Today

Online tracking and surveillance

VPNs

Tor

Online tracking & surveillance

Advertising & marketing

Data brokerage

Politics

Censorship

Repression

Tracking & surveillance

Cookies

Images

JavaScript

Browser fingerprinting

DPI

Websites can store include _____ and _____ cookies by asking your browser to store a cookie or by including content from a third-party site (e.g., Google Analytics) that does so. Browsers will commonly block third-part cookies when using "Private Browsing" mode, but you might need to tell your browser if you want to block these cookies all of the time. Just try inspecting the cookies that your browser holds about you on behalf of a site that shows ads... you might be a little surprised at what you find!

Images can also be used to track you: it's very common to have websites (even HTML emails!) to have a 1px × 1px transparent image that loads, e.g., https://tracker.example.com/tracker.png?userid=abd64cd8a0df.

JavaScript can be used to load all kinds of code that does all kinds of things, including...

_____. There are a lot of things that you can learn about a user's browser without JavaScript (IP, User-Agent, Language, etc.), but running JavaScript can access other details, e.g., screen height/width, installed fonts, installed plugins. Sure, lots of people run Firefox v90 on macOS, but how many have exactly my set of fonts, languages, timezone, etc.? The intersection of these various sets of people can be used to identify people surprisingly well... just try visiting https://amiunique.org ! Also maybe try https://coveryourtracks.eff.org just for fun.

Finally, we know that deep packet inspection is a real issue in some environments. Companies might want to see what you're doing on your work computer, but in many parts of the world, governments want to see what you're saying to others!

Postal analogy

Letters:

- *Cabinet noir**
- USPS *Mail Isolation and Tracking*†
- Remailing

---

* Chisholm, Hugh, ed. (1911). "Cabinet Noir". Encyclopædia Britannica (11th ed.). Cambridge University Press.

† "Postal Service Confirms Photographing All U.S. Mail", Nixon, *The New York Times*, 2 Aug 2013.

The *cabinet noir* was a feature of the French postal system since the 17th Century, later replicated by many other postal systems. Their job was to open letters that the government wanted to read, copy them and re-seal them before delivering them. This is analogous to deep-packet inspection. In fact, the analogy is very strong to rooms that contain fibre-splitting equipmment, as is alleged to be kept in places like Room 641A of the SBC Communications Building in San Francisco.

A less-invasive program for mail does exist in the form of the US Postal Service's — pubically-acknowledged! — *Mail Isolation and Tracking* program. In this program, the *outside* of every piece of mail in the US may be scanned for later use by law enforcement.

One way to try to hide metadata about who is mailing whom is to use a *remailer* service, which will receive your mail, open it and send its contents to someone else under new cover. However, if you're trying to hide criminal or other activity, you might want to consider that not very many people use such services, so merely using one might make you stand out a bit...

Put that box in another box...

A remailer network can work like a Matryoska doll set, putting packages inside of packages inside of packages, with each hop through the network removing one layer of packaging.

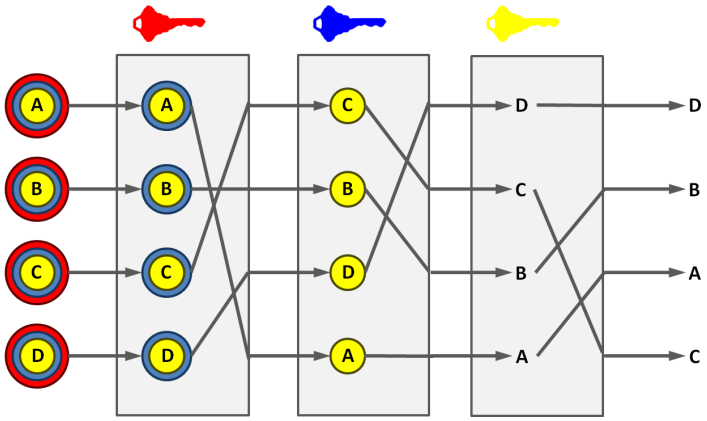
Private email

Network of *remailers*

Double-blind

Attacks

Timing, $n - 1$...



Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms", *Communications of the ACM*, 24(2), pp84-90, 1981. DOI: 10.1145/358549.358563

In the 1980s, some crypto researchers brought this concept of remailers into the world of electronic mail. They proposed sending encrypted email to a remailer — called a _____ — that would decrypt one layer of encryption and then send its contents on to the next mix, which would do the same, etc. Eventually, mail would go from senders to recipients, but an outside observer wouldn't be able to tell who was emailing whom.

One neat property of this system was that, even if an attacker controlled one of the mix nodes, they *still* wouldn't be able to see who was emailing whom. Each step in the process was (or, technically, still is) double-blind.

There are problems with the mix concept, however. Firstly, if a *global passive adversary* can see that Alice emails a mix, which emails a mix, which emails a mix, which emails Bob, we can have a pretty good guess about who's emailing whom. Thus, mixes have to _____ _____ to email in order to conceal identities. In particular, mixes would typically wait either a fixed amount of time or until they'd gathered up a batch of emails before sending them all on at the same time. Even *then*, however, if an attacker knew that a mix would wait for 50 emails, they could wait for you to send one, then send 49 themselves, then see where all 50 go (the $n - 1$ attack). They already know where 49 are going, so the remaining one must be yours!

Mixing lessons

Anonymity set

- size often unimportant!
- probable cause, reasonable doubt or mere suspicion?

Latency matters

The perfect vs the good

People who used anonymous remailers were very excited about the concept of an *anonymity set*, the set of _____. The theory was, if you can only prove that I *might've* been the one to send that email, but it *might've* been any of these other 49 people, I have reasonable doubt!

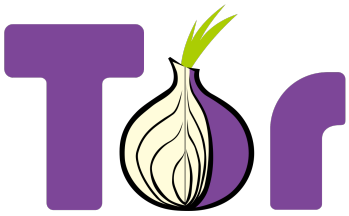
The problem with this thinking is twofold. Firstly, there are lots of circumstances in which the size of the anonymity set isn't terribly important. A criminal prosecutor might need to prove something beyond a reasonable doubt to put you in prison, but simply using a remailer at around the right time could give the police "reasonable suspicion" to have a conversation with you, or for your boss to have your work computer examined. In some circumstances, the adversary might not care much at all about the size of an anonymity set: they'd just as happily kick down 50 doors as one.

Secondly, there ain't that many people using anonymous remailers to begin with. If your adversary's threshold for action is mere suspicion, they can simply act against _____ _____!

Another key lesson from remailers is that latency is important. Nobody wants to use a high-latency service, which means that using the service makes you stand out, which makes even fewer people want to use the service, etc. If, on the other hand, a low-latency service can provide privacy properties for "regular" people as well as those seeking to evade censorship, it could be popular, which will provide even better privacy properties!

Tor

Dominant tool for:



- censorship resistance
- online privacy

Imperfect but usually "good enough"...

even against some strong adversaries!

Tor no longer uses the language of "anonymity". Anonymity is really hard, and it's also a word that feels devious ("what, do you have something to hide?"). Instead, Tor is typically described as a censorship resistance tool (a high-imporance use case) as well as a tool for online privacy (a high-volume use case).

Owing to its low-latency operation, Tor _____ _____. If there's a true Dolev-Yao attacker, or even just a passive attacker, who can see every message in the world, they can use correlation to figure out who's talking to whom. However, if the network can be large and popular enough, it becomes very difficult to actually become a global passive adversary.

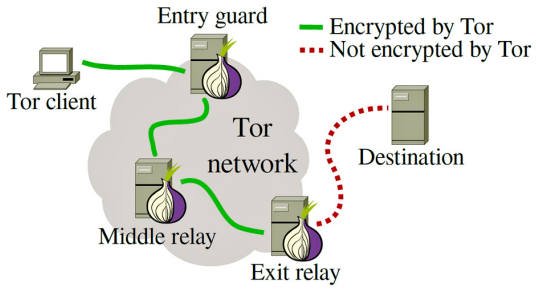
Tor mechanics

*Tor: The Onion Router**

(not called that any more)

Telescoping routing

Client builds *circuit* from *guard*, *relay* and *exit* nodes



Source: KitSploit

---

* "Tor: The Second-Generation Onion Router", Dingledine, Mathewson and Syverson, in *Proceedings of the 13th USENIX Security Symposium*, 2004. Available: usenix.org

Tor is _____: the _____ decides which nodes it wants to use. This is different from typical IP routing, where each router can decide which path a packet ought to take.

That said, Tor is an _____ that runs _____. Tor clients choose guard, relay and exit nodes from a directory of publically-visible Tor nodes.

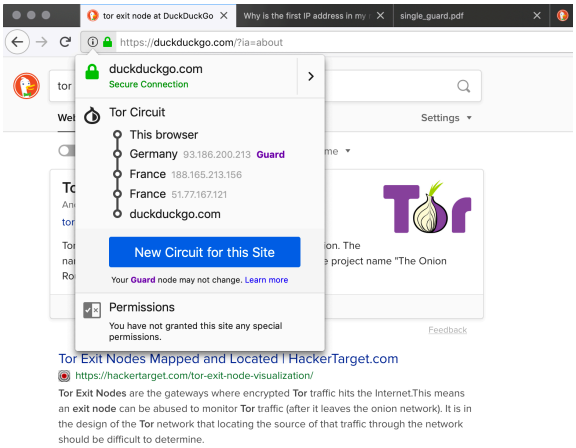
Tor nodes

Directory nodes

Guard nodes

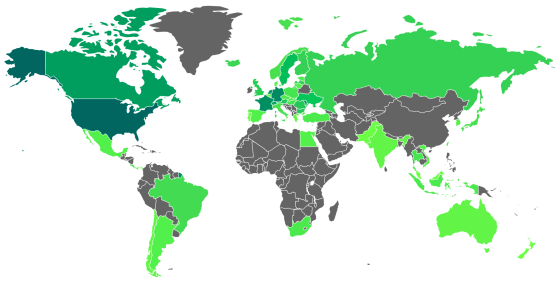
Relay nodes

Exit nodes



Ethical considerations

Dual-use technology



Exit nodes

Social contract

https://blog.torproject.org/tor-social-contract

What is a "bad day" for your users?

Tor is yet another of these dual-use technologies. It is used by people living under authoritarian governments who want to share uncensored news about the world, organize protests, etc. It is *also* used by people who want to share child sexual abuse imagery, organize terrorism, etc.

This dual-use nature of the technology can come home pretty quickly if you choose to run a Tor exit node. Running a Tor guard or relay node is a fairly safe thing to do: people open encrypted tunnels to you and you open encrypted tunnels to other Tor nodes. If you run an exit node, however, whatever stuff people want to do via Tor is exposed in the connections that you make to real web servers. If someone is retrieving uncensored news, it looks like you're retrieving it. If someone is sharing images, which can include awful things like child sexual abuse material (CSAM), it looks like you're sharing them. Thus, running a Tor exit node can be a risky thing to do.

Tor is largely run by a community of people — including lots of academics — who have stated objectives around advancing human rights, advocacy, research and other principles. Like other forms of technology, Tor reflects the goals of the people who make it; unlike many forms of technology, Tor's social contract is explicitly stated.

Also unlike many forms of technology, the risks of performing research studies on a live network can have _____. Thus, the *Privacy Enhancing Technologies* community thinks about research ethics much more explicitly than many people in computer science / engineering. In fact, other domains are only now starting to catch up.

Using Tor

What does telescoping routing buy you?

Proxy usage

- usability
- tracking vs surveillance

Tor Browser

What telescoping routing *does* buy you is reduced visibility from all but global adversaries (i.e., probably almost any adversary you might care about).

Counterintuitively, however, using a privacy-enhancing proxy often means *not* using features like TLS! The proxied mode of Tor needs to see your browser's traffic so that it can strip out lots of identifying information, etc. Thus, the current recommendation is to _____ _____.

Instead, the Tor Browser integrates Tor within its own version of Firefox. This allows you to avoid the mistake of _____ (something the Dread Pirate Roberts could've used!). It also does other fingerprint-reducing things like _____ _____; it's recommended to not install any plugins beyond the included blockers.

Hidden services

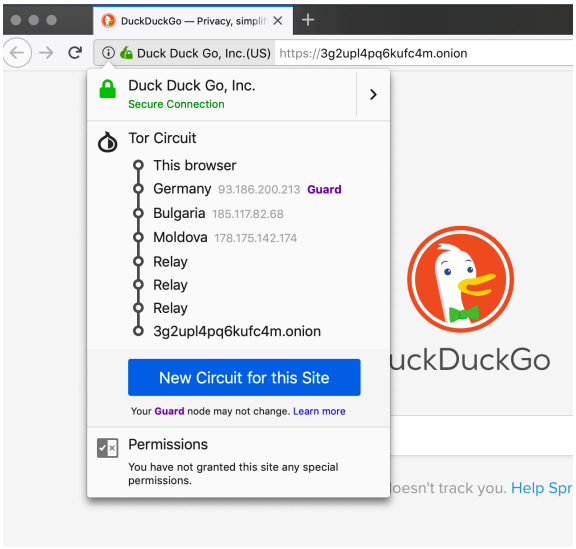
Rendezvous at a relay

Client, server *both* hiding

a.k.a., "onion" services

e.g.,
duckduckgo[...]wzczad.onion

a.k.a., "dark web"



"Onion" services

Web services:

- DuckDuckGo
- ProtonMail

SecureDrop:

- CBC
- Globe and Mail

Others:

Hidden Wiki
Sci-Hub (now dead)
CIA (yes, *that* CIA)
Facebook (!?)

... and untold other places, many of which are **not good**
(the now-defunct Silk Road being just the tip of the iceberg)

Onion services historically had shorter names like sml5wmpuq7ifq2mh.onion, but as time marched on, v3 onion services were required to use better cryptographic algorithms with longer data lengths (e.g., SHA-1 became SHA-3). Consequently, they now have names like a4zum5ydurvljrohxqp2rjjal5kro4ge2q2qizuonf2jubkhcr627gad.onion.

These names are a bit awkward to remember. As an alternative, the Freedom of the Press Foundation (also: http://fpfjxcrmw437h6z2xl3w4czl55kvkmxpapg37bbopsafdu7q454byxid.onion), which is the organization behind SecureDrop, maintains a set of "onion names" for SecureDrop sites. Unlike DNS, which requires you to tell somebody what site you want to visit, onion names are distributed to everyone's Tor Browser ahead of time (though only to the desktop version at present). This allows you to access the above-named onion service at the much-easier-to-remember theglobeandmail.securedrop.tor.onion.

Cats and mice

Blocking and Bridge nodes

Traffic analysis*†

Pluggable transport

---

* "Low-cost traffic analysis of Tor", Murdoch and Danezis, in *Proceedings of the 2005 IEEE Symposium on Security and Privacy (S&P'05)*, 2005. DOI: 10.1109/SP.2005.12

† "Users get routed: traffic correlation on tor by realistic adversaries", Johnson, Wacek, Jansen, Sherr and Syverson, in *CCS '13: Proceedings of the 2013 ACM SIGSAC conference on Computer & Communications Security*, 2013. DOI: 10.1145/2508859.2516651

If a government doesn't want people to use the uncensored Internet, they probably also don't want people to use Tor. Finding a Tor entry node relies on a directory authority run by one of a small number of Tor volunteers, but given that anyone can access any directory authorities, censors can also block them. That's why Tor incorporates bridge nodes for use in places (companies or countries) that block Tor directory authorities. You can request a bridge node through Tor, over HTTP or even via email, and the full list isn't published.

Tor's pluggable transport allows Tor to make its traffic profile look like another type of traffic. Does your country block Tor? Tor can make the traffic look like WebRTC (a videoconferencing protocol). Is WebRTC blocked? Tor can make it look like you're using a Microsoft website!

The game is afoot

Cats and mice continue

The story unfolds...

see, e.g., PETS Symposium

Summary

Online tracking and surveillance

Remailing

Tor