Measuring the Tor network

hiro arma

August 2024

The Tor Project




By Hamed Saber

Today's talk

(1) What is Tor, a brief introduction

(2) The Tor relay community—understanding and growing the Tor network

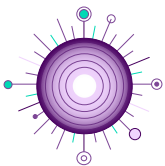
(3) The Tor network in metrics

(4) How to get involved

What do we mean by *Tor*?



Ecosystem of privacy preserving technologies



Non-profit organization



Global community of volunteers and supporters

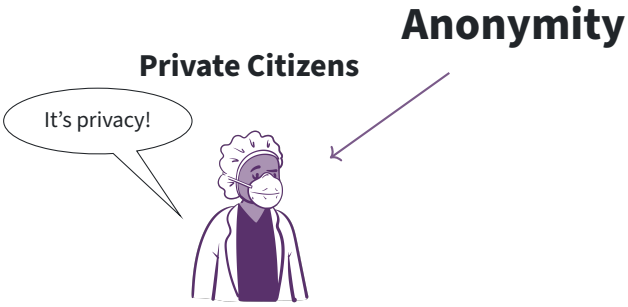
Communications Metadata



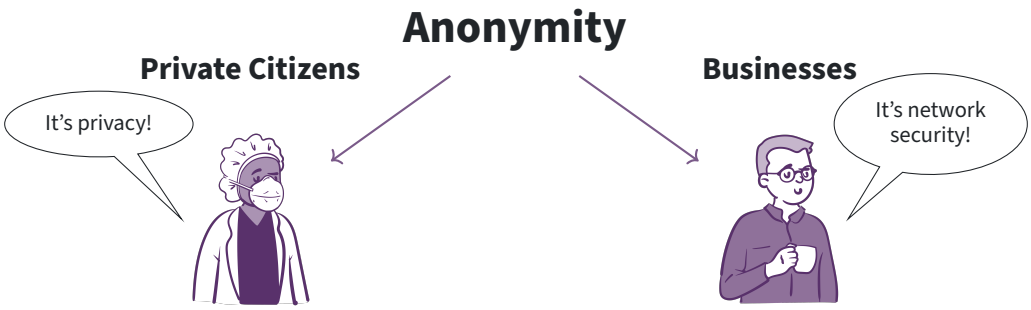
“We Kill People Based on Metadata.”

—Michael Hayden, former director, NSA

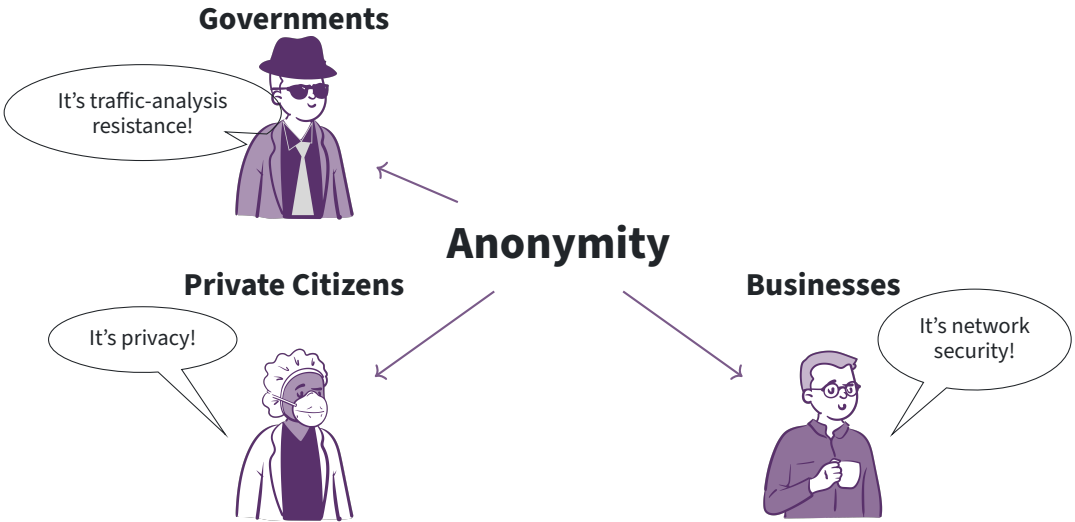
Anonymity is different for each use case



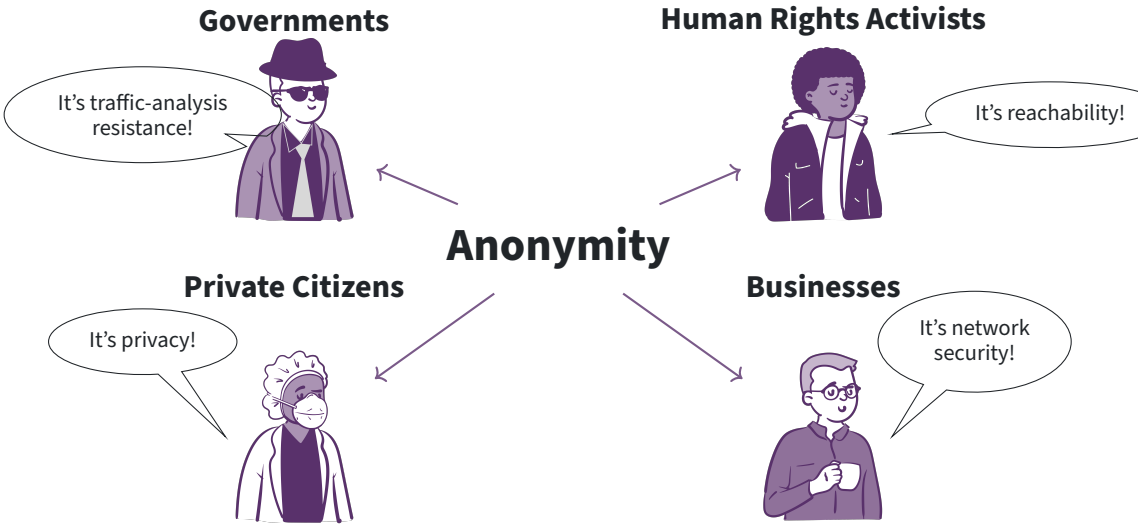
Anonymity is different for each use case



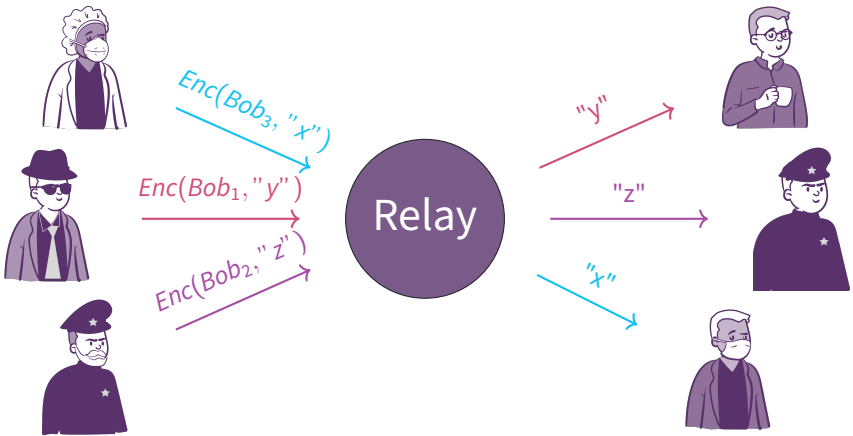
Anonymity is different for each use case



Anonymity is different for each use case

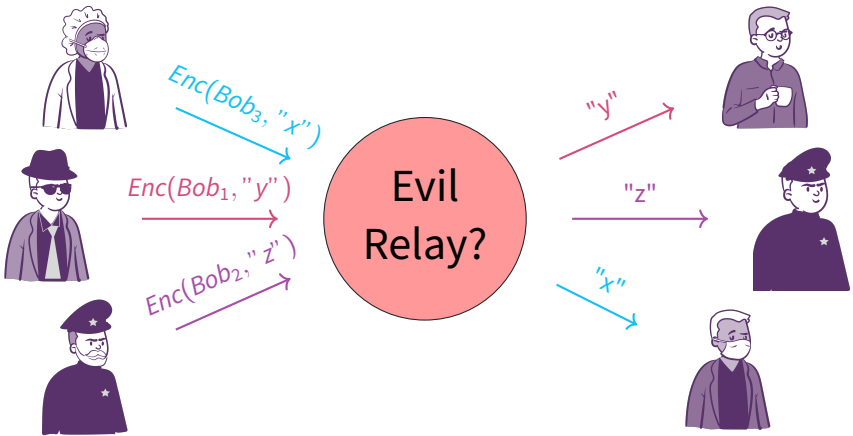


A Simple Design

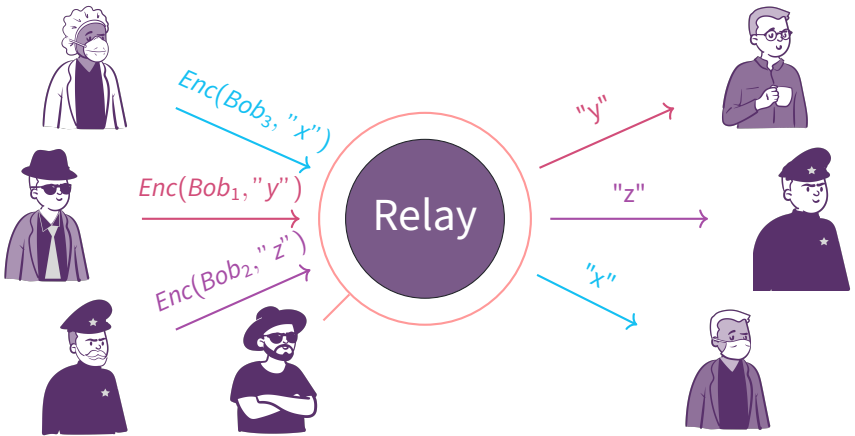


Equivalent to most commercial proxy/VPN providers

A Simple Design



A Simple Design



Timing analysis lets an observer match up connections

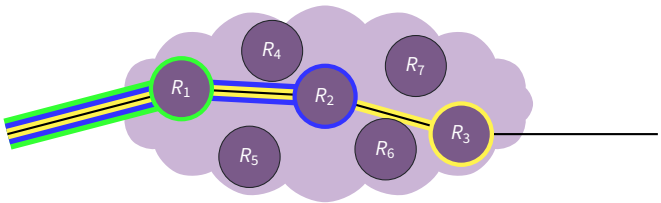
The Tor Design

Alice

Anonymity Network

Bob







Multiple relays so no single relay can link Alice to Bob

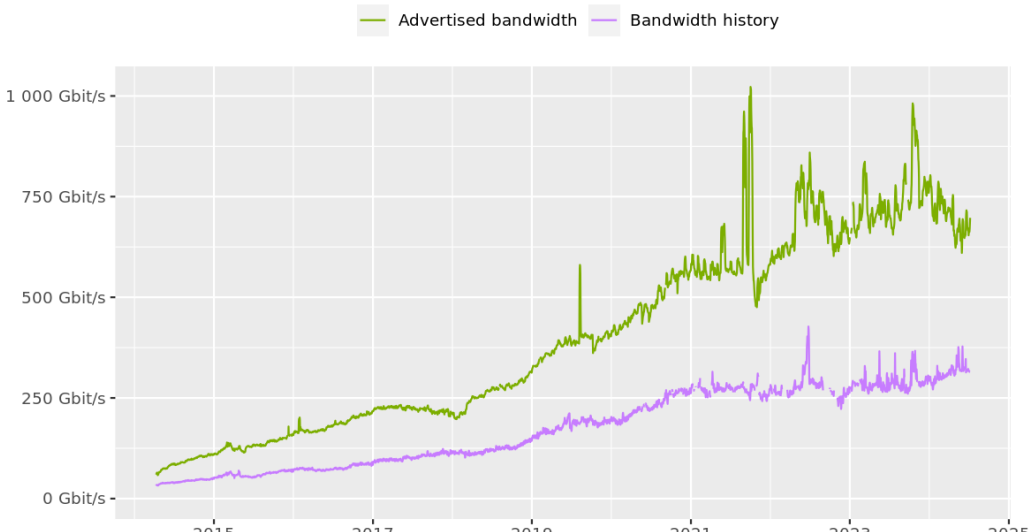
Transparency for Tor is key

- Open source / free software
- Public design documents and specifications
- Publicly identified developers

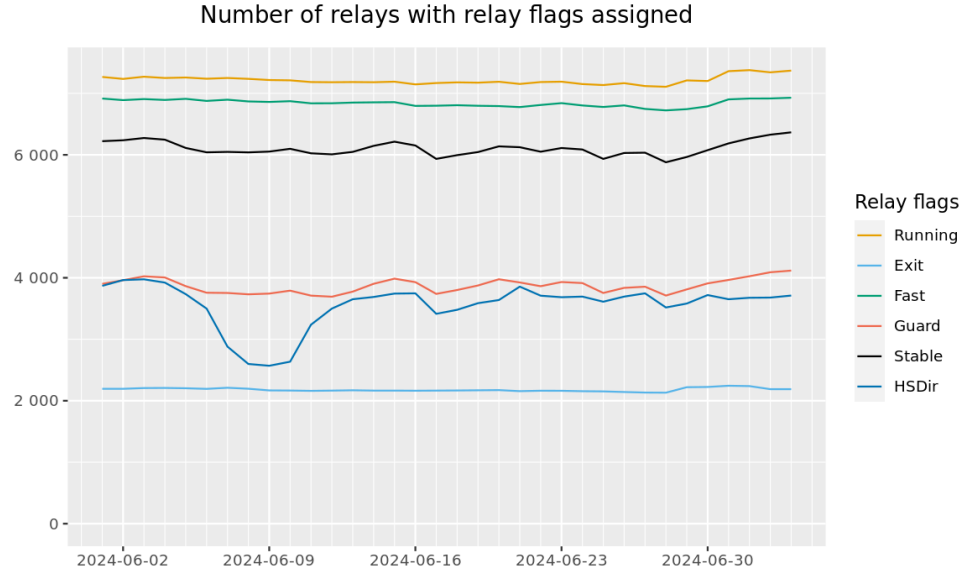
Transparency for Tor is key

- Open source / free software
- Public design documents and specifications
- Publicly identified developers
- Not a contradiction: privacy is about choice!

Total relay bandwidth



Understanding the Tor network



Understanding the Tor network

The Tor network is driven by a diverse **community**

Diversity within the Tor network is a critical component encopassing various desired properties

Types of hardware, software, configurations, location, users

Understanding the Tor network

Fostering a healthy community

- **direct** communication
- online and in person **meetups**
- **transparent policies**
  - community.torproject.org/policies/relays

Relay landscape, past/present/future

For the first decade, many relay operators had met me at a conference

Then the set of reasons to offer well-funded harmful relays expanded

(1) Need to improve applications, e.g. https by default

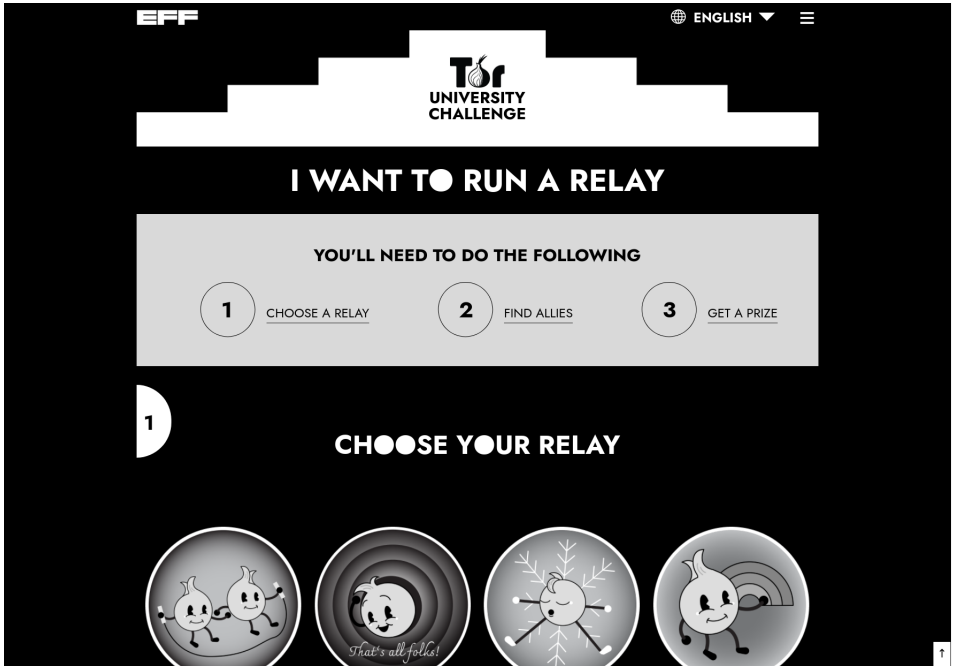
(2) One of the few tools we have for safety in this asymmetric situation is our social graph

Relay operator associations

Many Tor exit relays are run by nonprofits that people created to operate relays in their country

Successful in Switzerland, Netherlands, US, Germany, Sweden, France, Canada, Austria, Denmark, more

Even better is that they're part of our community: they're connected to hacker spaces and we met them at conferences like this one


EFF
ENGLISH
Tor
UNIVERSITY
CHALLENGE
I WANT TO RUN A RELAY
YOU'LL NEED TO DO THE FOLLOWING
1 CHOOSE A RELAY
2 FIND ALLIES
3 GET A PRIZE
1
CHOOSE YOUR RELAY
That's all folks!

Relay incentives

Gamification / badges / leader boards

Meetups, attentive operator support

Structural incentives, like better speed for relay operators, or everybody-a-relay

...Relay operator payments?


EL PROBLEMA
ES EL CAPITALISMO
Yo
una conversación
casual

The Tor network in metrics

In May 2020, we removed a group of Tor exit relays that were interfering with exit traffic by intercepting connections to a small number of cryptocurrency exchange websites

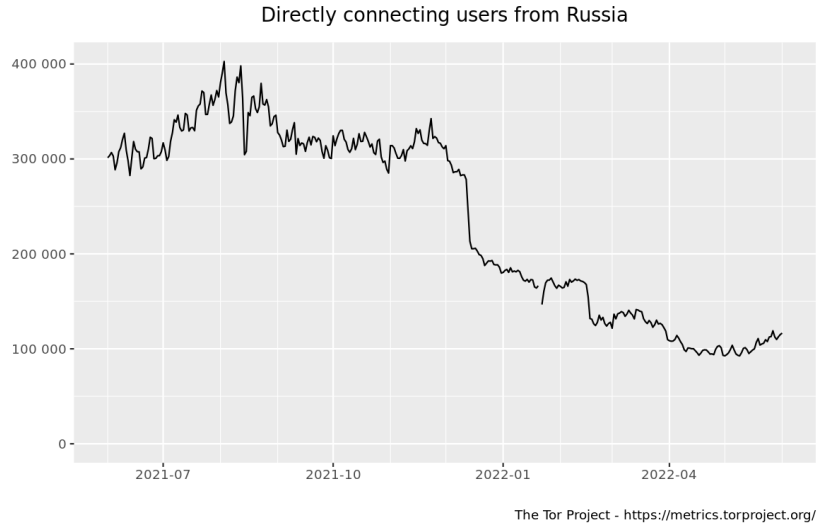
[*] blog.torproject.org/bad-exit-relays-may-june-2020

The Tor network in metrics

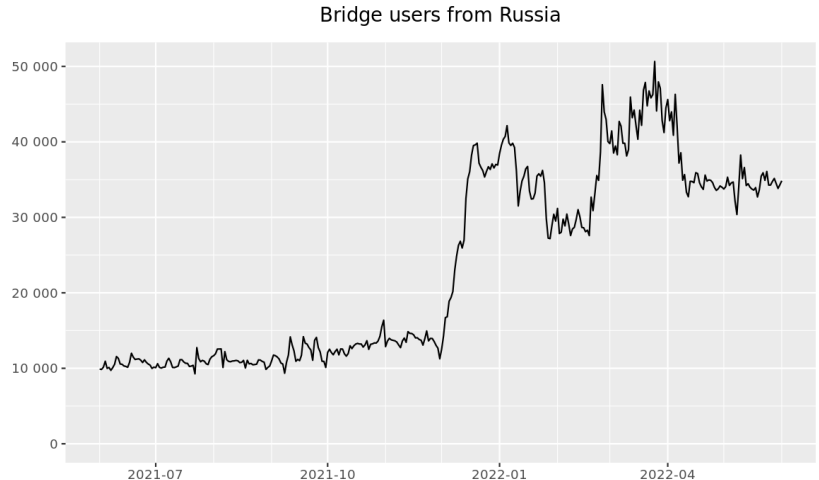
The Network Health team is dedicated to monitoring our network, encompassing a joint effort between various teams at Tor

The foundation of this work is the metrics that provide us with the status of the network

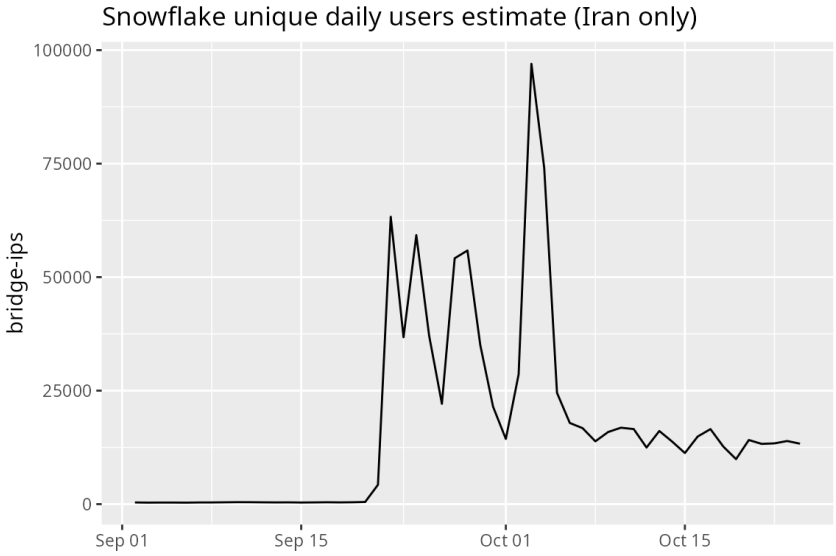
Start of the Ukraine invasion 2021-2022



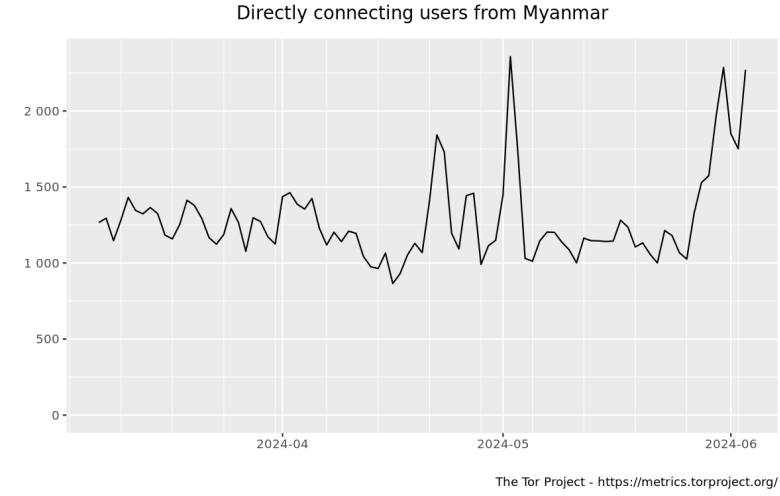
Start of the Ukraine invasion 2021-2022



Mahsa Amini protests in Iran 2022

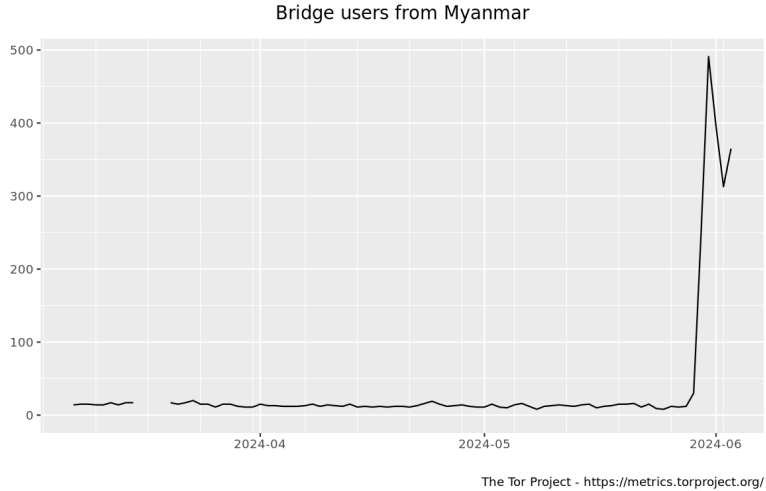


Internet censorship in Myanmar May 2024



gitlab.torproject.org/tpo/anti-censorship/censorship-analysis/-/issues/40043

Internet censorship in Myanmar May 2024



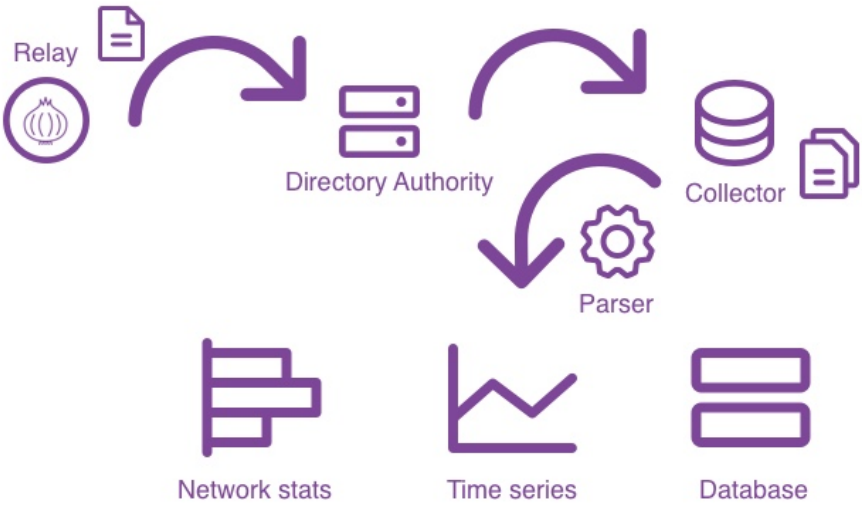
gitlab.torproject.org/tpo/anti-censorship/censorship-analysis/-/issues/40043

The Tor network in metrics

Do not collect data that cannot be public

- Do not collect in client data
  - Collect aggregated data from relays
- Do keep historical data (collector.torproject.org)
  - Network wide statistics
  - Status of individual nodes (bridges, relays)
- Accessible data and algorithm transparency

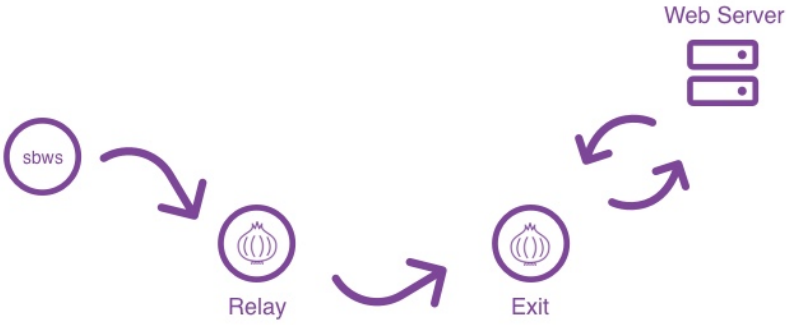
The Tor network in metrics



The Tor network in metrics

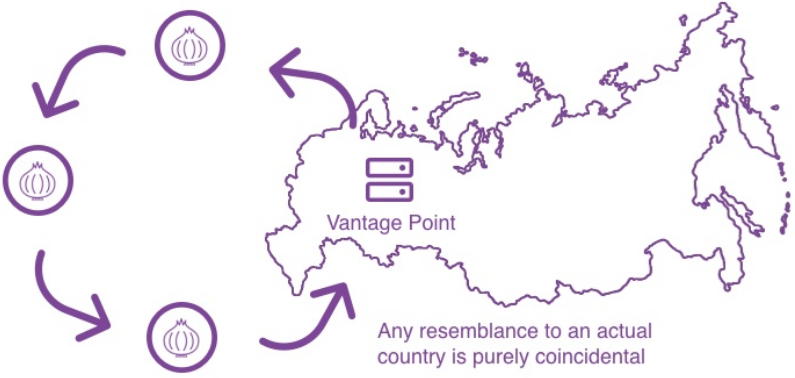
- Traffic/usage patterns
- Bandwidth measurements
- Performance measurements

Bandwidth probing



Bandwidth is measured from the bandwidth authority by creating a circuit to the relay and downloading a file from a web server

Performance measurements



Runs a Tor client in a censored region, spins up a web server and downloads 5MB files over the Tor network

metrics-api.torproject.org

Metrics data is open and accessible, but there is a learning curve

Our new API lowers the barrier for researchers to start working with Tor network data

Few endpoints (more coming up soon):

- metrics-api.torproject.org/details
- metrics-api.torproject.org/summary

metrics-api.torproject.org


```
{
    "nickname":"seele",
    "fingerprint":"000A10D43011EA4928A35F610405F92B4433B4DC",
    "last_seen":"2024-08-02 13:00:00",
    "running":true,
    "flags":["Fast","HSDir","Running","Stable","V2Dir","Valid"],
    "country_name":"United States of America",
    "as":"AS7018",
    "as_name":"AT&T Services, Inc.",
    "last_restarted":"2024-07-18 11:59:51",
    "bandwidth_rate":1048576,
    "bandwidth_burst":2097152,
    "observed_bandwidth":1062365,
    ...
}
```


gitlab.torproject.org/tpo/network-health/metrics/networkstatusapi/-/tree/main/docs

Get involved: run a relay!

More relays means better performance for users

And more good relays dilute any bad relays

Location matters, among many other diversity factors

Learn about exit policies, and learn about IP blocklists

Get involved: run a relay!

https://community.torproject.org/

Join the community, come to the relay operator meetups

If you can’t run a relay, run a bridge or a Snowflake!

Get Involved: Help with Network Health!

- Network data analysis
- **Tor Weather** – A tool for relay operators
- Get involved with the redesign of our metrics data pipeline
  - **Network Status API** – Access real-time network data.
  - **descriptorParser** – Parse network documents in Java.
  - **TagTor** – Visualize network data and metrics.
  - **tor_fusion** – Parse network documents in Rust.
- gitlab.torproject.org/tpo/network-health/team/-/wikis/home#how-to-get-involved