Tor: a brief intro

Roger Dingledine
The Tor Project
https://torproject.org/

What is Tor?

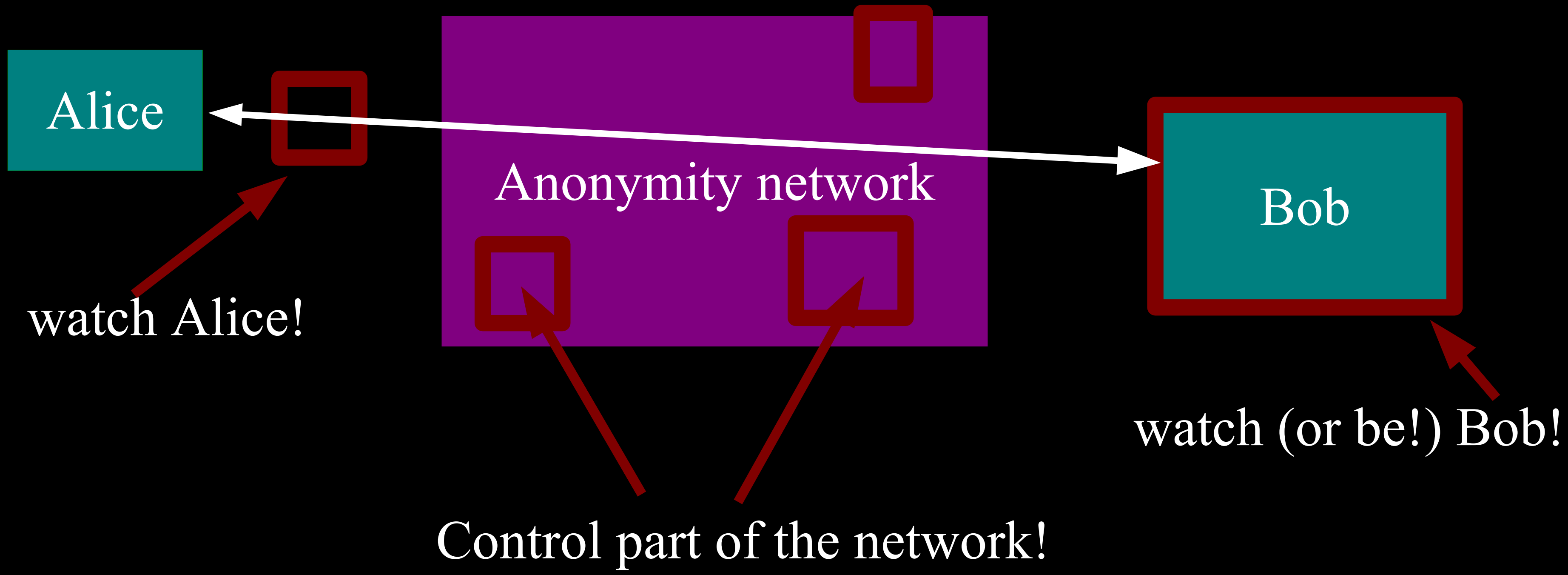
- Online anonymity 1) software, 2) network, 3) protocol
- Open source, freely available
- Community of researchers, developers, users, and relay operators
- Funding from US DoD, Electronic Frontier Foundation, Voice of America, Google, NLnet, Human Rights Watch, ...

The Tor Project, Inc.

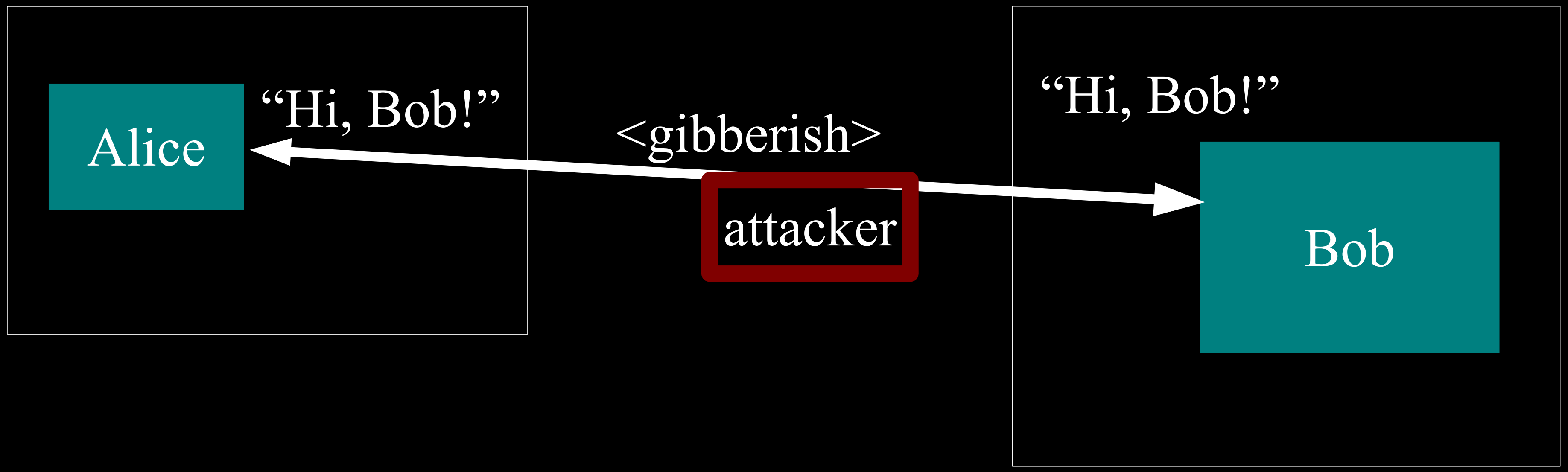
- 501(c)(3) non-profit organization dedicated to the research and development of tools for online anonymity and privacy

Estimated 500,000
daily Tor users

Threat model: what can the attacker do?



Anonymity isn't cryptography: Cryptography just protects contents.



Anonymity isn't just wishful thinking...

“You can't prove it was me!”

“Promise you won't look!”

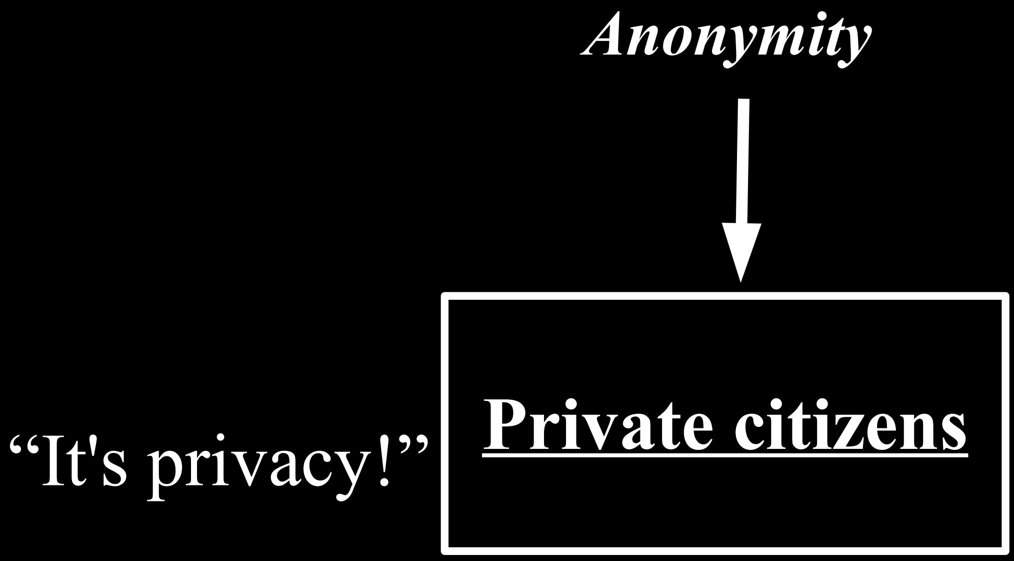
“Promise you won't remember!”

“Promise you won't tell!”

“I didn't write my name on it!”

“Isn't the Internet already anonymous?”

Anonymity serves different interests for different user groups.

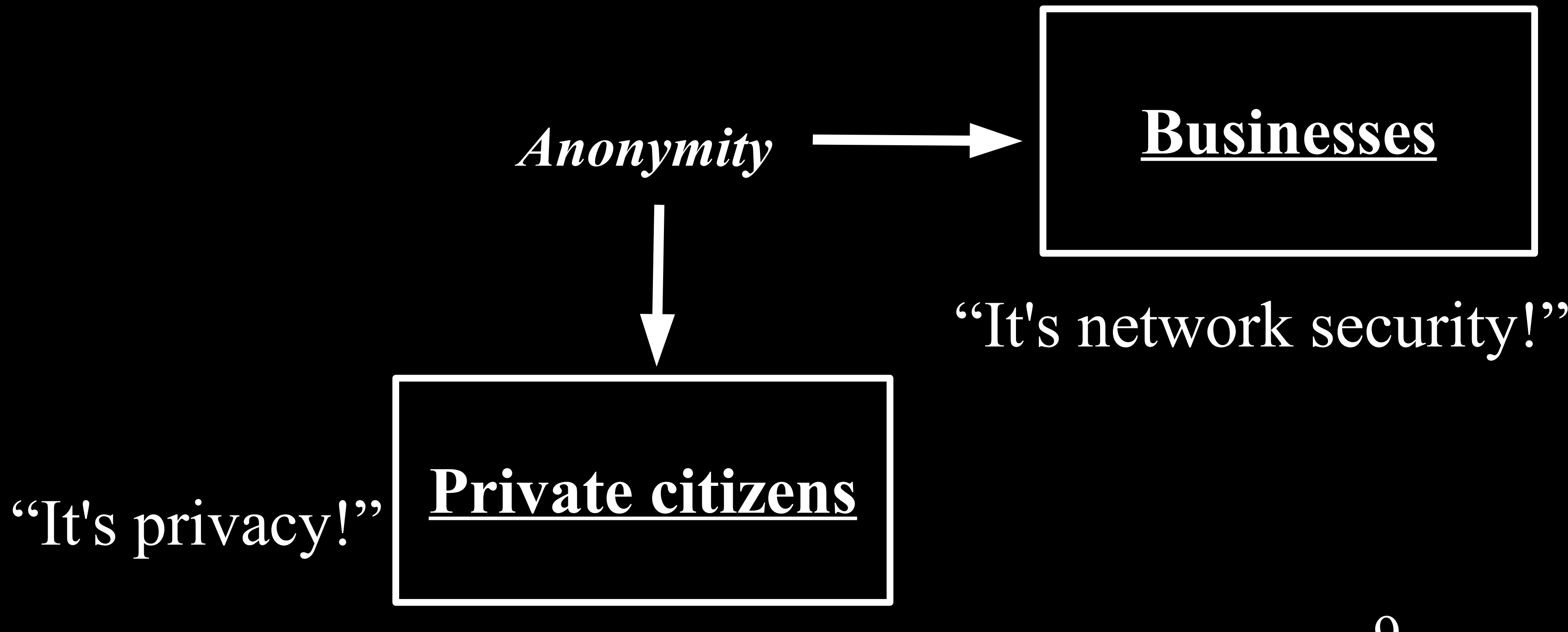


Anonymity

“It's privacy!”

Private citizens

Anonymity serves different interests for different user groups.



Anonymity

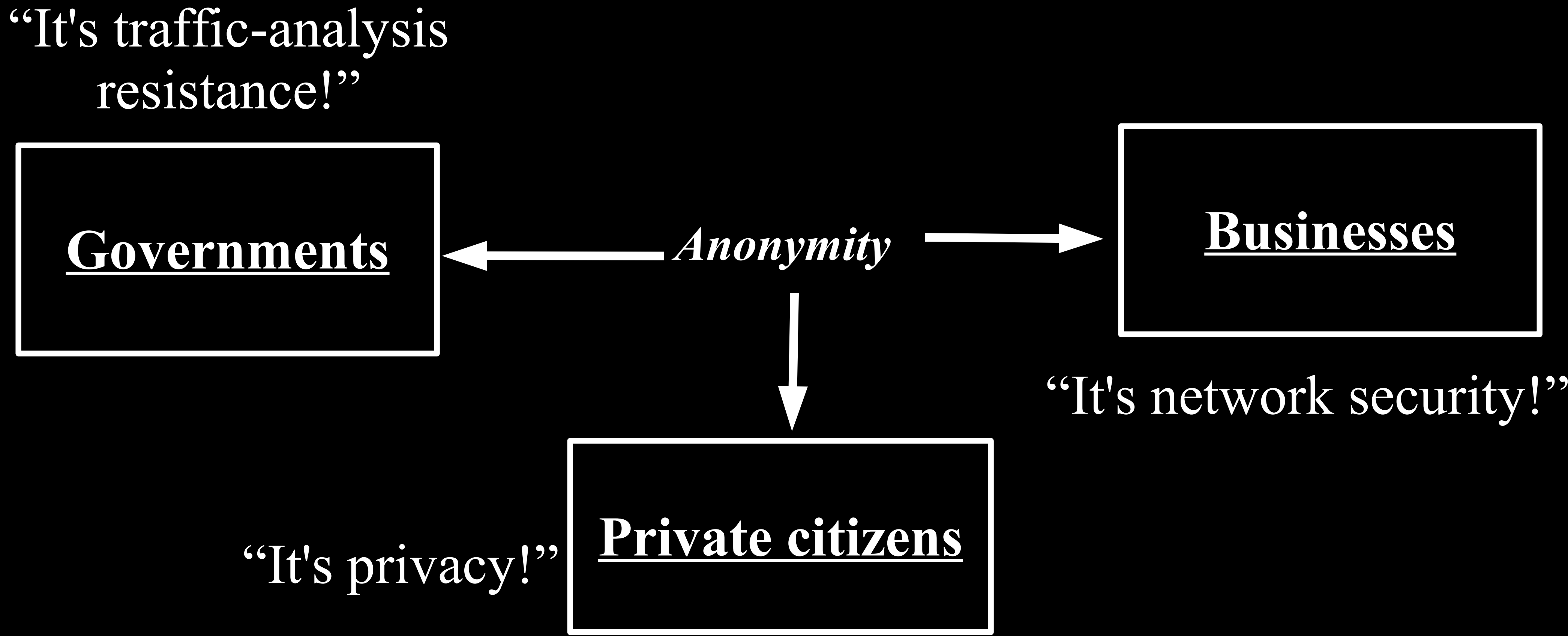
Businesses

“It's network security!”

Private citizens

“It's privacy!”

Anonymity serves different interests for different user groups.



“It's traffic-analysis resistance!”

Governments

Anonymity

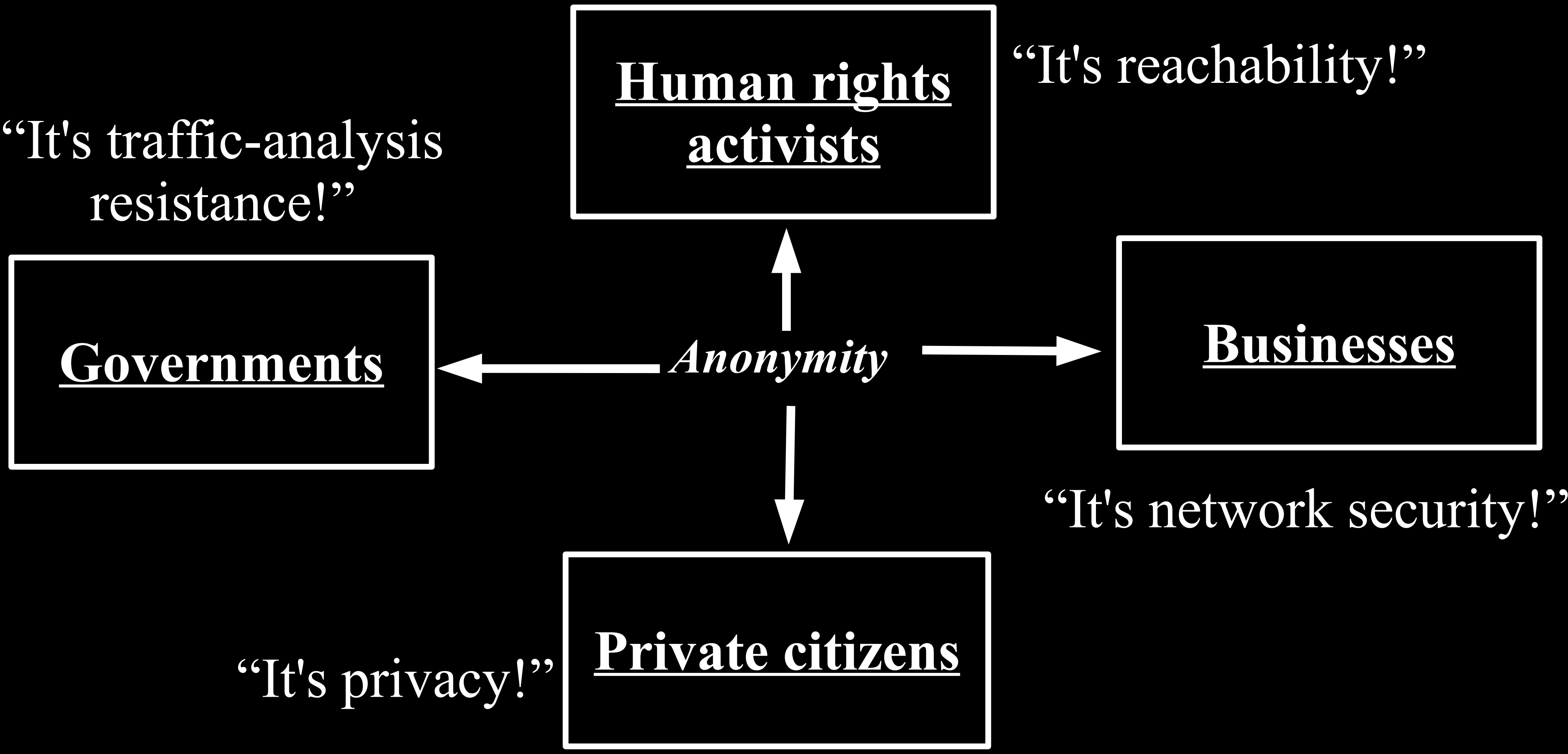
Businesses

“It's network security!”

Private citizens

“It's privacy!”

Anonymity serves different interests for different user groups.



Human rights activists

“It's reachability!”

“It's traffic-analysis resistance!”

Governments

Anonymity

Businesses

“It's network security!”

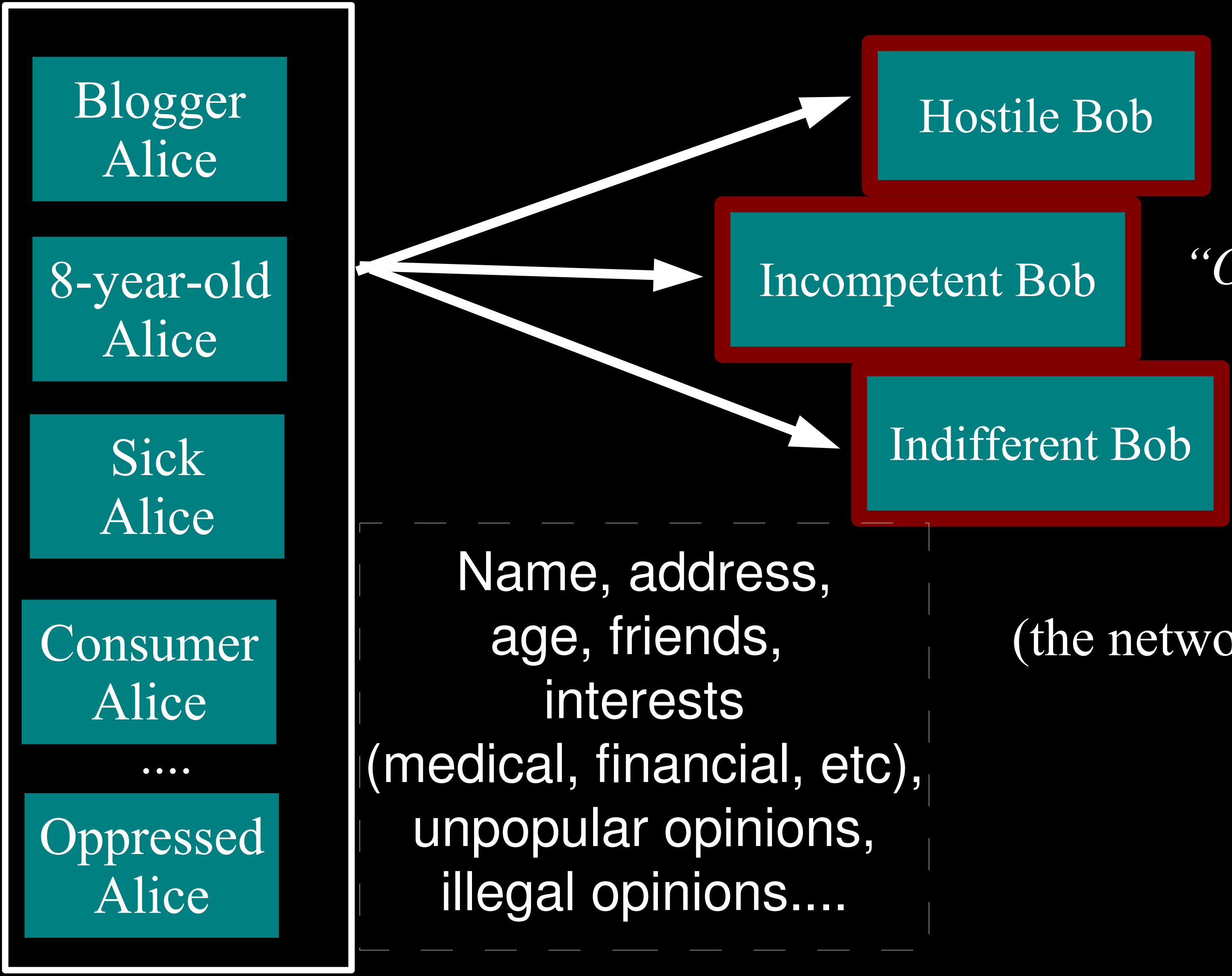
Private citizens

“It's privacy!”

Five pieces to today's talk

- 1) Who uses Tor and why?
- 2) The Tor design in a nutshell
- 3) Tor and censorship
- 4) We have data
- 5) Performance questions

Regular citizens don't want to be watched and tracked.



Blogger Alice

8-year-old Alice

Sick Alice

Consumer Alice

....

Oppressed Alice

Hostile Bob

Incompetent Bob

Indifferent Bob

Name, address, age, friends, interests (medical, financial, etc), unpopular opinions, illegal opinions....

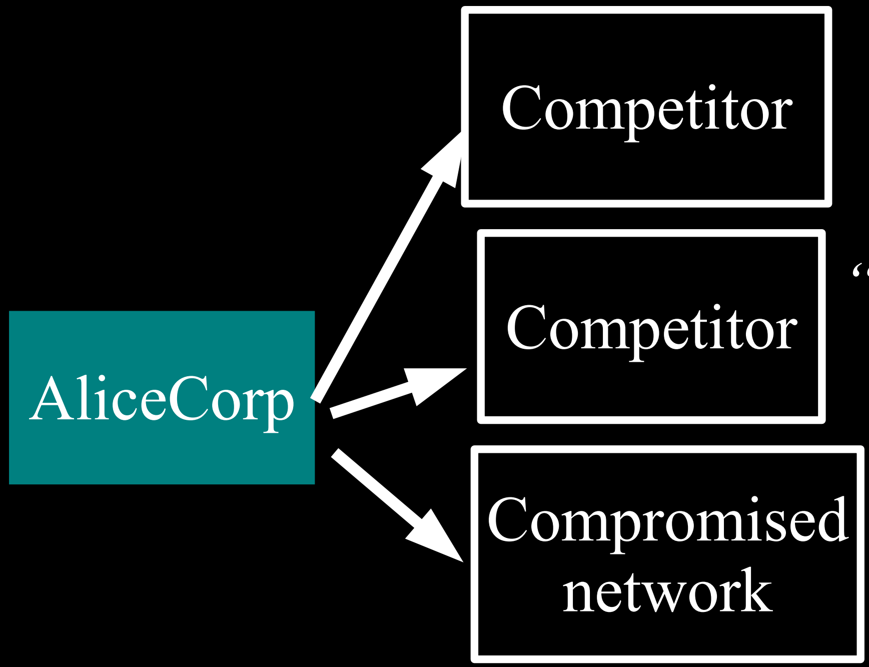
"I sell the logs."

"Oops, I lost the logs."
The AOL fiasco

*"Hey, they aren't **my** secrets."*

(the network can track too)

Businesses need to keep trade secrets.



AliceCorp

Competitor

Competitor

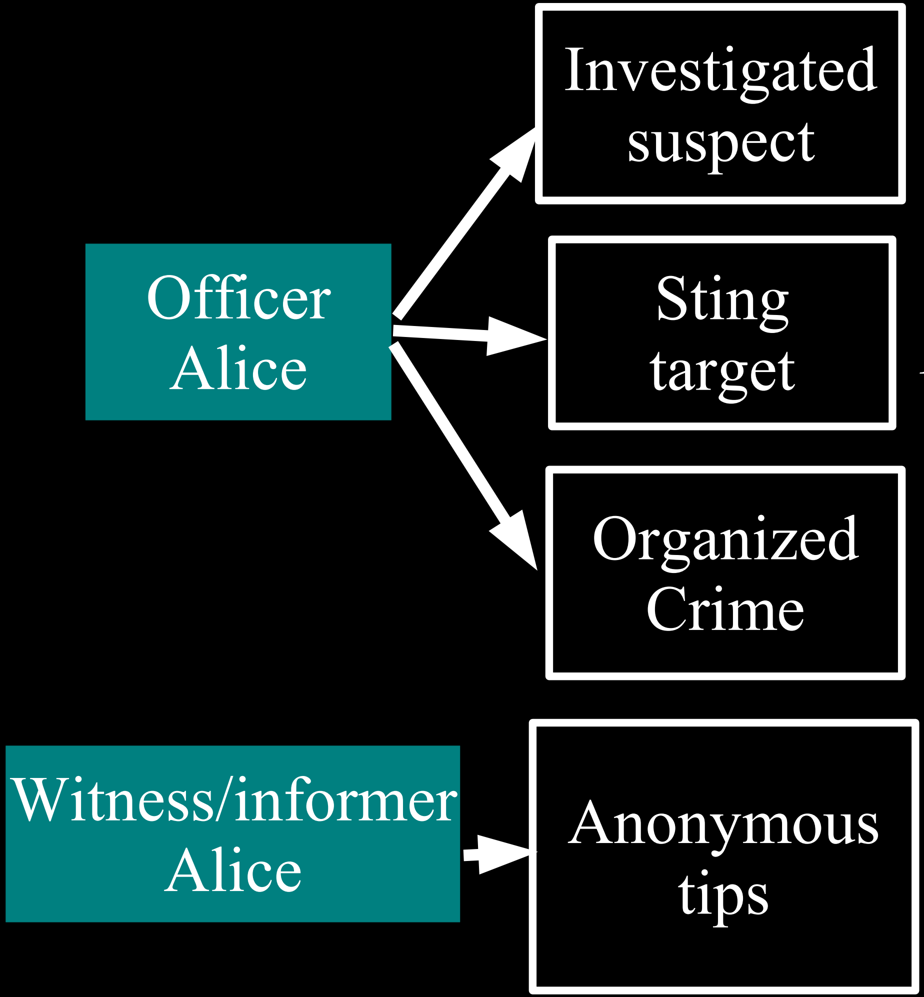
Compromised network

“Oh, your employees are reading our patents/jobs page/product sheets?”

“Hey, it's Alice! Give her the 'Alice' version!”

“Wanna buy a list of Alice's suppliers? What about her customers? What about her engineering department's favorite search terms?”

Law enforcement needs anonymity to get the job done.



Officer Alice

Investigated suspect

Sting target

Organized Crime

Witness/informer Alice

Anonymous tips

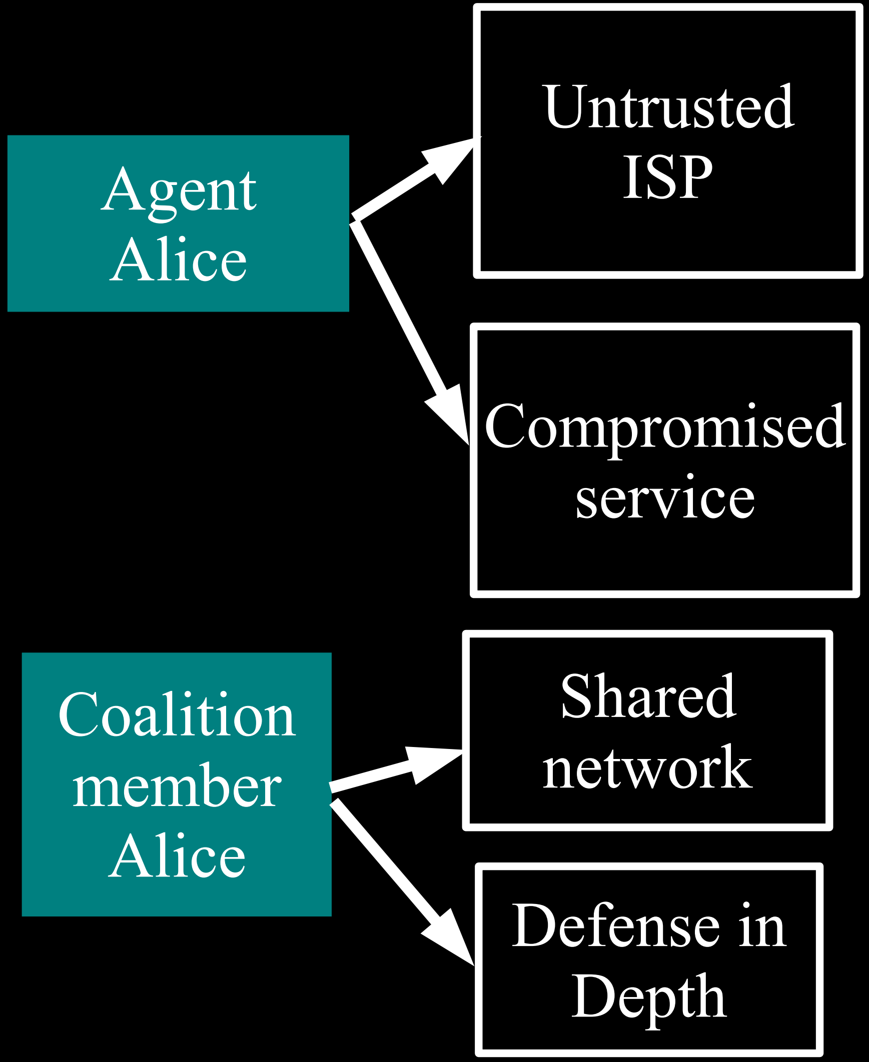
"Why is alice.localpolice.gov reading my website?"

"Why no, alice.localpolice.gov! I would never sell counterfeits on ebay!"

"Is my family safe if I go after these guys?"

"Are they really going to ensure my anonymity?"

Governments need anonymity for their security



Agent Alice

Untrusted ISP

Compromised service

Coalition member Alice

Shared network

Defense in Depth

“What will you bid for a list of Baghdad IP addresses that get email from .gov?”

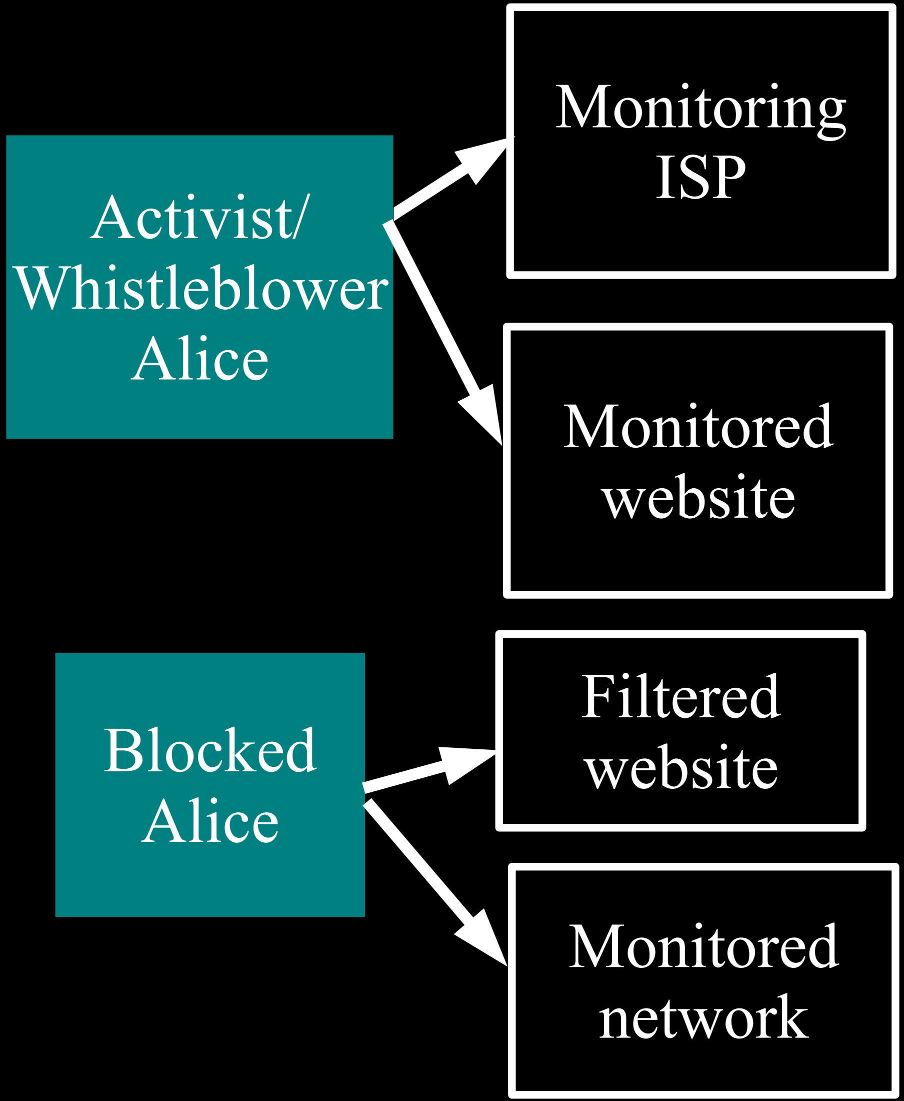
“Somebody in that hotel room just checked his Navy.mil mail!”

“What does *FBI Google for?”*

“Do I really want to reveal my internal network topology?”

“What about insiders?”

Journalists and activists need Tor for their personal safety



Activist/ Whistleblower Alice

Monitoring ISP

"Did you just post to that website?"

Monitored website

"Where are the bloggers connecting from?"
"I run livejournal and track my users"
"Of course I tell China about my users"

Blocked Alice

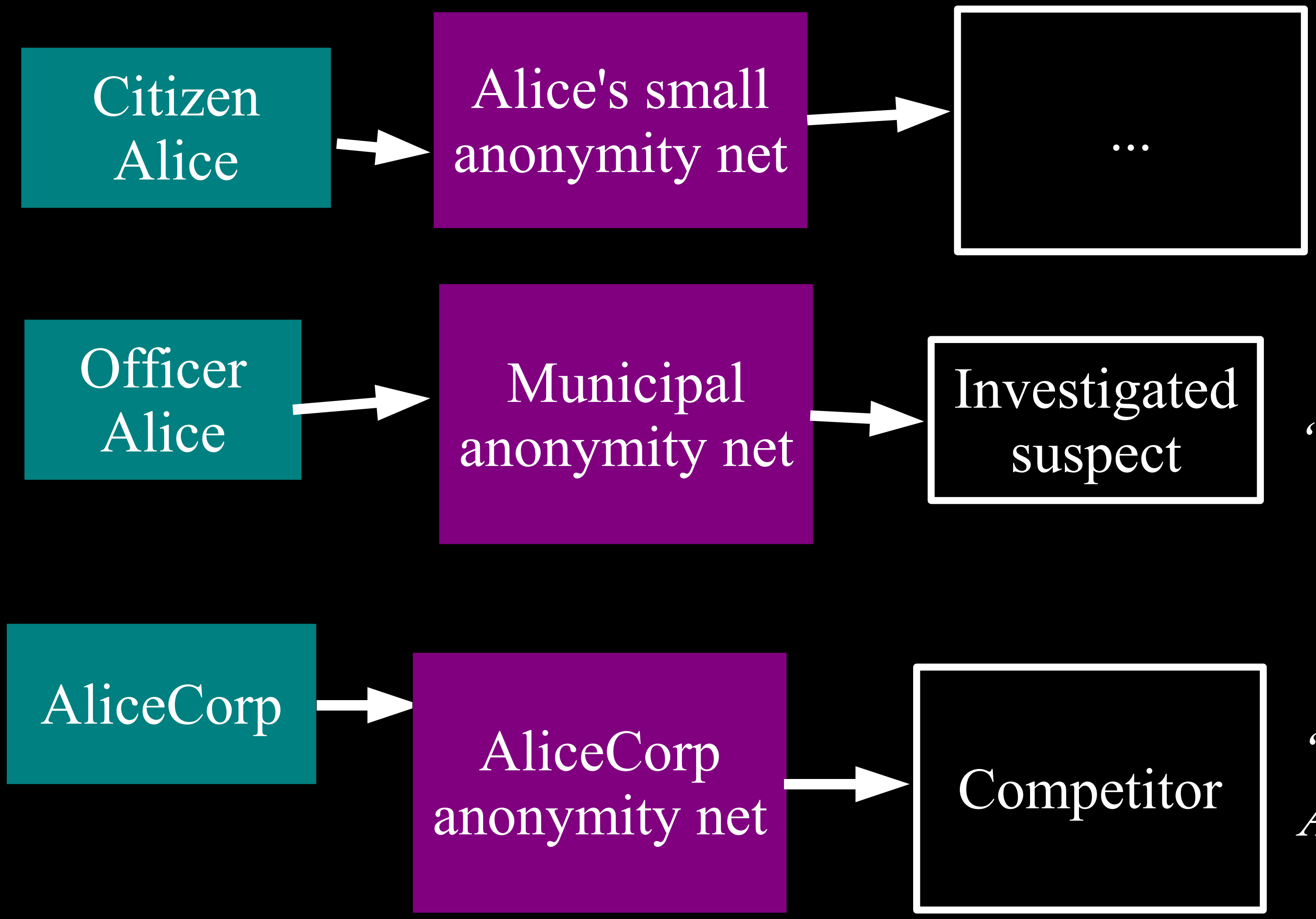
Filtered website

"What does the Global Voices website say today?"
"I want to tell people what's going on in my country"

Monitored network

"I think they're watching. I'm not even going to try."

You can't get anonymity on your own: private solutions are ineffective...



Citizen Alice → Alice's small anonymity net → ...

"One of the 25 users on AliceNet."

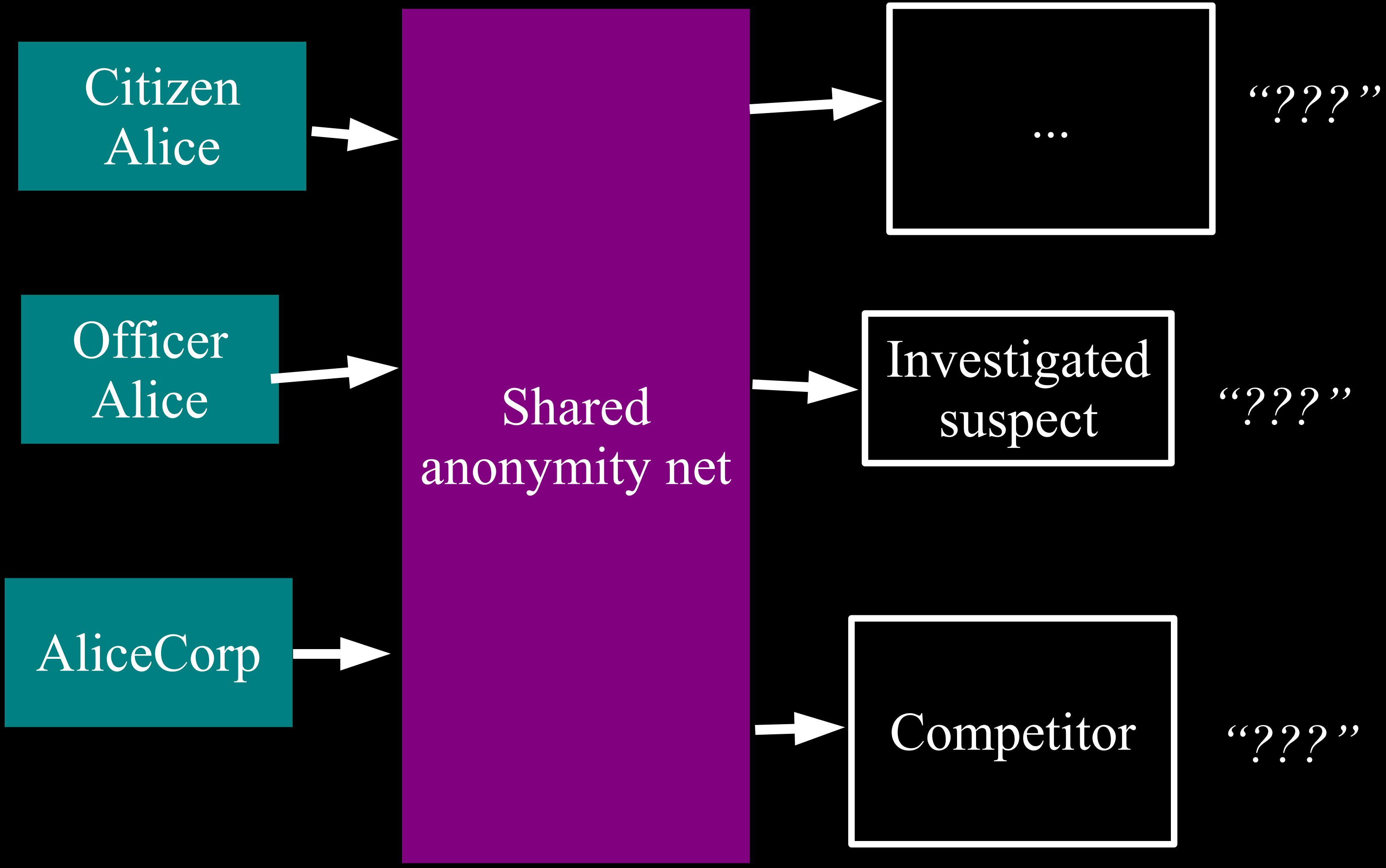
Officer Alice → Municipal anonymity net → Investigated suspect

"Looks like a cop."

AliceCorp → AliceCorp anonymity net → Competitor

*"It's **somebody** at AliceCorp!"*

… so, anonymity loves company!



Citizen Alice

Officer Alice

AliceCorp

Shared anonymity net

…

"???"

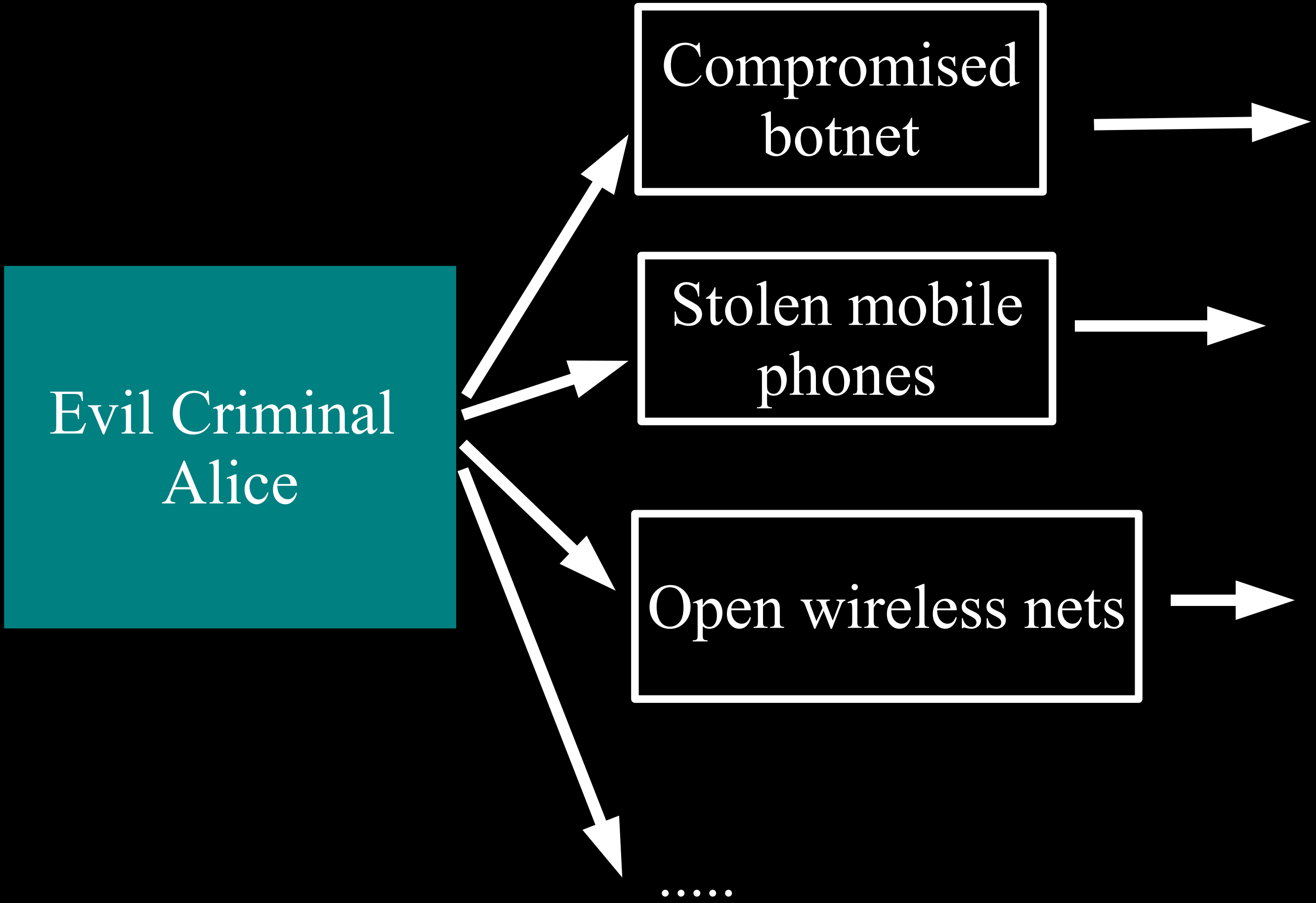
Investigated suspect

"???"

Competitor

"???"

Yes, bad people need anonymity too.
But they are *already* doing well.



Evil Criminal Alice

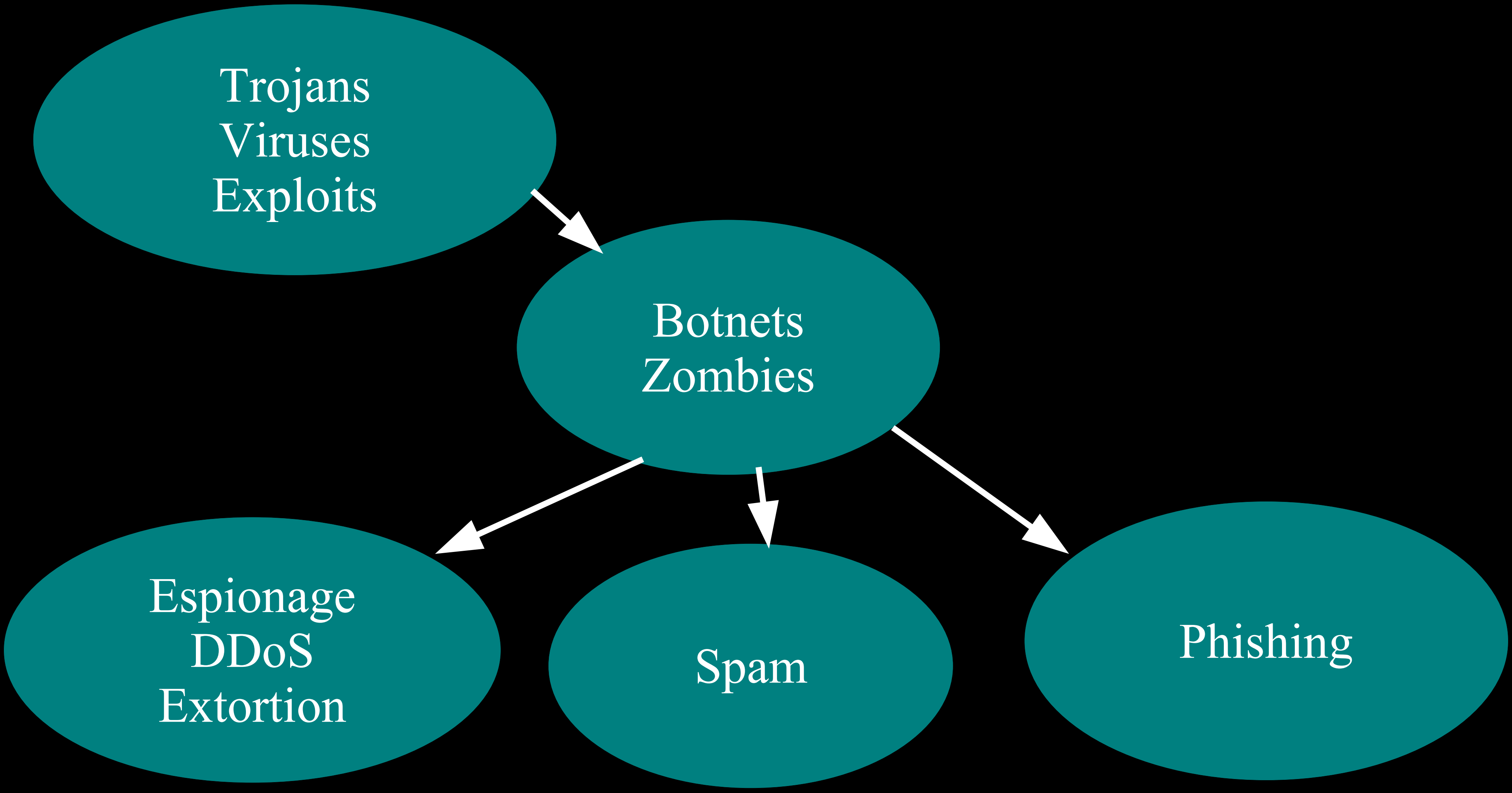
Compromised botnet

Stolen mobile phones

Open wireless nets

.....

Current situation: Bad people on the Internet are doing fine



Trojans
Viruses
Exploits

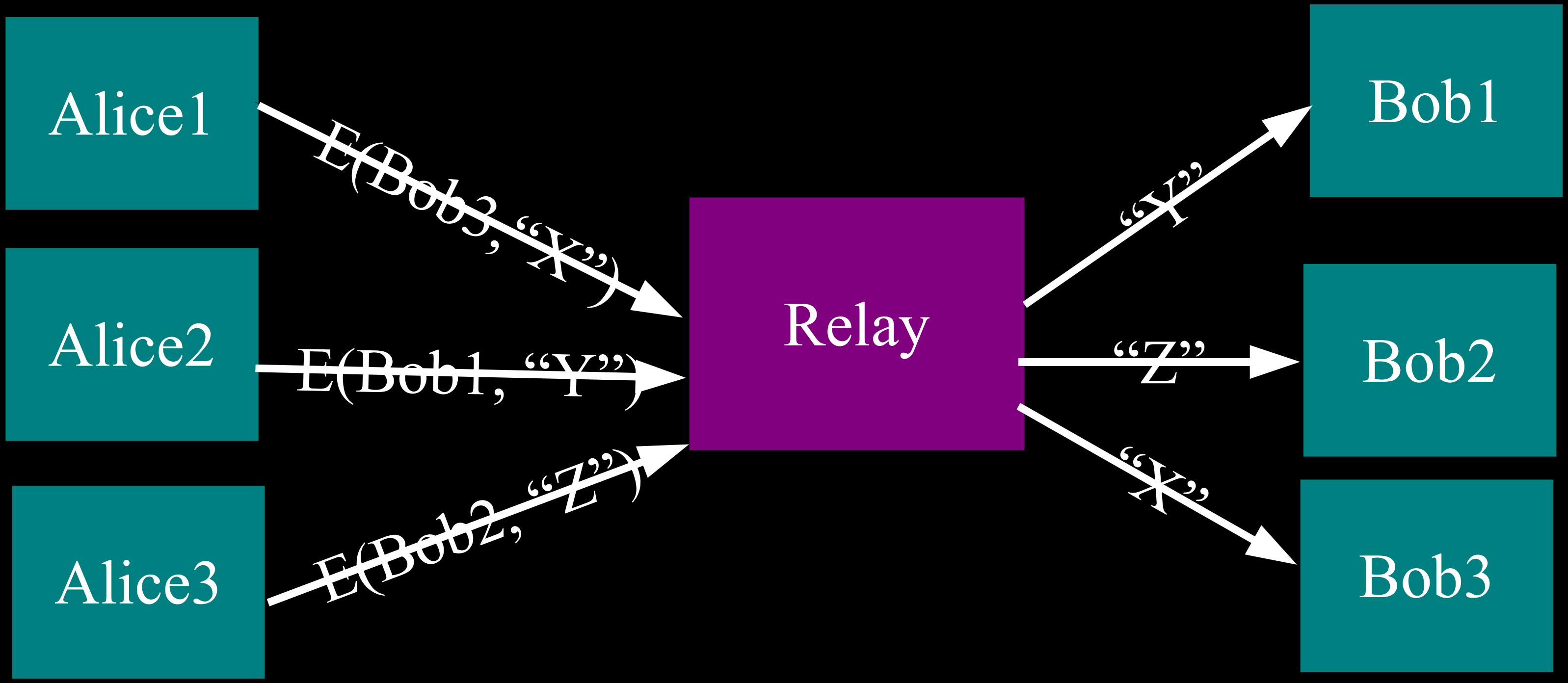
Botnets
Zombies

Espionage
DDoS
Extortion

Spam

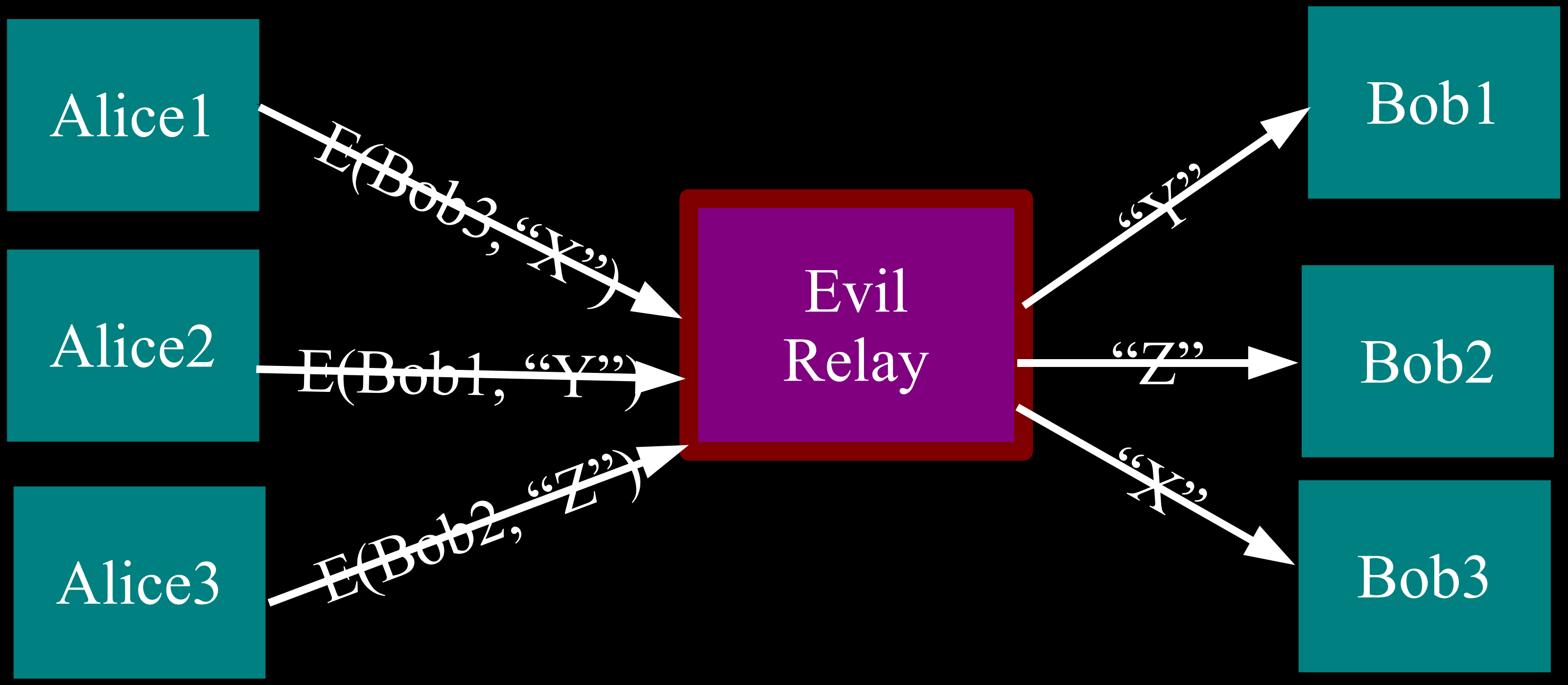
Phishing

The simplest designs use a single relay to hide connections.

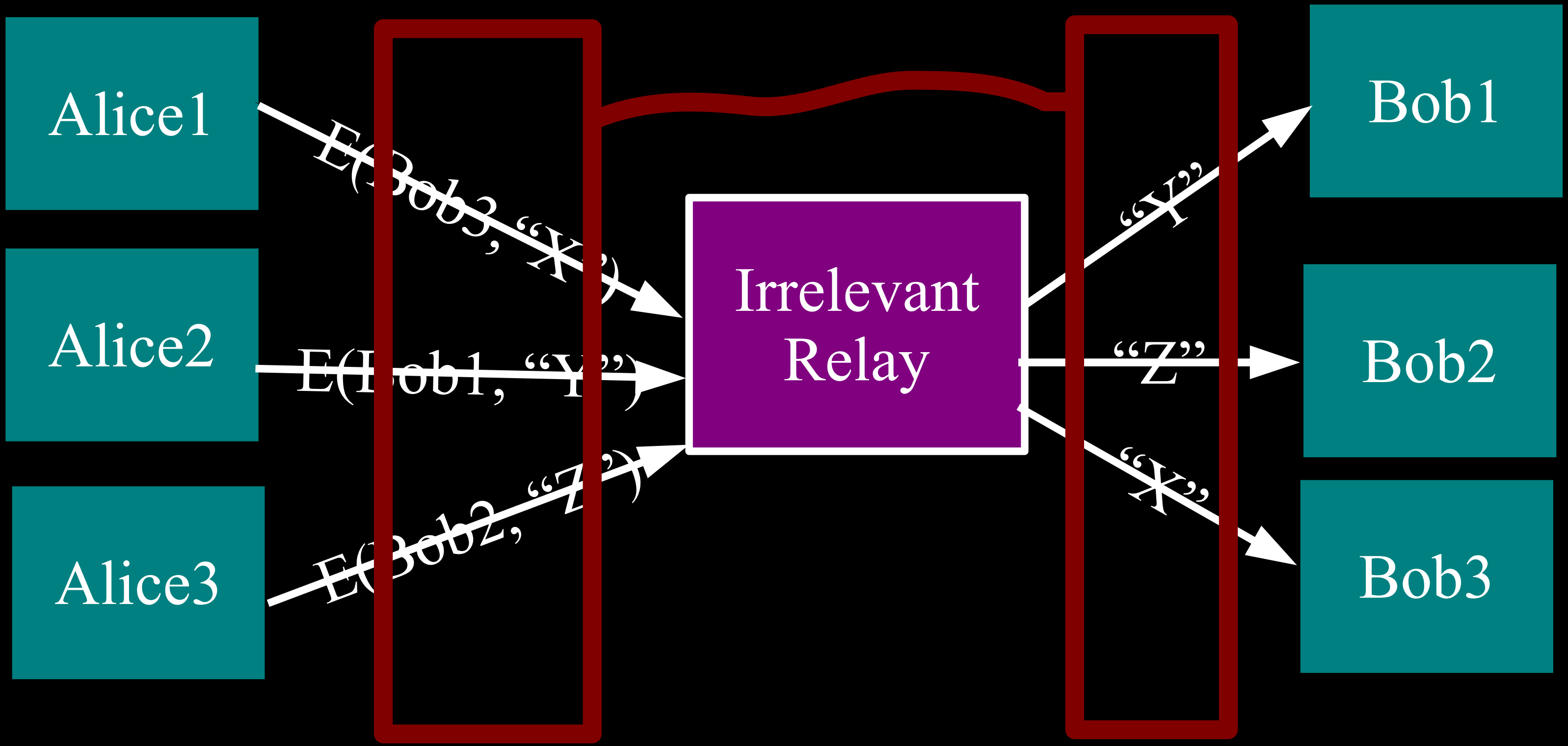


(example: some commercial proxy providers)

But a single relay (or eavesdropper!) is a single point of failure.

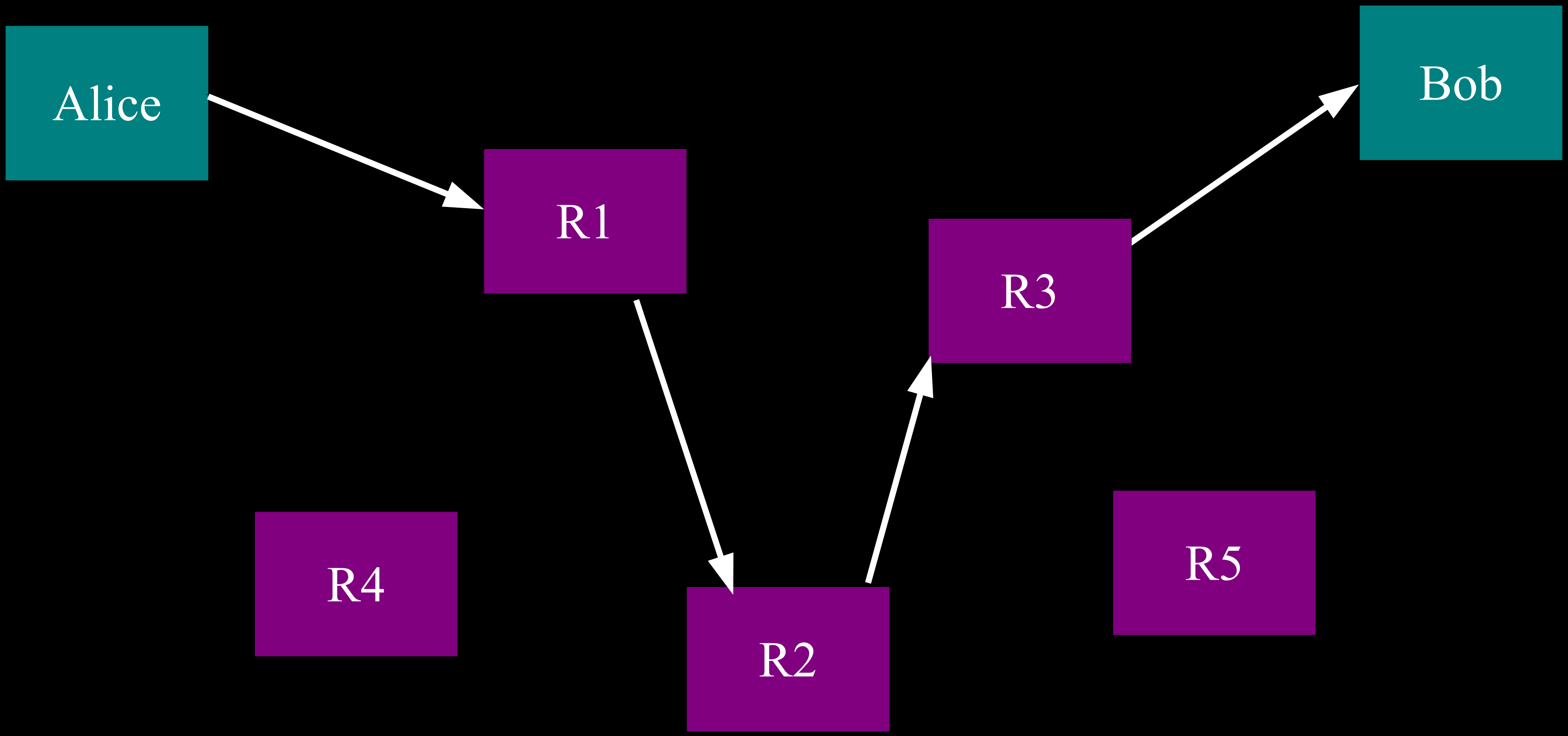


... or a single point of bypass.

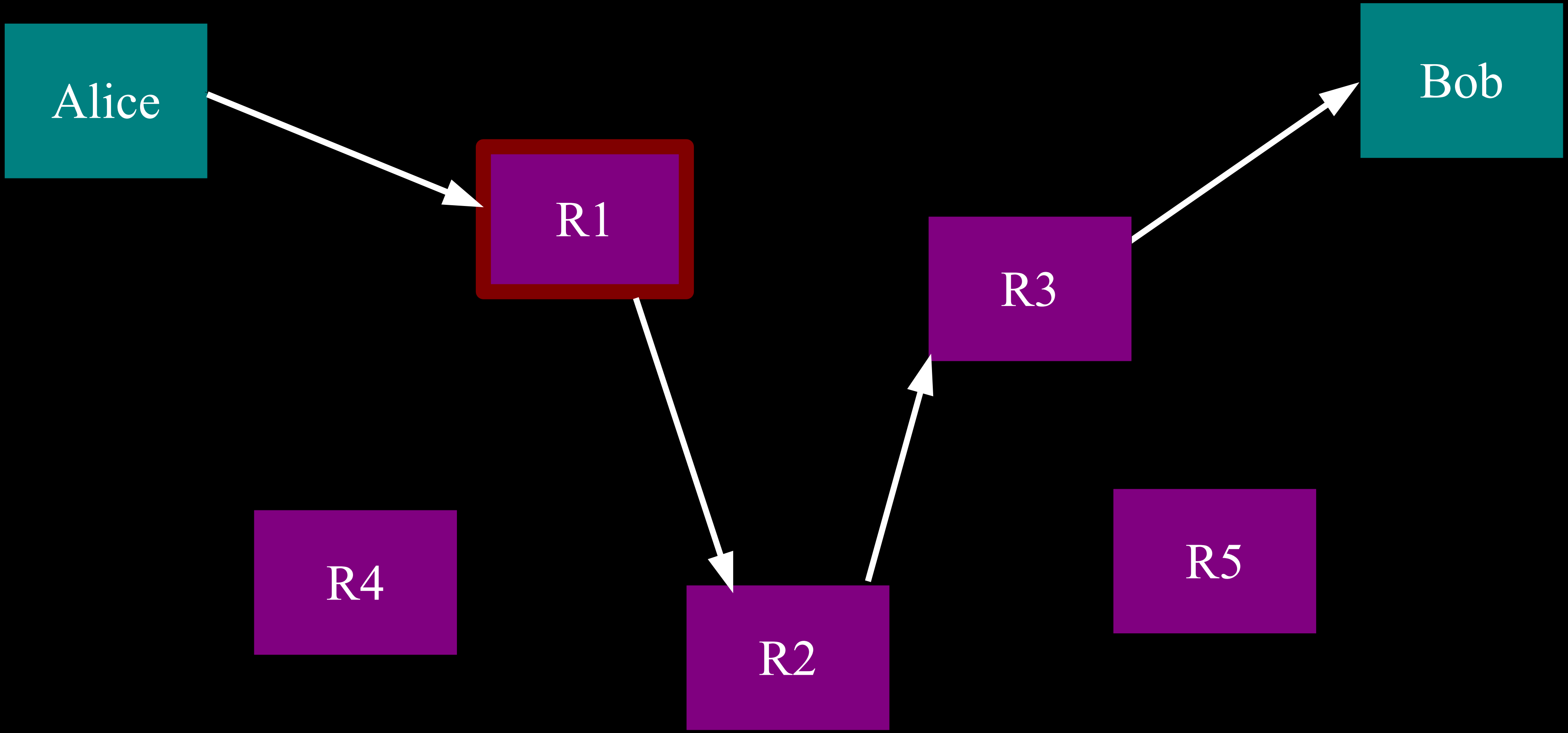


Timing analysis bridges all connections through relay ⇒ An attractive fat target

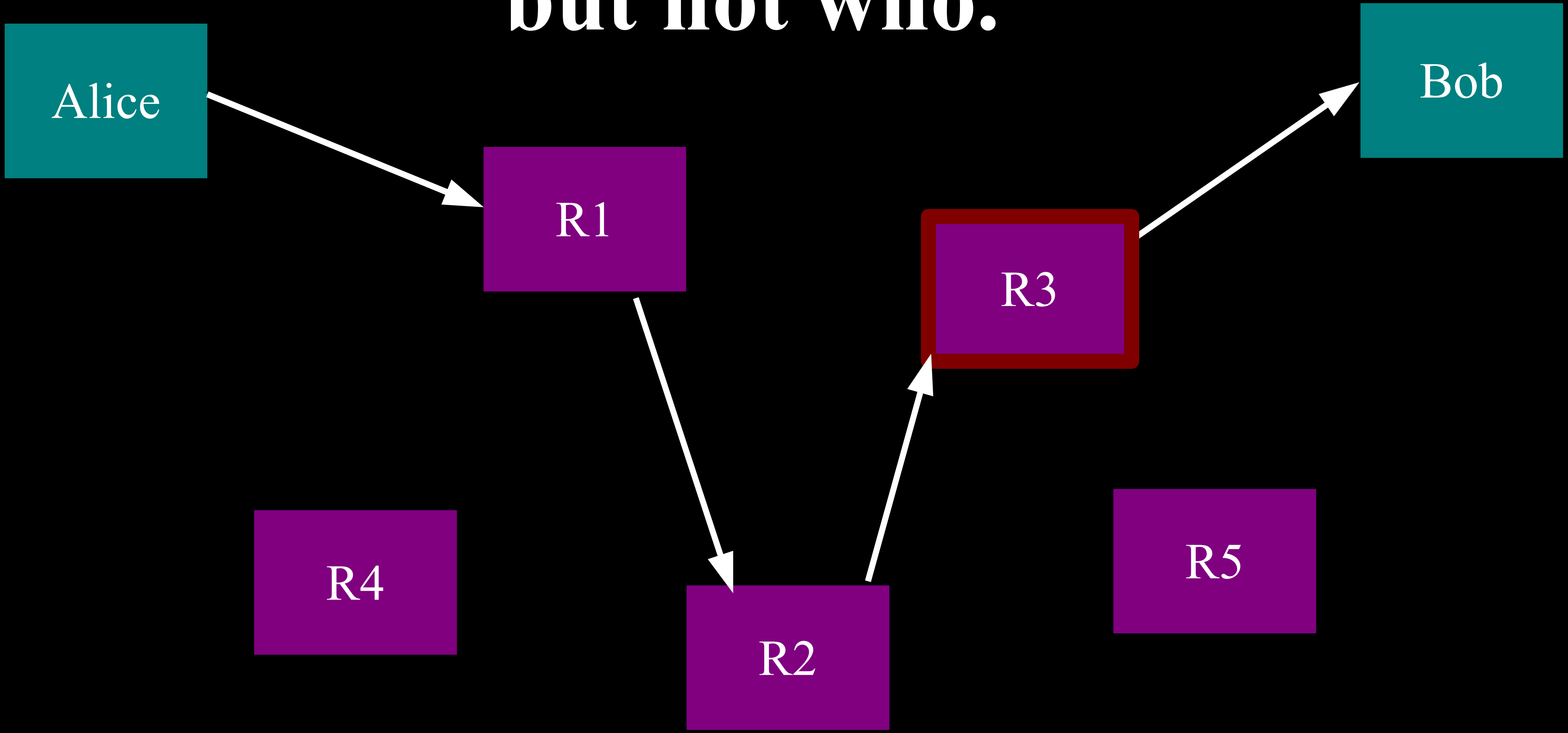
So, add multiple relays so that no single one can betray Alice.



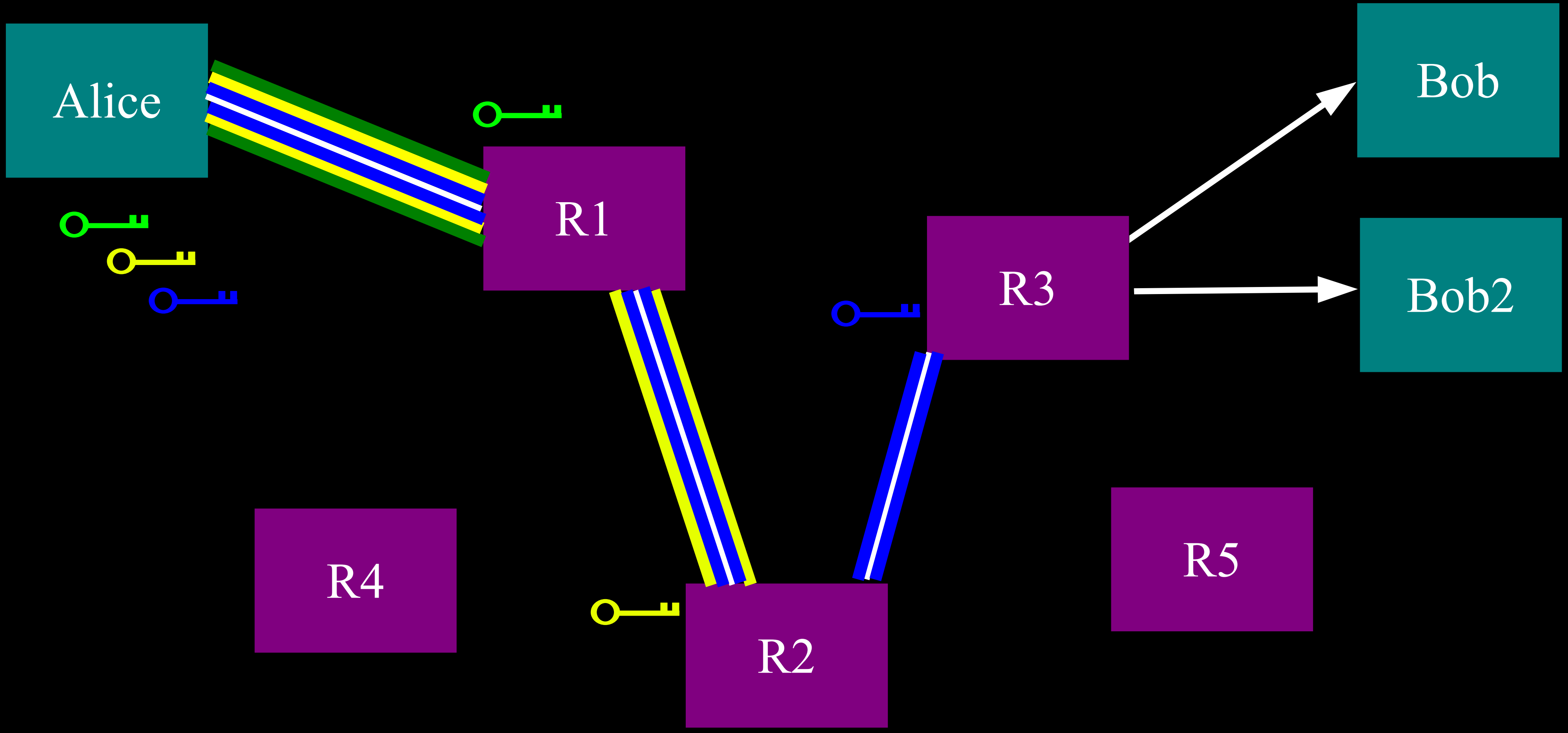
A corrupt first hop can tell that Alice is talking, but not to whom.



A corrupt final hop can tell that somebody is talking to Bob, but not who.



Alice makes a session key with R1 ...And then tunnels to R2...and to R3



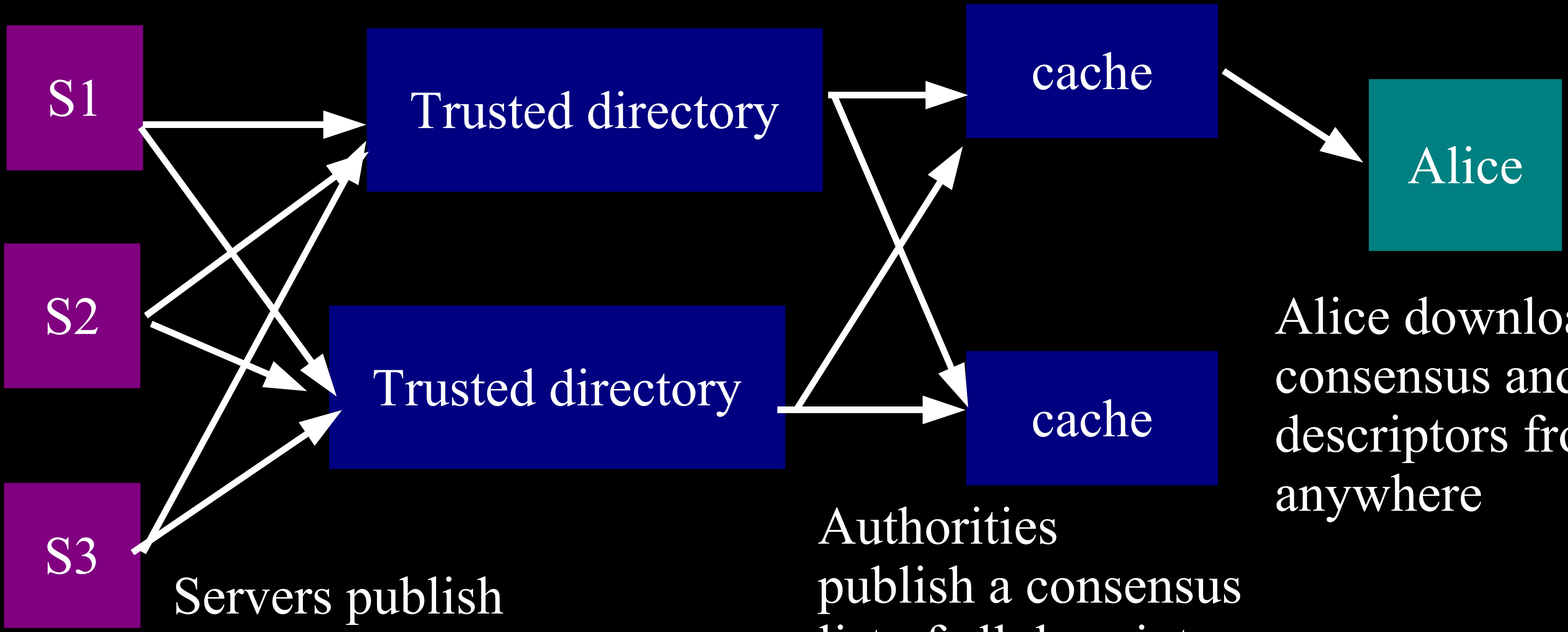
What we spend our time on

- Performance and scalability
- Maintaining the whole software ecosystem
- Blocking-resistance (circumvention)
- Basic research on anonymity
- Reusability and modularity
- Advocacy, education, and trainings around the world
- Metrics, data, and analysis

Relay versus Discovery

- There are two pieces to all these “proxying” schemes:
- a **relay** component: building circuits, sending traffic over them, getting the crypto right
- a **discovery** component: learning what relays are available

The basic Tor design uses a simple centralized directory protocol.



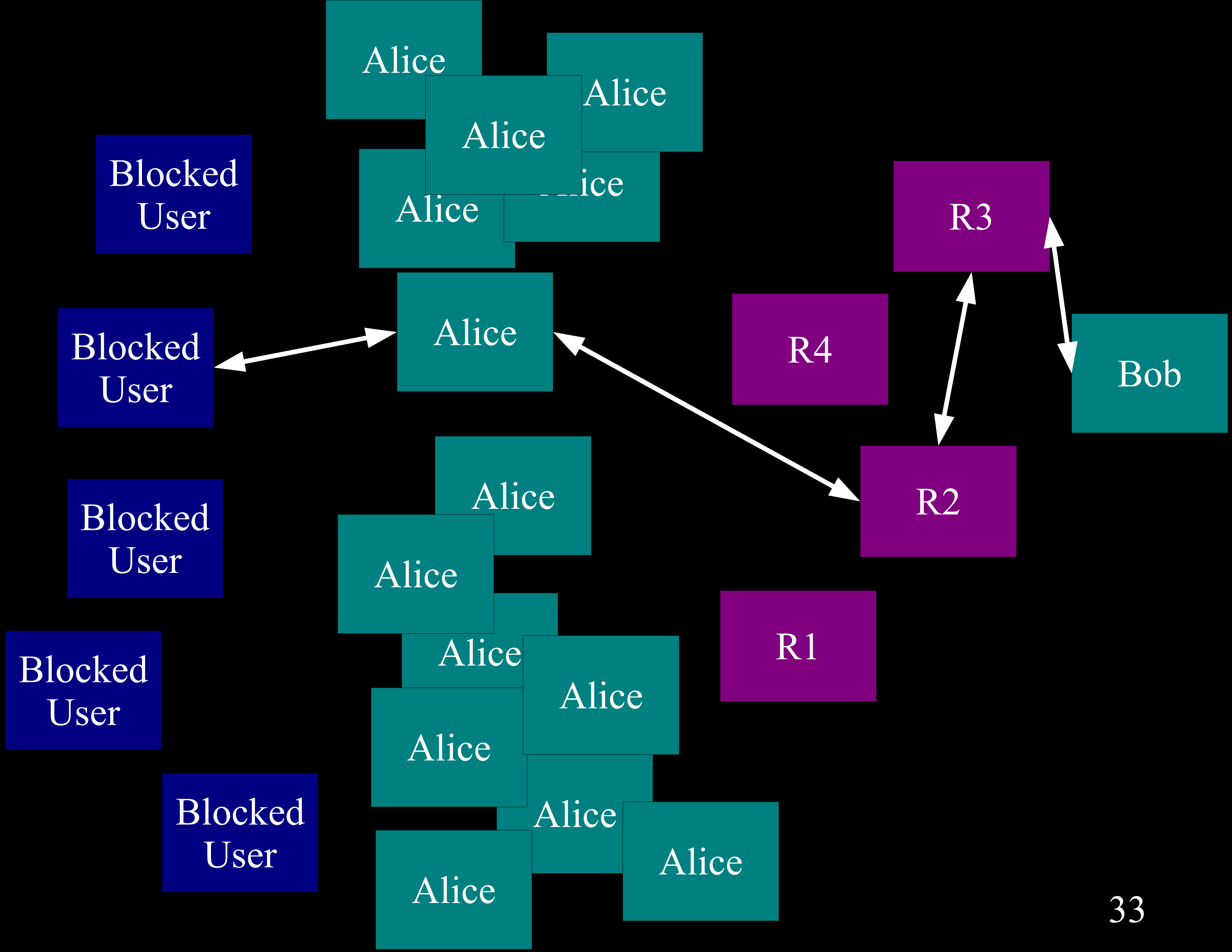
Servers publish self-signed descriptors.

Authorities publish a consensus list of all descriptors

Alice downloads consensus and descriptors from anywhere

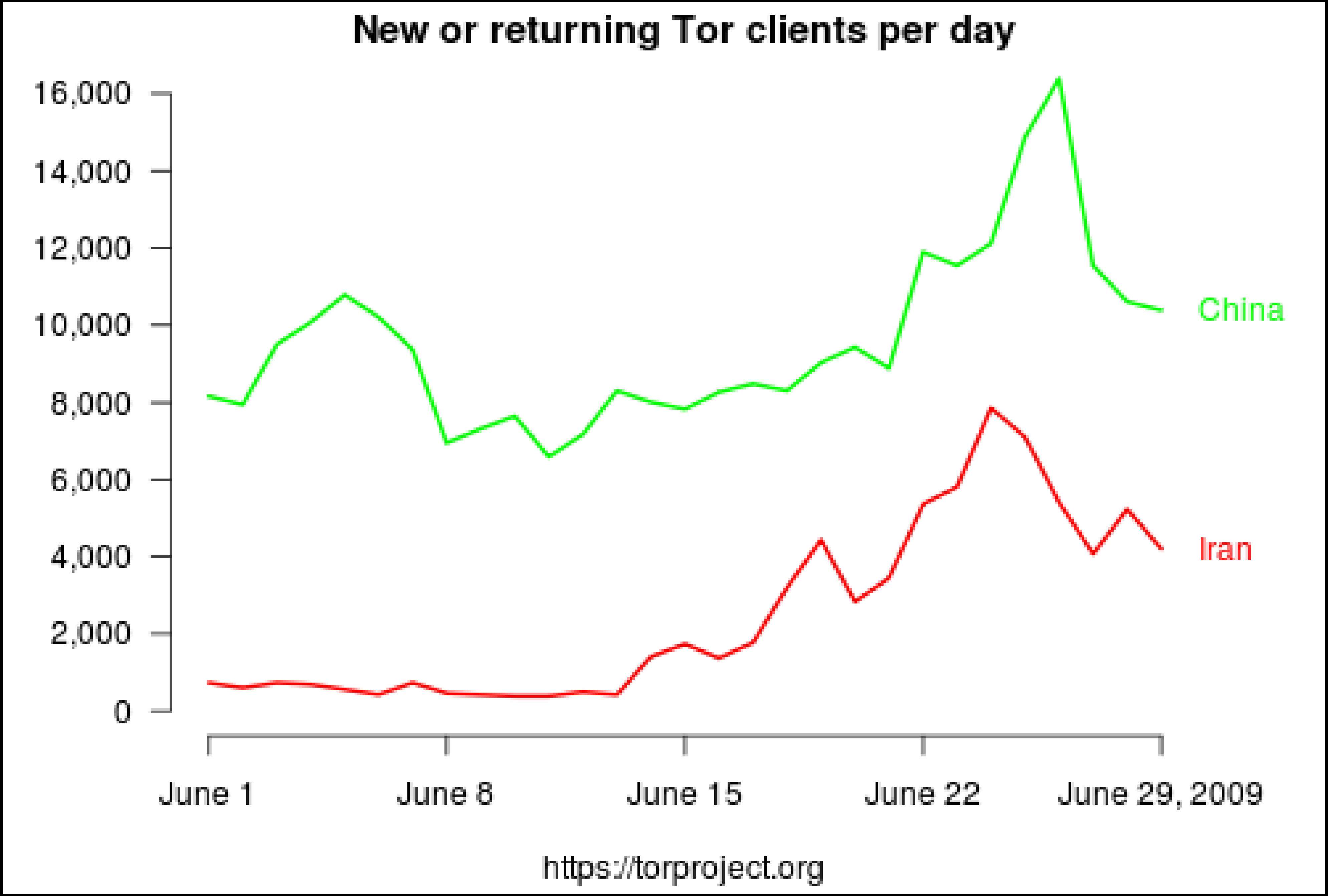
Attackers can block users from connecting to the Tor network

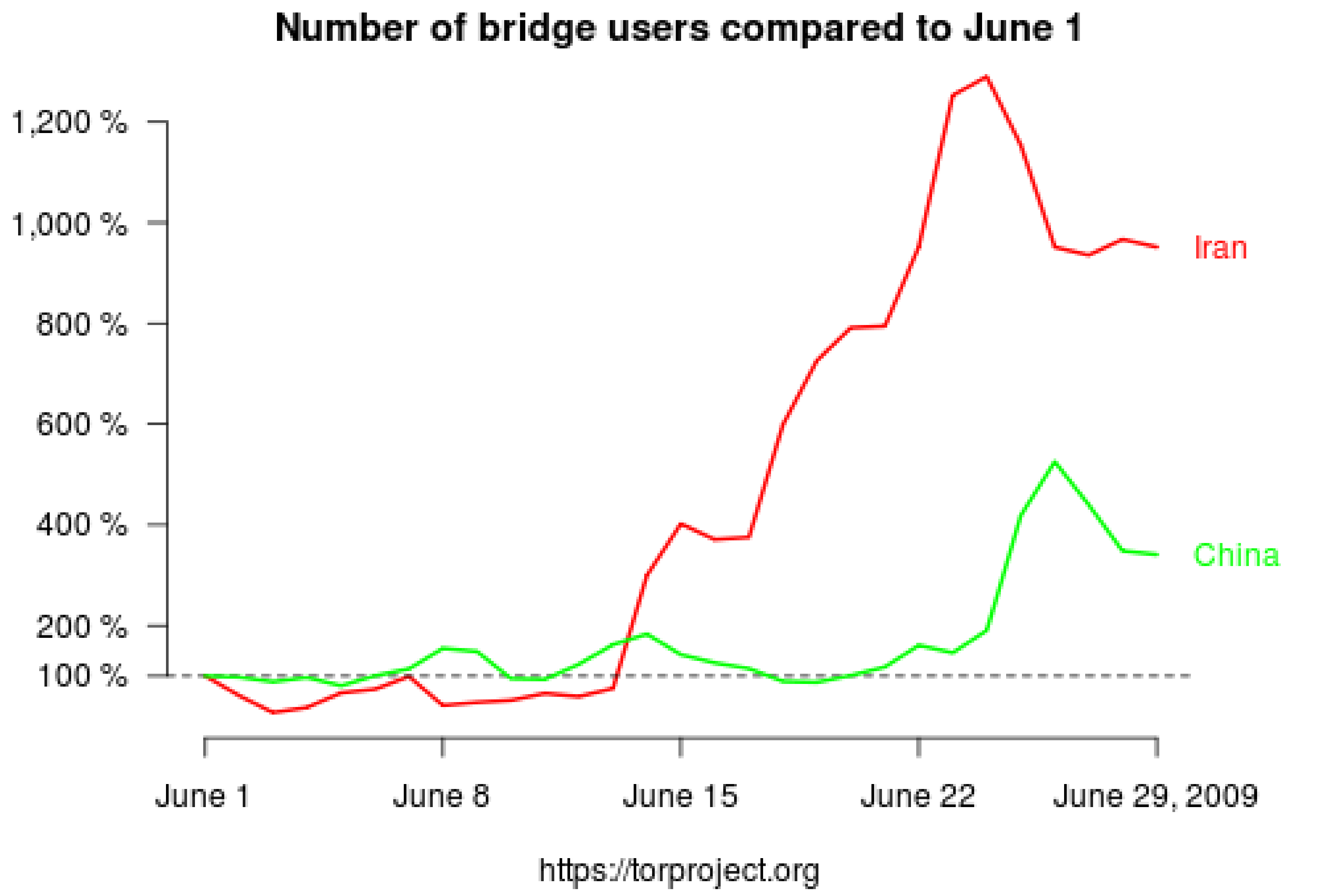
- By blocking the directory authorities
- By blocking all the relay IP addresses in the directory
- By filtering based on Tor's network fingerprint
- By preventing users from finding the Tor software

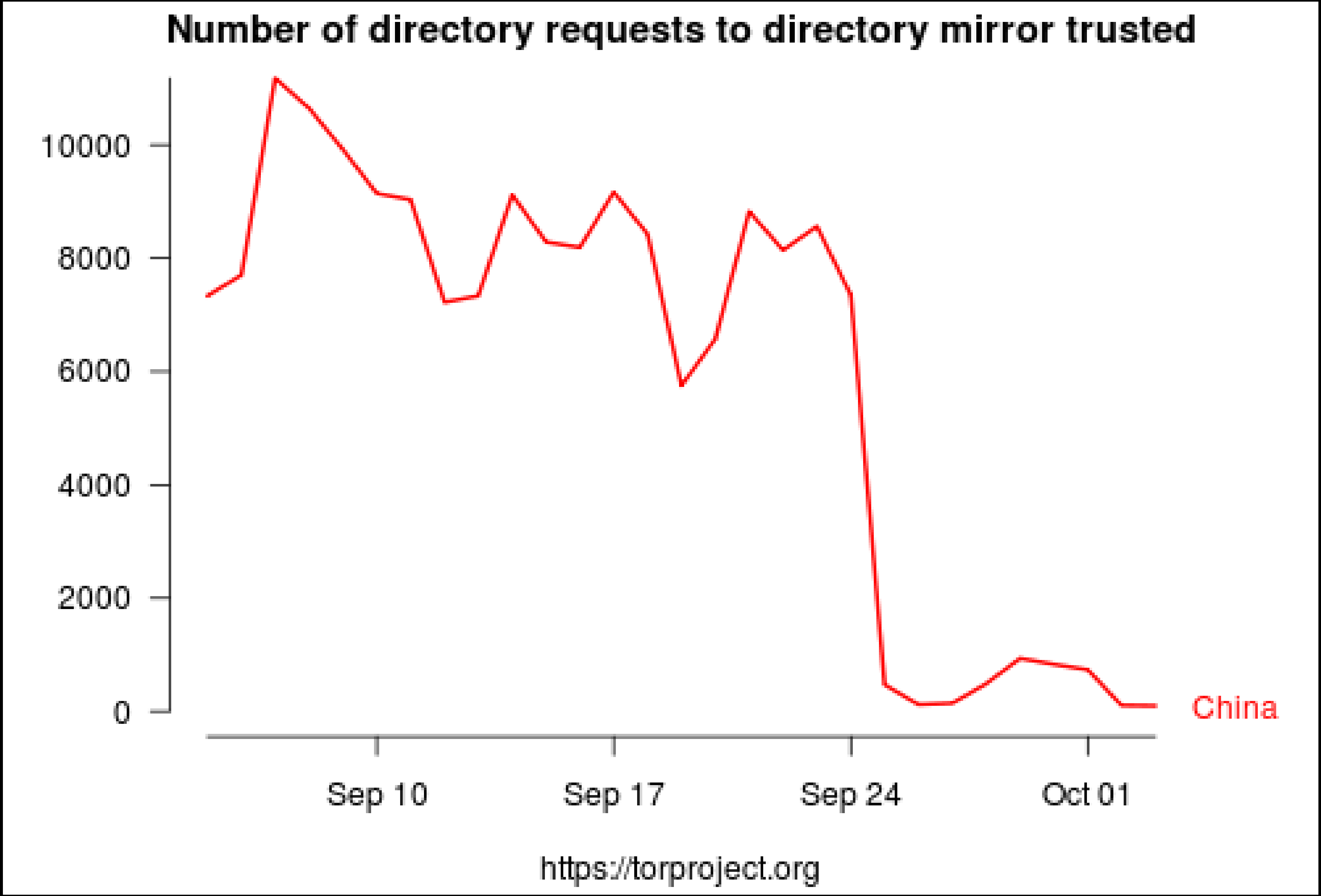

Alice
Alice
Alice
Alice
Alice
Blocked User
R3
Blocked User
Alice
R4
Bob
R2
Alice
Blocked User
Alice
R1
Alice
Blocked User
Alice
Alice
Alice
Blocked User
Alice
Alice

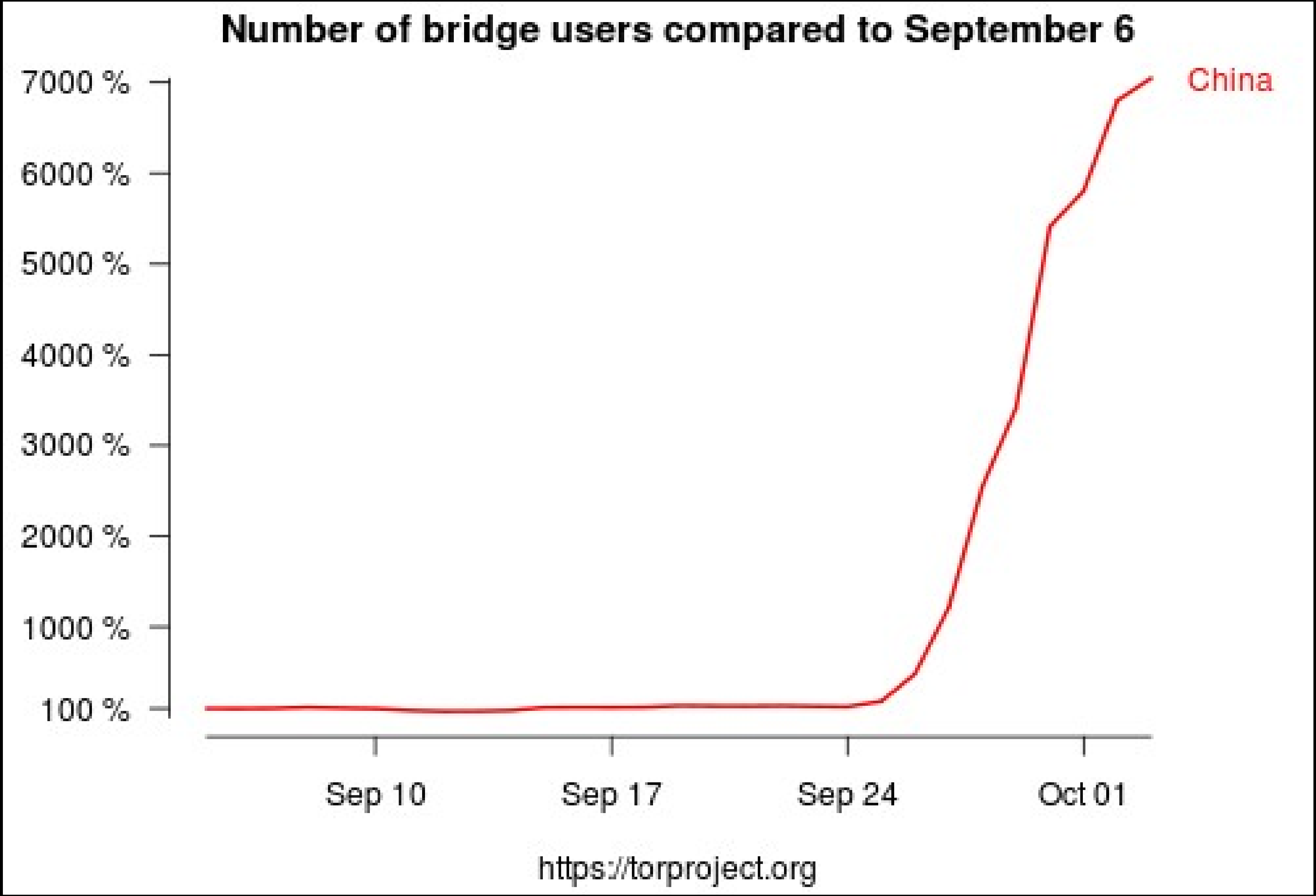
“Bridge” relays

- Hundreds of thousands of Tor users, already self-selected for caring about privacy.
- Rather than signing up as a normal relay, you can sign up as a special “bridge” relay that isn't listed in any directory.
- No need to be an “exit” (so no abuse worries), and you can rate limit if needed
- Integrated into Vidalia (our GUI) so it's easy to offer a bridge or to use a bridge


New or returning Tor clients per day
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
China
Iran
June 1
June 8
June 15
June 22
June 29, 2009
https://torproject.org


Number of bridge users compared to June 1
1,200 %
1,000 %
800 %
600 %
400 %
200 %
100 %
Iran
China
June 1
June 8
June 15
June 22
June 29, 2009
https://torproject.org


Number of directory requests to directory mirror trusted
10000
8000
6000
4000
2000
0
Sep 10
Sep 17
Sep 24
Oct 01
China
https://torproject.org



https://torproject.org

Javascript, cookies, history, etc

- Javascript refresh attack
- Cookies, History, browser window size, user-agent, language, http auth, ...
- Mostly problems when you toggle from Tor to non-Tor or back
- Mike Perry's Torbutton Firefox extension tackles many of these

Flash is dangerous too

- Some apps are bad at obeying their proxy settings.
- Adobe PDF plugin. Flash. Other plugins. Extensions. Especially Windows stuff: did you know that Microsoft Word is a network app?

Choose how to install it

- Tor Browser Bundle: standalone Windows exe with Tor, Vidalia, Firefox, Torbutton, Polipo, e.g. for USB stick
- Vidalia bundle: Windows/OSX installer
- Tor VM: Transparent proxy for Windows
- “Net installer” via our secure updater
- Amnesia Linux LiveCD

Only a piece of the puzzle

- Assume the users aren't attacked by their hardware and software
  - No spyware installed, no cameras watching their screens, etc
- Users can fetch a genuine copy of Tor?

Publicity attracts attention

- Many circumvention tools launch with huge media splashes. (The media loves this.)
- But publicity attracts attention of the censors.
- We threaten their *appearance* of control, so they must respond.
- We can control the pace of the arms race.

How to scale the network?

- The clients need to learn info about the relays they can use. Eventually this means partial network knowledge, and non-clique topology.
- Everybody-a-relay, and the anonymity questions that come with that.

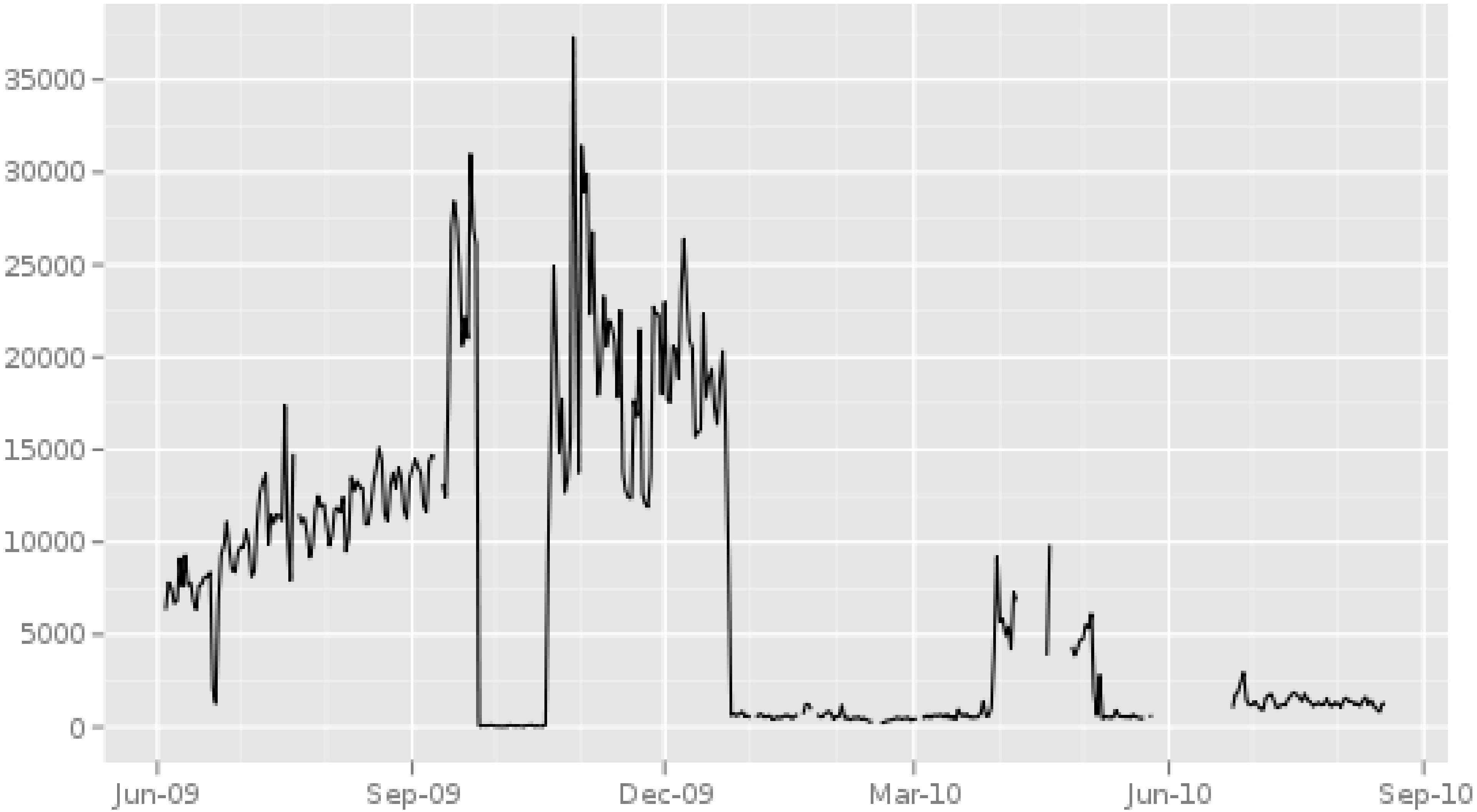
Advocacy and education

- Unending stream of people (e.g. in DC) who make critical policy decisions without much technical background
- Worse, there's a high churn rate
- Need to teach policy-makers, business leaders, law enforcement, journalists, ...
- Data retention? Internet driver's license?

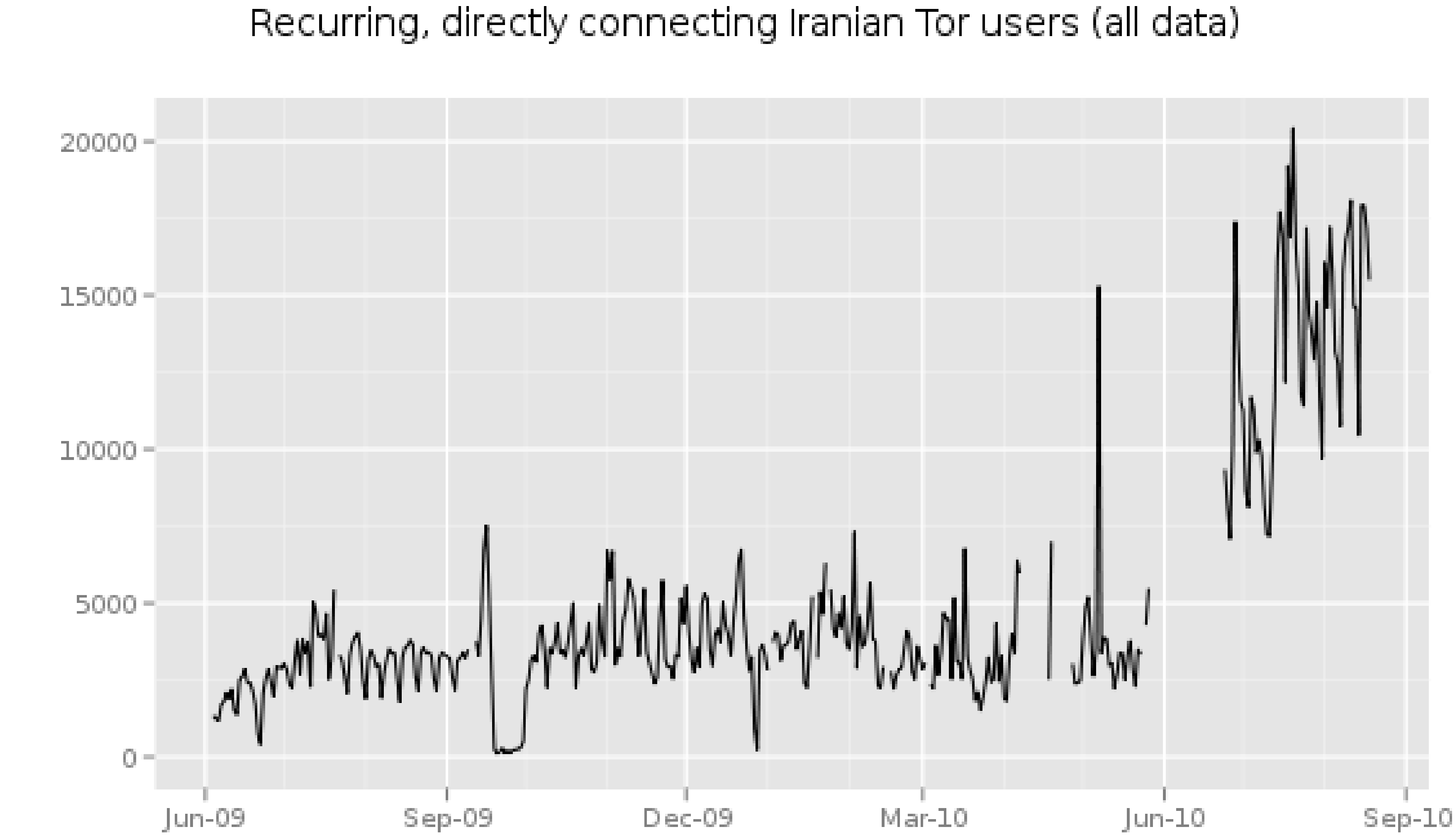
Our NSF EAGER

- 1) Invent and deploy new privacy-preserving algorithms to collect data about the Tor network, its performance, and its users
- 2) Publish this data, plus tools to analyze it
- 3) Figure out what else to measure and do it
- 4) Work with other research groups to make sure they get the data they need to solve the problems Tor actually has

Recurring, directly connecting Chinese Tor users (all data)

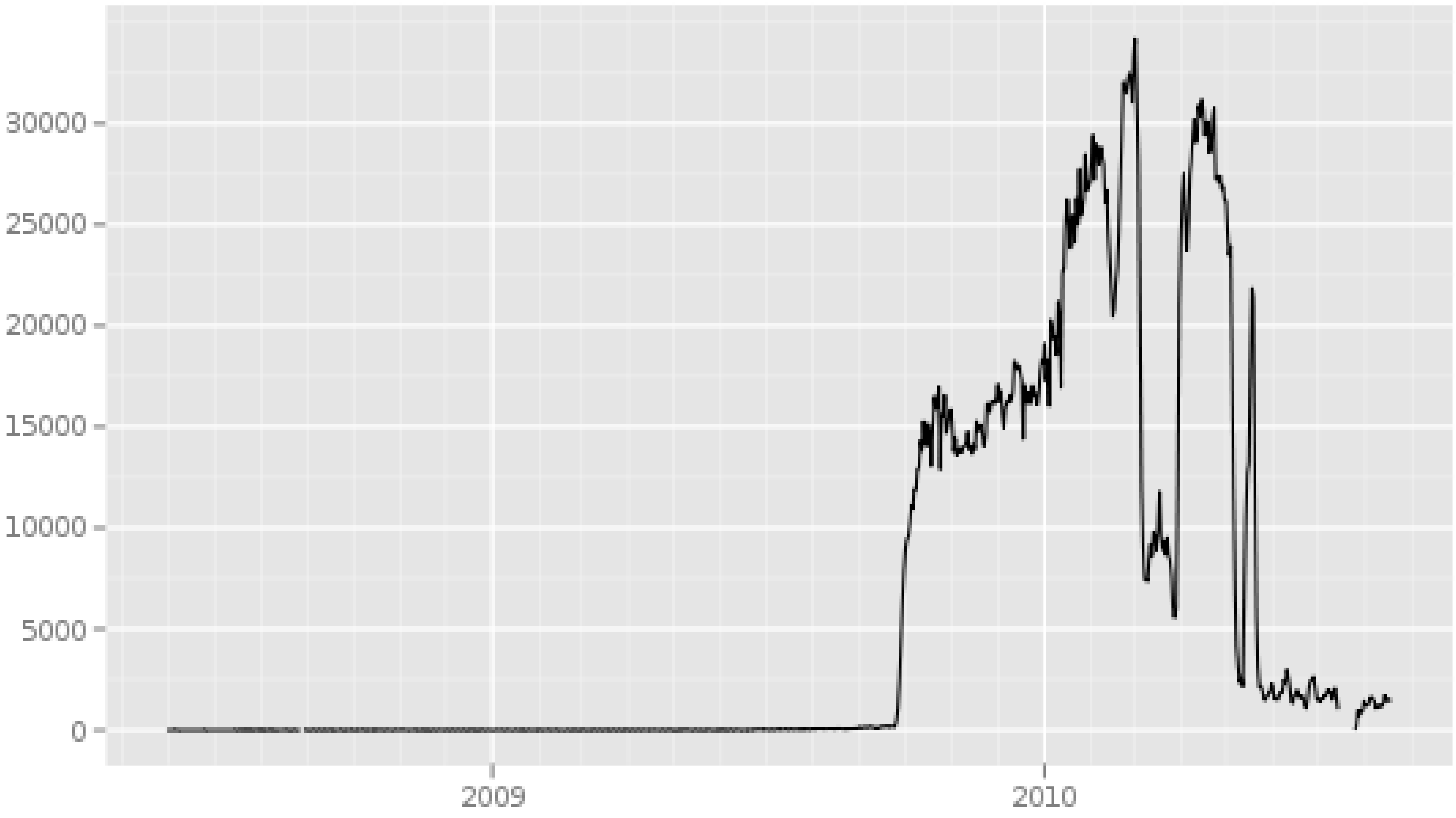


The Tor Project - https://metrics.torproject.org/



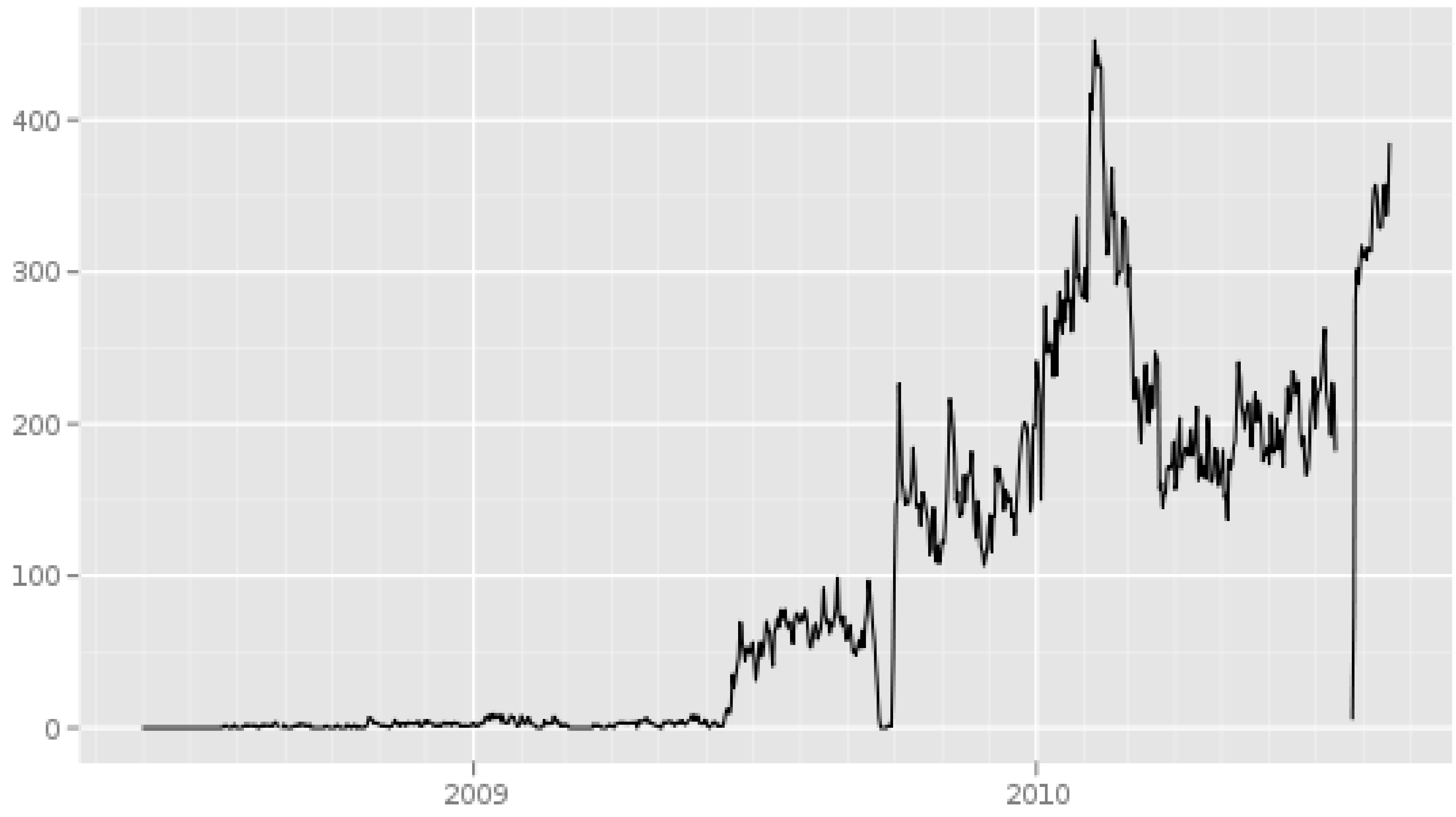
The Tor Project - https://metrics.torproject.org/

Chinese Tor users via bridges (all data)

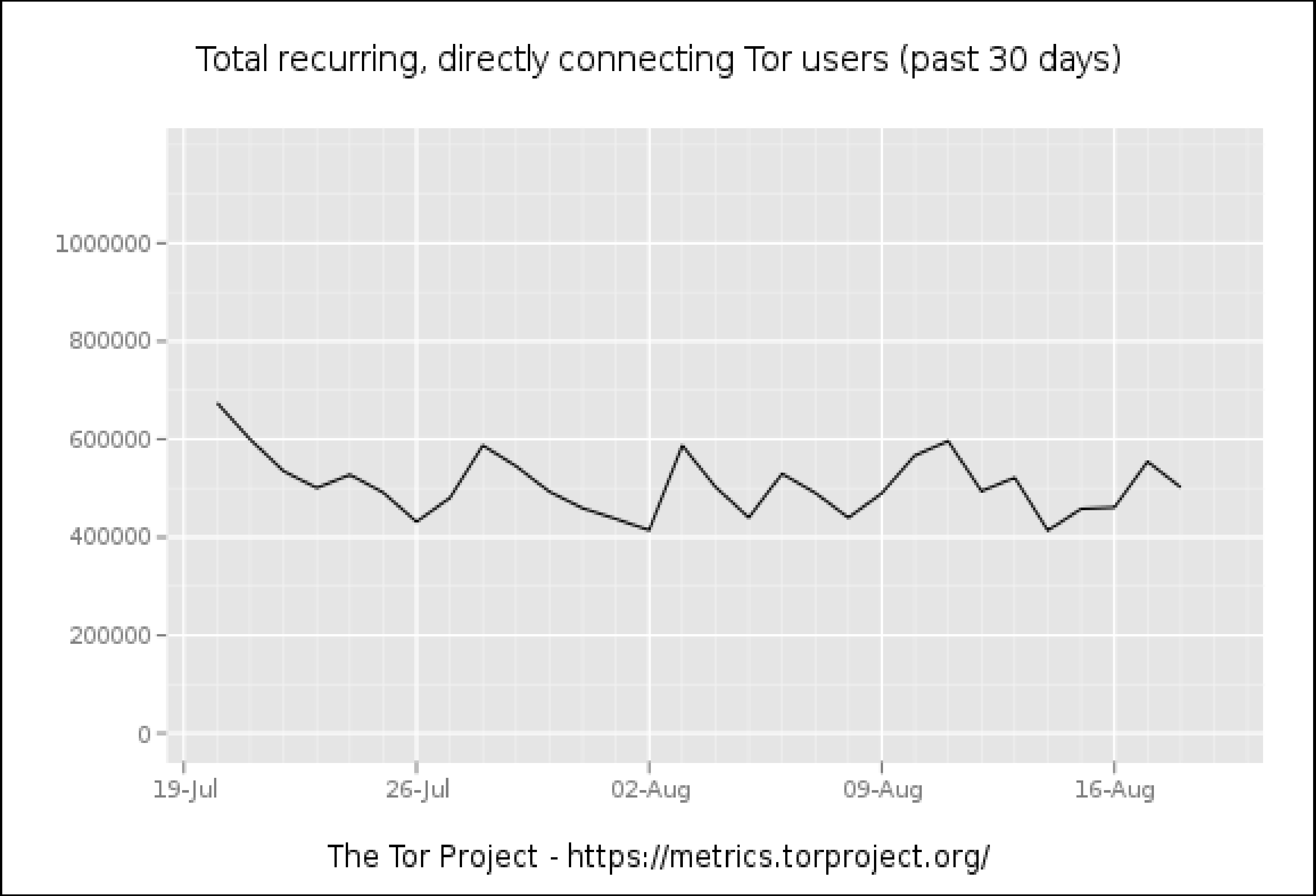


The Tor Project - https://metrics.torproject.org/

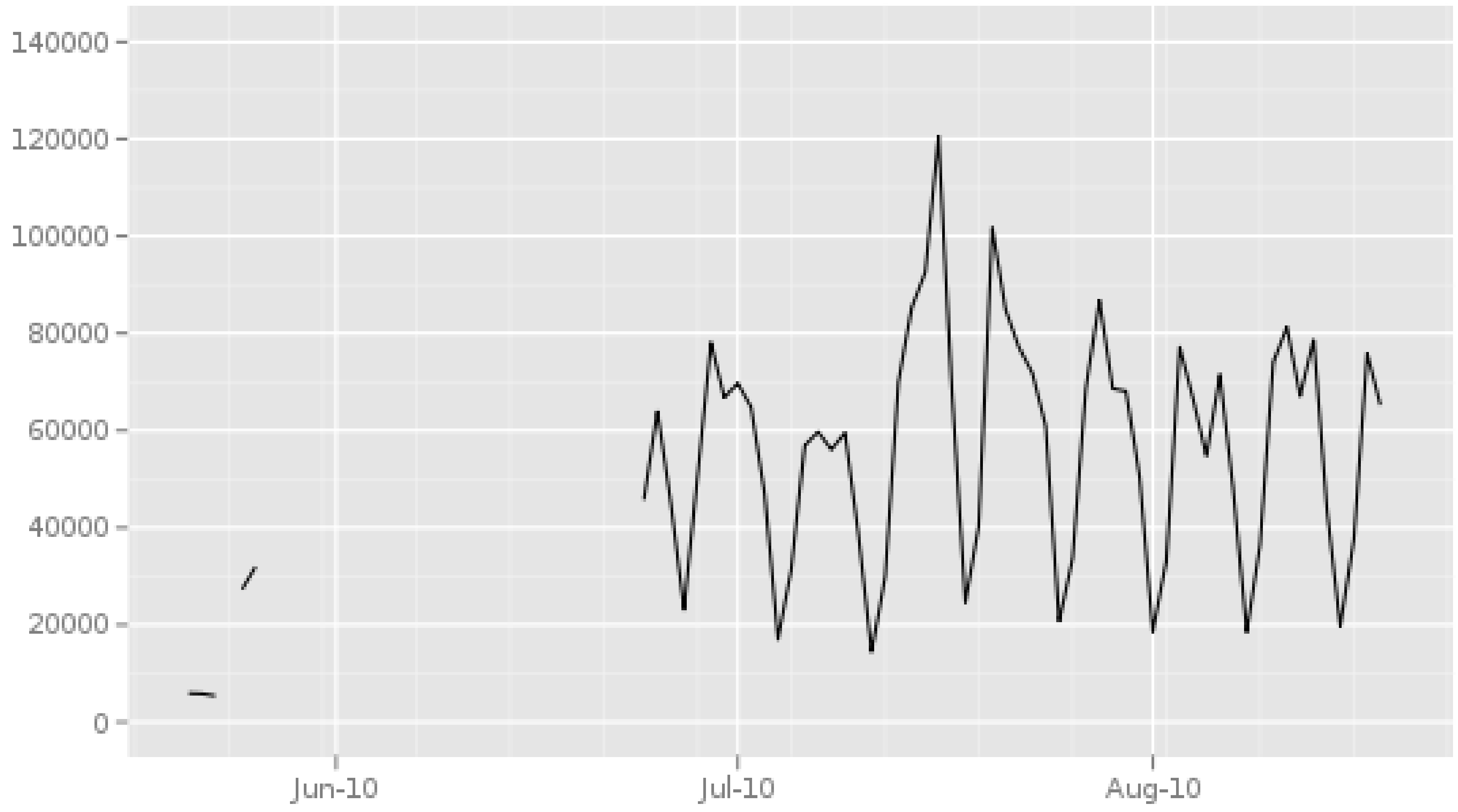
Iranian Tor users via bridges (all data)



The Tor Project - https://metrics.torproject.org/

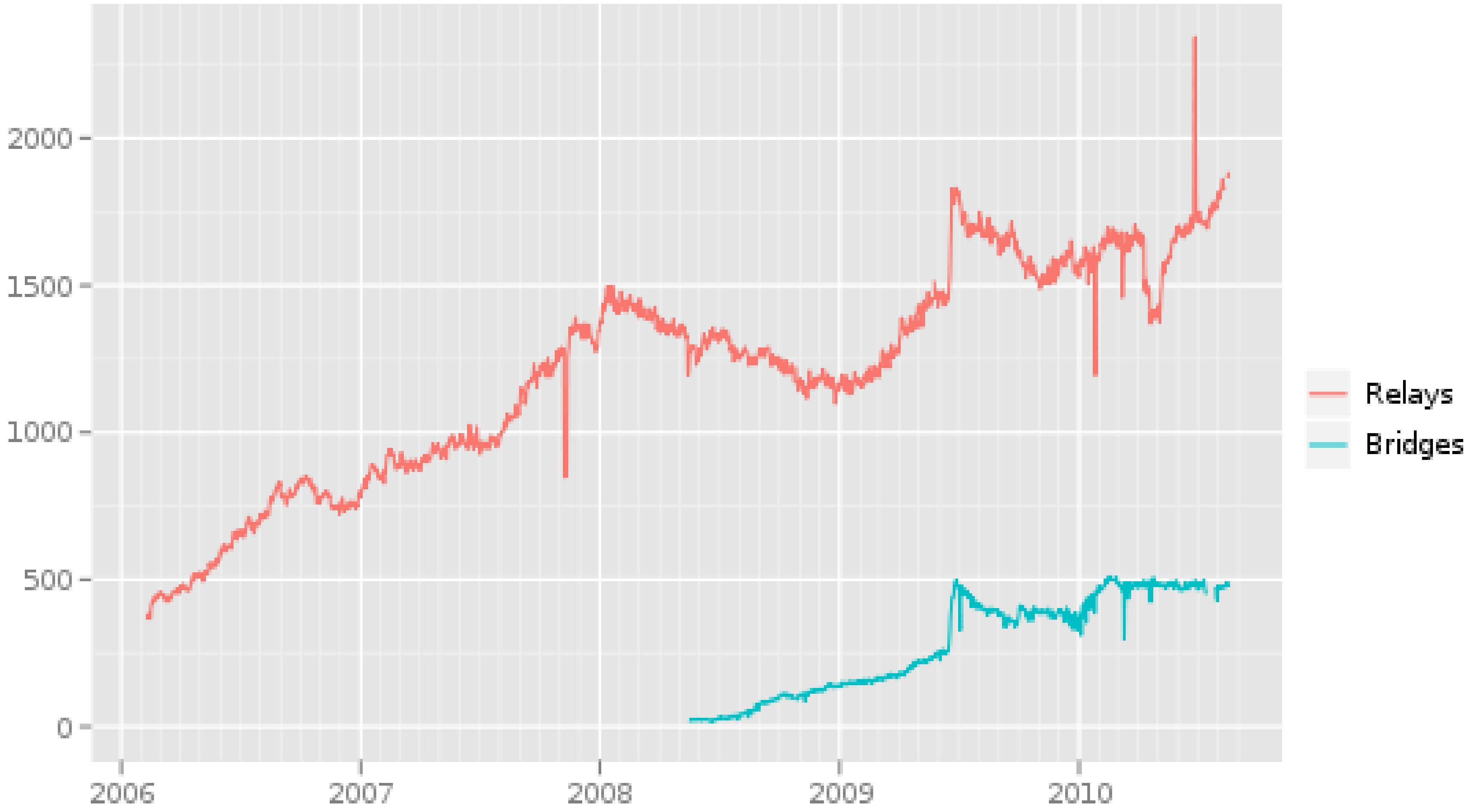

Total recurring, directly connecting Tor users (past 30 days)
1000000
800000
600000
400000
200000
0
19-Jul
26-Jul
02-Aug
09-Aug
16-Aug
The Tor Project - https://metrics.torproject.org/

Recurring, directly connecting South Korean Tor users (past 90 days)

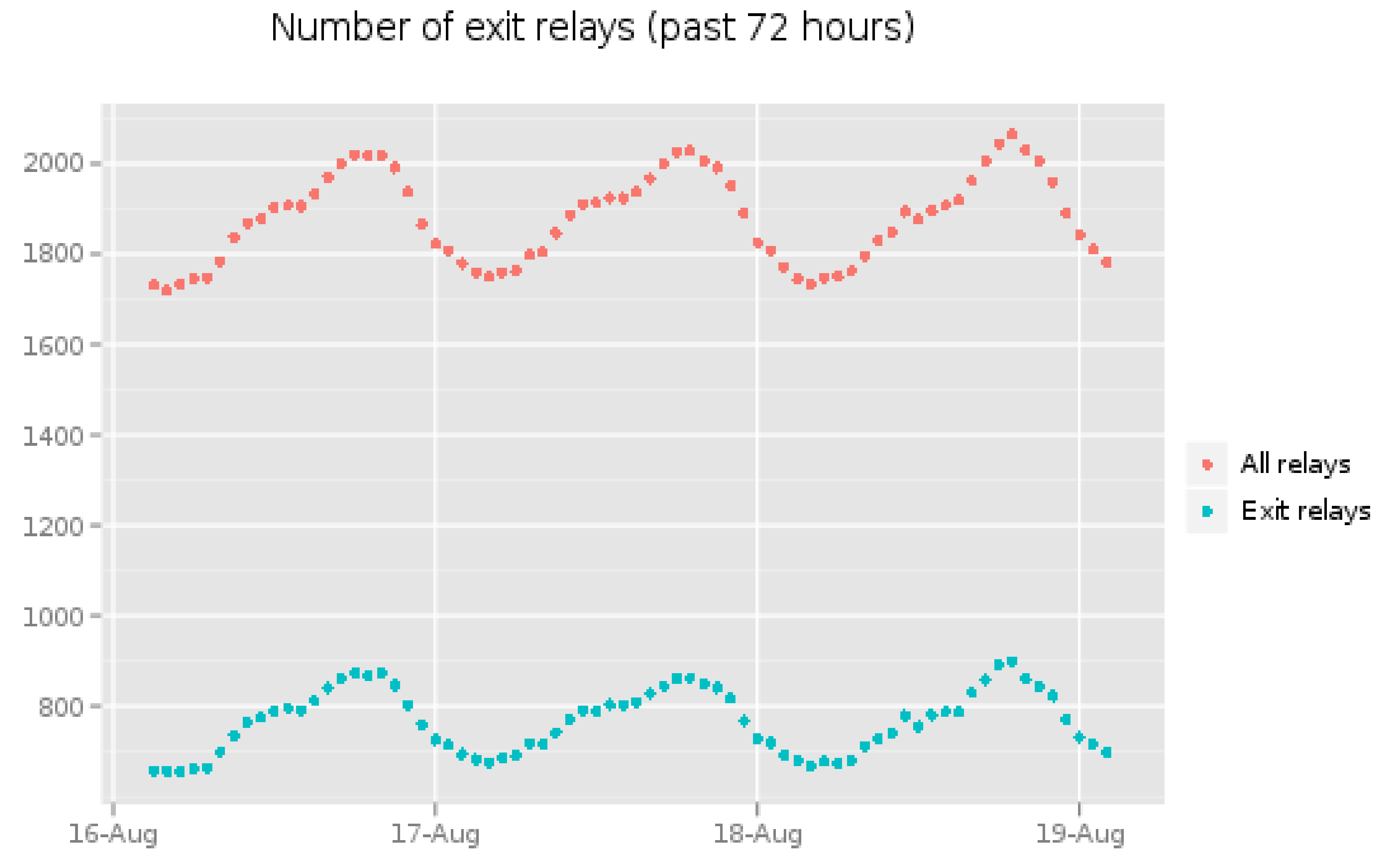


The Tor Project - https://metrics.torproject.org/

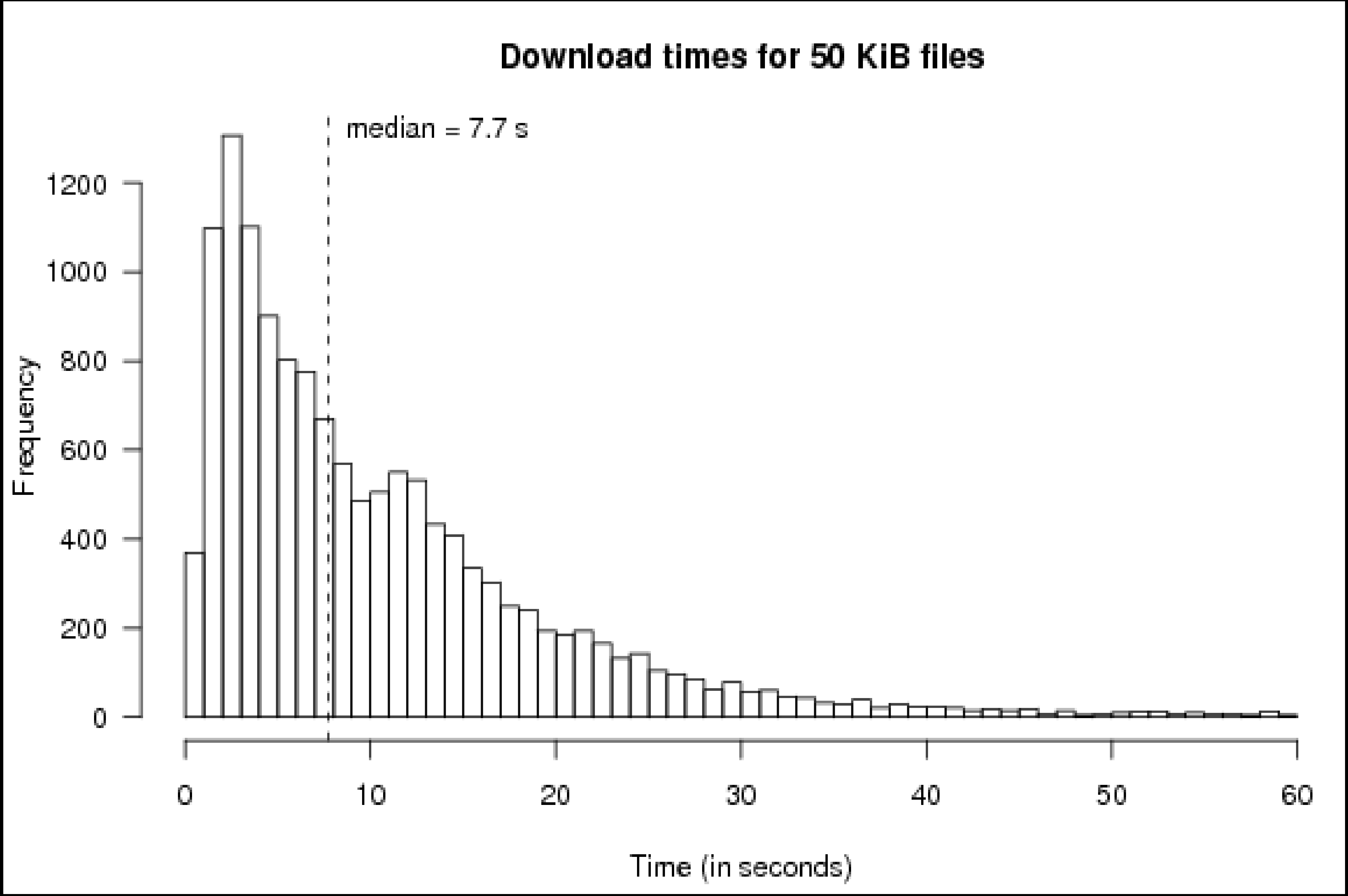
Number of relays and bridges (all data)



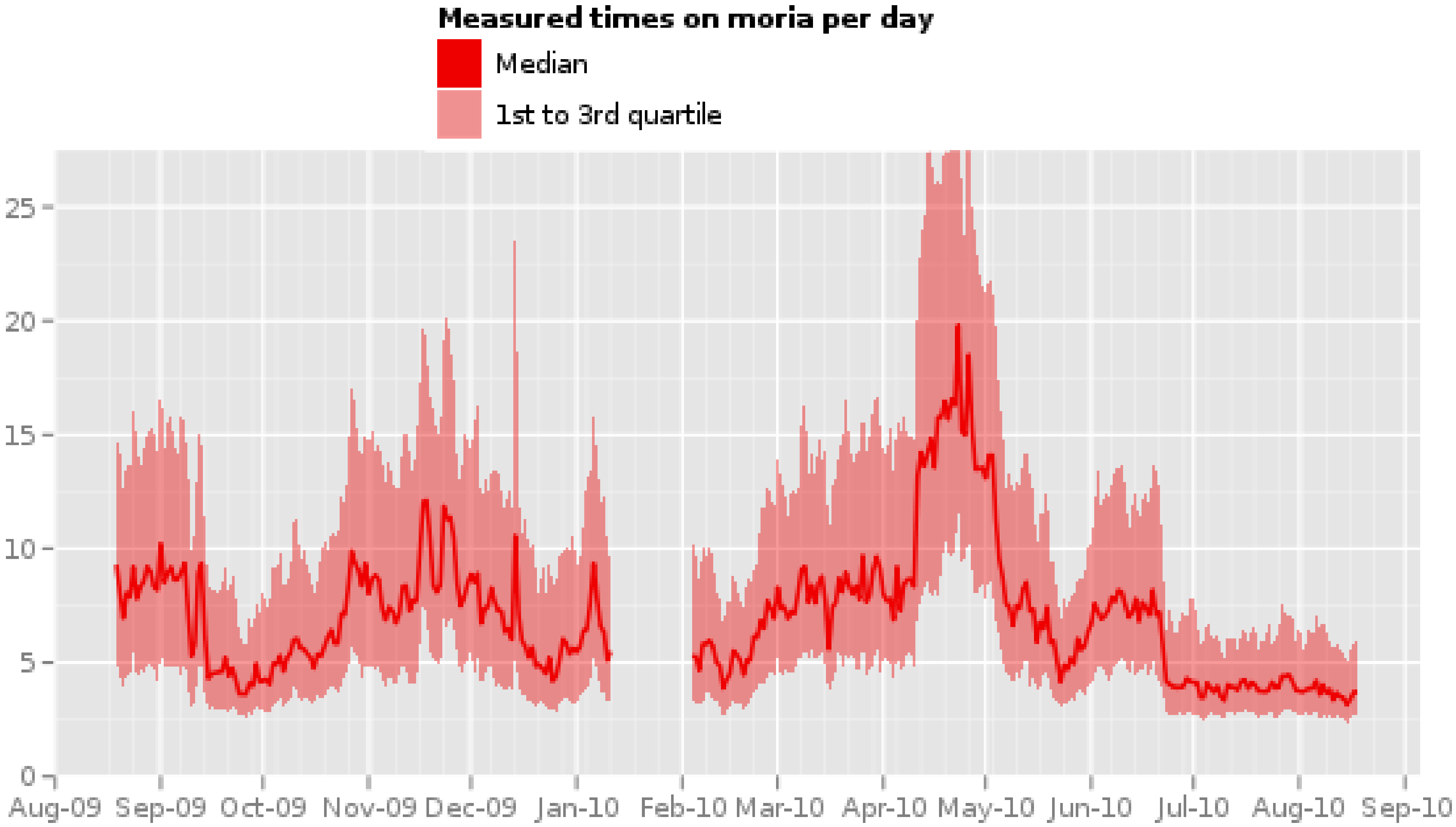
The Tor Project - https://metrics.torproject.org/



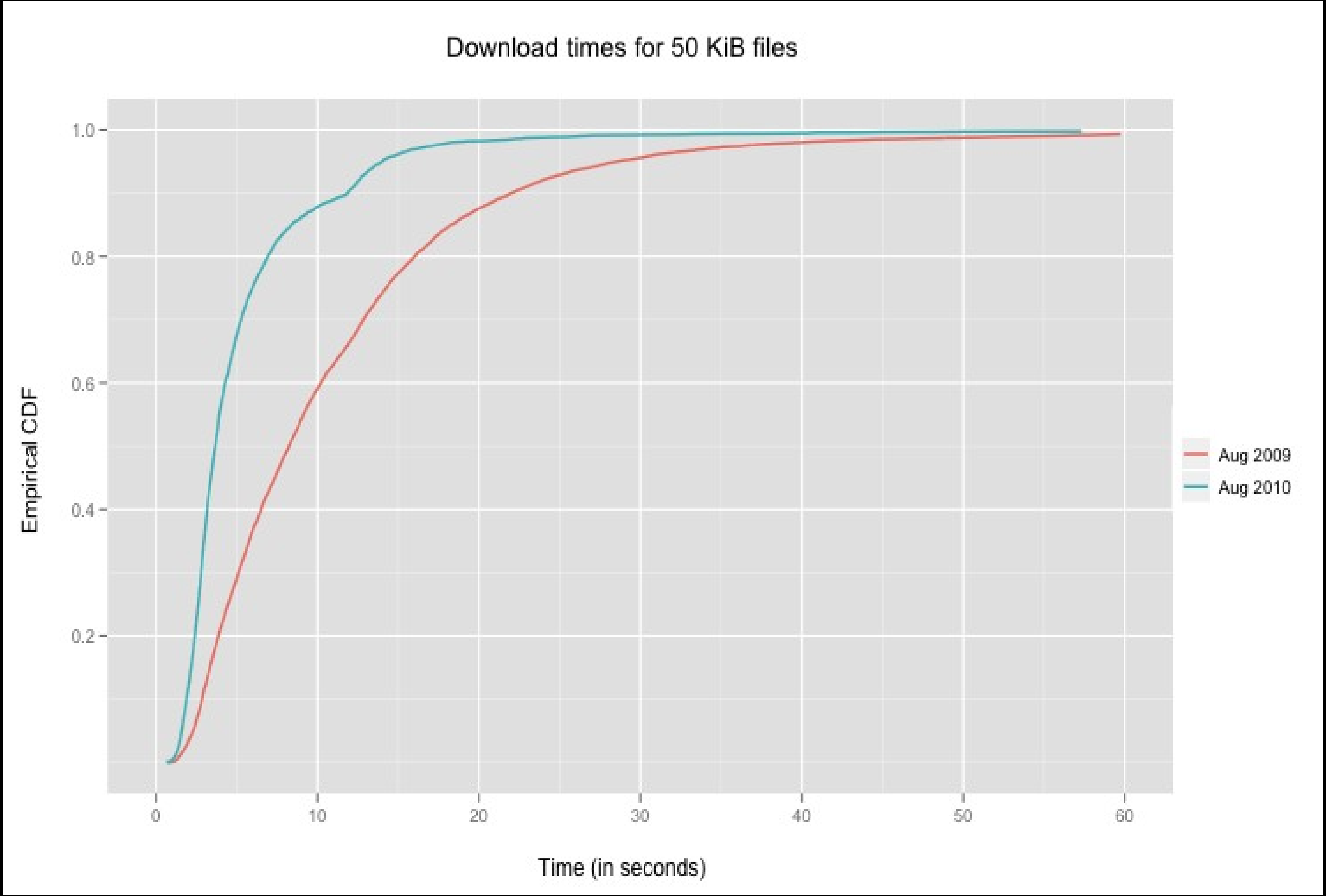
The Tor Project - https://metrics.torproject.org/


Download times for 50 KiB files
median = 7.7 s
Frequency
1200
1000
800
600
400
200
0
0
10
20
30
40
50
60
Time (in seconds)

Time in seconds to complete 50 KiB request



The Tor Project - https://metrics.torproject.org/


Download times for 50 KiB files
Empirical CDF
1.0
0.8
0.6
0.4
0.2
0
10
20
30
40
50
60
Time (in seconds)
Aug 2009
Aug 2010

Six performance problems

- Tor's congestion/flow control is not good
- Some users bulk-transfer over Tor
- Not enough capacity (run a relay!)
- Load balancing isn't right
- Not just high latency, but high variability
- High directory downloading overhead