Tor and circumvention: Lessons learned

Roger Dingledine
The Tor Project
https://torproject.org/
Today's plan

• 0) Crash course on Tor
• 1) History of Tor censorship attempts
• 2) Attacks on low-latency anonymity
• 3) Tor performance issues
• 4) Next research questions
What is Tor?

Online anonymity 1) open source software, 2) network, 3) protocol

Community of researchers, developers, users, and relay operators

Funding from US DoD, Electronic Frontier Foundation, Voice of America, Google, NLnet, Human Rights Watch, NSF, US State Dept, SIDA, ...
The Tor Project, Inc.

501(c)(3) non-profit organization dedicated to the research and development of tools for online anonymity and privacy
Estimated 400,000? daily Tor users
Threat model: what can the attacker do?

- **Alice**: Anonymity network
  - watch Alice!

- **Anonymity network**: Control part of the network!

- **Bob**: watch (or be!) Bob!
Anonymity isn't encryption: Encryption 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!”

Private citizens
Anonymity serves different interests for different user groups.

- “It's privacy!”
- “It's network security!”

- Private citizens
- Businesses
Anonymity serves different interests for different user groups.

“\textit{It's traffic-analysis resistance!}”

Governments

Anonymity

Private citizens

“\textit{It's privacy!}”

Businesses

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

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

(Blogger Alice
8-year-old Alice
Sick Alice
Consumer Alice
Oppressed Alice)
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.

“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?”

“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

“I run livejournal and track my users”

Filtered website

“I want to tell people what's going on in my country”

Monitored network

“Did you just post to that website?”

“Of course I tell China about my users”

“Where are the bloggers connecting from?”

“What does the Global Voices website say today?”

 Blocked Alice

Monitored website

“I think they're watching. I'm not even going to try.”
You can't get anonymity on your own: private solutions are ineffective...

“One of the 25 users on AliceNet.”

“Looks like a cop.”

“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.
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
Number of relays

The Tor Project - https://metrics.torproject.org/
Time in seconds to complete 50 KiB request

Measured times on all sources per day

- Median
- 1st to 3rd quartile

The Tor Project - https://metrics.torproject.org/
Directly connecting users from all countries

The Tor Project - https://metrics.torproject.org/
Today's plan

• 0) Crash course on Tor
• 1) History of Tor censorship attempts
• 2) Attacks on low-latency anonymity
• 3) Tor performance issues
• 4) Next research questions
Smartfilter/Websense (2006)

- Tor used TLS for its encrypted connection, and HTTP for fetching directory info.
- Smartfilter just cut all HTTP GET requests for “/tor/...”
Iran/Saudi Arabia/etc (2007)

- Picked up these Smartfilter/Websense rules by pulling an update
- The fix was to tunnel directory fetches inside the encrypted connection
- When Iran kicked out Smartfilter in early 2009, Tor's old (non-TLS) directory fetches worked again!
Iran throttles SSL (June 2009)

- We made Tor's TLS handshake look like Firefox+Apache.
- So when Iran freaked out and throttled SSL bandwidth by DPI in summer 2009, they got Tor for free.
Tunisia (summer 2009)

• As of the summer of 2009, Tunisia used Smartfilter to filter every port but 80 and 443
• And if they didn't like you, they could block 443 just for you
• You could use a Tor bridge on port 80, but couldn't bootstrap into the main network
• So we set up a Tor directory authority doing TLS on port 80
China (September 2009)

• China grabbed the list of public relays and blocked them

• They also enumerated one of the three bridge buckets (the ones available via https://bridges.torproject.org/)

• But they missed the other bridge buckets.
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
How do you find a bridge?

1) [https://bridges.torproject.org/](https://bridges.torproject.org/) will tell you a few based on time and your IP address
2) Mail bridges@torproject.org from a gmail address and we'll send you a few
3) I mail some to a friend in Shanghai who distributes them via his social network
4) You can set up your own private bridge and tell your target users directly
Number of directory requests to directory mirror trusted

https://torproject.org
China (March 2010)

- China enumerated the second of our three bridge buckets (the ones available at bridges@torproject.org via Gmail)
- We were down to the social network distribution strategy, and the private bridges
Chinese users via bridges

The Tor Project - https://metrics.torproject.org/
Iran (January 2011)

- Iran blocked Tor by DPI for SSL and filtering our Diffie-Hellman parameter.
- Socks proxy worked fine the whole time (the DPI didn't pick it up)
- DH p is a server-side parameter, so the relays and bridges had to upgrade, but not the clients
Directly connecting users from the Islamic Republic of Iran

The Tor Project - https://metrics.torproject.org/
Egypt (January 2011)

• When Egypt unplugged its Internet, no more Tor either.
Libya (March-July 2011)

- Libya might as well have unplugged its Internet.
- But they did it through throttling, so nobody cared.
Directly connecting users from Libya

The Tor Project - https://metrics.torproject.org/
Syria (June 2011)

• One ISP briefly DPIed for Tor's TLS renegotiation and killed the connections.
• A week later, that ISP went offline. When it came back, no more Tor filters.
• Who was testing what?
Directly connecting users from the Syrian Arab Republic

The Tor Project - https://metrics.torproject.org/
Iran (September 2011)

- This time, DPI for SSL and look at our TLS certificate lifetime.
- (Tor rotated its TLS certificates every 2 hours, because key rotation is good, right?)
- Now our certificates last for a year
- These are all low-hanging fruit. How do we want the arms race to go?
Directly connecting users from the Islamic Republic of Iran

The Tor Project - https://metrics.torproject.org/
October 2011 advances?

- Iran DPIs for SSL, recognizes Tor, and throttles rather than blocks?
- China DPIs for SSL, does active follow-up probing to see what sort of SSL it is?
Directly connecting users from India

The Tor Project - https://metrics.torproject.org/
Attacker's goals

Little reprisal against passive consumers of information.
Producers and distributors of information in greater danger.
Censors (actually, govts) have economic, political, social incentives not to block the whole Internet.
But they don't mind collateral damage.
What we're up against

Govt firewalls used to be stateless. Now they're buying fancier hardware.

Burma vs Iran vs China

New filtering techniques spread by commercial (American) companies :(  

How to separate “oppressing employees” vs “oppressing citizens” arms race?
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.
Using Tor in oppressed areas

Common assumption: risk from using Tor increases as firewall gets more restrictive.
But as firewall gets more restrictive, more ordinary people use Tor too, for more mainstream activities.
So the “median” use becomes more acceptable?
Trust and reputation

See January 2009 blog post by Hal Roberts about how some circumvention tools sell user data.

Many of these tools see circumvention and privacy as totally unrelated goals.
I CAN HAZ FREEDOM?

Tor
TorProject.org
Today's plan

• 0) Crash course on Tor
• 1) History of Tor censorship attempts
• 2) Attacks on low-latency anonymity
• 3) Tor performance issues
• 4) Next research questions
Snooping on Exit Relays (1)

- Lots of press in 2007 about people watching traffic coming out of Tor. (Ask your lawyer first...)
- Tor hides your location; it doesn't magically encrypt all traffic on the Internet.
- Though Tor does protect from your local network.
Snooping on Exit Relays (2)

- https as a “premium” feature
- Should Tor refuse to handle requests to port 23, 109, 110, 143, etc by default?
- Torflow / setting plaintext pop/imap “traps”
- Need to educate users?
- Active attacks on e.g. gmail cookies?
- Some research on exit traffic properties is legitimate and useful. How to balance?
Who runs the relays? (1)

- At the beginning, you needed to know me to have your relay considered “verified”.
- We've automated much of the “is it broken?” checking.
- Still a tension between having lots of relays and knowing all the relay operators
Who runs the relays? (2)

- What if your exit relay is running Windows and uses the latest anti-virus gadget on all the streams it sees?
- What if your exit relay is in China and you're trying to read BBC?
- What if your exit relay is in China and its ISP is doing an SSL MitM attack on it? (What if China owns a CA?)
Who runs the relays? (3)

- What happens if ten Tor relays show up, all on 149.9.0.0/16, which is near Washington DC?
- “EnforceDistinctSubnets” config option to use one node per /16 in your circuit
- At most 2 relays on one IP address
- How about ASes? IXes? Countries?
Tor Browser Bundle traces

• We want to let you use Tor from a USB key without leaving traces on the host
• “WINDOWS/Prefetch” trace
• Windows explorer's “user assist” registry entry
• Vista has many more?
Application-level woes (1)

- 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 tackles many of these
Application-level woes (2)

- Some apps are bad at obeying their proxy settings.
Traffic confirmation

- If you can see the flow into Tor and the flow out of Tor, simple math lets you correlate them.
Countermeasures?

- Defensive dropping (2004)? Adaptive padding (2006)?
- Traffic morphing (2009), Johnson (2010)
- Tagging attack, traffic watermarking
Tor gives three anonymity properties

• **#1**: A local network attacker can't learn, or influence, your destination.
  − Clearly useful for blocking resistance.
• **#2**: No single router can link you to your destination.
  − The attacker can't sign up relays to trace users.
• **#3**: The destination, or somebody watching it, can't learn your location.
  − So they can't reveal you; or treat you differently.
Tor's safety comes from diversity

- #1: Diversity of relays. The more relays we have and the more diverse they, the fewer attackers are in a position to do traffic confirmation. (Research problem: measuring diversity over time)
- #2: Diversity of users and reasons to use it. 40000 users in Iran means almost all of them are normal citizens.
Website fingerprinting

- If you can see an SSL-encrypted link, you can guess what web page is inside it based on size.
- Does this attack work on Tor? “maybe”
- Considering multiple pages (e.g. via hidden Markov models) would probably make the attack even more effective.
Low-resource routing attacks

- Bauer et al (WPES 2009)
- Clients use the bandwidth as reported by the relay
- So you can sign up tiny relays, claim huge bandwidth, and get lots of traffic
- Fix is active measurement.
Long-term passive attacks

- Matt Wright's predecessor attack
- Øverlier and Syverson, Oakland 2006
- The more circuits you make, the more likely one of them is bad
- The fix: guard relays
Denial of service as denial of anonymity

- Borisov et al, CCS 2007
- If you can't win against a circuit, kill it and see if you win the next one
- Guard relays also a good answer here.
Epistemic attacks on route selection

- Danezis/Syverson (PET 2008)
- If the list of relays gets big enough, we'd be tempted to give people random subsets of the relay list
- But, partitioning attacks
Congestion attacks (1)

- Murdoch-Danezis attack (2005) sent constant traffic through every relay, and when Alice made her connection, looked for a traffic bump in three relays.
- Couldn't identify Alice – just the relays she picked.
Congestion attacks (2)

- Hopper et al (2007) extended this to (maybe) locate Alice based on latency.
- Chakravarty et al (2008) extended this to (maybe) locate Alice via bandwidth tests.
- Evans et al (2009) showed the original attack doesn't work anymore (too many relays, too much noise) – but “infinite length circuit” makes it work again?
Profiling at exit relays

- Tor reuses the same circuit for 10 minutes before rotating to a new one.
- (It used to be 30 seconds, but that put too much CPU load on the relays.)
- If one of your connections identifies you, then the rest lose too.
- What's the right algorithm for allocating connections to circuits safely?
Declining to extend

- Tor's directory system prevents an attacker from spoofing the whole Tor network.
- But your first hop can still say “sorry, that relay isn't up. Try again.”
- Or your local network can restrict connections so you only reach relays they like.
Attacks on Tor

- Pretty much any Tor bug seems to turn into an anonymity attack.
- Many of the hard research problems are attacks against all low-latency anonymity systems. Tor is still the best that we know of – other than not communicating.
- People find things because of the openness and thoroughness of our design, spec, and code. We'd love to hear from you.
Today's plan

• 0) Crash course on Tor
• 1) History of Tor censorship attempts
• 2) Attacks on low-latency anonymity
• 3) Tor performance issues
• 4) Next research questions
Total relay bandwidth

- Advertised bandwidth
- Bandwidth history

The Tor Project - https://metrics.torproject.org/
Time in seconds to complete 5 MiB request

Measured times on all sources per day

- Median
- 1st to 3rd quartile

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

• Not enough capacity
• Bulk downloaders
• Multiplexing circuits over one TCP flow
• ExperimenTor / Shadow
• Flow control, N23. Slow first hop?
• Drop relays with less than x bandwidth
Today's plan

- 0) Crash course on Tor
- 1) History of Tor censorship attempts
- 2) Attacks on low-latency anonymity
- 3) Tor performance issues
- 4) *Next research questions*
BridgeDB needs a feedback cycle

- Measure how much use each bridge sees
- Measure bridge blocking
- Then adapt bridge distribution to favor efficient distribution channels
- (Need to invent new distribution channels)
Measuring bridge reachability

• Passive: bridges track incoming connections by country; clients self-report blockage (via some other bridge)

• Active: scan bridges from within the country; measure remotely via FTP reflectors

• Bridges test for duplex blocking
Other components

Traffic camouflaging

Super-encrypt so no recognizable bytes?
Shape like HTTP?

We're working on a modular transport API

Need “obfuscation” metrics?
Other discussion points

- Can bridges just be proxies?
- Secure update (Diginotar/Iran)
- Usability work
- Hidden services