Tor: a brief intro

Roger Dingedine
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?

- Alice
  - watch Alice!

- Anonymity network
  - Control part of the network!

- Bob
  - watch (or be!) Bob!
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.

“It's privacy!”
Anonymity serves different interests for different user groups.

- Private citizens: "It's privacy!"
- Businesses: "It's network security!"
Anonymity serves different interests for different user groups.

“It's traffic-analysis resistance!”

Governments

Anonymity

Private citizens

“It's privacy!”

Businesses

“It's network security!”
Anonymity serves different interests for different user groups.

- **Governments**: “It's traffic-analysis resistance!”
- **Human rights activists**: “It's reachability!”
- **Private citizens**: “It's privacy!”
- **Businesses**: “It's network security!”

Anonymity
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.

Hostile Bob

Incompetent Bob

Indifferent Bob

"I sell the logs."

"Oops, I lost the logs."

The AOL fiasco

"Hey, they aren't my secrets."

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

(the network can track too)
Businesses need to keep trade secrets.

“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

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

“What about insiders?”

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

“What does FBI Google for?”

“Somebody in that hotel room just checked his Navy.mil mail!”
Journalists and activists need Tor for their personal safety

"Did you just post to that website?"

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

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

"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 → Officer Alice → Municipal anonymity net → Investigated suspect → AliceCorp anonymity net → Competitor

"One of the 25 users on AliceNet."

"Looks like a cop."

"It's somebody at AliceCorp!"
... so, anonymity loves company!
Yes, bad people need anonymity too. But they are already doing well.
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.

- S1: Trusted directory
- S2: Trusted directory
- S3: Servers publish self-signed descriptors.
- Alice downloads consensus and descriptors from anywhere.
- Authorities publish a consensus list of all descriptors.
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
“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

https://torproject.org
Number of bridge users compared to September 6

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/
Recurring, directly connecting Iranian Tor users (all data)

The Tor Project - https://metrics.torproject.org/
Chinese Tor users via bridges (all data)

The Tor Project - https://metrics.torproject.org/
Total recurring, directly connecting Tor users (past 30 days)

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/
Number of exit relays (past 72 hours)

The Tor Project - https://metrics.torproject.org/
Download times for 50 KiB files

median = 7.7 s
Time in seconds to complete 50 KiB request

Measured times on moria per day

- Median
- 1st to 3rd quartile

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