IMC Hackathon 2025

Housekeeping

- Restrooms: Immediately outside this room
- 14:15 Coffee/beverages
- 17:45-19:30 Dinner

Wi-Fi Options:

eduroam

- Connection guide: https://kb.wisc.edu/helpdesk/page.php?id=25020

Discovery-Guest Discovery Building Guest Wi-Fi

- Go to http://discovery.wisc.edu to be redirected to the Wi-Fi authentication page.
- Scroll to the bottom and click Accept to connect.

Schedule

- 13:00-14:00 Introduction to datasets and team formation
- 14:00-19:00 Time for analysis!
- 19:00-20:00 Share outs

Introductions

- Arturo Filastò (Founder, ED & CTO), Open Observatory of Network Interference (OONI) Foundation
- Romain Fontugne (Deputy Director), Internet Initiative Japan Research Laboratory
- Zachary Bischof (Senior Research Scientist), Internet Outage Detection and Analysis @ Georgia Tech's Internet Intelligence Lab
- Zenyep Arslan, Dioptra Research Group, Sorbonne University
- Lai Yi Ohlsen (Senior Product Manager), David Belson (Head of Data Insight), Cloudflare
- Pavlos Sermpezis (Director and Tech Lead), Chris Ritzo (Event Support), Measurement Lab


CLOUDFLARE

Cloudflare Radar

Cloudflare Radar is a hub that showcases global Internet traffic, attack, and technology trends and insights. Cloudflare Radar is powered by data from Cloudflare's global network, as well as aggregated and anonymized data from Cloudflare's 1.1.1.1 public DNS Resolver.

In some cases Cloudflare Radar uses data from PeeringDB (interconnection meta-information) and APNIC (visible ASNs customer population measurements).

With Radar, you can access trends and insights, like the adoption of new technologies, browsers or operating systems. Radar also keeps up to date with relevant events around the world to provide information on Internet activity patterns.

https://radar.cloudflare.com/ is built on top of the **Radar API** (specification, documentation).

speed.cloudflare.com

Cloudflare provides a speed test at speed.cloudflare.com, which is powered by an open Javascript library. It receives about ~200k tests per day, on average.

You can view the results of these tests in aggregate through Radar on the Internet Quality page.

The raw results are available through the Radar API *and* M-Lab's Public Dataset in BigQuery.

More info about how Cloudflare's speed test works: [blog post]

BTW!

Cloudflare announced that we're hiring 1,111 interns (as a nod to 1.1.1.1)

You can read more about the program here:
https://blog.cloudflare.com/cloudflare-1111-intern-program/

Also, we did an Internet Measurement, Transparency, and Resiliency blog takeover – see the posts at https://blog.cloudflare.com/tag/research/

Radar Data Explorer

The Radar Data Explorer (https://radar.cloudflare.com/explorer) is a web-based interface designed to simplify the process of building, and viewing the results of, more complex API queries.

It also provides example API calls and responses for the queries built using the interface.

More information is available at
https://blog.cloudflare.com/radar-data-explorer-ai-assistant/#building-a-query-directly

A free Cloudflare account is required to generate an API token that is needed to make API queries. If you do not currently have a Cloudflare account, or have not yet generated the necessary token, please follow these steps…

Creating a Cloudflare account

- Go to https://dash.cloudflare.com/sign-up
- If you want to login with your Google/Apple/Github account, click the appropriate link.
- Otherwise, enter your email address and a password in the appropriate fields, check the box to verify that you are a human, and click “Sign up”.
- Click the link in the “[Action required] Verify your email address” email sent to the address associated with the account.

Creating an API token

- Login to the Cloudflare dash at https://dash.cloudflare.com/login
- Click the profile icon in the upper right corner, and click “Profile” in the drop-down menu.
- On the Profile page, click “API Tokens” from the left-side navigation bar.
- Click the “Create Token” button in the “API Tokens” card on the “User API Tokens” page.
- Click the “Use template” button next to “Read Cloudflare Radar data”
- Under “Account Resources”, choose the account associated with your login from the drop-down menu.
- Click the “Continue to summary” button at the bottom of the page.
- Click “Create Token” on the next page
- If token creation was successful, it will be displayed in a text box. Copy it to a password manager or other secure location, as it will not be shown again. You can confirm that the token is valid and active by executing the curl command shown on the page.

Possible Project

[Cloudflare Radar - Project 1] Aggregate and visualize metrics across platforms

- Radar provides a number of metrics that can help assess the occurrence and impact of a potential Internet disruption. (These include traffic metrics for HTTP requests, Netflows, and DNS; TCP resets & timeouts; bandwidth & latency; and announced IP address space.) However, these are largely all from a single platform. In order to corroborate observed disruptions, it would be helpful to have an aggregated view that includes metrics from other platforms.

OONI

The **Open Observatory of Network Interference Foundation (OONI)** is an **NGO** that develops **open source software** and **open data** to **empower decentralized efforts** in **documenting internet censorship** around the world

Internet shutdowns: UN report details ‘dramatic’ impact on people’s lives and human rights

23 June 2022

https://www.ohchr.org/en/press-releases/2022/06/internet-shutdowns-un-report-details-dramatic-impact-peoples-lives-and-human

“Shutdowns effectively deepen digital divides between and within countries,” the report warns. At a time when substantial development aid is justifiably directed towards enhancing connectivity in less developed countries, some of the beneficiaries of that assistance are themselves deepening the digital divide through shutdowns. At least 27 of the 46 least developed countries have implemented shutdowns between 2016 and 2021, most of which have received support to increase connectivity.

The report urges States to refrain from imposing shutdowns, to maximize Internet access and remove the multiple obstacles standing in the way of communication. The report also urges companies to speedily share information on disruptions and ensure that they take all possible lawful measures to prevent shutdowns they have been asked to implement.

United Nations A/HRC/50/55

General Assembly

Distr.: General
13 May 2022

Original: English

Human Rights Council
Fiftieth session
Agenda items 2 and 3
Annual report of the United Nations High Commissioner for Human Rights and reports of the Office of the High Commissioner and the Secretary-General

Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development

Internet shutdowns: trends, causes, legal implications and impacts on a range of human rights

Technical multi-stakeholder report on Internet shutdowns: The case of Iran amid autumn 2022 protests

OONI, IODA, M-Lab, Cloudflare, Kentik, Censored Planet, ISOC, Article19, 2022-11-29

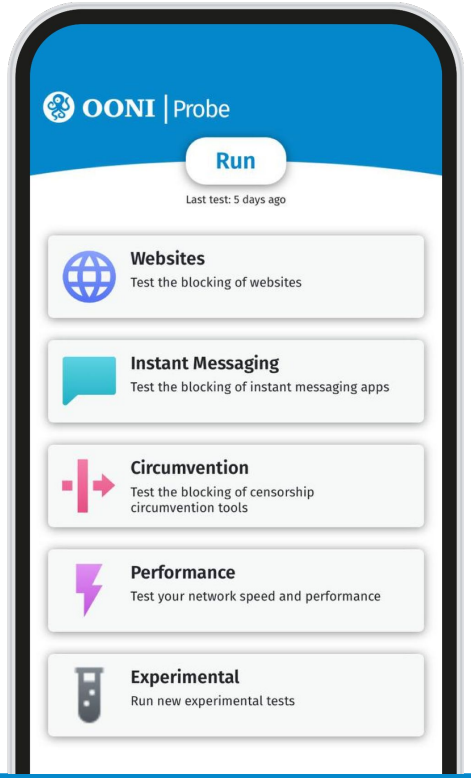
- **Coordinators:** OONI, ISOC
- **Contributors:** OONI, IODA, Measurement Lab (M-Lab), Cloudflare, Kentik, Censored Planet, ISOC, Article19
- **Facilitators:** European Commission, United States

This report shares empirical technical findings on the recent Internet shutdown events that emerged in Iran following the death of Jhina (Mahsa) Amini in September 2022. The report is intended to be the first among a series of multi-stakeholder reports aimed at shedding light on what is becoming a widespread and increasingly sophisticated practice of certain governments around the world. The stakeholders participating in the report share a concern about the global trend in Internet shutdowns, but are contributing to the analysis only on the basis of their technical expertise.

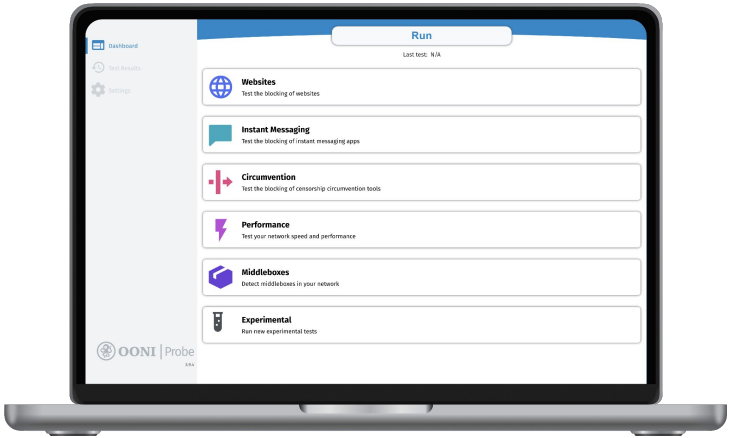
Over the years, Iranian authorities have followed a pattern of blocking social media apps, numerous websites and even resorting to shutting networks entirely, implementing overall pervasive levels of Internet control. The latest shutdown events that emerged in Iran in September 2022 amid ongoing protests follow the same pattern, but show a wider range of technical strategies to prevent censorship circumvention.

- Summary of Key Findings

https://ooni.org/post/2022-iran-technical-multistakeholder-report



Hundreds of thousands of people all over the world have **installed OONI Probe on their mobile phone or computer** to collect **evidence of internet censorship in their country.**



All our software is developed in the open and **released as free software**:
https://github.com/ooni/

OONI | Probe

Open Source Software

OONI Explorer is the world's largest open data repository on internet censorship worldwide.

From **2012 to 2024** we have collected and published **2 Billion network measurements** from **241 countries** and territories covering **27 thousand networks**.

OONI | Explorer

Open Data

Censorship Findings (57)

Search for findings | All Findings | Most Recently Ended

Nepal

July 19, 2025 UTC - ongoing

published on July 28, 2025 UTC

Nepal blocked Telegram

This report shares OONI data on the blocking of Telegram in Nepal in July 2025.

Read More

Jordan

May 13, 2025 UTC - ongoing

published on June 3, 2025 UTC

Jordan blocked 12 news media websites

This report shares OONI data on the blocking of 12 news media websites in Jordan in May 2025.

Read More

Türkiye

March 28, 2025 UTC - ongoing

published on June 13, 2025 UTC

Türkiye blocked the opposition campaign Boykotyap amid protests

This report shares OONI data on the blocking of the the opposition campaign Boykotyap in Türkiye in March 2025.

This report shares OONI data on the blocking of TikTok in Albania in March 2025.

OONI Explorer Findings

OONI Data

OONI data can be accessed through the public OONI Explorer web interface. OONI Explorer has a component called the Measurement Aggregation Toolkit (MAT), which can be used to produce plots of measurements over time. OONI Explorer is backed by OONI API, links for the relevant API endpoints used to produce the charts can be found by clicking on the JSON Data or CSV Data on the MAT pages.

Data Access for IMC Hackathon

OONI

For the IMC hackathon we are also offering direct access to the underlying clickhouse database, the schema of which is available in the relevant documentation page: https://docs.ooni.org/data/oonidata-analysis-db/.

In order to query the clickhouse database you may use the following credentials to access the OONI jupyter notebook server:

URL: https://notebook.ooni.org/

user: imc2025
password: aGoodUserIsAGentleUser

Please note that this is a shared instance, so please try to limit resource usage as much as possible. If you need to run particularly memory or CPU heavy computation it's recommended you export a dump of the data and run these computations on your own machine.

Documentation

Other relevant documentation on accessing and analysing OONI data can be found at the following links:

- OONI Explorer MAT: https://ooni.org/support/ooni-explorer/#measurement-aggregation-toolkit-mat
- Interpreting OONI Data: https://ooni.org/support/interpreting-ooni-data/
- OONI base data format specifications: https://github.com/ooni/spec/tree/master/data-formats
- OONI Test specifications: https://github.com/ooni/spec/tree/master/nettests
- OONI Database schema: https://docs.ooni.org/data/oonidata-analysis-db/
- Fetching OONI data from Amazon S3: https://docs.ooni.org/data/

If you run into any issues, an OONI team will be at IMC, but you may also reach out to the OONI team and community on the public slack channel #ooni-dev on https://slack.ooni.org/ (you can sign up for an account by entering your email address and then join the #ooni-dev channel).

Welcome to the IMC Hackathon 2025! 👷

OONI

Below are few instructions that we encourage you to read in order to make everybodies time best.

While BigTech don't take coin for our queries,
we still have limited resources,
so please be kind on Miss DB.

In practice this means the following:

- When experimenting, always apply some kind of `LIMIT` to your queries (eg. `LIMIT 42`)
- When running queries, apply sane `WHERE` constraints (eg. `WHERE measurement_start_time > '2025-01-01' AND measurement_start_time > '2025-02-01'`)
- Use common sense: full scans make it bad for everyone, we don't need to enforce that by monetary incentives, because a "Good User is a Gentle User"

Thank you :)

Your friendly system administrator

~ A.

Docs

Relevant documentation on accessing and analysing OONI data can be found at the following links:

- OONI Explorer MAT: https://ooni.org/support/ooni-explorer/#measurement-aggregation-toolkit-mat
- Interpreting OONI Data: https://ooni.org/support/interpreting-ooni-data/
- OONI base data format specifications: https://github.com/ooni/spec/tree/master/data-formats
- OONI Test specifications: https://github.com/ooni/spec/tree/master/nettests
- OONI Database schema: https://docs.ooni.org/data/oonidata-analysis-db/ Fetching OONI data from Amazon S3: https://docs.ooni.org/data/

Project Ideas

OONI

[OONI - Project 1] Analyze and visualize internet censorship data

OONI created the Measurement Aggregation Toolkit (MAT) which enables you to track internet censorship around the world and create your own custom charts based on real-time OONI network measurement data. As part of this challenge, we invite you to analyze OONI data and create your own data visualization(s) based on the questions you would like to answer.

For example, this could involve exploring the following questions based on OONI data:

- Which circumvention technologies (among the ones measured by OONI) are most effective nowadays at a given location? Is there a trend (increase or decrease) in such effectiveness?
- Where in the world are located the sites that appear blocked from a given location?
- Which censorship techniques are adopted in X country? A novel visualization could depict blocking techniques in use, sorted by degree of sophistication. Live filtering according to blocking techniques could also be useful in terms of visualization overlays.

Project Ideas

[OONI - Project 2] Encrypted Client Hello (ECH) blocking

OONI has a test for measuring Encrypted Client Hello (ECH) called echcheck (https://github.com/ooni/spec/blob/master/nettests/ts-039-echcheck.md). There are reports of ECH being blocked in Russia when used in conjunction with the cloudflare outer_sni cloudflare-ech.com (see: https://github.com/net4people/bbs/issues/417).

It would be interesting to study the evolution of this block in Russia, but also investigate if we see a signal of it getting blocked in other regions as well.

Tips

- Relevant DB columns in obs_web table are: tls_echconfig, tls_outer_server_name, tls_server_name, tls_failure

Project Ideas

[OONI - Project 3] Automatic detection of blocking signals

This project is about applying automated methods to identify changes in blocking in a particular country, ASN pair. It may involve looking at absolute volume of measurements, as we anecdotally have seen users start running OONI Probe more frequently when something starts to be blocked in a certain country, or making use of some of the features in the fastpath or analysis_web_measurements tables.

Tips

- Experiment with different changepoint detection algorithms to identify shifts in blocking patterns within a hostname, probe_cc, probe_asn tuple (eg. CUSUM, BOCPD, etc.)
- For the analysis table use the (dns|tcp|tls)_(blocked|ok) metrics grouped by probe_cc, probe_asn, resolver_asn on the analysis_web_measurements table

Project Ideas

[OONI - Project 4] Examine TLS certificate diversity across countries/networks

This is about examining when a specific domain name presents a TLS certificate that is different or issued by a different authority depending on the vantage point that carried out the measurements.
While this can be a signal of censorship, in the event of an attempted TLS MiTM, it may also happen as a result of a different TLS certificate being used depending on the location of the server hosting the site.

Tips

- Look in the obs_web table for the tls_end_entity_* columns

Project Ideas

[OONI - Project 5] Investigate misbehaving OONI Probes

OONI has multiple probe apps, with different versions, running our test engine with different versions. OONI is currently working on a project, which makes use of **Anonymous credentials**, to reduce the impact of bad data.

Can you tell if any of our apps/engines is misbehaving, compared with the rest?
For example, providing constantly different results for the same tests and networks?

Tips

- Fields to look for: software_name, software_version, engine_name, engine_version and the outcomes from the fastpath (eg. msm_failure, anomaly, confirmed) or obs_web tables

OONI is sponsoring the hackathon

We have some **SurPrizes** at the end for the following categories

- **The mad scientists**, for the most inventive and out of the ordinary use of data
- **Ghost in the network**, for finding something that nobody had seen before

We spent 42.70 EUR for the prizes, if you would like to contribute to OONI and help us cover these costs: https://ooni.org/donate


Measurement
Lab

History of M-Lab

M-Lab founded in 2009 to answer the question:

How do we measure Internet performance, at scale?

M-Lab Today

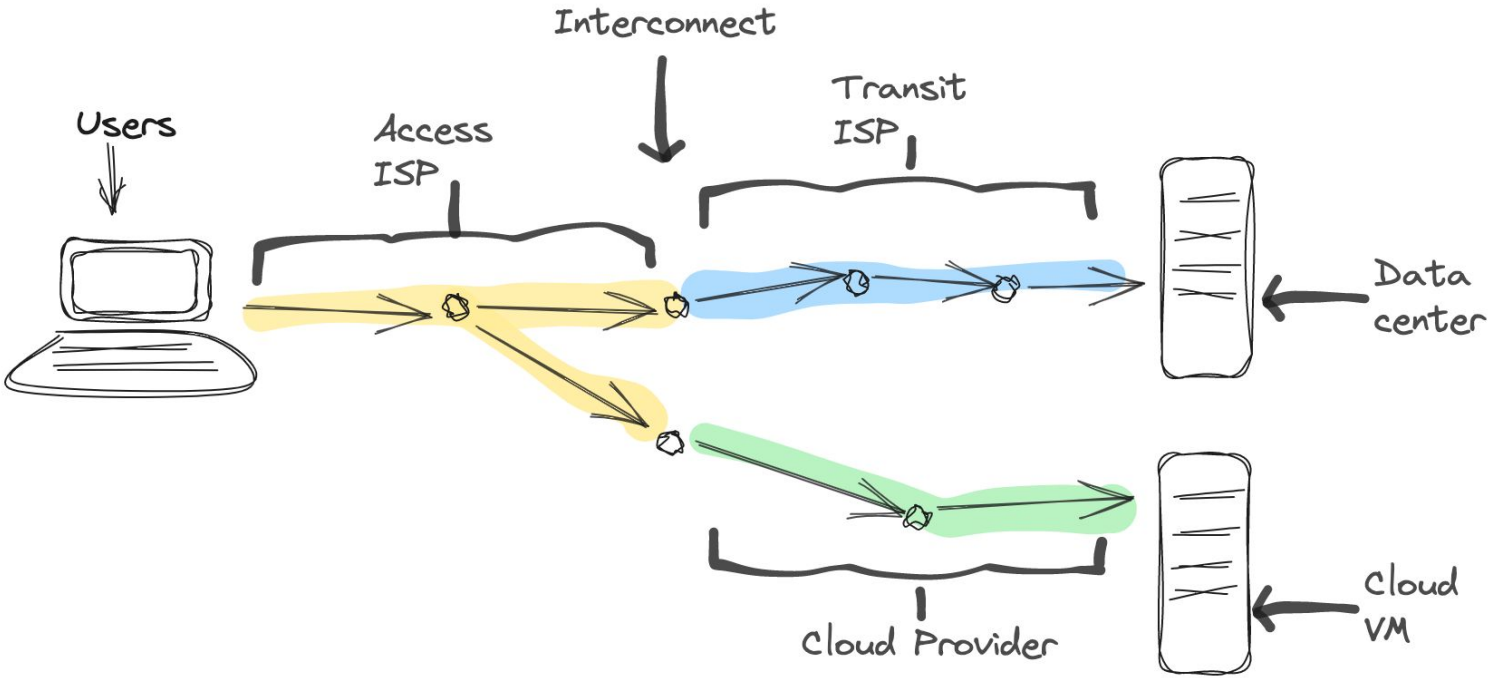
- Provides an open, verifiable measurement platform for global network performance
- Hosts one of the largest open Internet performance datasets

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

M-Lab's platform

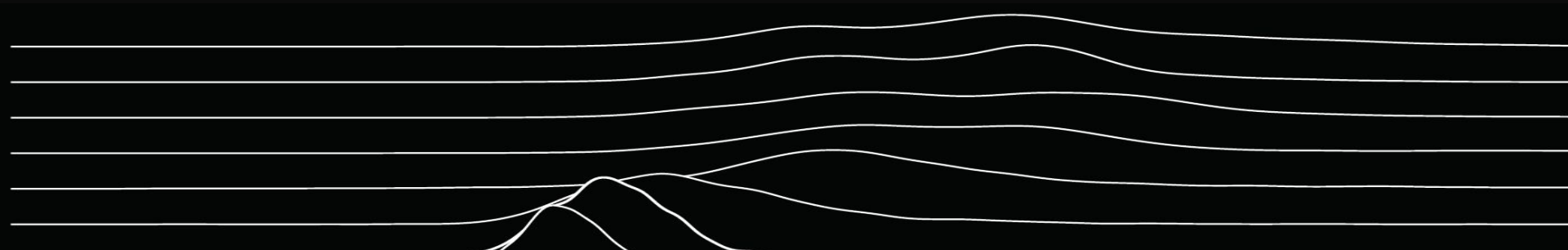
- 500+ servers globally
  - https://www.measurementlab.net/status/
- 40+ countries / 100+ metros,
- 50+ transit providers

How M-Lab Measures the Internet



Measurement Lab

Datasets



M-Lab measurements

MLAB @ CS&S Code for Science & Society

- On the M-Lab platform, we host the server-side of “experiments” or “measurement services”.
- When clients run these measurements, they test against M-Lab servers.
- Every measurement is publicly archived and published in BigQuery.

NDT (Network Diagnostic Tool)
Tests your connection speed, and provides a sophisticated diagnosis of problems limiting speed.

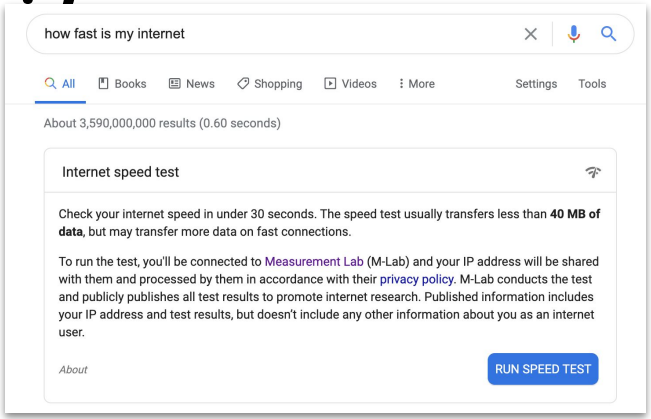
Neubot DASH
DASH is designed to measure the quality of tested networks by emulating a video streaming player.

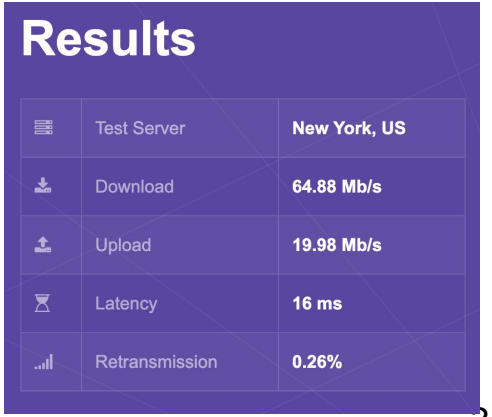
Reverse Traceroute
Measures the network path back to a user from selected network endpoints.

WeHe
Wehe uses your device to exchange Internet traffic recorded from real, popular apps like YouTube and Spotify, and attempts to tell you whether your ISP is giving different performance to an app's network traffic.

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”





Network Diagnostics Tool (NDT)

	ndt3	ndt4	ndt5	ndt7
	2009-07 to 2019-11	2016-07 to 2019-11	2019-11 to present	2020-05 to present
TCP Kernel Instrumentation				
web100				
TCP_INFO				
Download Congestion Control Algorithm [1]				
TCP RENO				
TCP CUBIC				
TCP BBR				
Running over port				
3001				
3010				
80/443				
Protocol				
RAW TLV				
Websocket				

Sidecar Services

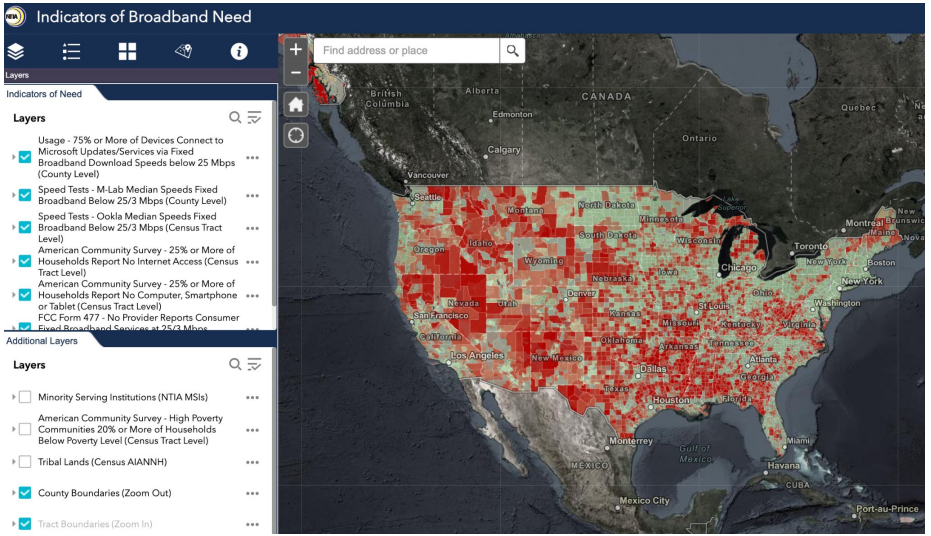
- **Traceroute** collects network path information from our server back to the client
- **PCAP** collects packet headers for all incoming TCP flows
- **TCP Info** collects statistics for every open TCP socket

How M-Lab Data is Used

Measurement Lab

@ CS&S Code for Science & Society

NDT data is integrated into NTIA's Indicators of Broadband Need map, as well as the National Broadband Availability Map.

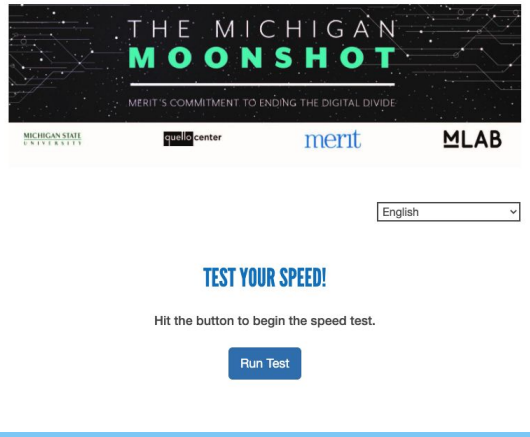


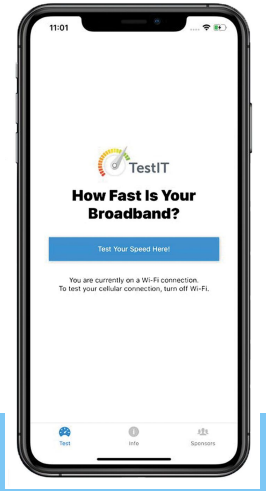
How M-Lab Data is Used

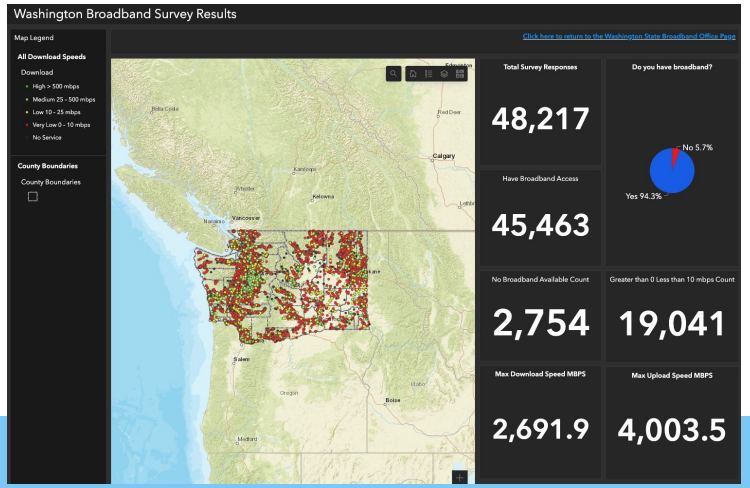
Measurement Lab

@ CS&S Code for Science & Society

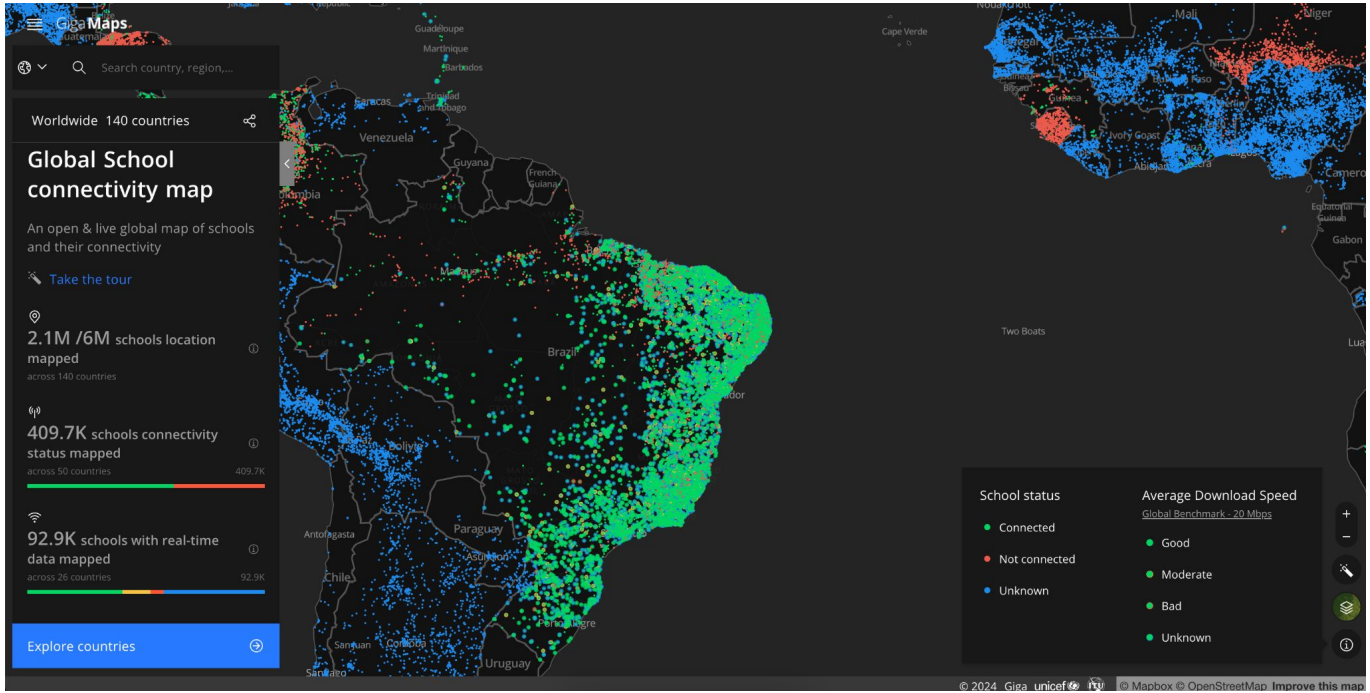
Digital inclusion efforts such as MERIT's Michigan Moonshot Project, the State of Washington's Department of Commerce State Broadband Survey and the National Association of Counties use NDT to collect information about their constituents/communities Internet connection to advocate for their needs







How M-Lab Data is Used



Ideas for Today (Review our Public Doc for Complete Details)

Measurement Lab

@ CS&S Code for Science & Society

- [M-Lab - Project 1] Load balancing M-Lab: Smarter M-Lab server deployment
  Goal: Model the traffic load of M-Lab platform and optimize for server deployments
- [M-Lab - Project 2] Mapping M-Lab platform's global connectivity
  Goal: Map/characterize how M-Lab servers are connected to the Internet
- [M-Lab - Project 3] Detecting Internet shutdowns from user reactions
  Goal: Analyze user behavior upon Internet shutdowns. Design methodology to automatically detect Internet shutdowns based on the analysis.
- [M-Lab - Project 4] "CensorScope": Dashboards for Internet shutdown and throttling detection
  Goal: Use M-Lab data to derive insights for supporting detection/characterisation of Internet censorship or shutdown events

Ideas for Today

Measurement Lab

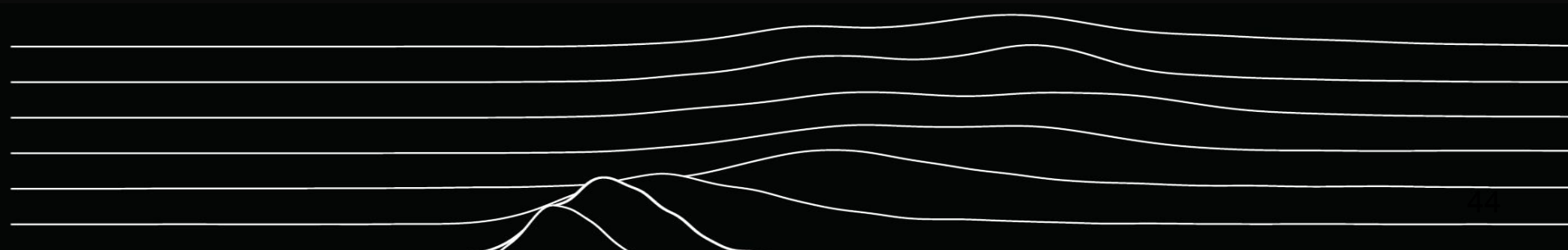
@ CS&S Code for Science & Society

- [M-Lab - Project 5] Internet Quality Barometer (IQB) in action
  Goal: Implement an instance of the IQB framework for a region using multiple datasets
- [M-Lab - Project 6] “State of the Net”: Interactive Internet performance reports
  Goal: Characterize the Internet performance of a region or network using open measurement datasets.
- [M-Lab - Project 7] “TCP Showdown”: Comparing MSAK and NDT Measurements
  Goal: How do M-Lab measurements differ based on the TCP congestion control algorithm and number of streams used?
- [M-Lab - Project 8] Internet Topology Visualization: Mapping Network Paths from Traceroute Data
  Goal: How can we effectively visualize and understand the physical and logical structure of internet routing at scale?

Measurement Lab

@ CS&S Code for Science & Society

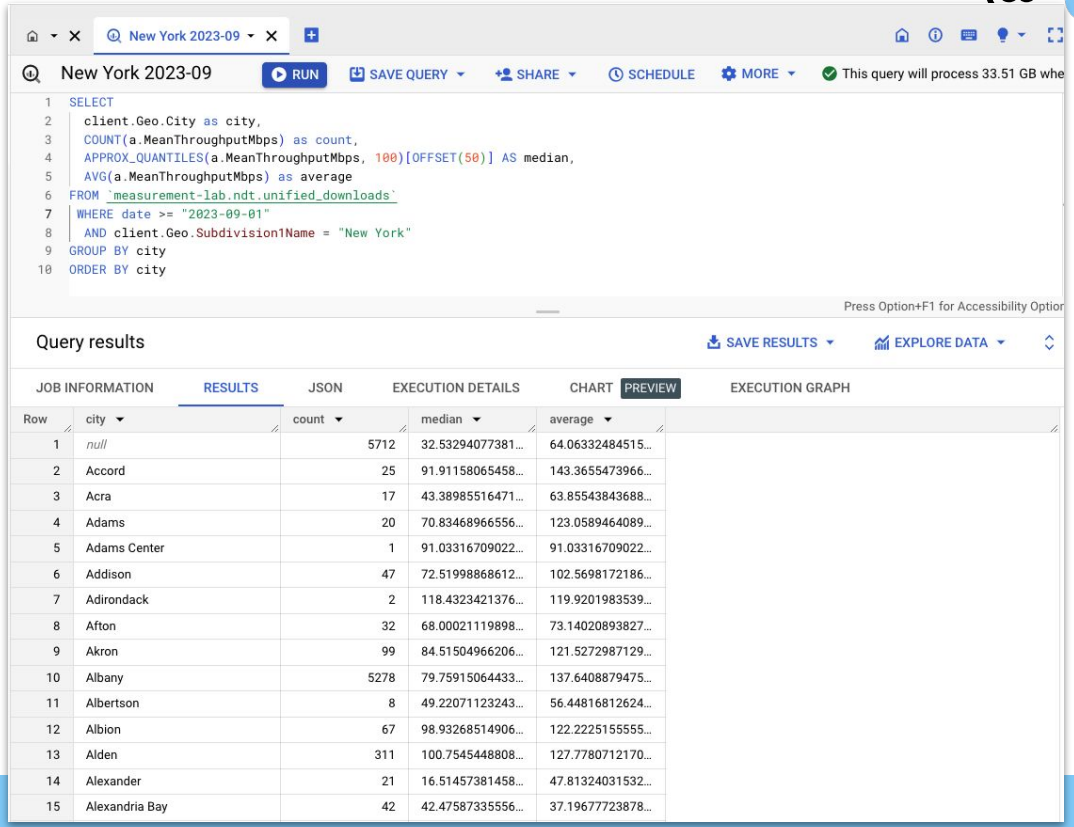
Accessing the Data



M-Lab Data in BigQuery

Measurement Lab

@ CS&S Code for Science & Society

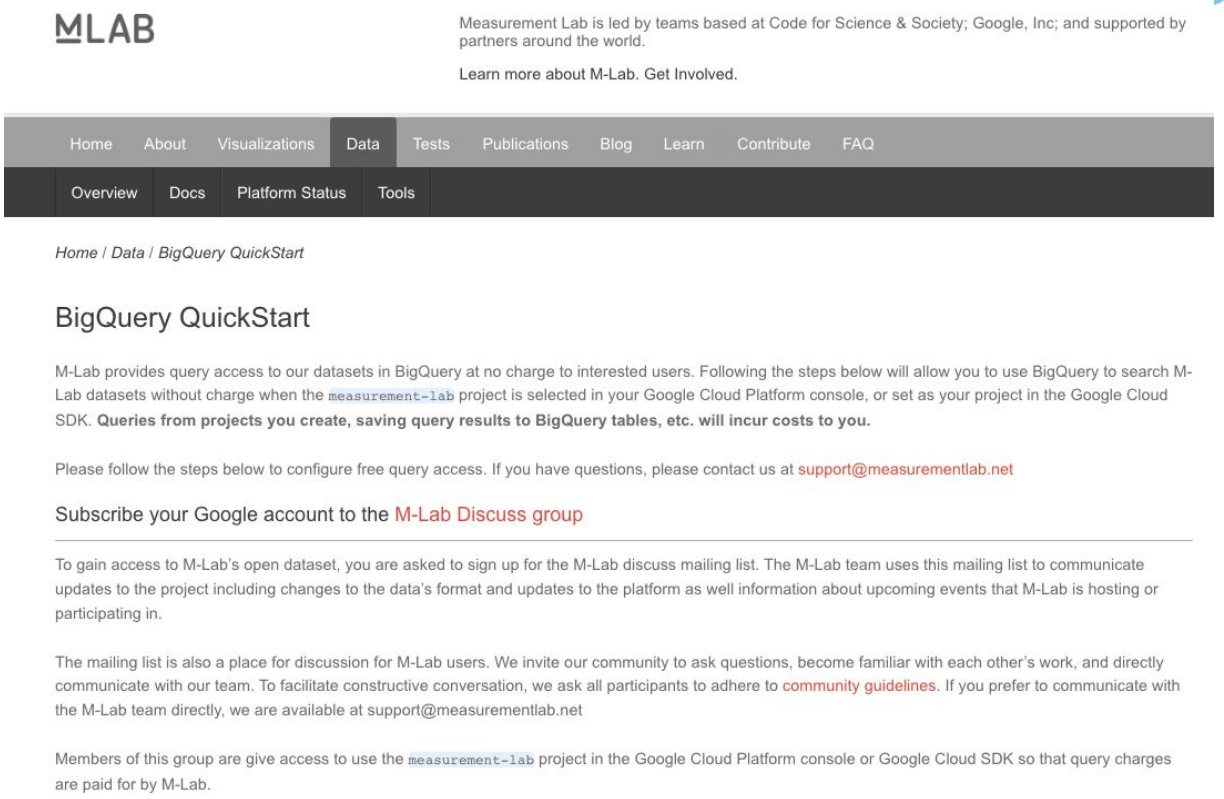


Row	city	count	median	average
1	*null*	5712	32.53294077381...	64.06332484515...
2	Accord	25	91.91158065458...	143.3655473966...
3	Acra	17	43.38985516471...	63.85543843688...
4	Adams	20	70.83468966556...	123.0589464089...
5	Adams Center	1	91.03316709022...	91.03316709022...
6	Addison	47	72.51998868612...	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...	122.2225155555...
13	Alden	311	100.7545448808...	127.7780712170...
14	Alexander	21	16.51457381458...	47.81324031532...
15	Alexandria Bay	42	42.47587335556...	37.19677723878...

Accessing M-Lab Data

Measurement Lab

@ 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 | About | Visualizations | Data | Tests | Publications | Blog | Learn | Contribute | FAQ

Overview | Docs | Platform Status | Tools

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 give 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.

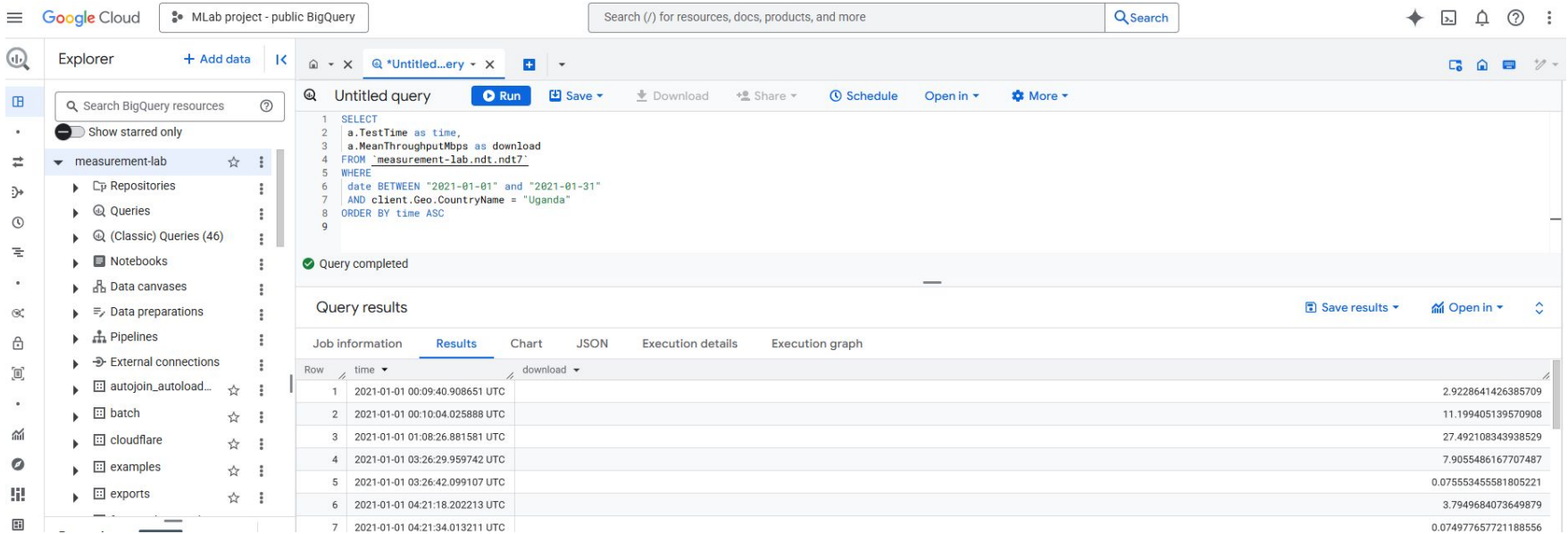
Accessing M-Lab Data

Measurement Lab @ CS&S Code for Science & Society

- To access M-Lab's public data, you need to sign up for the discuss@ list
  - https://groups.google.com/a/measurementlab.net/g/discuss?pli=1
- Then you can access the data through BigQuery
  - Tutorial: https://www.measurementlab.net/data/docs/bq/quickstart/
  - BigQuery: https://console.cloud.google.com/bigquery?project=measurement-lab
  - Schema (for ndt7 dataset): https://www.measurementlab.net/tests/ndt/ndt7/
    - find details for other datasets https://www.measurementlab.net/data/
- Tools & Languages/Frameworks
  - BigQuery - SQL
  - CoLab - Python, Pandas (e.g., see IMC 2023 examples & Optima shutdown example)

Accessing M-Lab Data

Measurement Lab @ CS&S Code for Science & Society



Accessing M-Lab Data

Measurement Lab @ CS&S Code for Science & Society

- BigQuery examples

```
SELECT
 a.TestTime as time,
 a.MeanThroughputMbps as download
FROM `measurement-lab.ndt.ndt7`
WHERE
 date BETWEEN "2021-01-01" and
"2021-01-31"
 AND client.Geo.CountryName = "Uganda"
ORDER BY time ASC
```

```
SELECT
  a.TestTime,
  a.MeanThroughputMbps,
  a.MinRTT,
  a.LossRate,
  client.Geo.city as ClientCity,
  client.Network.ASNumber as ClientASNumber,
  client.Network.ASName as ClientASName,
  server.Geo.City as ServerCity,
  server.Network.ASNumber as ServerASNumber,
  server.Network.ASName as ServerASName
FROM
  `measurement-lab.ndt.unified_downloads`
WHERE
  date >= "2022-07-01" AND date <= "2022-07-01"
  AND client.Geo.city = "Baltimore"
  AND client.Geo.CountryCode = "US"
ORDER BY a.TestTime
```

Sorbonne University
Dioptra Research Group

About Dioptra research group

Dioptra is an Internet cartography research group within the Networks and Performance Analysis (NPA) team at the LIP6 computer science laboratory in Paris, France.

We are located on the campus of the Faculty of Science of Sorbonne University, which jointly runs LIP6 with the French National Centre for Scientific Research (CNRS). We are also associated with the LINCS laboratory.

IP Route Survey (IPRS) Dataset

IPRS is an initiative to continuously monitor IP-level routing across the internet. This is done through the regular collection of traceroute-style measurements from multiple vantage points towards a significant portion of the internet's routable address blocks. IPRS consists of distributed route traces from, currently, 10 vantage points to all routable IPv4 prefixes.The survey is conducted by the Dioptra research group at Sorbonne University's LIP6 computer science laboratory.

IPRS is similar to CAIDA's Archipelago (Ark) data, consisting of multipath route traces. The data is available in the **iprs1** schema, designed to be consistent with **scamper1** schema used for M-Lab's existing large collection of traceroutes. We hope that the compatible formats will make it easier for researchers to use both datasets.

Possible Projects

The IPRS dataset is available via M-Lab.

[M-Lab - Project 8] Internet Topology Visualization: Mapping Network Paths from Traceroute Data

Goal: How can we effectively visualize and understand the physical and logical structure of internet routing at scale? This project aims to transform raw traceroute measurements into interactive graphical and geographic visualizations that reveal network topology, routing patterns, and infrastructure dependencies across the internet.

What:
Analyze large-scale traceroute data from M-Lab's IPRS dataset to:

- Extract and process network topology from multipath traceroute measurements stored in BigQuery
- Geocode network infrastructure by mapping IP addresses to real-world geographic location approximations
- Visualize routing patterns using interactive web-based maps that show how internet traffic flows between continents, countries, and cities
- Identify network characteristics including:
  - Major internet exchange points (IXPs) and hub locations
  - Route redundancy and multipath routing behavior
  - ISP backbone infrastructure and peering relationships

Datasets:

- M-Lab IPRS (Internet Path Routing Study) BigQuery tables
- IP Geolocation: IPInfo API for router location mapping

Internet Initiative Japan

Internet Yellow Pages

Why IYP?

Internet Initiative Japan

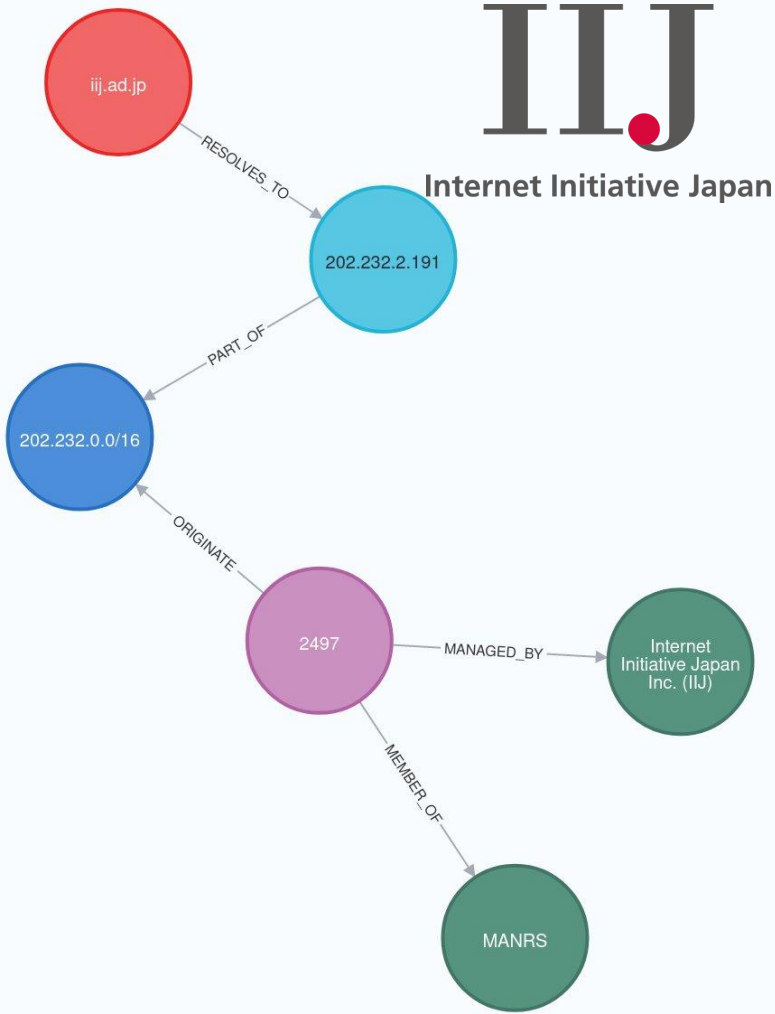
Our Goal:

- Provide a place for **sharing knowledge** and connect data from different sources
- **Open** to anyone, easily accessible, easy to contribute to
- **Structured**: not a repo with tons of data dumps
- **Extensible**: no fixed database schema

Internet Yellow Pages

Knowledge graph for Internet:

- 60+ datasets (PeeringDB, CAIDA, RIPE, APNIC, Cloudflare, OONI, BGPKit, BGP.Tools, IHR, …)
- Available online at: https://iyp.iijlab.net
- Or download snapshots at: https://ihr-archive.iijlab.net



IYP: Benefits

Internet Initiative Japan

- Great for sneak peek in numerous datasets
- Cut time to insights
- Also good for extracting/filtering a dataset
- List of available datasets: https://github.com/InternetHealthReport/internet-yellow-pages/blob/main/documentation/data-sources.md

Steep learning curve

Internet Initiative Japan

Take a look at Cypher examples:

- IYP Tutorial: https://docs.google.com/document/d/1cl3oUY-65TluosIjEBgOT3NNXq_UP-ZGQD4waqVvdhs/edit?usp=sharing
- IYP Gallery: https://github.com/InternetHealthReport/internet-yellow-pages/blob/main/documentation/gallery.md
- APNIC blog article: https://blog.apnic.net/2023/09/06/understanding-the-japanese-internet-with-the-internet-yellow-pages/
- IYP Notebooks: https://github.com/InternetHealthReport/iyp-notebooks
- Ask me for help!

IYP instance for hackathon

Internet Initiative Japan

A public instance of IYP is available at: https://iyp.iijlab.net

This is sufficient for retrieving small datasets and executing most common queries.

For heavier analysis please consider installing IYP locally.

MCP interface available at neo4j://iyp-bolt.ihr.live:7687

Project Ideas

Internet Initiative Japan

[IIJ - Project 1] Visualization of IYP data

Create visualizations for IYP data. Fetch data from IYP and plot it with your favorite visualization tools.

[IIJ - Project 2] Your research with IYP

Reproduce your own research with IYP and extend it with the numerous datasets available in IYP.

[IIJ - Project 3] Adding data to IYP

Start a local IYP instance and add your favorite dataset to IYP. This would connect the imported dataset with the 60+ datasets already integrated into IYP, hence enabling new analysis.

[IIJ - Project 4] Web centralization

- Reproduce (and extend if possible) key results from this recent study on web centralization: https://cs.stanford.edu/~gakiwate/papers/sigcomm25-centralization.pdf

Reverse Traceroute

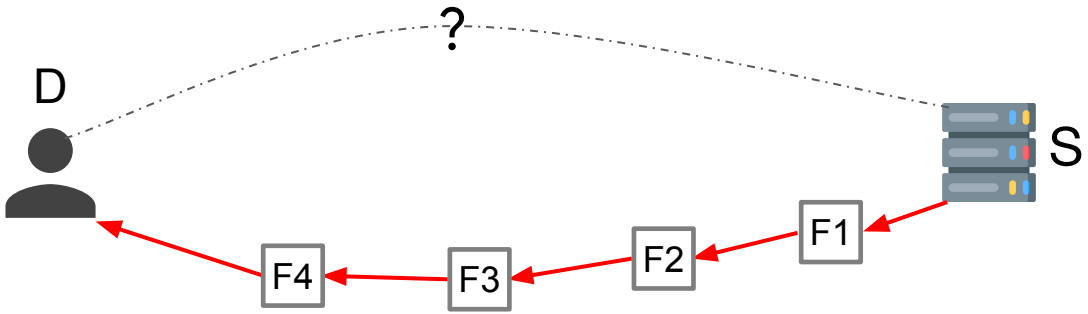
Reverse Traceroute

Traceroute shows the path from a server to a client. Reverse Traceroute (RevTr) extends that visibility by measuring the path back from the client to the server. Because M-Lab servers cannot initiate traceroutes from the client side, RevTr reconstructs reverse paths using a combination of techniques.

RevTr is integrated with M-Lab, and about 25% of NDT speed tests now include reverse traceroute measurements alongside. The results are publicly available in BigQuery, providing Internet-scale visibility into reverse paths and enabling studies of asymmetry, interconnection, and coverage.

Measuring the reverse path

Problem: We want to map D -> S.
S -> D can be measured with traceroutes, but how can we measure the **reverse path**?

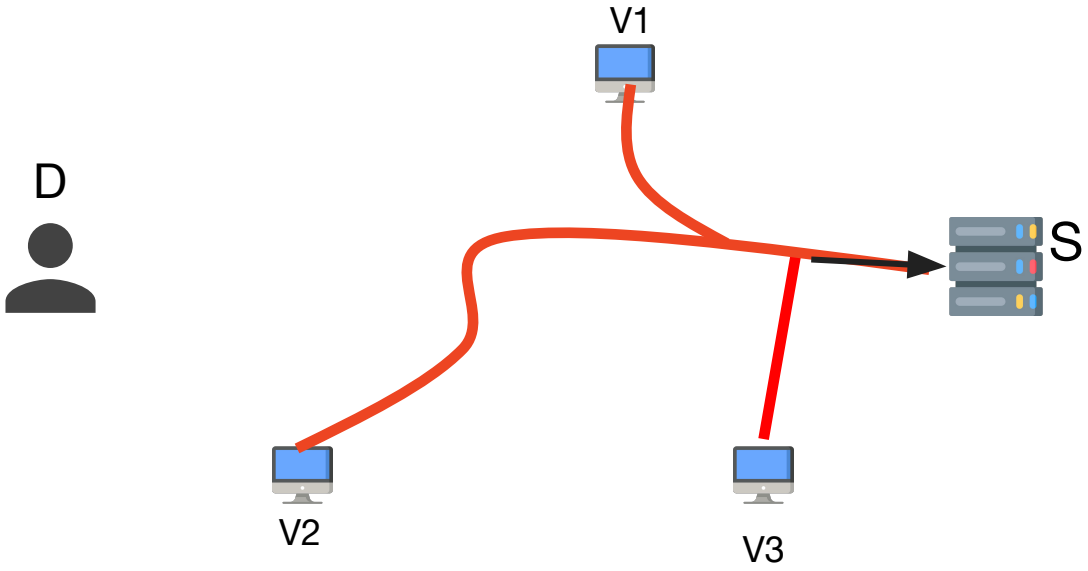


Solution 1:
Running traceroute from the client?

Solution 2:
Reverse Traceroute

Vermeulen et al. Internet Scale Reverse Traceroute 2022 ACM IMC
Katz-Bassett et al. Reverse Traceroute 2010 NSDI

What is Reverse Traceroute?



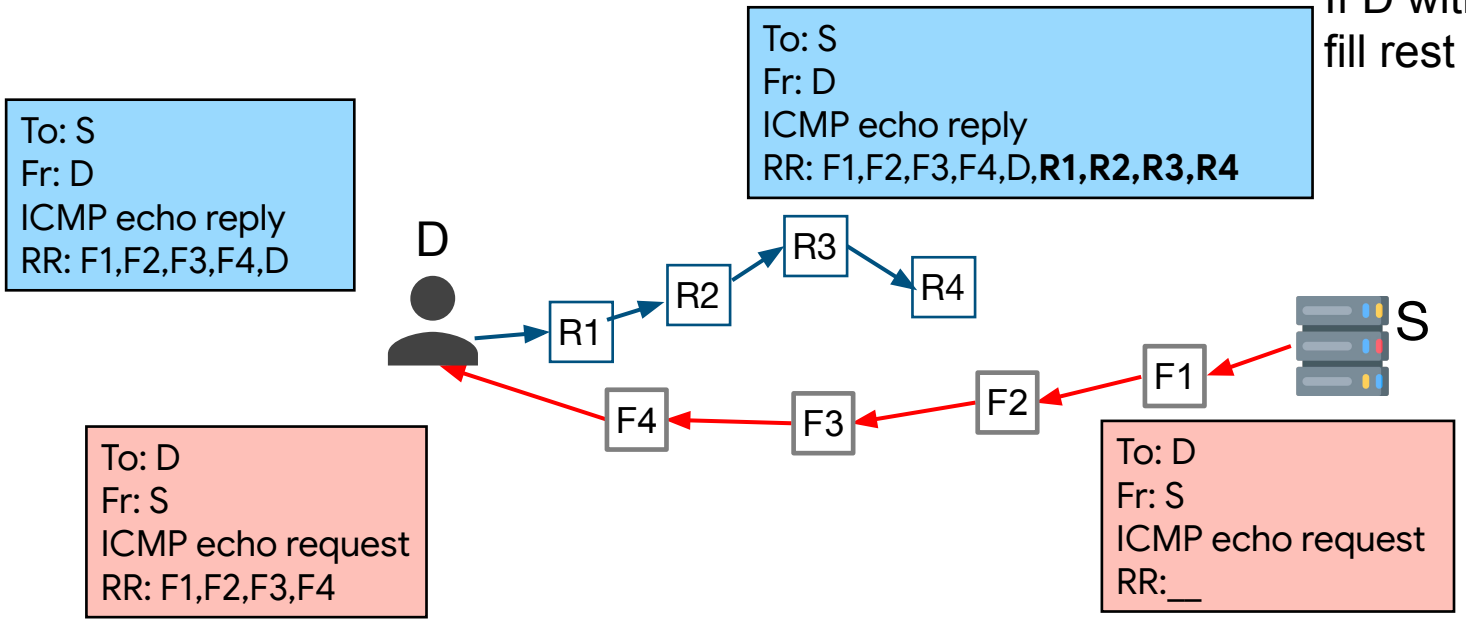
Traceroute from many (exterior) vantage points to S.
Gives **atlas of paths** to S.
If we intersect one, we know rest of path (because of destination-based routing).

Vermeulen et al. Internet Scale Reverse Traceroute 2022 ACM IMC
Katz-Bassett et al. Reverse Traceroute 2010 NSDI

What is Reverse Traceroute?

How do we find reverse hops?

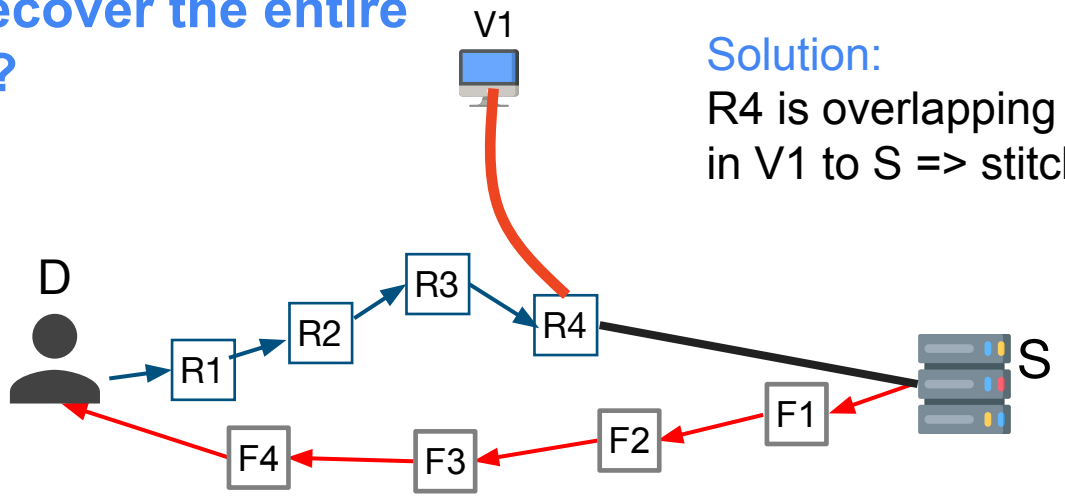
Solution:
IP Options are reflected in reply via the **Record Route IP Option (RR)**.
Record first 9 routers
If D within 8, reverse hops fill rest of slots!



Vermeulen et al. Internet Scale Reverse Traceroute 2022 ACM IMC
Katz-Bassett et al. Reverse Traceroute 2010 NSDI

What is Reverse Traceroute?

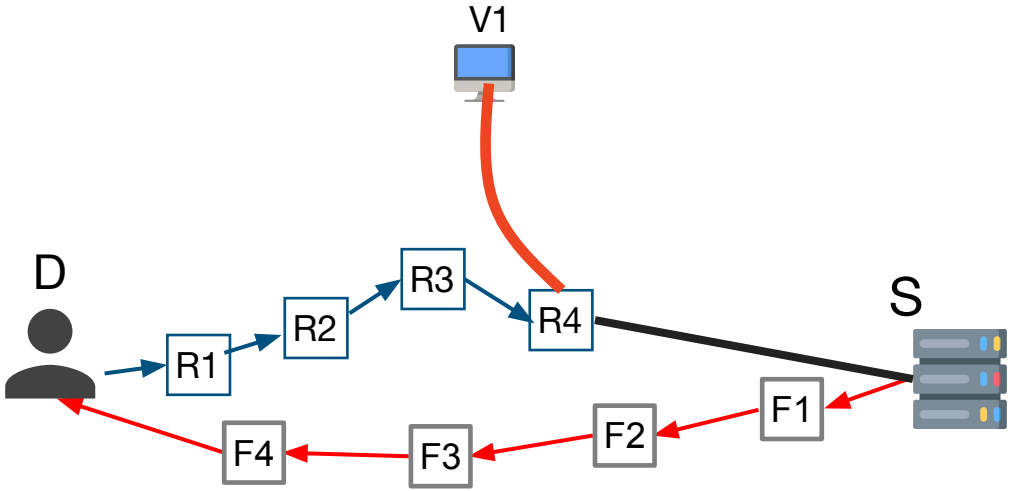
How do we recover the entire reverse path?



Solution:
R4 is overlapping with an IP address in V1 to S => stitch the path back.

Vermeulen et al. Internet Scale Reverse Traceroute 2022 ACM IMC
Katz-Bassett et al. Reverse Traceroute 2010 NSDI

What is Reverse Traceroute?



The details are described in our blog post, which also explains how to access and use the dataset.

RevTr is integrated with M-Lab, and about 25% of NDT speed tests now include reverse traceroute measurements alongside. The results are publicly available in BigQuery! (only works on IPv4 though)

Vermeulen et al. Internet Scale Reverse Traceroute 2022 ACM IMC
Katz-Bassett et al. Reverse Traceroute 2010 NSDI

Project Ideas

Reverse Traceroute

[RevTr - Project 1] General Study

- **Asymmetry analysis:** Using the notebook we shared as part of the blog post, identify cases of severe path asymmetry. How do these affect performance (e.g., by comparing with other measurements from the same city to the same site)? Are the asymmetries persistent? Could faster routes exist, (e.g., by using PeeringDB as a proxy for PoP locations while keeping the same AS path)?
- **Coverage gaps:** Find parts of the network with very low response rates and suggest new Reverse Traceroute sources that could improve coverage.
- **Visualization:** Design new ways to visualize forward and reverse paths at scale.

Project Ideas

Reverse Traceroute

[RevTr - Project 2] WeHeY Integration with RevTr

WeHeY is a system for diagnosing whether an ISP is shaping or throttling traffic. The key idea is to compare performance to two different servers: if both tests share the same access link but diverge upstream, then differences in performance can reveal throttling/shaping inside the ISP.

For this to work, WeHeY needs two guarantees:

1. **Overlap at the client side.** Both tests must share the same path through the client's ISP, so any differences can be attributed to that network.
2. **Divergence upstream.** Beyond the ISP, the tests should be split into different paths, so that comparisons isolate the ISP's influence.

The challenge is that with forward traceroutes alone, it is often unclear whether server pairs truly meet these conditions. Paths may look disjoint but still overlap in the reverse path.

Project Ideas

Reverse Traceroute

[RevTr - Project 2] WeHeY Integration with RevTr

Project idea: Use RevTr data to improve WeHeY's precision and server selection. Reverse visibility can:

- **Reveal hidden overlaps** by showing that two candidate servers share more of the upstream path than expected, avoiding false positives.
- **Confirm clean divergence** by verifying that beyond the ISP boundary, the reverse paths split as intended.
- **Enable smarter server selection** by choosing pairs that truly only share the client's ISP, maximizing the diagnostic power of each comparison.

Participants could quantify how often RevTr changes the overlap/divergence picture, and test whether these insights improve WeHeY's accuracy at scale.

Project Ideas

Reverse Traceroute

[RevTr - Project 3] Poiroot Coverage with RevTr

Poiroot is a system for attributing BGP path changes to the AS that caused them. Suppose traffic from AS A to a destination used to follow one sequence of ASes (the **OldPath**) and now follows a different sequence (the **NewPath**). Poiroot's goal is to figure out which AS was responsible for that change.

To do this, Poiroot cannot rely only on the end-to-end OldPath and NewPath. It also needs to know:

1. **The new paths used by each AS that was on the OldPath.**
2. **The old paths used by each AS that is on the NewPath.**

Without this supporting information, attribution is ambiguous.

In other words, the "before" and "after" paths of the ASes that join or leave the end-to-end path provide the crucial evidence for whether the decision was made by the source, the departing AS, or the newly added AS. Poiroot uses these extra views to isolate the specific AS responsible for the path change.

Project Ideas

Reverse Traceroute

[RevTr - Project 3] Poiroot Coverage with RevTr

Project idea: Evaluate how much RevTr helps fill in these gaps. Reverse Traceroute expands visibility into return paths from many vantage points, which may capture the “before” and “after” routes that Poiroot needs. Hackathon participants could:

- Measure how often RevTr provides the additional paths required for (1) and (2).
- Quantify how attribution success rates improve with RevTr data.
- Identify where coverage remains insufficient and propose where new RevTr sources would add the most value.
- Characterize the prevalence of induced path changes.

IODA

Internet Intelligence Lab

Georgia Tech

What is IODA?

- **Internet Outage Detection and Analysis (IODA)** is an open-source project by the Internet Intelligence Lab at Georgia Tech
- It provides measurements of the connectivity of Internet infrastructure at the country, subnational region and AS level
- Publicly available: *https://ioda.live*

- IODA can be used to identify instances of large-scale disruptions of Internet connectivity

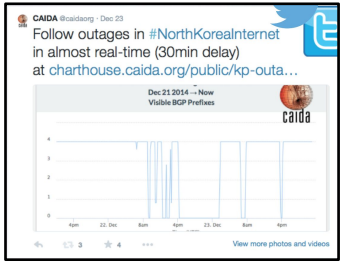
Internet Intelligence Lab | Georgia Tech

IODA

History of IODA

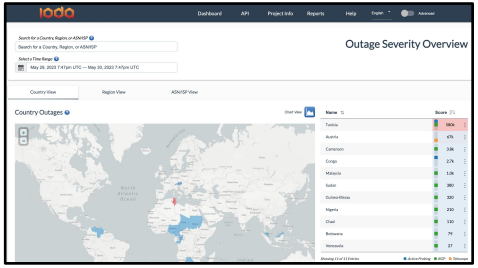
Arab Spring 2011

Experimental research at CAIDA on how to measure the Internet, using Arab Spring as a case study



Dashboard v1 2014

An open-source, publicly available dashboard that provides Internet infrastructure connectivity measurements in near real time



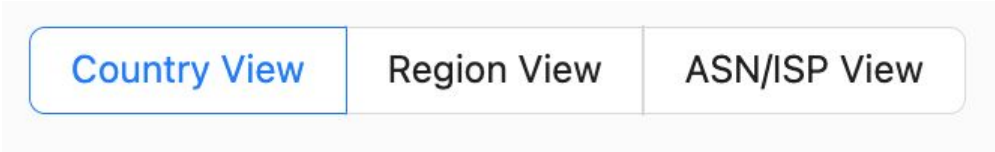
2022 - Present

Ongoing research to improve geographic granularity, measure throttling; user-centered design; community engagement

Internet Intelligence Lab | Georgia Tech

IODA

IODA Levels of Measurement: Country, Region, ISP



Country

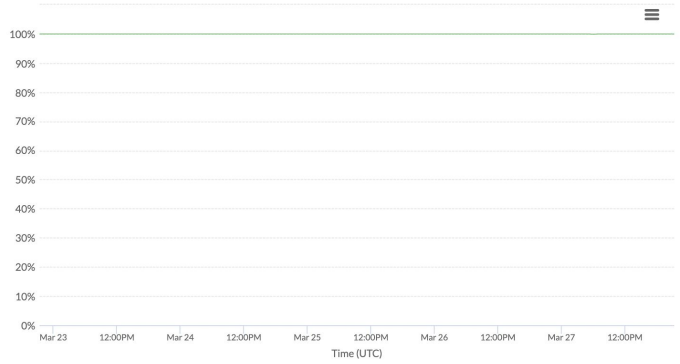
Region

- First-level of subnational administrative divisions (e.g., state, province, etc.)

Internet Service Provider / Autonomous System

New feature! Can now combine to view connectivity of an AS in a specific country/region

IODA signals: BGP



Normal BGP Signal Behavior

Disrupted/ Abnormal BGP Signal Behavior

IODA's BGP signal is calculated by monitoring all updates from Routeviews and RIPE RIS collectors

Data is updated at 5-minute intervals

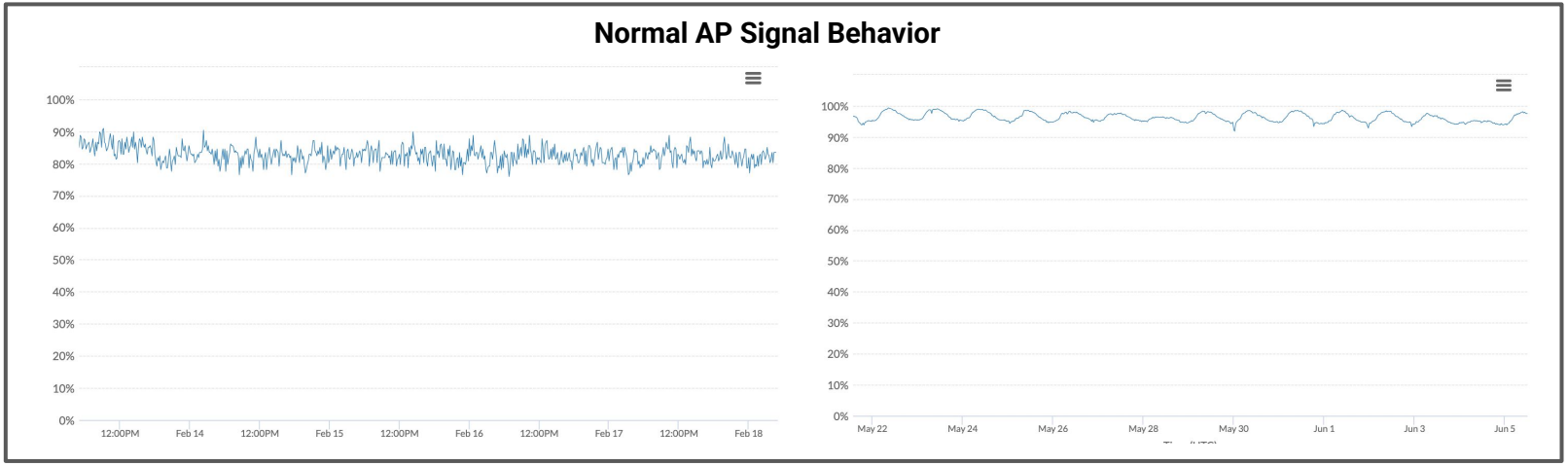
During each interval, IODA calculates the number of "full-feed" peers that observe each prefix

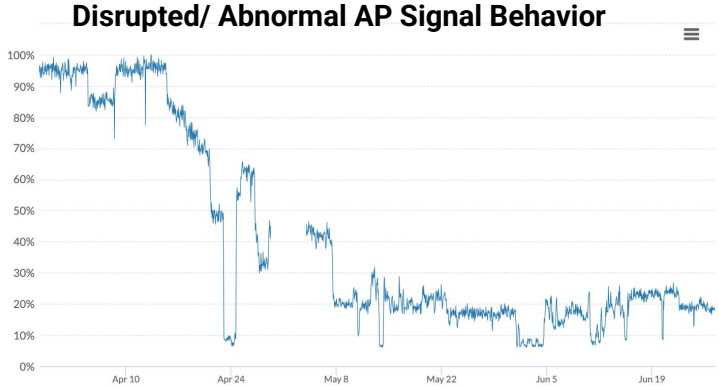
If >50% of full-feed peers observe a prefix, the prefix is considered "visible"

Internet Intelligence Lab | Georgia Tech

IODA

IODA signals: Active Probing





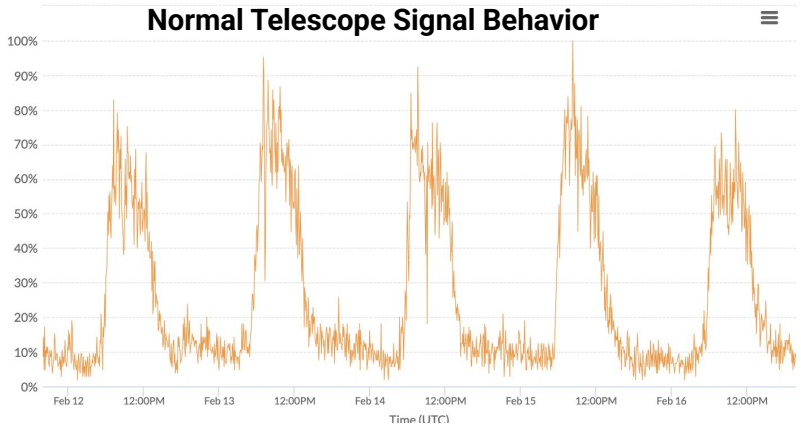
Using an approach similar to Trinocular, IODA continuously monitors network reachability via ICMP pings

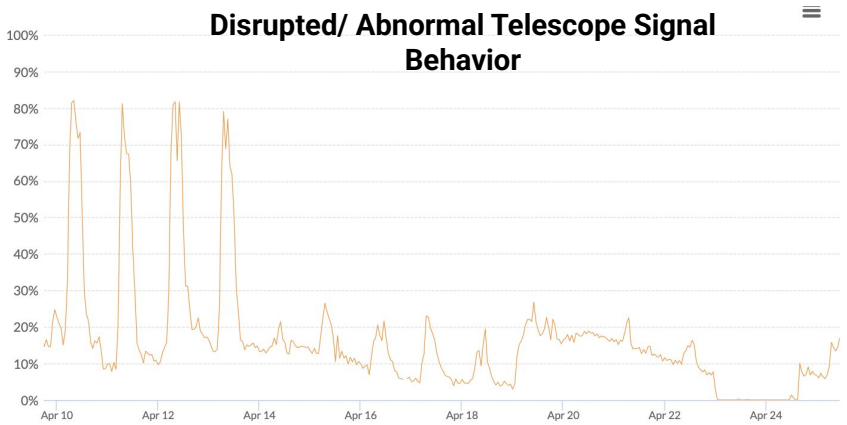
If hosts stop responding, the active probing signal will drop, indicating a possible disruption in connectivity

IODA

Internet Intelligence Lab | Georgia Tech

IODA signals: Telescope



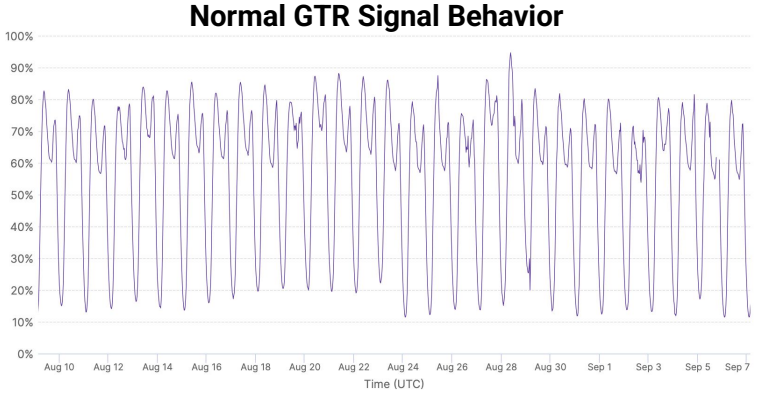


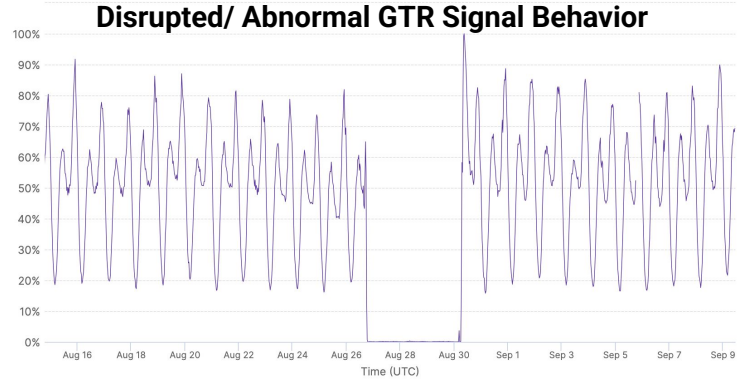
IODA processes Telescope traffic data (unsolicited network traffic captured through dedicated infrastructure)

IODA counts the number of “legitimate” packets and computes the number of unique source IPs per minute, aggregated by country, region, and ASN

Sudden drops below what is normally observed may indicate an outage

IODA signals: Google Transparency Report





Signals for multiple Google products (e.g., Search, Maps, Spreadsheets, etc.)

Each product signal is a normalized value of the number of visits to that Google product, approximately geolocated to the country of the user

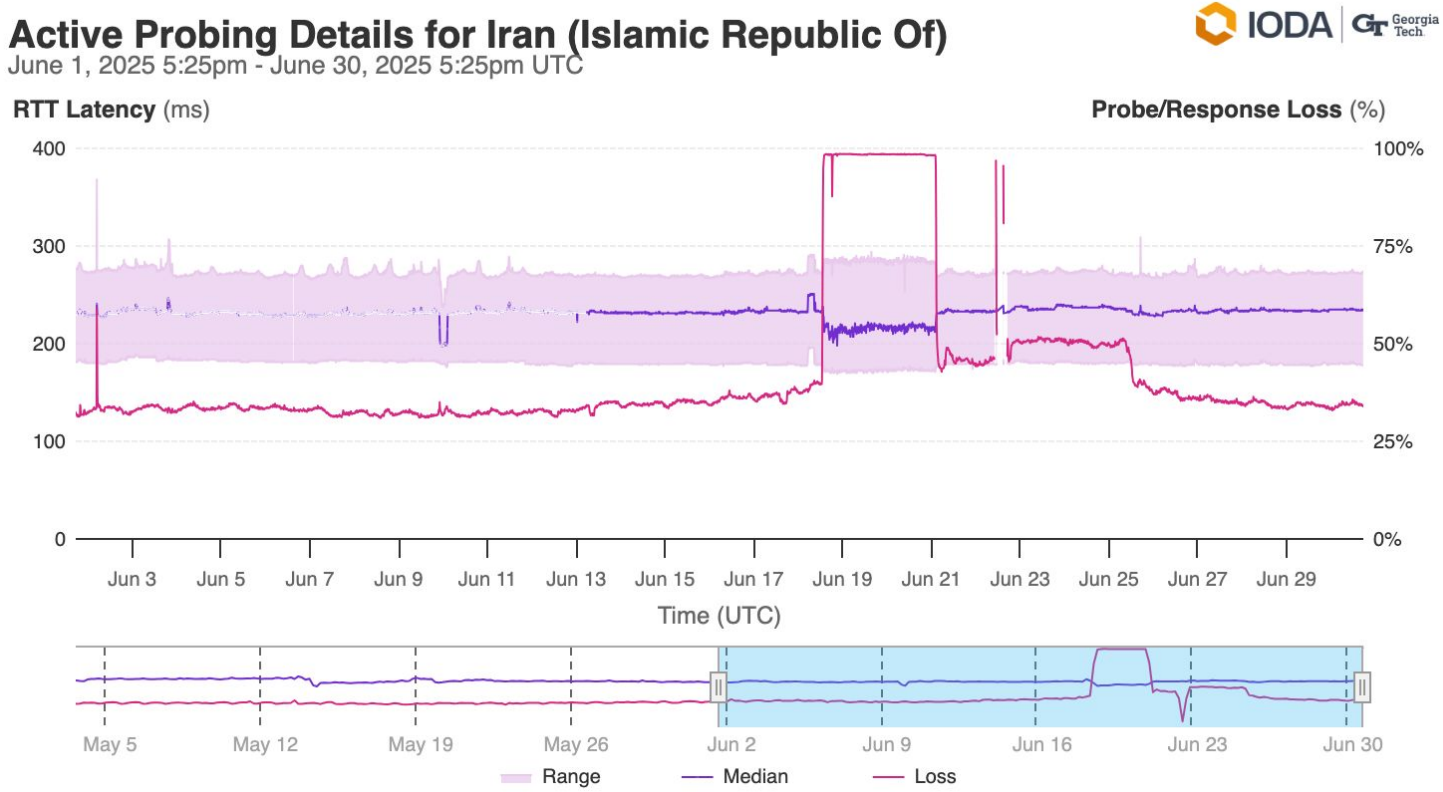
Signal values recorded at 30 minute intervals (with a lag of ~60-90 minutes)

Only available at country view

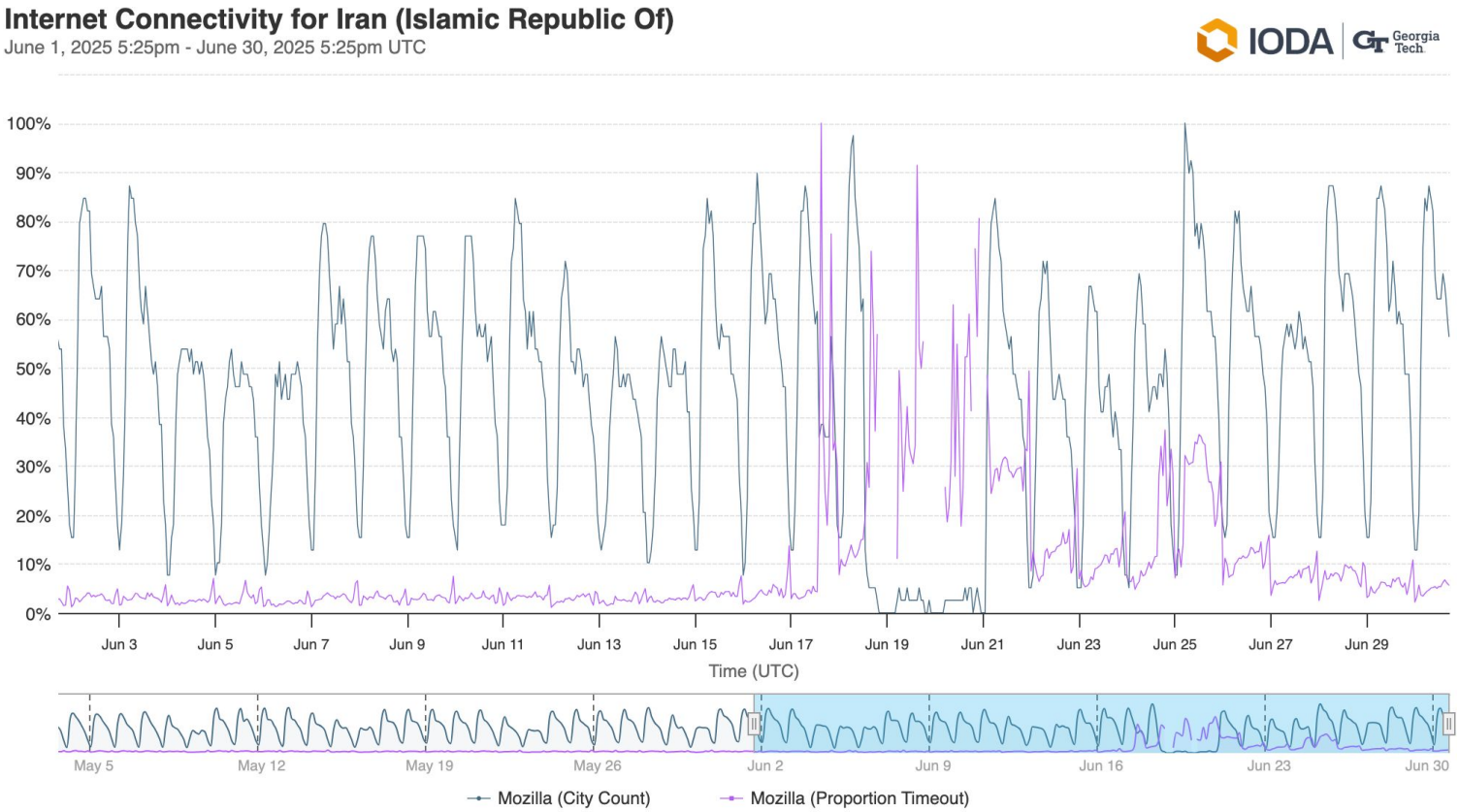
Internet Intelligence Lab | Georgia Tech

IODA

IODA signal: Active probing loss/delay *NEW!*



IODA signal: Mozilla Telemetry (under construction)



IODA Alerts, Events, Overall Outage Scores

Outage Detection	Definition	Data
Alerts	IODA detects that a signal demonstrates an abnormal drop or recovery	Time, signal, actual value, base value
Events	IODA summarizes alerts into an outage event with a severity score	start. end, duration, score
Overall Outage Scores	Events are summarized at the country, region, or AS/ISP level and visualized on a map or time series.	overall outage score, signals associated with the outage, signal level outage score

IODA API

- All signal and event data can be obtained via IODA's public API
  - Requests must be broken into <90-day intervals

- API URL: https://api.ioda.inetintel.cc.gatech.edu/v2/

Project Ideas

[IODA - Project 1] Improved automated outage detection

Improve on IODA's current (simplistic) outage detection algorithm (described here: https://ioda.inetintel.cc.gatech.edu/resources?tab=glossary).

[IODA Project 2] Investigating Internet disruptions in IODA with Internet Yellow Pages

Leverage IYP to investigate disruptions seen by IODA. Create a workflow to identify Internet infrastructure or ASes potentially impacted as collateral damage during an event.

[IODA Project 3] Look for new types of events

Leverage IODA's AP loss/delay or Mozilla Telemetry data to spot new types of disruptions

[IODA Project 4] Cross reference IODA outages with service degradations

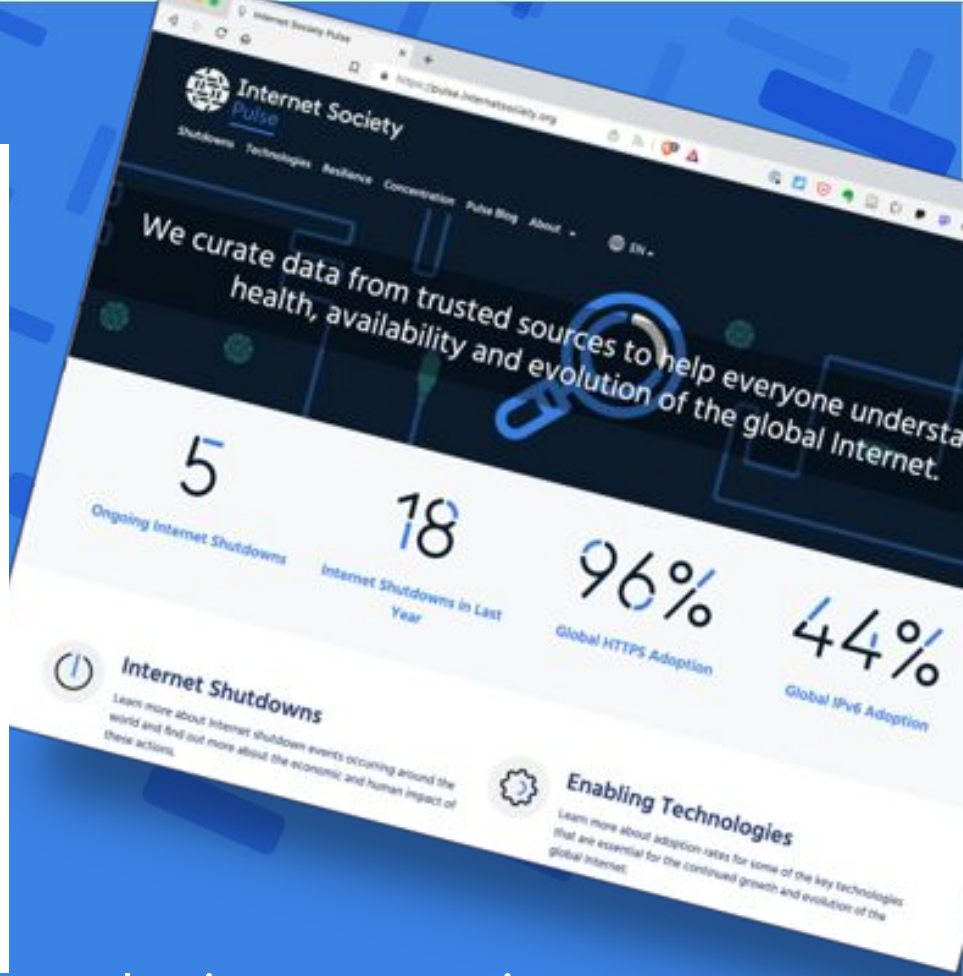
Combine IODA data with performance or path data from other datasets (e.g., Cloudflare Radar, M-LAB, etc.) to study the impact of major network outages.


Internet
Society
Pulse

- Launched December 2020.
- We curate Internet measurement data from trusted sources to help everyone gain deeper, data-driven insight into the Internet.

Trusted data from multiple sources:

- **Benefit**: Helps to assess whether efforts to ensure that the Internet remains open, globally connected, secure, and trustworthy are working.
- **Benefit:** Allows policymakers, researchers, journalists, network operators, civil society groups, and others to better understand the health, availability, and evolution of the Internet.



pulse.internetsociety.org

Pulse Data Partners

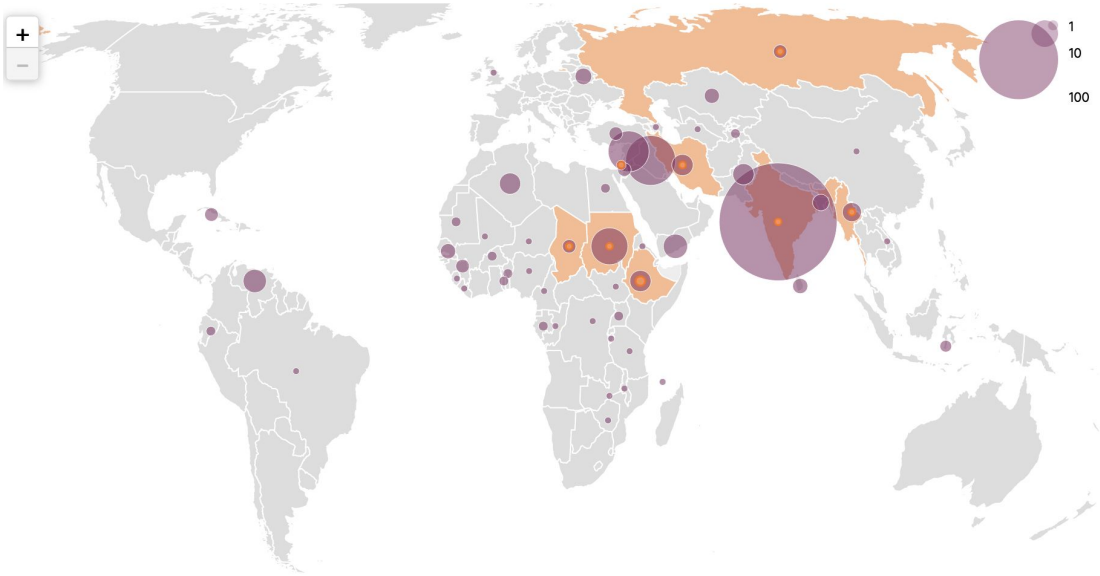
- Data is provided by our trusted data partners

Global Shutdown Trends

Shutdowns tend to occur in response to several factors:

- Civil unrest and protests
- Armed conflict
- Elections
- National or regional exams

Shutdowns tend to be performed at the national level, although in India shutdowns are often ordered by regional governments.



Location of Shutdowns from 10/2019 to 3/2024

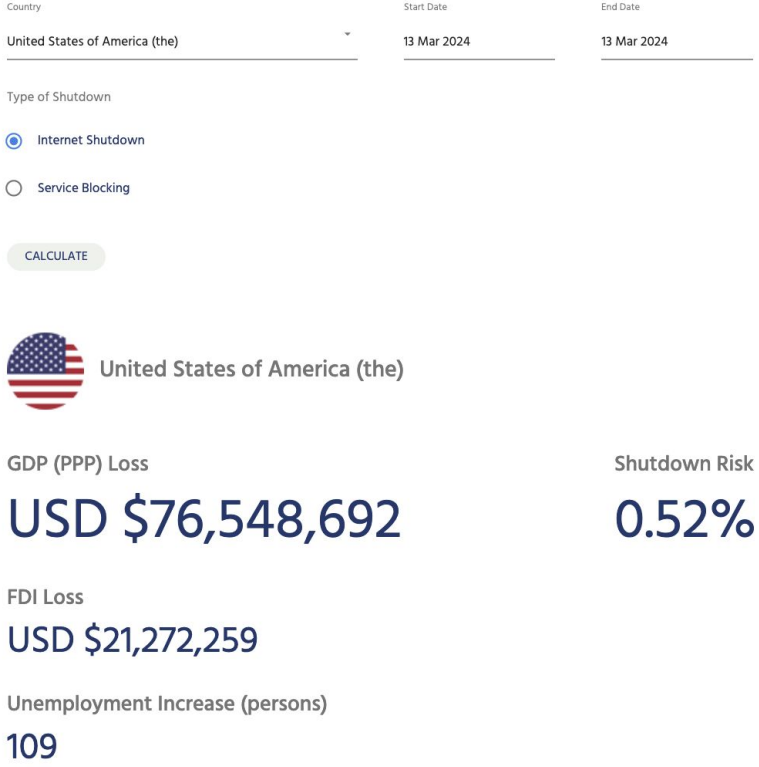
NetLoss: Calculating the Cost of Shutdowns

Our NetLoss tool allows users to estimate the economic cost of an Internet shutdown in a country or territory.

NetLoss helps Internet advocates make the point to governments that shutting down the Internet is harmful to their economy.

Important notes:

- NetLoss uses an economic framework to estimate the impact of Internet shutdowns on a range of economic, social, and other outcomes and uses econometric tools to provide a rigorous estimate of the economic impact of a given shutdown. But it is an estimate.
- Estimates the cost of national shutdowns, not regional shutdowns.



Enabling Technologies

HTTPS

96%

Current percentage of top 1000 websites globally that support HTTPS.

IPv6

48%

Current percentage of top 1000 websites globally that support IPv6.

TLS 1.3

81%

Current percentage of top 1000 websites globally that support TLS 1.3.

DNSSEC

31%

Percentage of ccTLD registries with operational DNSSEC and global DNSSEC validation rate (data sources: DNS, APNIC)

ROA Coverage

43%

Percentage of address space covered by ROA (data source: APNIC)

HTTP/3

23%

HTTP/3 adoption (data source: Mozilla Firefox Telemetry)

Concentration

- **Market Concentration:** The concentration of providers in a given market
- **Country Market Shares:** The jurisdiction of providers in a given market.

Gini HHI

All Top 10,000 sites Top 1,000 sites

Data Center | Gini **0.67**
Data center providers supply hardware and software infrastructure to serve websites on the Internet.

Top Level Domain | Gini **0.77**
Top Level Domains are the highest level in the hierarchical Domain Name System (DNS) of the Internet.

SSL Certificate | Gini **0.66**
SSL certificate authorities are trusted institutions that issue SSL certificates for verifying the owner of a website and encrypting web traffic with SSL/TLS.

DNS Server | Gini **0.72**
DNS (domain name system) servers manage mappings between Internet domain names and their associated records such as IP addresses.

CDN | Gini **0.86**
Content Delivery Networks (CDNs) are geographically distributed networks of proxy servers and their data centers.

Web Hosting | Gini **0.64**
A web hosting service provides hardware and software infrastructure to enable webmasters to make their websites accessible via the Internet.

IXP Tracker

Active Internet Exchange Points

The total number of IXPs in operation in Kenya, as of October 2024.

8

Active IXPs

33.00 %

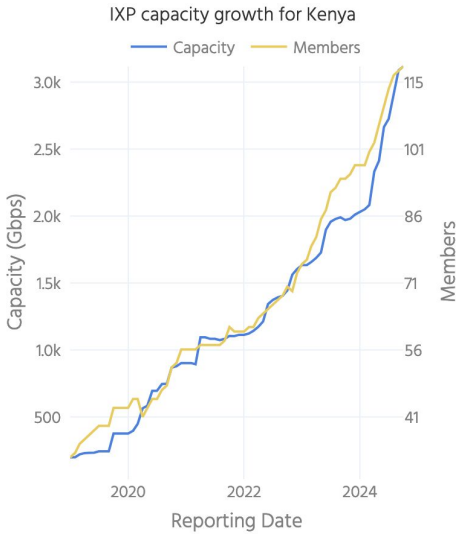
Proportion of the local Internet that can be reached through IXPs in this country.

IXPs in Kenya

IXP Name	▲ Location
Asteroid Mombasa	Mombasa
Asteroid Nairobi	Nairobi
Kenya Internet Exchange Point - Nairobi - KIXP - Nairobi	Nairobi
KIXP - Mombasa icolo	Mombasa
KIXP - Nairobi icolo	Nairobi
KIXP - Nairobi PAIX	Nairobi
LINX Nairobi	Nairobi
Nairobi Internet Exchange - Nairobi-IX	Nairobi

IXP capacity growth over time in Kenya

The total of IXPs over time, shown along with the growth in combined capacity.



Country Reports: Open Internet

Internet Use
Individuals using the Internet as a percentage of the total population

44%
Regional
Rank: 41

71%
Asia avg.

Internet Shutdowns
Intentional disruptions of Internet communications, making them unavailable for a specific population, location, or type of access

1
Ongoing

0
Last 12 month

Read more about Internet Shutdowns

Internet Resilience Score
A resilient Internet connection is one that maintains an acceptable level of service in the face of faults and challenges to normal operation

45%
Regional
Rank: 32

46%
Asia avg.

See details

IXP Operator Market
A measure of the diversity and concentration of the local market for Internet Exchange Point operations

Retail ISP Diversity
Diversity of retail Internet providers improves resilience and user choice

Very Good

Transit Provider Diversity
More diversity in routes to the global Internet improves connection resilience

Poor

Internet Freedom
Freedom on the Net measures Internet freedom in 70 countries

Not Free

See details on freedomhouse.org

Popular Content Locality
A measure of how much locally popular web content is hosted in-country or in-region

44%
Regional
Rank: 25

33%
Africa avg.

See details

The Internet Resiliency Index (IRI)

The framework collates around 30 sets of public metric data that relate to **four pillars** of a resilient Internet:

.

Infrastructure

The existence and availability of physical infrastructure that provides Internet connectivity.

Performance

The ability of the network to provide end-users with seamless and reliable access to Internet services.

Security

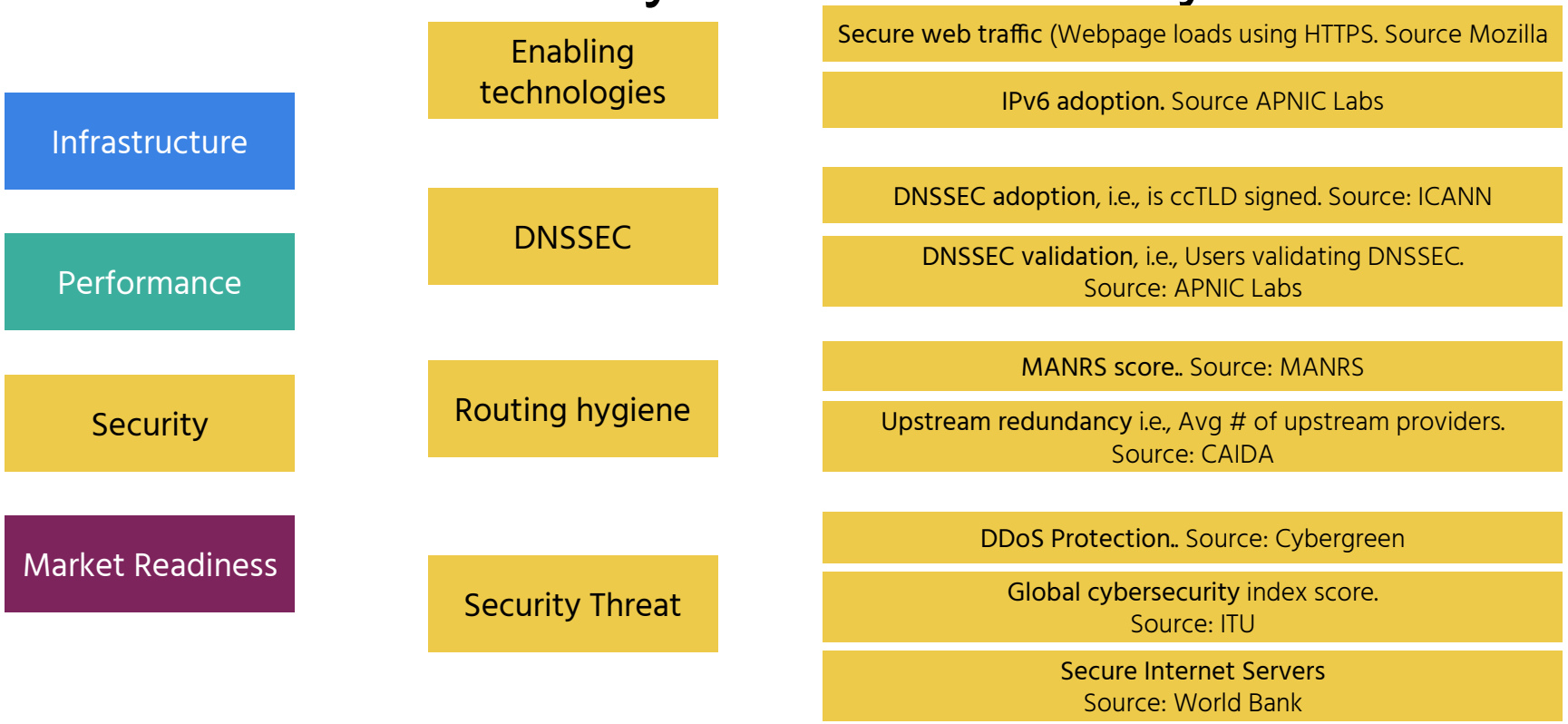
The ability of the network to resist intentional or unintentional disruptions through the adoption of security technologies and best practices.

Market Readiness

The ability of the market to self-regulate and provide affordable prices to end-users by maintaining a diverse and competitive market.

Methodology: https://pulse.internetsociety.org/wp-content/uploads/2023/07/Internet-Society-Pulse-IRI-Methodology-July-2023-v2.0-Final-EN.pdf

The Internet Resiliency Index — Security



IRI API

v1

- GET /v1/affordability/
- GET /v1/cybersecurity-index/
- GET /v1/data-centers/
- GET /v1/ddos-full/
- GET /v1/ddos/
- GET /v1/dnssec-validation/
- GET /v1/dnssec/
- GET /v1/domains/
- GET /v1/egdi/
- GET /v1/electricity/
- GET /v1/exit-points/
- GET /v1/fibre/
- GET /v1/hegemony/

Project ideas

- **Visualizing the Internet Resilience Index (IRI) data** - Using the IRI API get access to the historical data and create visualization
- **Detecting IXPs on traceroute data** - For a study on content locality, we want to identify whether there is an IXP on the path between the user and the content.
- **Internet fragmentation** - come up with a metric or list of metrics to measure Internet fragmentation and a way to visualize it.
- **Internet concentration** - expand the work from Kashaf et al. to develop a country-level Internet concentration index.
- **IXP Saturation** – define a set of metrics that can help us calculate the saturation of a market in terms of IXP.