given access to use the [measurement-lab](#) project in the Google Cloud Platform console or Google Cloud SDK so that query charges are paid for by M-Lab.

Accessing M-Lab Data

The screenshot shows a web-based SQL interface. The query is as follows:

```
1 SELECT
2   client.Geo.City as city,
3   COUNT(a.MeanThroughputMbps) as count,
4   APPROX_QUANTILES(a.MeanThroughputMbps, 100)[OFFSET(50)] AS median,
5   AVG(a.MeanThroughputMbps) as average
6 FROM `measurement-lab.ndt.unified_downloads`
7 WHERE date >= "2023-09-01"
8   AND client.Geo.Subdivision1Name = "New York"
9 GROUP BY city
10 ORDER BY city
```

The results table is titled "Query results" and has the following columns: Row, city, count, median, average, and an empty column. The data is as follows:

Row	city	count	median	average	
1	null	5712	32.53294077381...	64.06332484515...	
2	Accord	25	91.91158065458...	143.3655473966...	
3	Acra	17	43.38985516471...	63.85543843688...	
4	Adams	20	70.83468966556...	123.0589464089...	
5	Adams Center	1	91.03316709022...	91.03316709022...	
6	Addison	47	72.51998868612...	102.5698172186...	
7	Adirondack	2	118.4323421376...	119.9201983539...	
8	Afton	32	68.00021119898...	73.14020893827...	
9	Akron	99	84.51504966206...	121.5272987129...	
10	Albany	5278	79.75915064433...	137.6408879475...	
11	Albertson	8	49.22071123243...	56.44816812624...	
12	Albion	67	98.93268514906...	122.225155555...	
13	Alden	311	100.7545448808...	127.7780712170...	
14	Alexander	21	16.51457381458...	47.81324031532...	
15	Alexandria Bay	42	42.47587335556...	37.19677723878...	

M-Lab Tutorial in CoLab



M-Lab Data @ Splintercon ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

+ Code + Text

Quick start for use of NDT data to find evidence of throttling and/or shutdown events

Written by Lai Yi Ohlsen (laiyi@measurementlab.net) Last updated 2023-12-07

How to use

1. Get access to M-Lab data: <https://www.measurementlab.net/data/docs/bq/quickstart/>
2. Make a copy of this CoLab
3. Be logged into the email you signed up to the discuss@ list
4. In your copy, configure the parameters to fit your research interests
5. Run all cells
6. When asked, approve access to BigQuery

Parameters

Change the variables in the cell below to fit your research interests.

- Begin: Your preferred start date. Consider 1 week or 1 month before beginning of event, so you can compare
- End: Your preferred end date. If you are looking right after a shutdown or throttling event, this might be today. Ideally you'd be able to look at one week or more after the event to see how activity compared to before.
- Country: Country of interest.

MLAB



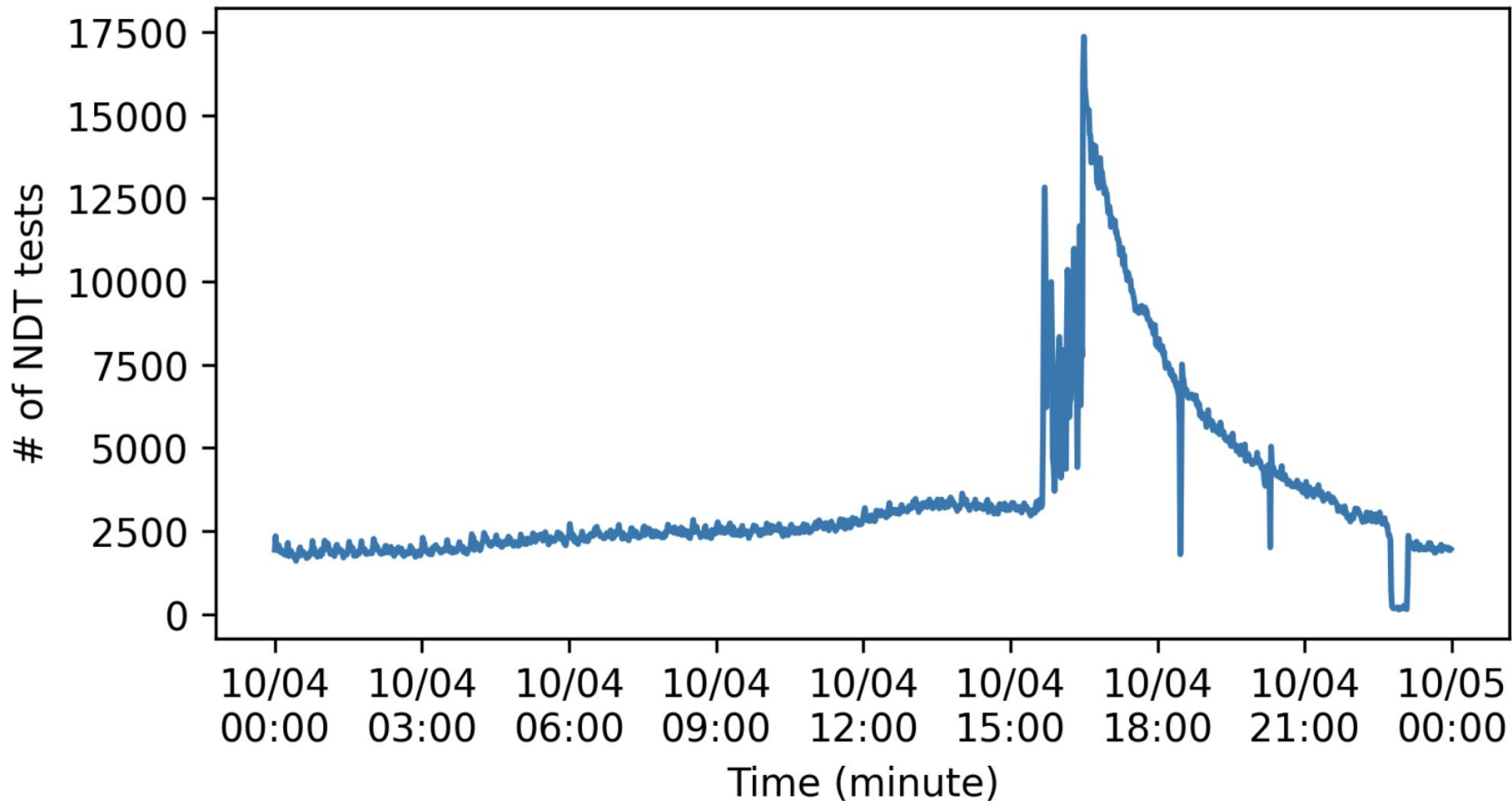
CS&S

Code for
Science &
Society

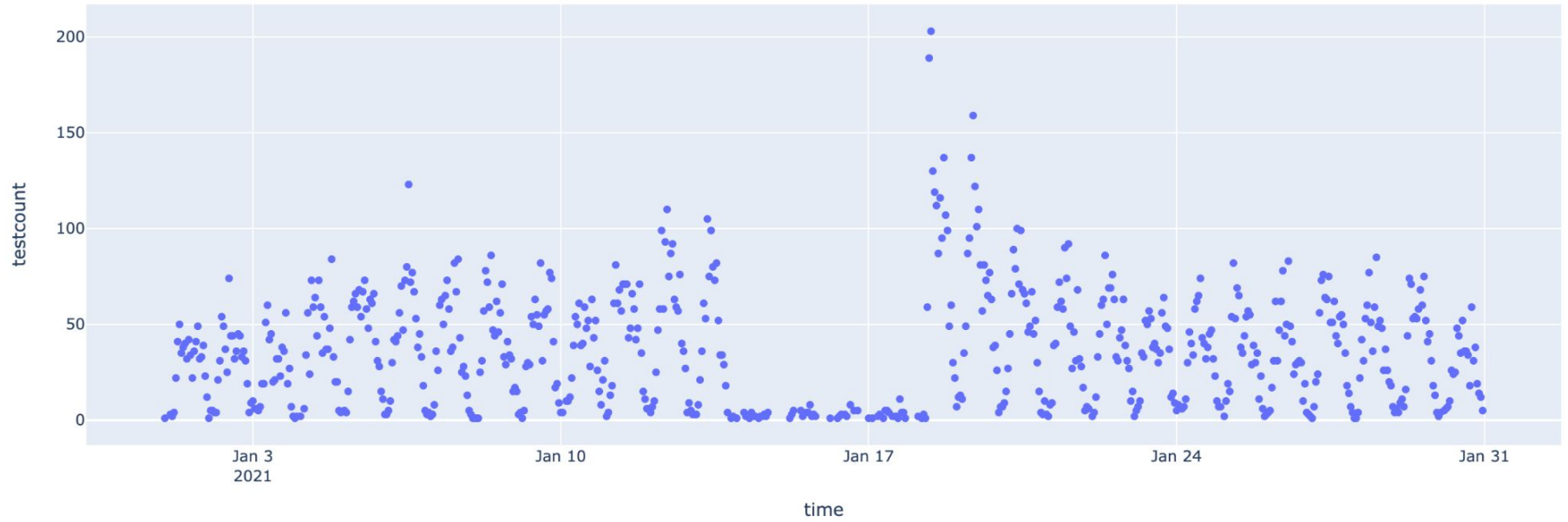


<https://bit.ly/3ThxT1x>

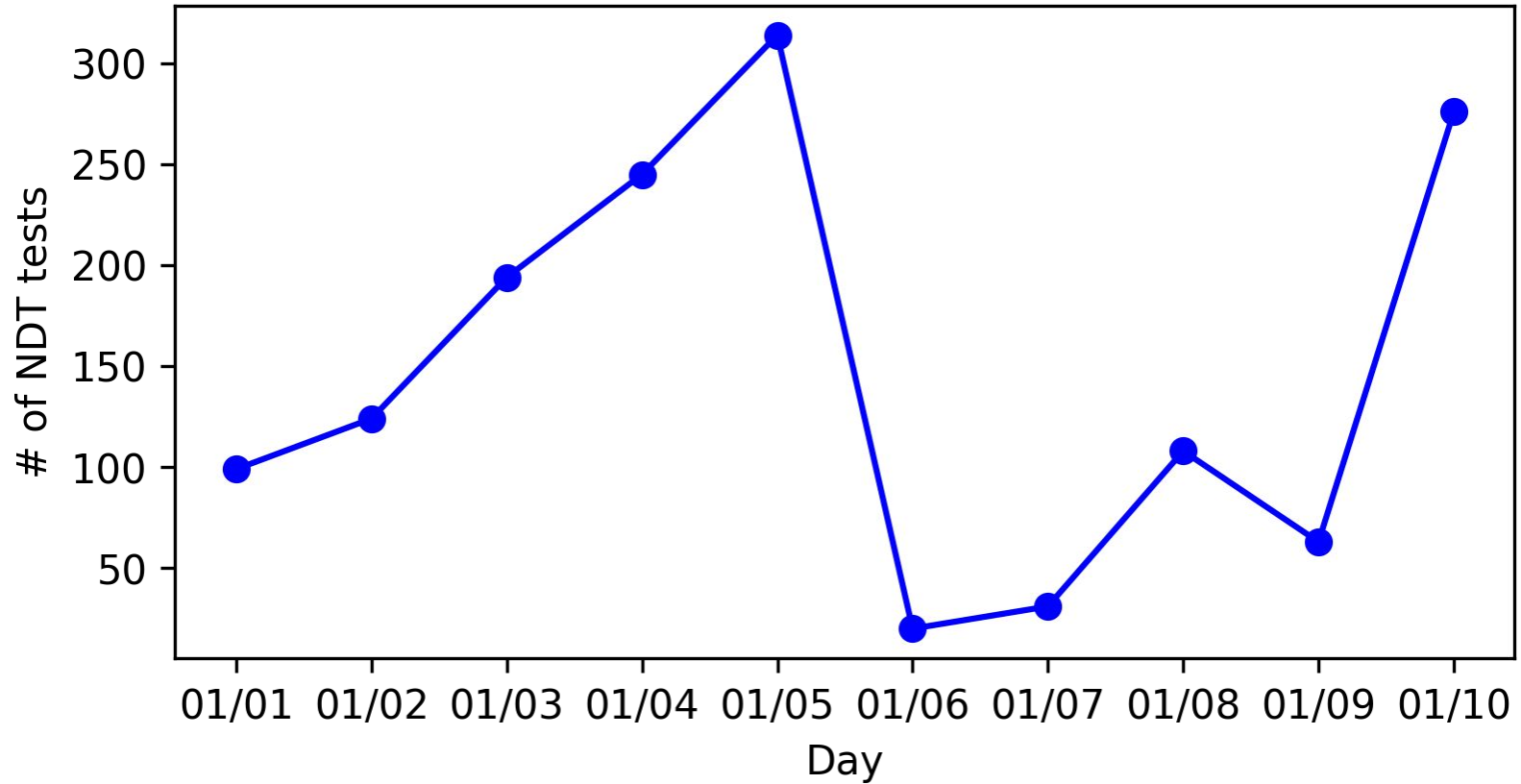
NDT tests per minute during Facebook outage Fall 2021



NDT TestCount in Uganda between 2021-01-01 and 2021-01-30



NDT tests per day during 2022 Kazakh unrest



MLAB

@ **CS&S** Code for
Science &
Society

Brainstorming

Brainstorming - Data Analysis

- Our traceroute data has yet to be sufficiently explored.
 - Can we observe any large scale shifts in the pathways taken between clients and M-lab servers over time?
 - (This will become even more interesting with more revtr data)
- Can we see evidence of shutdowns via traceroute data? How do network interference events affect the paths being taken?

Brainstorming - Data Collection

- What is the ideal server topology to measure the network interference events and Internet fragmentation behaviors?
- What metrics are most useful to collect during a crowd-sourced measurement campaign to document a network interference event?

Brainstorming - Data Publication

- Are there private datasets that could be published anonymously to help document network interference events?
 - From big content providers such as Meta, Netflix and/or smaller scale tool developers providing network interference solutions.

Brainstorming - Methodology



- In some ways M-Lab as a platform and methodology assumes a globally connected Internet - how should Splinternets be measured?