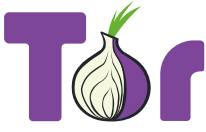
OSINT work at Tor

Georg Koppen, The Tor Project

July 18, 2025

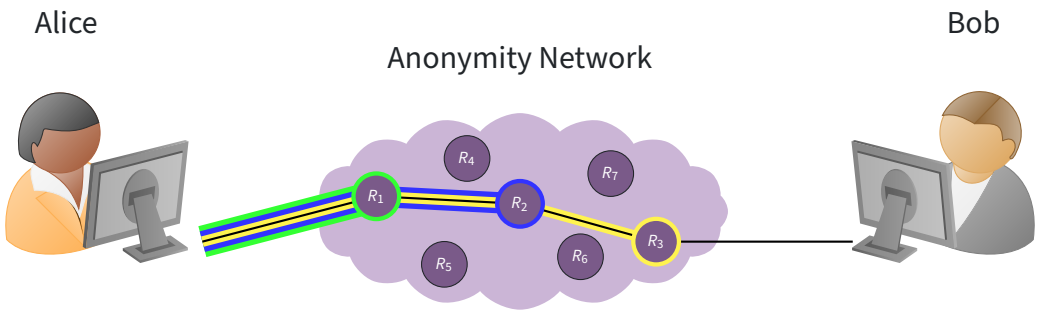
Bornhack 2025



What is Tor?

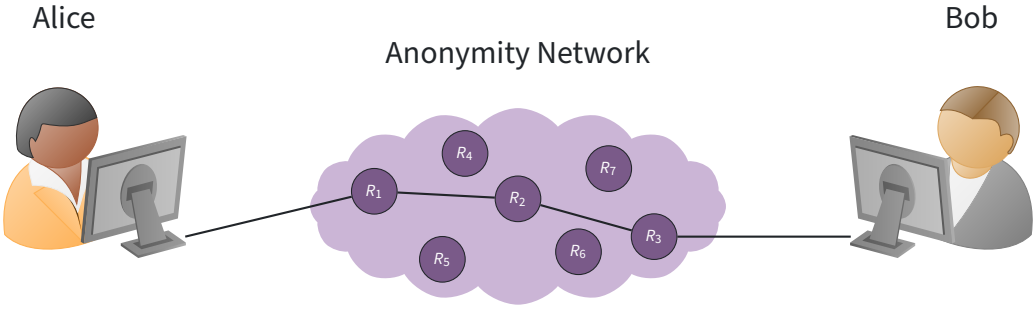
- Online anonymity, and censorship circumvention
  - Free software
  - Open network
  - Open specifications
- Community of researchers, developers, users, and relay operators
- U.S. 501(c)(3) non-profit organization

The Tor design



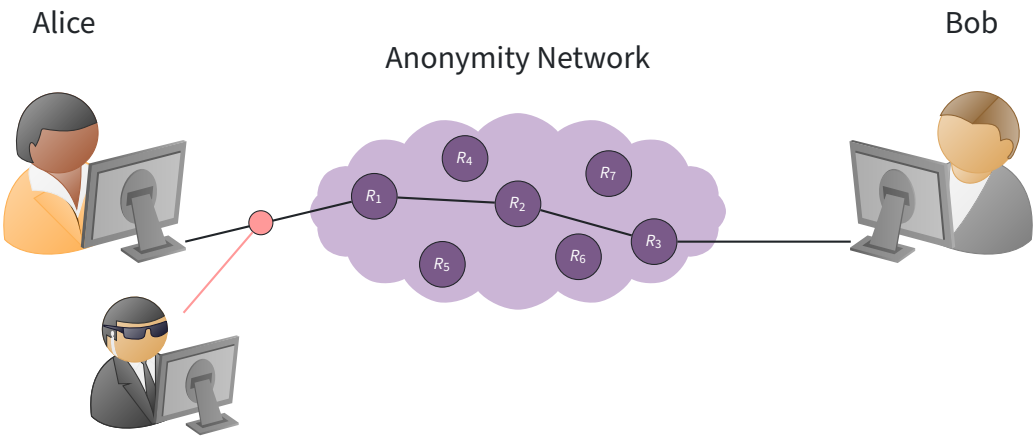
Multiple relays, so no single relay can link Alice to Bob

Threat model

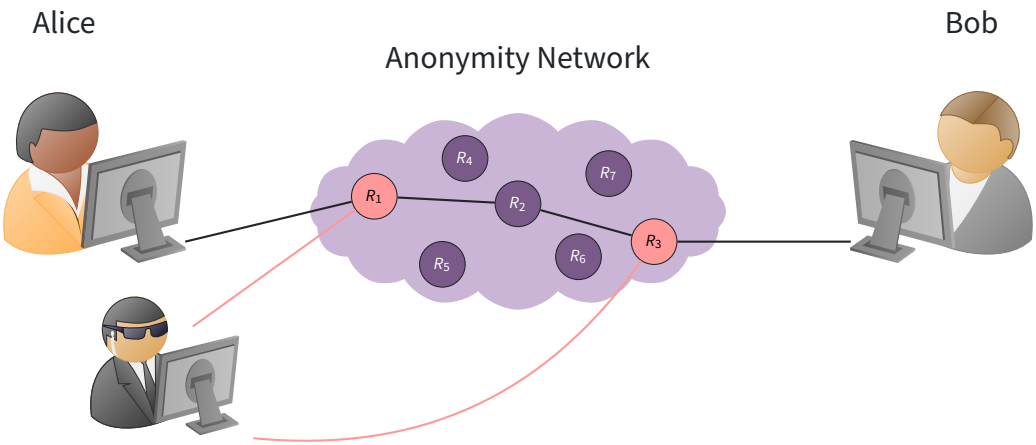


What can the attacker do?

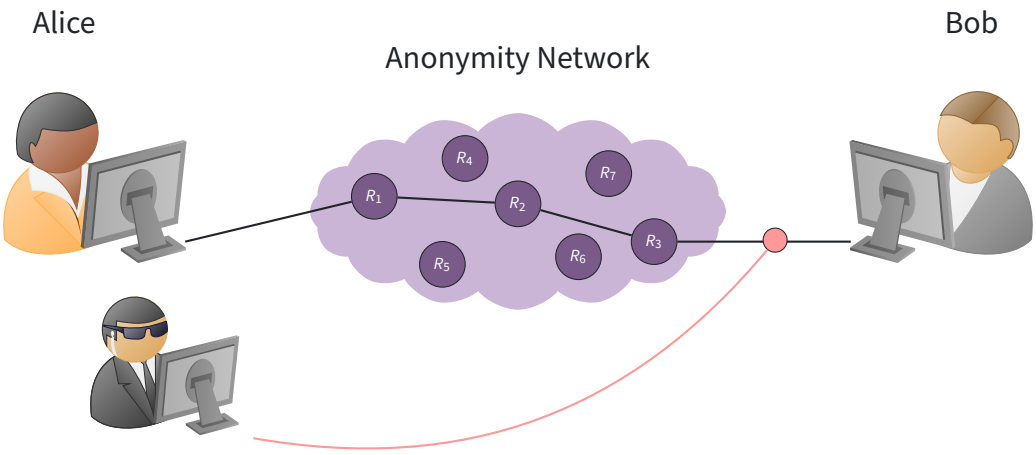
Threat model



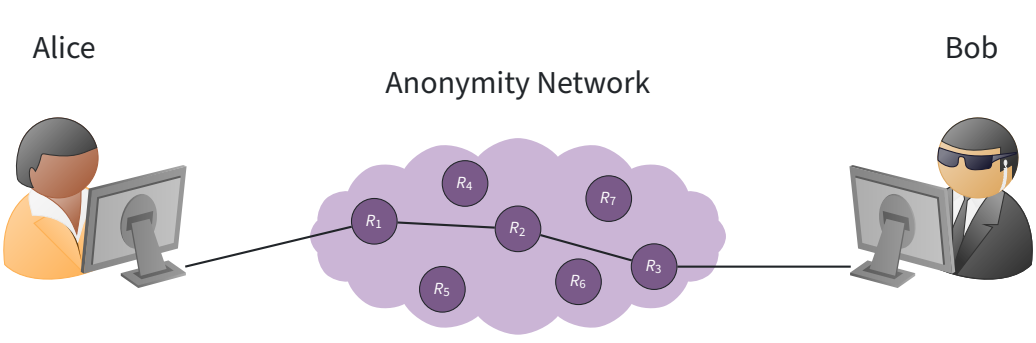
Threat model



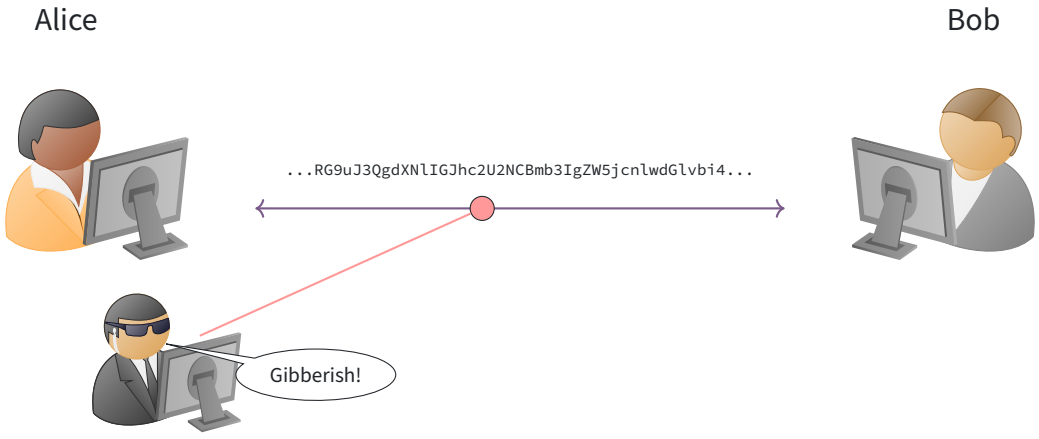
Threat model



Threat model

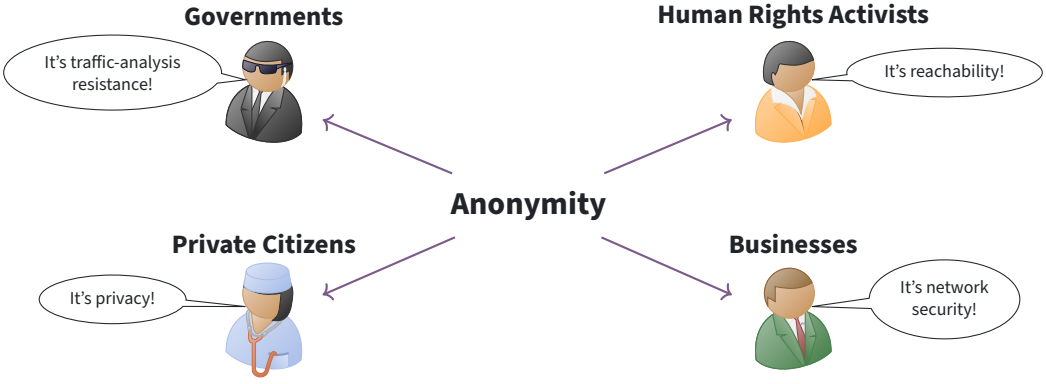


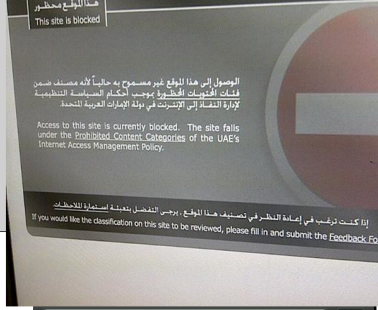
Anonymity isn’t encryption



Encryption just protects contents

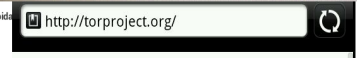
Anonymity is different for each use case





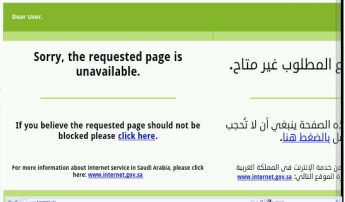
Access Denied

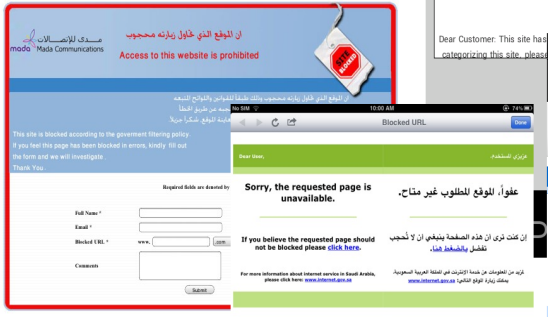
our request was denied because of its content categorization: "Computers/Internet;Proxy Avoida

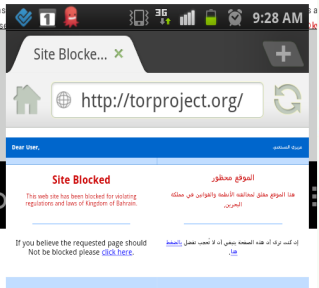


عزيزي العميل: تم حجب هذا الموقع بناء على القوانين والقوانين
بأن هناك خطأ يرجى إرسال رسالة على البريد الالكتروني unblock.kw@kw.zain.com مع ذكر عنوان الموقع الذي تم حجبه.

Dear Customer: This site has
categorizing this site, please







Du var på vej ind på en ulovlig hjemmeside

Vi vil meget gerne hjælpe dig med at finde den film eller serie, du søger.

Søg med FilmFinder →

Hvis du er på udkig efter musik, bøger eller møbler

Gå til SHARE WITH CARE →

SHARE WITH CARE

Hjemmesiden er blevet blokeret, fordi den er dømt ulovlig ved en dansk domstol. Brug Share With Care til at finde det, du leder efter, lovligt. På den måde passer du både på dig selv og på kulturen. **Læs mere om Share WIth Care**

What is OSINT?

Open-Source Intelligence (OSINT), broadly defined, involves gathering and analyzing publicly accessible information to produce actionable insights. (Wikipedia)

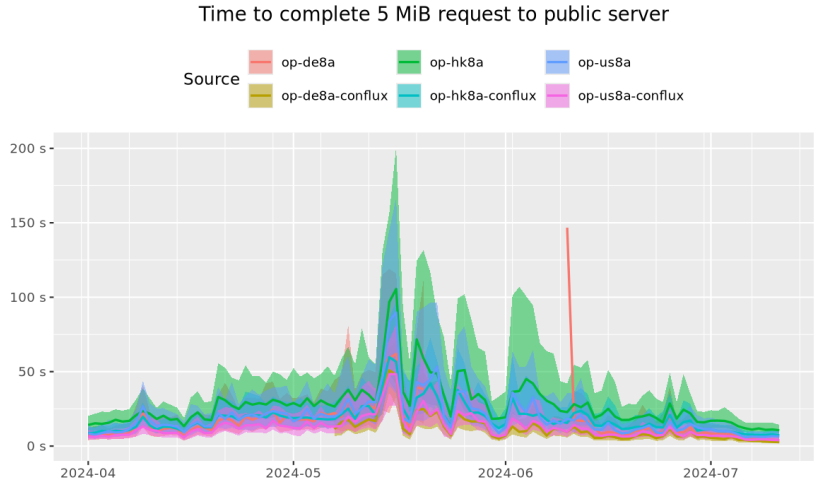
But what does that mean in a Tor context?

Network health investigations

From time to time we need to do network health investigations which aim at:

- Performance issues (potentially caused by denial of service (DoS) attempts)
- Attackers targeting Tor relays/users
- Network mysteries

Network health investigations

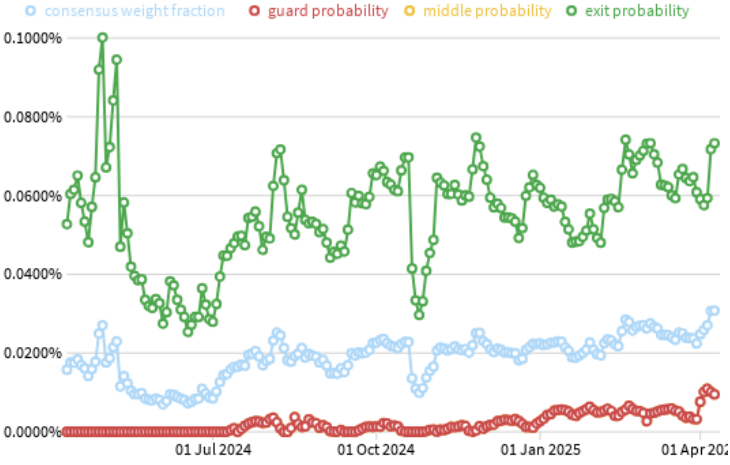


Network health investigations

From the tor-relays@ mailing list:

> *I've noticed a new kind of possible attack on some of my relays, as early as Dec.23 which causes huge spikes of outbound traffic that eventually maxes out RAM and crashes Tor. The newest one today lasted for 5 hours switching between two of the three relays on the same IP.*

Network health investigations



https://gitlab.torproject.org/tpo/network-health/analysis/-/issues/95

Accessible information sources

- Collection of data and metrics produced by Tor itself
  - Tor network and relay statuses
  - Output of scanning tools
  - …
- Collection of data provided by external tools
  - Internet infrastructure information
  - Social media posts
  - …

The world of descriptors and metrics

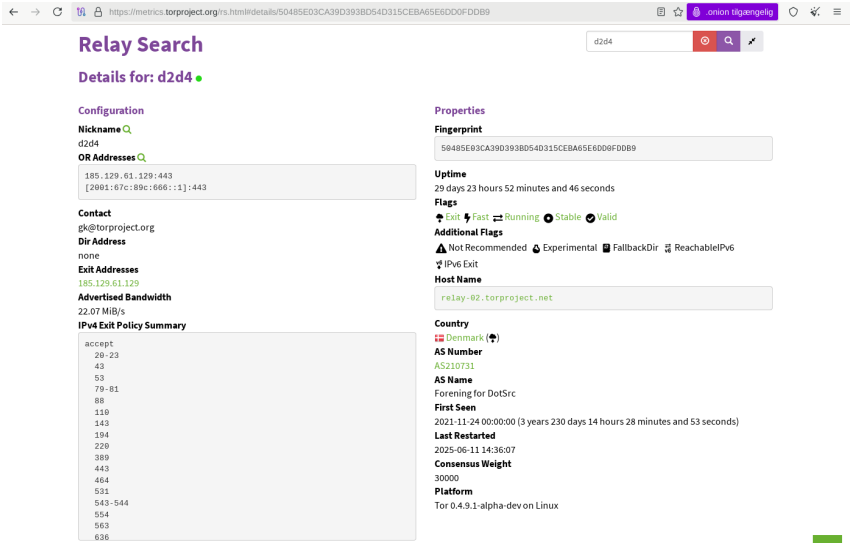
Descriptor Type	Description
Relay server descriptor	Information that relays publish about themselves
Network status vote	View of a Directory Authority on known relays
Network status consensus	Hourly consensus created by Directory Authorities
Network status entry	Relay information provided by Directory Authorities
Exit list	Exit relays and their IP addresses while using them
OnionPerf analysis file	Tor network performance measurement data

Table 1: Tor descriptor types

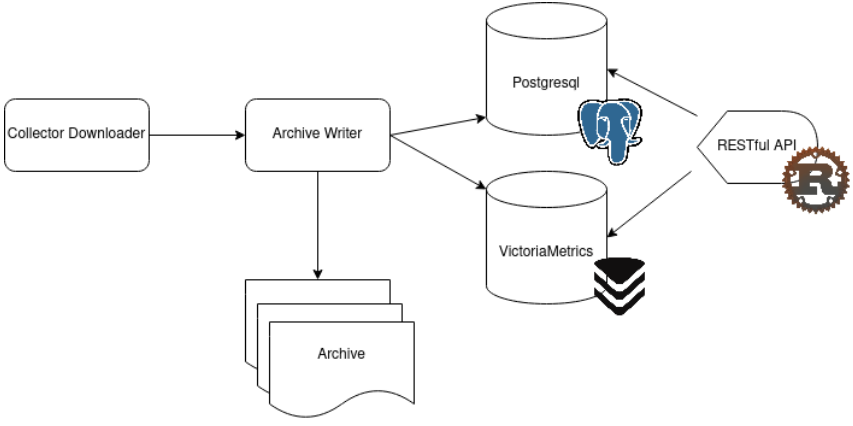
Current metrics tooling

- CollecTor
  - collects descriptors and makes them available at https://collector.torproject.org/archive/
  - it's loads of monthly compressed tarballs 😕
- Onionoo
  - web-based protocol which allows someone to learn about currently running Tor relays
  - Onionoo clients send HTTP GET requests to Onionoo server which reponds with JSON-formatted replies
  - for a specification and all the endpoints available, see: https://metrics.torproject.org/onionoo.html

Relay search



Metrics pipeline 2.0



External data sources

Non-Tor data sources are important in investigations to enrich and refine data provided by the Tor eco system itself. Their overall usefulness depends on the actual type of investigation, though. Examples are:

- Censys (internet scan data)
- SecurityTrails (historical domain data)
- WHOIS (domain/IP address ownership data)
- BGP routing information
- Snapshots of websites (e.g. archive.org)

How to do network health investigations?

Guidelines for how to do network health investigations at Tor are meant to be for a diverse audience:

- volunteers and Tor community members wanting to help
- Tor Project staff
- other organizations or interested third parties (media/governments etc.)

How to do network health investigations?

Tor's mission statement:

> *To advance human rights and freedoms by creating and deploying free and open source anonymity and privacy technologies, supporting their unrestricted availability and use, and furthering their scientific and popular understanding.*

How to do network health investigations?

How do we deal with attackers that don't care about the human rights of those involved or targeted in their attacks or about the Tor network?

- risk of subverting Tor's ethical foundation
- "risk of ruthlessness" ("ruthlessness" being: *"both a feature of action and a quality of thought and feeling that rejects all scruples, doubts, hesitation, and remorse in pursuing some ultimate purpose or serving some paramount principle"* [Cherniss 2021, p.2])

How to do network health investigations?

Our proposal is modeled after the *Berkeley Protocol on Digital Open Source Investigations*, but with adaptations as:

- the Berkeley Protocol is *"conducting online research of alleged violations of international criminal, human rights, and humanitarian law"* but the context at Tor is different
- evidence collection covered by the Berkeley Protocol (but not in Tor's context) has to pass scrutiny in courts
- researching human rights abuses, breaches of international criminal and humanitatian law comes with potentially serious personal risks involved, contrary to Tor network health investigations

How to do network health investigations?

We created a *Framework for running Tor Network Health open source investigations* which is

- following the guiding principles of the Berkeley Protocol (chapter II. Principles)
  - professional principles (accountability, objectivity, operational security...)
  - methodological principles (accuracy, data minimization, preservation...)
  - ethical principles (dignity, humility, independence, transparency...)

How to do network health investigations?

- adopting the investigation cycle (chapter VI. Investigation Process):

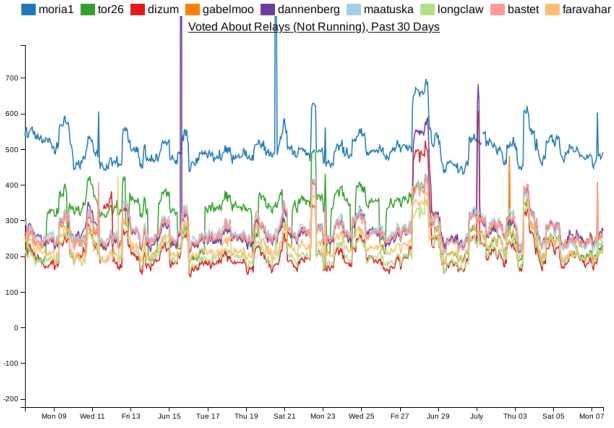
How to do network health investigations?

Reporting findings can be done via different venues, depending e.g. on the type of investigation:

- Gitlab bug tracker (optionally making the ticket confidential)
- public mailings lists, IRC/Matrix, our forum
- private bad-relays@ mailing list

Hands on

tor26 voted "not running" for a lot of relays but then dropped again. Why?



https://gitlab.torproject.org/tpo/network-health/analysis/-/issues/92

Hands on

There are different ways to start investigating this:

- CollecTor has the votes and consensuses... but it's monthly compressed tarballs
- PostgreSQL dumps for the win!
  - Votes, consensuses and server descriptors can be found at: https://people.torproject.org/~gk/bornhack2025/
  - The table schemes are collected in our metrics-sql-tables repository

References

- Cherniss, Joshua L.: Liberalism in Dark Times. The Liberal Ethos in the Twentieth Century. Princeton, Oxford 2021.

How can you help?

- Hack on some of our cool projects.
- Find, and maybe fix, bugs in Tor.
- Test Tor on your platform of choice.
- Work on some of the many open research projects.
- Run a Tor relay or a bridge!
- Teach others about Tor and privacy in general.

Questions?