Web Tracking

CS 642 Guest Lecture

Yohan Beugin

April 7, 2026 • UW–Madison

Computer Sciences
SCHOOL OF COMPUTER, DATA & INFORMATION SCIENCES
UNIVERSITY OF WISCONSIN–MADISON

MADS&P

About Me 🇫🇷

Education

- *Diplôme d'Ingénieur (B.S & M.S. in Engineering Sciences)* from Centrale Lyon
- *M.S. in Computer Science and Engineering* from Penn State
- (currently) *Ph.D. in Computer Sciences* at UW-Madison working with Prof. Patrick McDaniel

Systems Security & Privacy Research

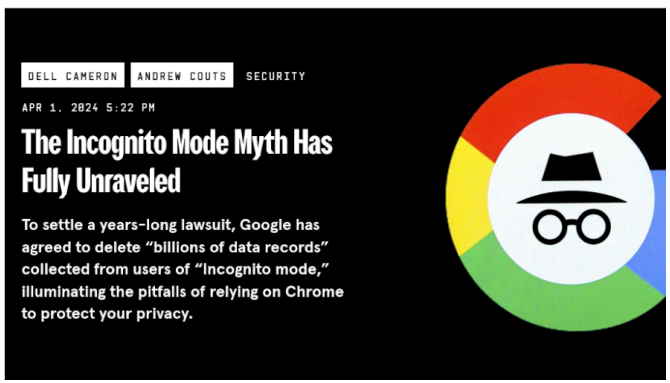
- **PhD:** Measuring and Modeling Web Tracking Risk in Web Proposals
  - Privacy analyses of Topics API cited by Apple and Firefox at W3C
  - Recent systematization work on web tracking
  - Creator and maintainer of the Privacy Sandstorm research portal

yohan@beugin.org https://yohan.beugin.org

Motivation

From Behavorial Advertising to a Massive Corporate Surveillance Network

- Industry with $150+ billion/year in revenues
- Technologies underpinning it all: tracking, profiling, and targeting



Related Publication

[1] SoK: Advances and Open Problems in Web Tracking - Yash Vekaria*, **Yohan Beugin***, Shaoor Munir, Gunes Acar, Nataliia Bielova, Steven Englehardt, Umar Iqbal, Alexandros Kapravelos, Pierre Laperdrix, Nick Nikiforakis, Jason Polakis, Franziska Roesner, Zubair Shafiq, Sebastian Zimmeck (* = equal contribution)- *In submission*, 2026

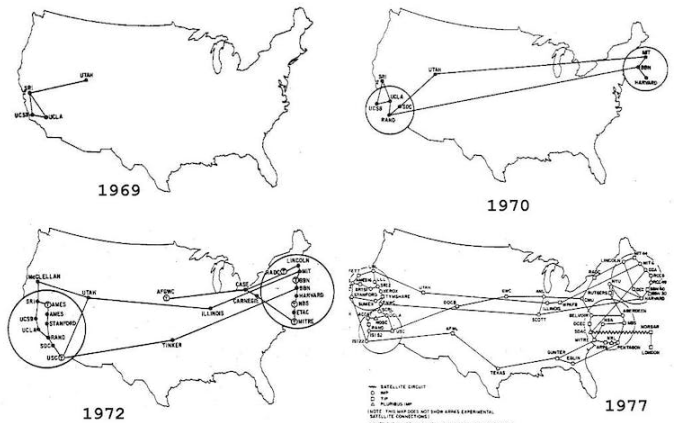
Background

- Technological Context
- The World Wide Web
- Web Browser

Technological Context

Internet's Origin

- Network for military, academic, and research institutions
- Academic and federal funding



1980s & 1990s

- AOL dial up subscription and ads
- Access to chat rooms, media, and shopping (sales team)
- 1989: Tim Berners-Lee invented the World Wide Web at CERN

Technological Context

> *"The Internet has yet to fulfill its promise of commercial success. Why? Because there is no business model."*
> Ed Horowitz (CEO of ViacomCBS, 1996)

Early 2000s: How to Finance Search Engines?

- AdWords: ads alongside search results (auctions, quality metrics)
- AdSense: space offering for ads on publishers' websites
- DoubleClick: acquired in 2007 by Google for $3.1 billion

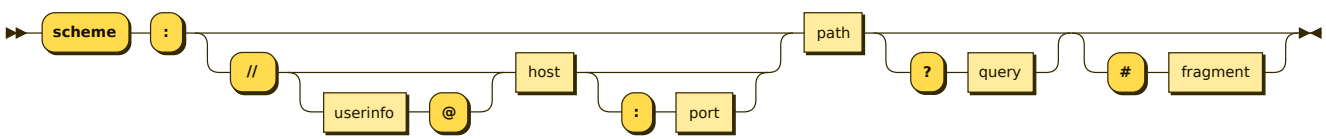
Software engineers + economists = creation of a new marketplace

The World Wide Web

“A Network of Links”

- HyperText Markup Language (HTML)
- Uniform Resource Locator (URL)

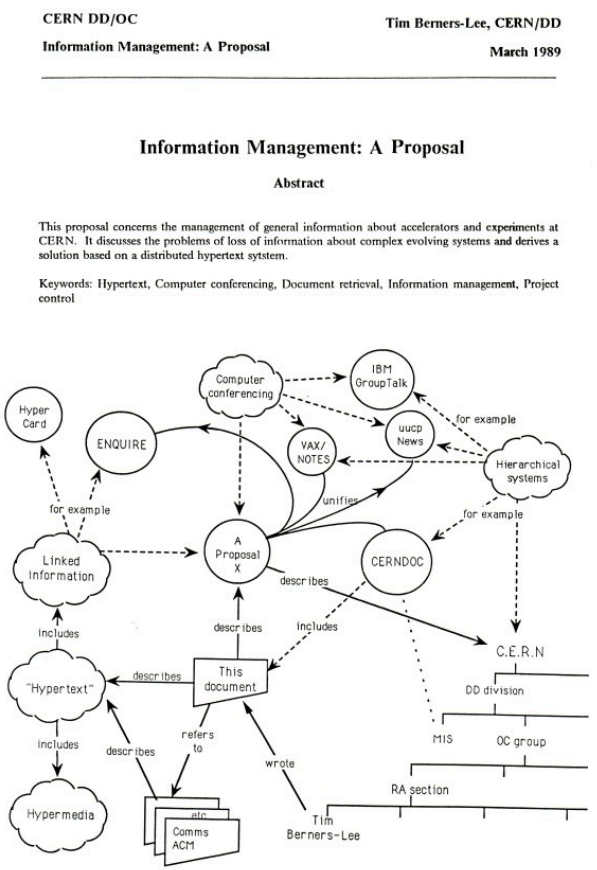
```
scheme ":" ["//" authority] path ["?" query] ["#" fragment]
```



URI Syntax Tree

Open Standards & Open Source

- HyperText Transfer Protocol (HTTP) Server
- Client Software (Browser), Engines, Languages, etc.


CERN DD/OC
Tim Berners-Lee, CERN/DD
Information Management: A Proposal
March 1989

Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

First Proposal (1989)

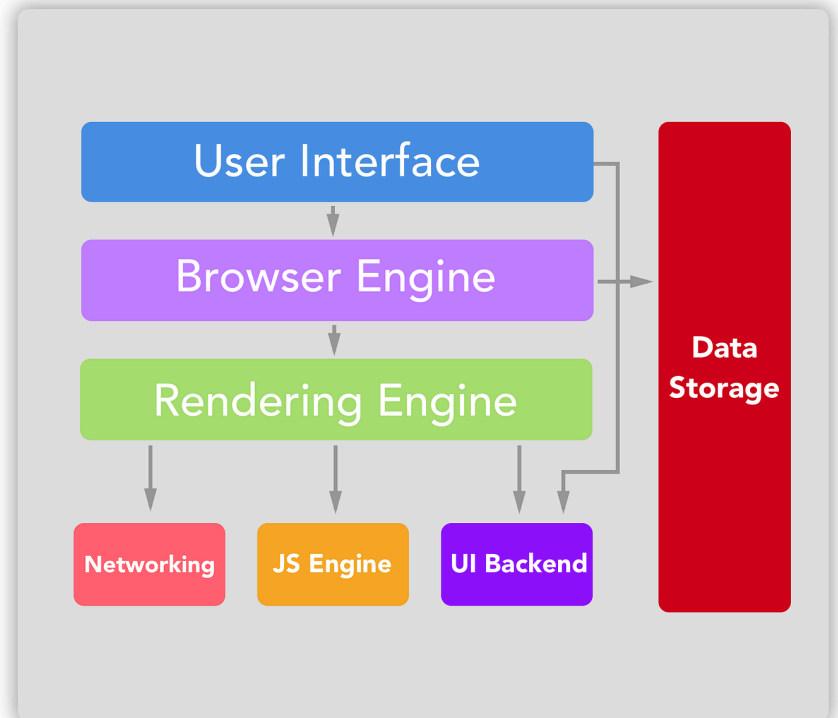
Web Browser

Inner Workings

- Render HTML
- Cache resources
- Check TLS certificates
- Enforce access policies and permissions
- Handle extensions, etc.

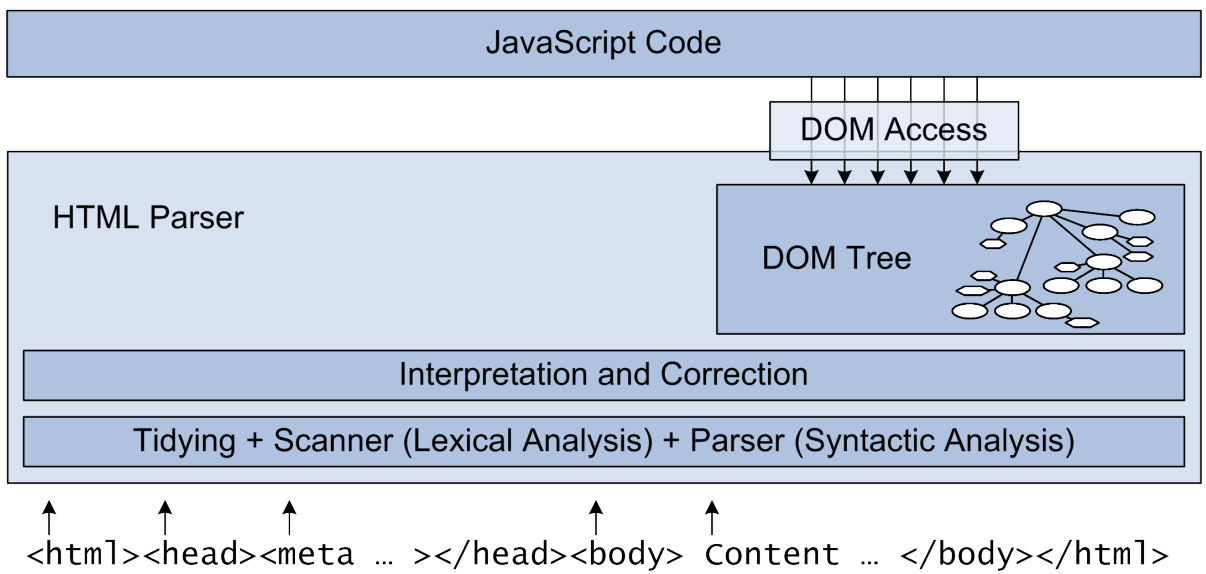
Browser	Rendering Engine	JavaScript Engine
	Blink	v8
	Blink	v8
	Gecko	SpiderMonkey
	WebKit	JavaScriptCore

Major Browsers with Open Source Engines



Web Browser: Rendering HTML

1. `GET text/html` document from server
2. Parse and fix errors
3. Interpret HTML document
4. `GET` other resources (CSS, JS, etc.)
5. Build Document Object Model (DOM)
6. Apply the CSS styling rules
7. Render the HTML document
8. Start executing JavaScript code
9. Listen for events and execute code
10. Discard and start over on new page



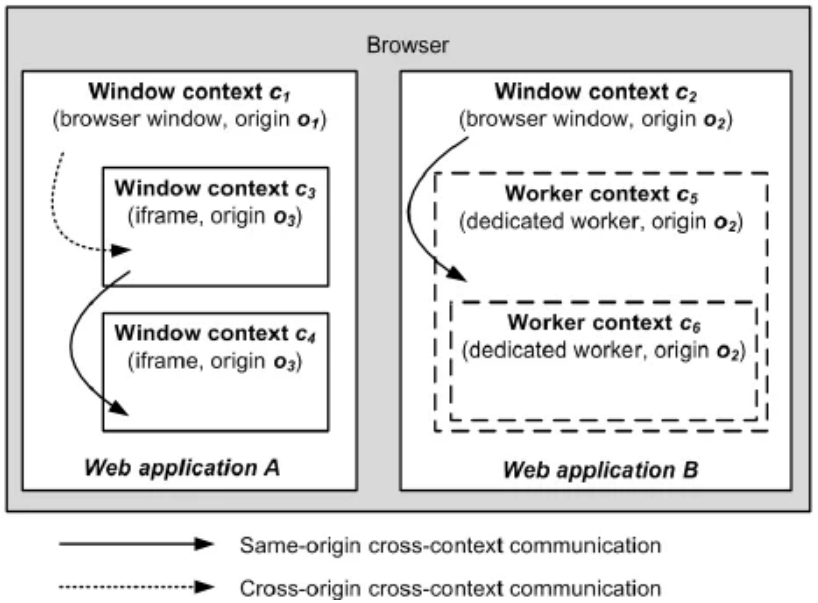
Web Browser: Context-Origin Boundaries

Same-Origin Policy (Code's Privileges)

- **Origin:** `scheme + host + port`
- Related origins grouped by eTLD+1 and public suffix lists
- Code from different origins cannot interfere maliciously with each other's state.

Context Isolation (State Separate From Others)

- **Context:** browser tab + window + mainframe (+ iframes)
- Code in one context cannot arbitrarily interfere with a document in another context, especially if origins differ.



Only if two browsing contexts share the same origin can they interact freely.

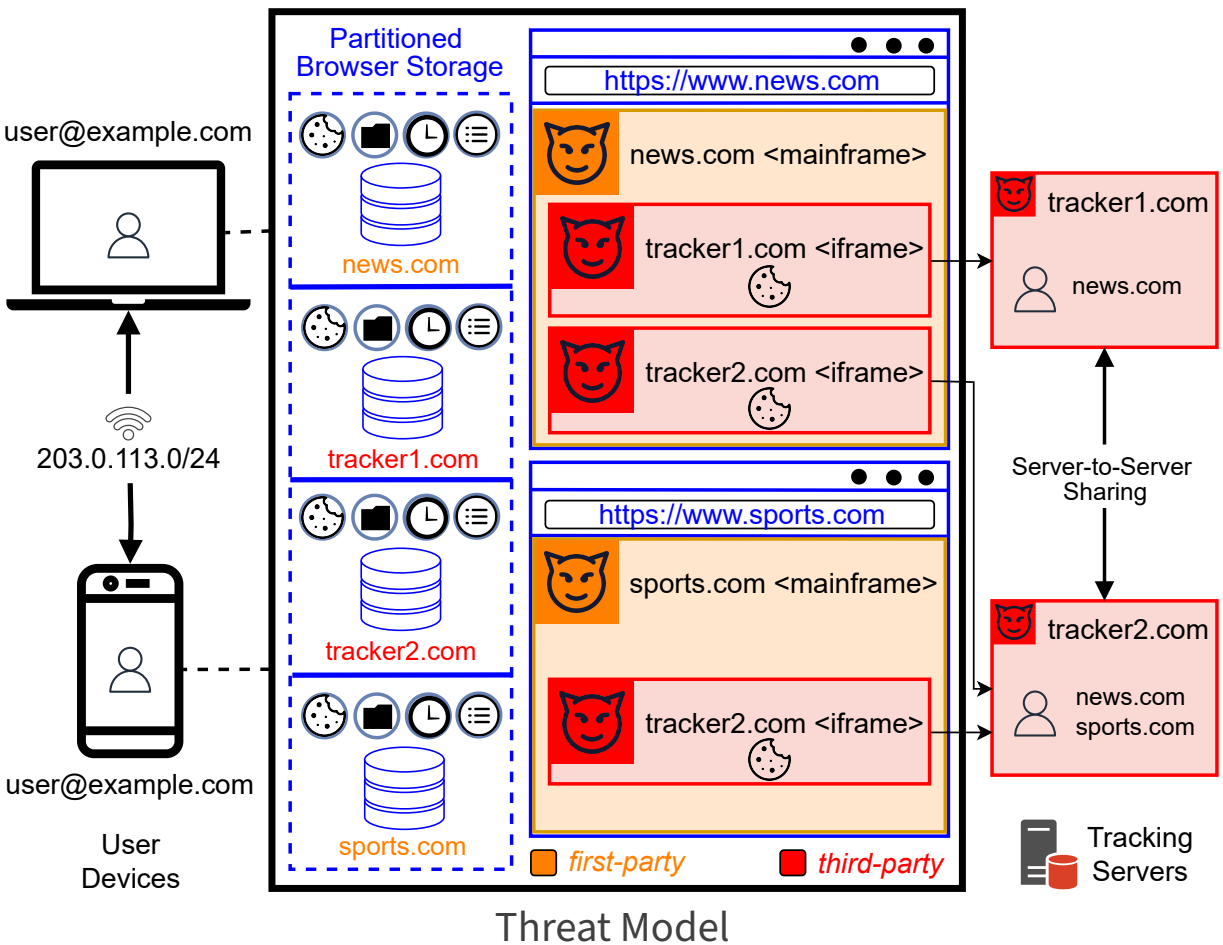
Policies

- **Script Execution:** inclusion of external script into document of another origin means trust.
- **Browser Storage:** partitioned by origin, except for cookies (scoped by domain and path)

Techniques & Mitigations

- Threat Model
- Web Tracking
  - Stateful
  - Stateless
  - Cross-device

Threat Model



Threat Model

Actors

1. Users
2. Browsers
3. First-party websites
4. Embedded third-parties

Adversary Goals

- Same-site
- Cross-site
- Cross-device

Adversary Capabilities

- Inclusion (1st or 3rd-party)
- Collection
- Storage
- Sharing

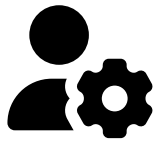
Web Tracking

Purposes

Analytics

Ad (re)targeting

Conversion measurement



Fraud and bot detection

Definition

Refers to the process of collecting, storing, and sharing information about a user, user-agent, or device online.

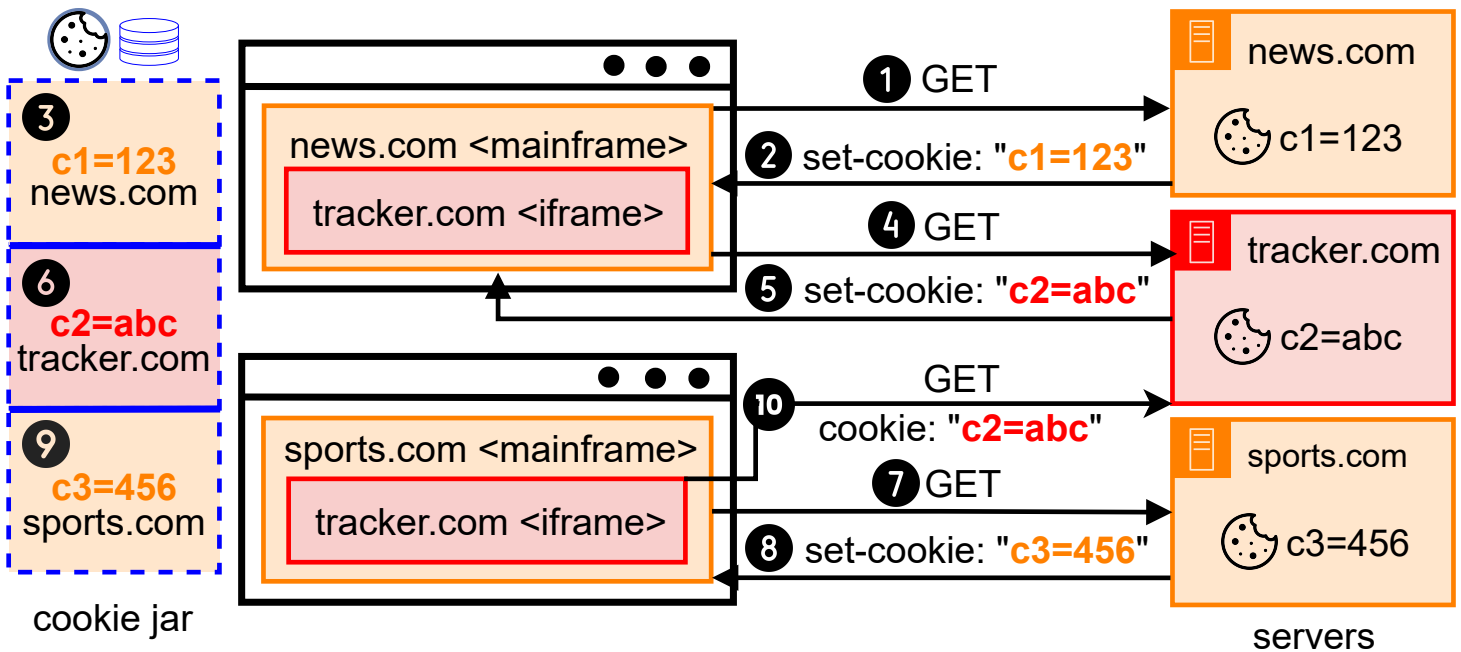
Types

Stateful tracking (1st & 3rd Party)

Stateless tracking

Cross-device tracking

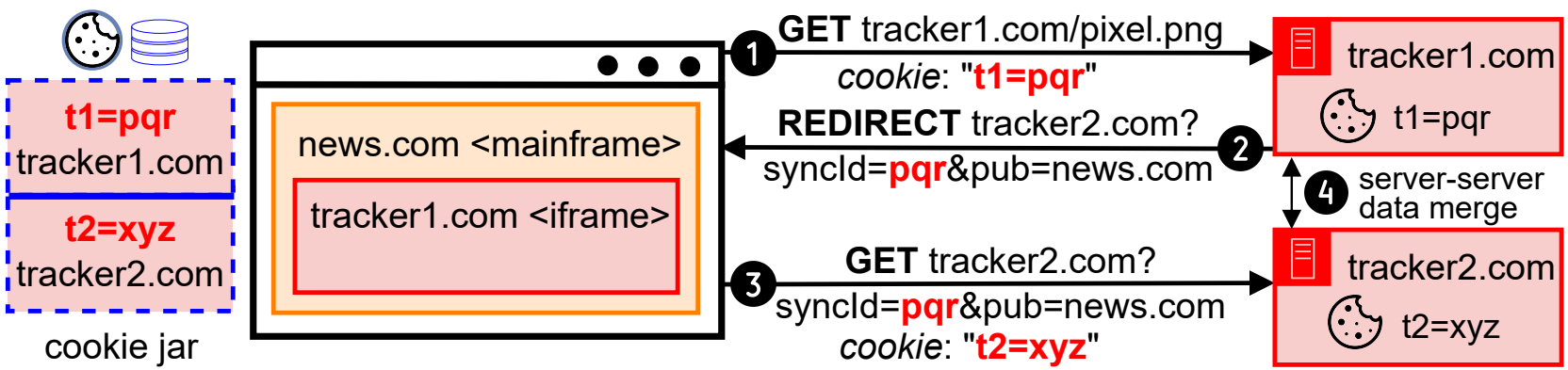
Cookie-based Tracking



Context Dependent

- First-party: only re-identified to the visited website
- Third-party: enables cross-site tracking
- Browsers mediate which resources access specific cookies

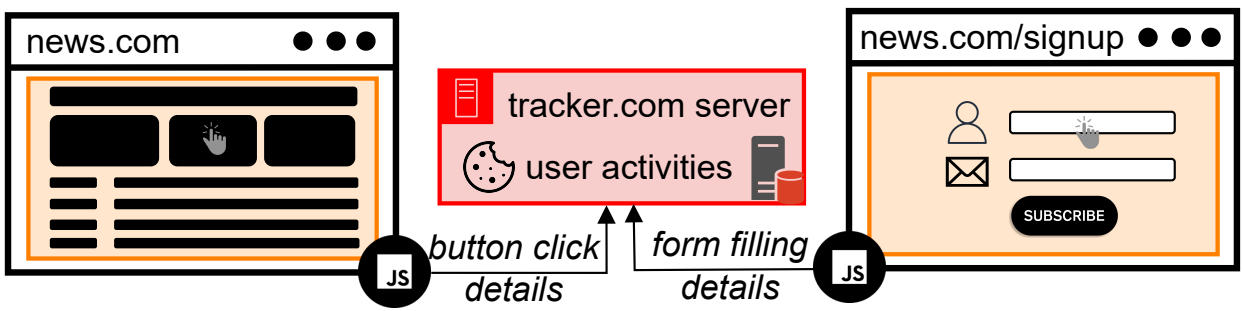
Cookie Syncing



How to Share Cookies With Others?

- By synchronizing user-identifiers
- Commonly done through URL parameters
- Server-to-server data sharing
- Also used to sync first-party cookies with other third-parties

Tracking Tags



From tracking pixels to more complex tags

- Historically: 1x1 image elements pointing to tracking endpoint
- Allows same-site and cross-site tracking
- Modern tracking tags rely on JavaScript to collect more fine-grained information
- Can still be used in first-party context

Defenses Against Third-Party Cookies & Trackers

1. Clearing Third-Party Cookies

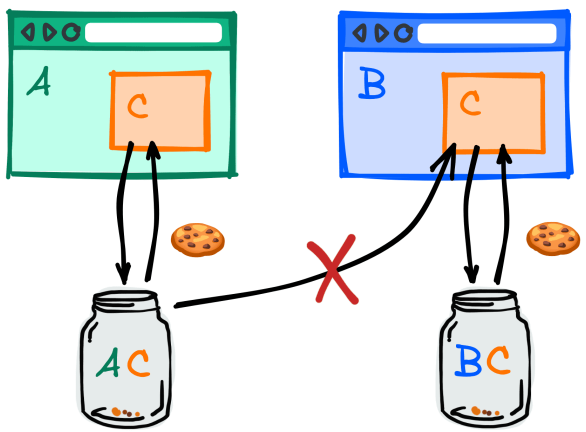
- Respawning, evercookie, or supercookie risk
- Network and storage partitioning efforts

2. Blocking Third-Party Cookies

- Brave: partitioned ephemeral storage
- Firefox and Safari: blocking approaches
- Chrome: opt-in restrictions only

3. Blocking Trackers

- Filter lists (maintenance burden, evasion)
- Block network requests and JS scripts

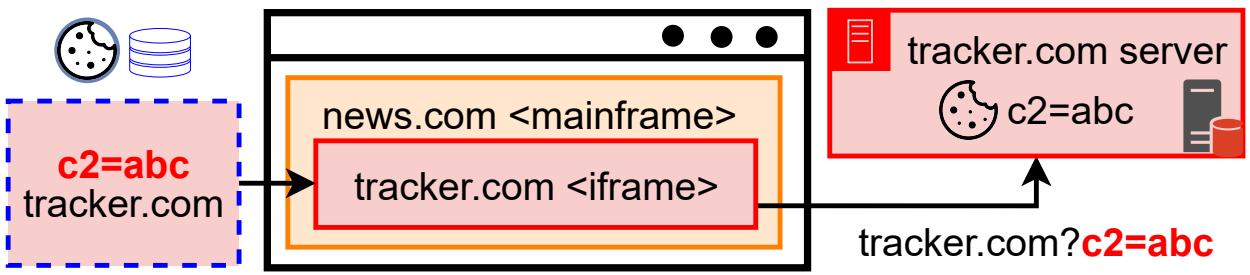


A,B – top-level sites
C – embedded site

Partitioned Cookies

Can be effective, but trackers have also migrated to alternative techniques or circumventions, including first-party stateful tracking.

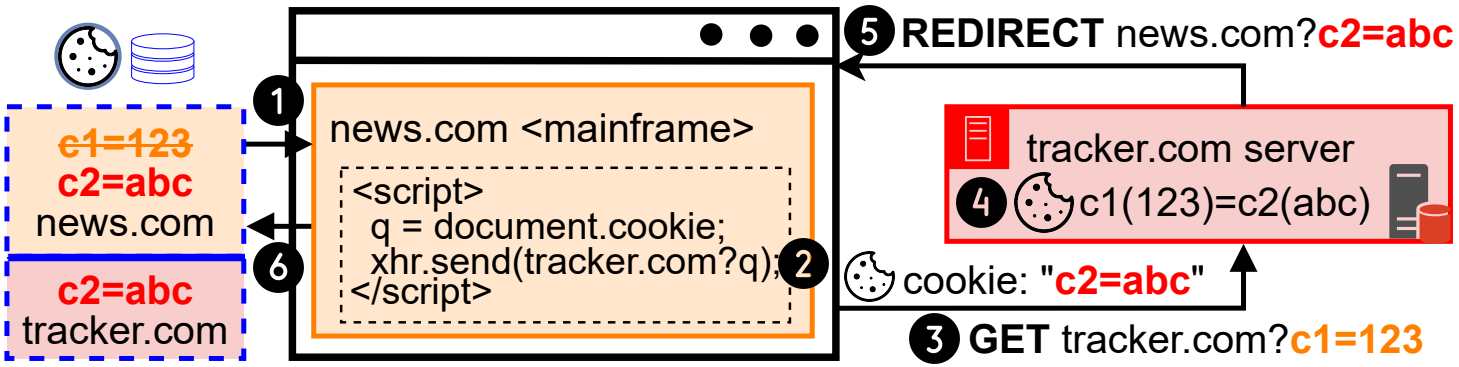
Navigational Tracking



How to Work Around Partitioned Cookies?

- Share identifier through link decorations
- Also used to bounce identifiers stored in first-party cookies

Bounce Tracking



1. A 3rd-party inline script on `news.com` reads 1st-party identifiers stored under `news.com`
2. And includes them in the request to `tracker.com`.
3. Browser redirects to `tracker.com` either by a user click or automatically.
4. Once loaded, `tracker.com` reads the identifier from the URL or merges it with an existing cookie, rewriting the URL to its final destination, embedding the consolidated identifier.
5. The browser navigates to `news.com`; the tracker's script extracts identifier from the decorated URL
6. And stores it in `news.com`'s 1st-party storage, completing the **cross-context** (if redirected to same first-party) or **cross-site** (if redirected to a different domain) **linkage**.

Further Defenses Against Stateful Tracking

It is harder to block first-party storage (cookies), so more targeted countermeasures are needed.

Limiting Lifetime of First-Party Storage set by Scripts

- **Safari:** expirations post no user interaction for 7 days + heuristics
- **Brave:** 7 days lifetime limit

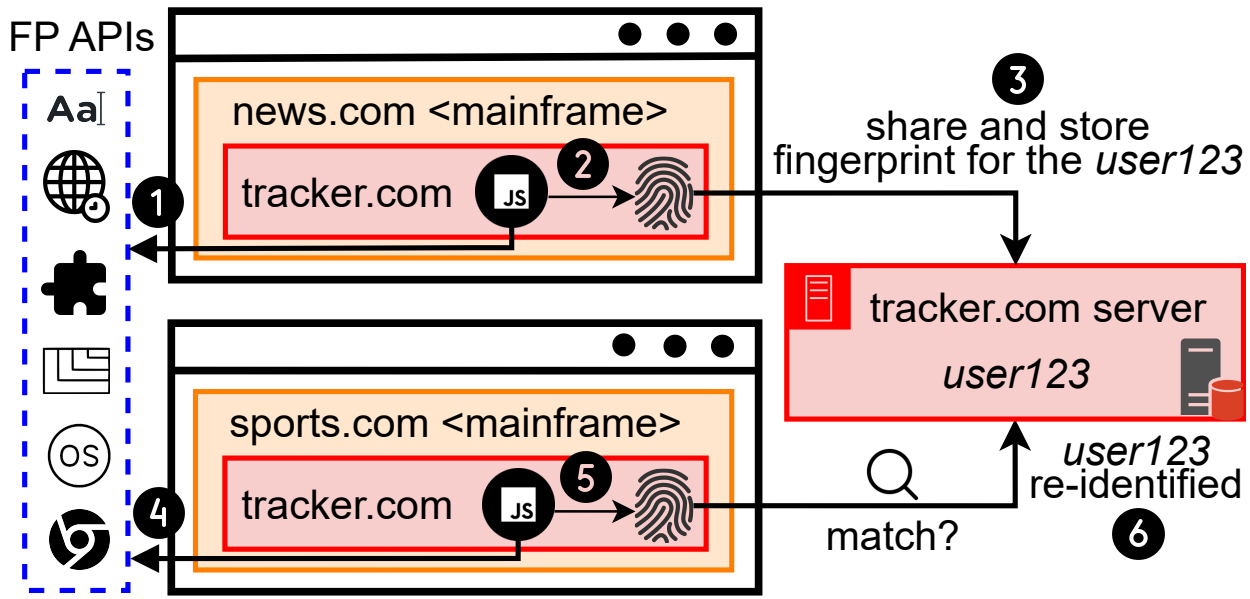
Removing or Limiting the Persistence of Identifiers Passed in URL Parameters

- **Brave, Firefox**: remove URL parameters known to be used by trackers
- **Safari**: limits storage lifetime to 24h if link decoration detected

Limiting First-Party Storage Set During a Bounce

- Distinguishing between legitimate redirects and bounce visits
- **Brave, Firefox**: use blocklists
- **Safari, Chrome**: use heuristics based on site behavior

Browser Fingerprinting



Unlike Stateful Tracking, Fingerprinting:

- Does not rely on a stored state (ID),
- Is hard to detect and block,
- Is difficult to evade (stable fingerprints over time).

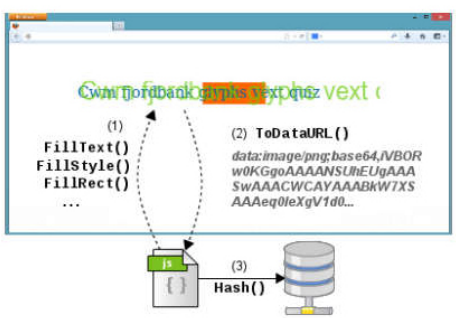
Defenses Against Browser Fingerprinting

Normalizing Exposed Information

- Making all users appear the same (Tor).
- Reducing identifying information (User-Agent).
- Refusal or removing APIs (Battery Status).

Adding Randomness

- Keep changing across page loads.
- Site-specific noise added to APIs output.
- 2D Canvas, WebGL, AudioBuffer, etc.



Canvas Fingerprinting

There has been a lack of unified effort to tackle browser fingerprinting so far.

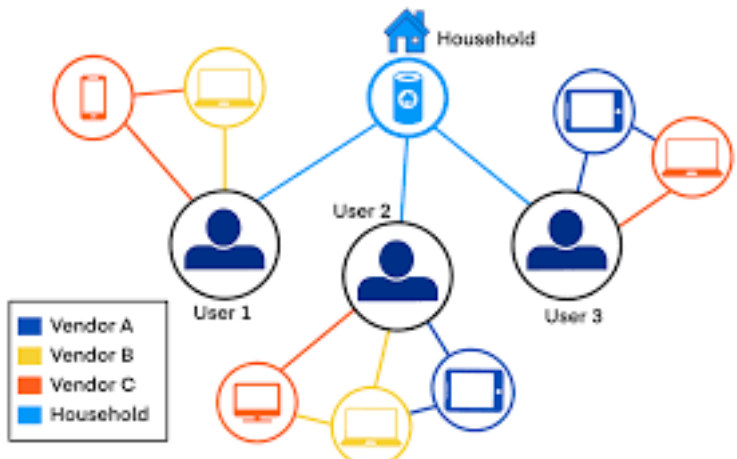
Cross-device Techniques & Mitigations

1. Deterministic

- Account information
- Username, email, phone number, etc.

2. Probabilistic

- IP addresses
- URL browsing patterns
- OS and hardware characteristics
- Typing behavior



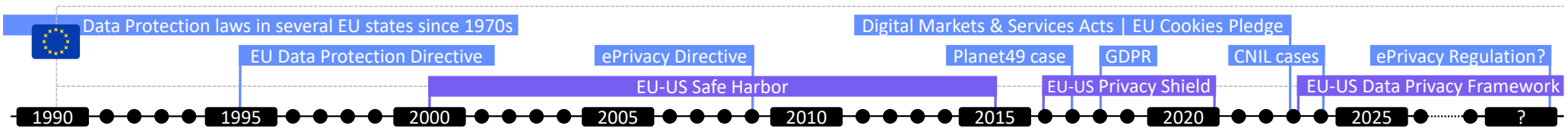
Cross-device Graph

Mitigations are inherently limited by users being logged in across devices into the same account(s). Traditional tracking defenses apply otherwise.

Regulations

- In the EU
- In the US
- Elsewhere
- Policy-Oriented Solutions

Regulatory Actions in the EU

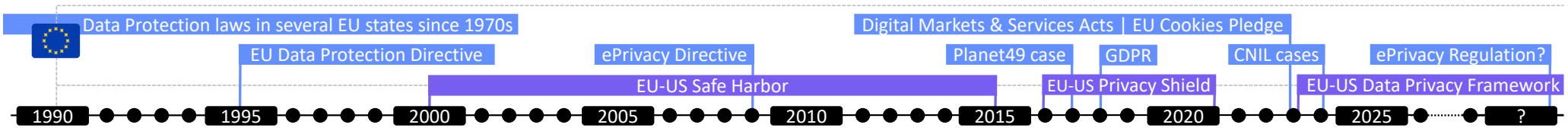


ePrivacy Directive (2002, amended in 2009)

- Requires a valid user's consent before *"storing of information, or the gaining of access to information already stored, in the terminal equipment"* article 5(3)
- Right to privacy and freedom in electronic communications
- Future: will it be turned into a regulation?

The **EU-US Privacy Shield** was invalidated on July 16, 2020 by the European Union Court of Justice.

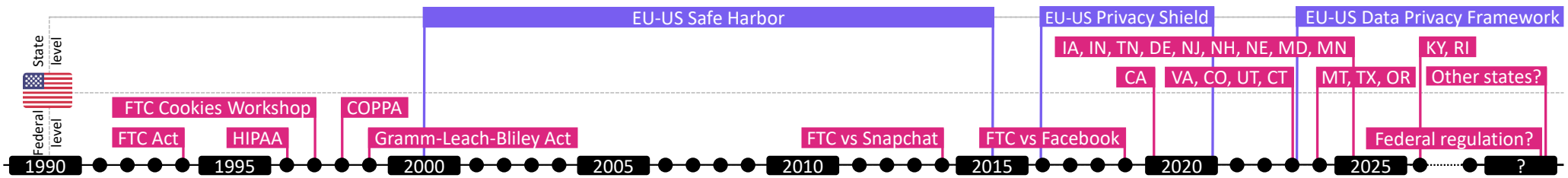
Regulatory Actions in the EU



General Data Protection Regulation (2018)

- Data Breach Notification: within 72h
- Rights: to be informed, to access, to rectify, to be forgotten, to restrict processing, to ask for data portability, and to object
- Privacy by Design: privacy designed into products from the very beginning
- Data Protection Officer: *"[if] core activities consist of processing operations which require regular and systematic monitoring of data subjects on a large scale"*
- Fines: up to 20 million of euros or up to 4% of a company's annual worldwide revenue

Regulatory Actions in the US



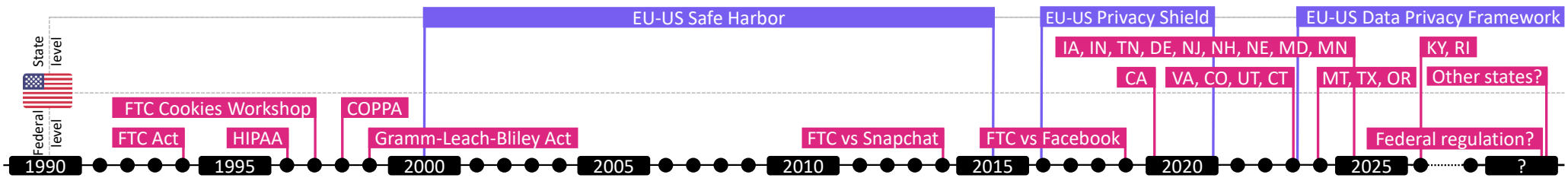
At the Federal Level

- **Children's Online Privacy Protection Act (COPPA)**: verifiable parental consent for $<$ **13 years old**.
- **Health Insurance Portability and Accounting Act (HIPAA)**: **portability (not privacy)**.
- **Gram Leach Bliley Act**: banks and financial institutions.
- **Fair Credit Reporting Act**: credit info, SSN.

FTC can consider privacy policies misrepresenting a business's data handling practices as unfair or deceptive, affecting commerce per 15 U.S.C. §45(a)(1). This led to **a body of common law of privacy**.

In the US: data collected for business purposes may be used in civil and criminal proceedings.

Regulatory Actions in the US 🇺🇸



At the State Level

- **California Consumer Privacy Act (CCPA)**: CA residents have the rights to be informed when and how their data is collected, to access, correct, and delete it.
- **Stop Hacks and Improve Electronic Data Security Act (NY SHIELD)**: data security requirements for NY residents.
- …

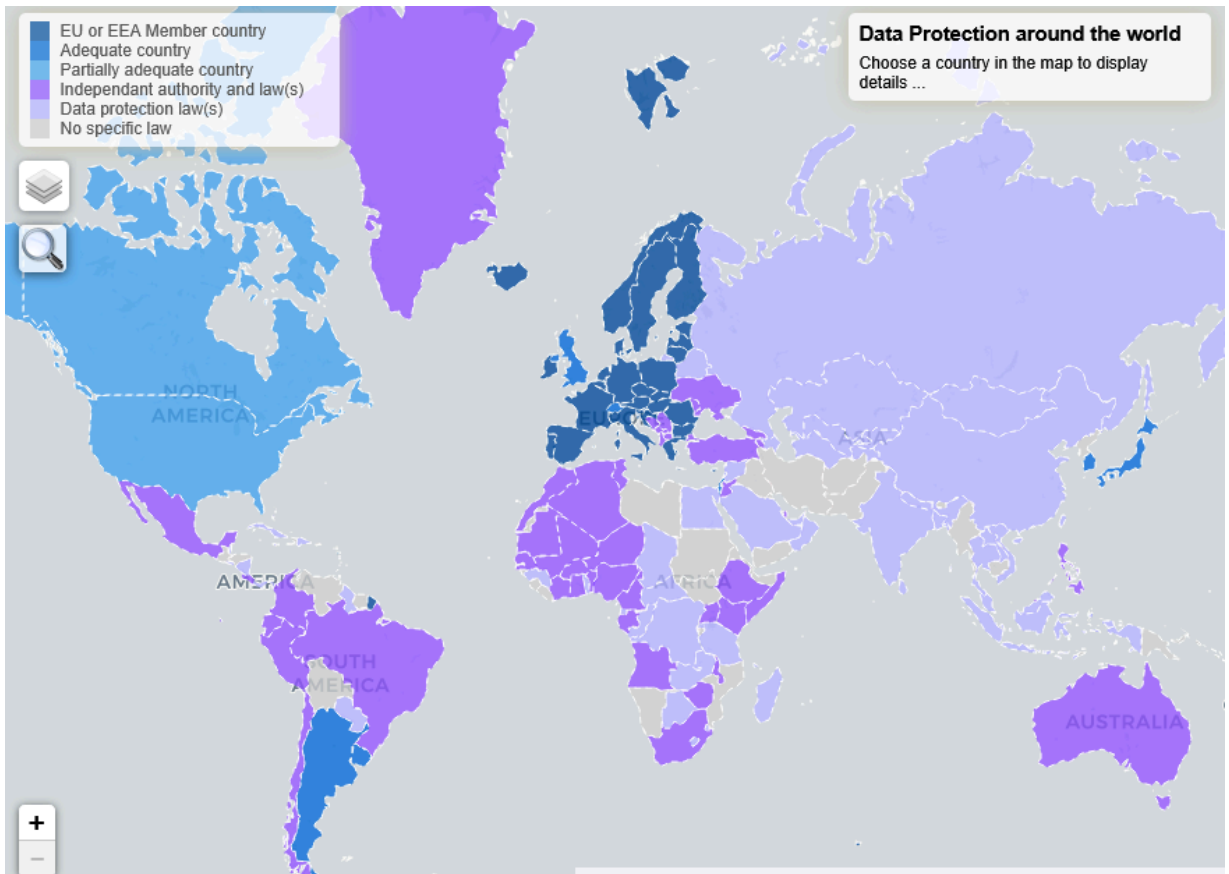
Wisconsin does not currently have a law (like CCPA) that specifically governs personal data.

Elsewhere

Many more regulations

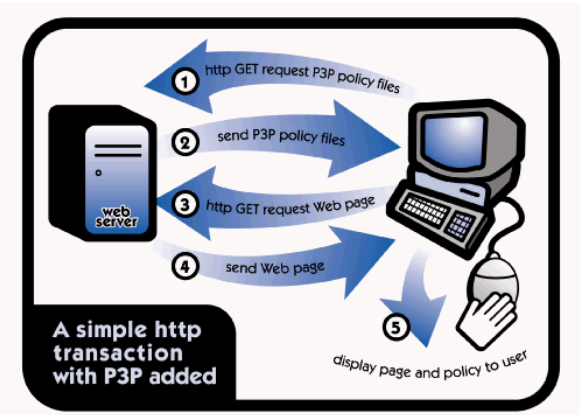
- Data Protection Act in the UK (2018)
- Lei Geral de Proteção de Dados (LGPD) in Brazil (2020)
- India's Digital Personal Data Protection Act (DPDA, 2023)
- …

Interconnected world: regulations somewhere can improve privacy elsewhere (GDPR and CCPA "effects").



Data Protection around the world (CNIL)

Policy-Oriented Solutions



Platform for Privacy Preferences
(P3P - 2002)

- Standardized protocol
- Fine-grained format
- Low number of adopters and implementers

Do Not Track
(DNT - 2009)

- Binary opt out signal
- Low adoption
- CalOPPA: services only have to declare if they respect it or not

Global Privacy Control
(GPC - 2020)

- Slow adoption
- Compliance required in California and Colorado
- Applicable in the EU?

Takeaways

Archetypal cat-and-mouse game

- Browsers are powerful, but unreliable, gatekeepers.
- Regulations alone are not enough (slow enforcement vs. new techniques).

Purely reactive approaches are insufficient

- Need for collaboration between regulators and measurement community (agile and evidence-driven auditing).
- Default privacy-first solutions needed for users to control their privacy.

yohan@beugin.org https://yohan.beugin.org

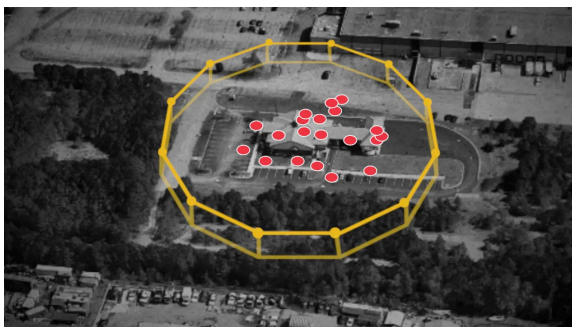
Additional Slides

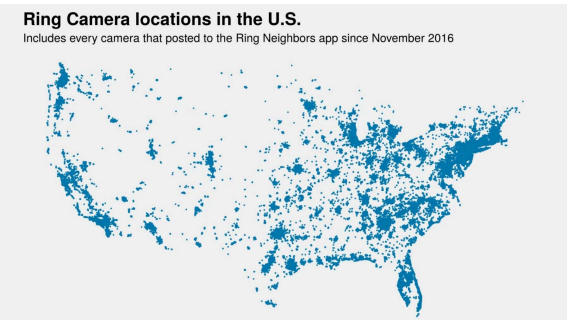
- More Technological Context
- Global Timeline Overview
- Onion Routing & Tor
- What Can Users Do?

More Technological Context

Mobile & Internet of Things (IoT)

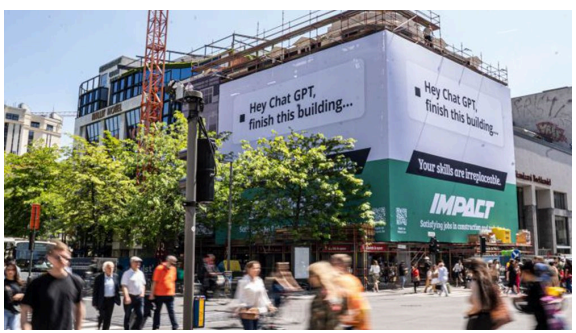
- Smartphones and apps
- Mobile tracking and advertising
- More sensors data and more regularly

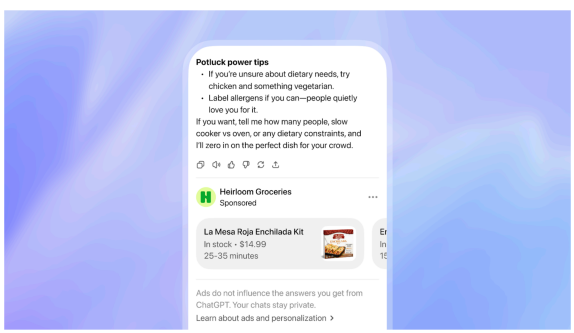




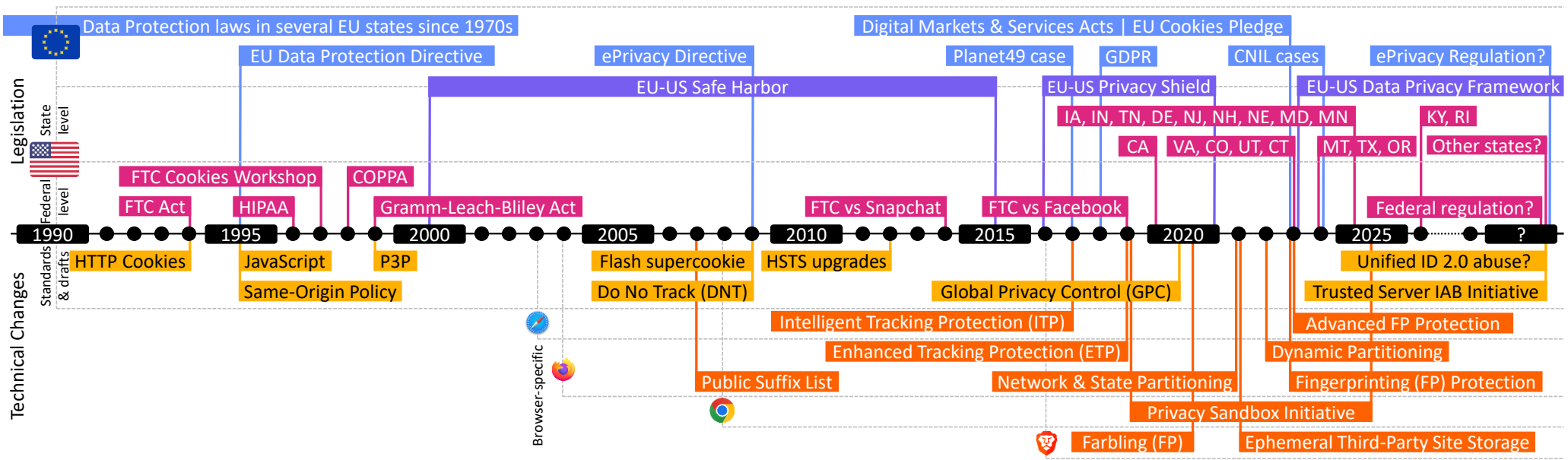
Generative AI, LLMs

- Latest disruptive technology
- Business model?
- Where do the data flow?





Global Timeline Overview

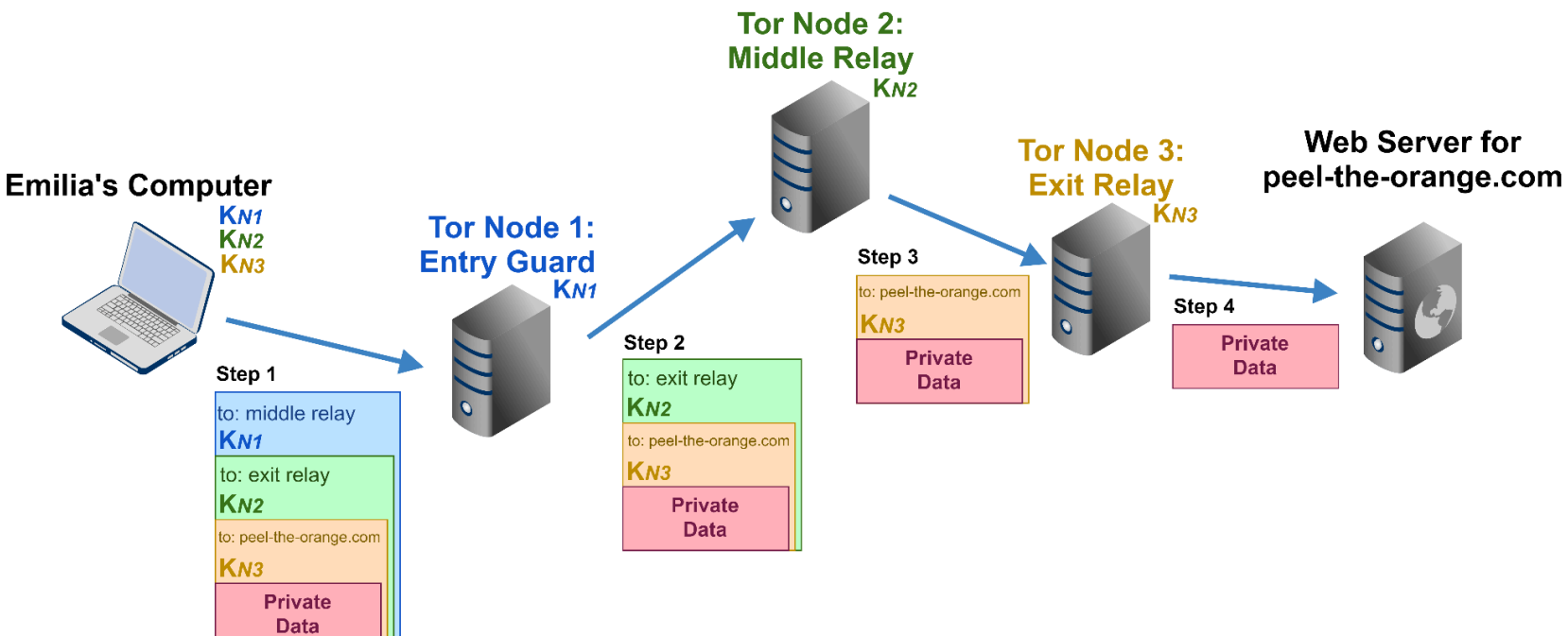


Overview of Regulatory and Technical Changes

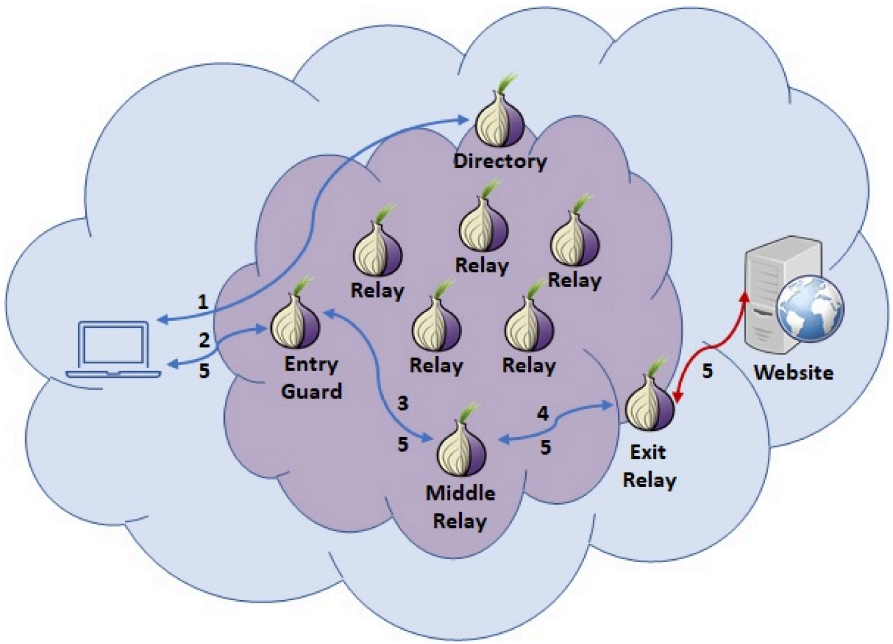
Onion Routing

Objectives

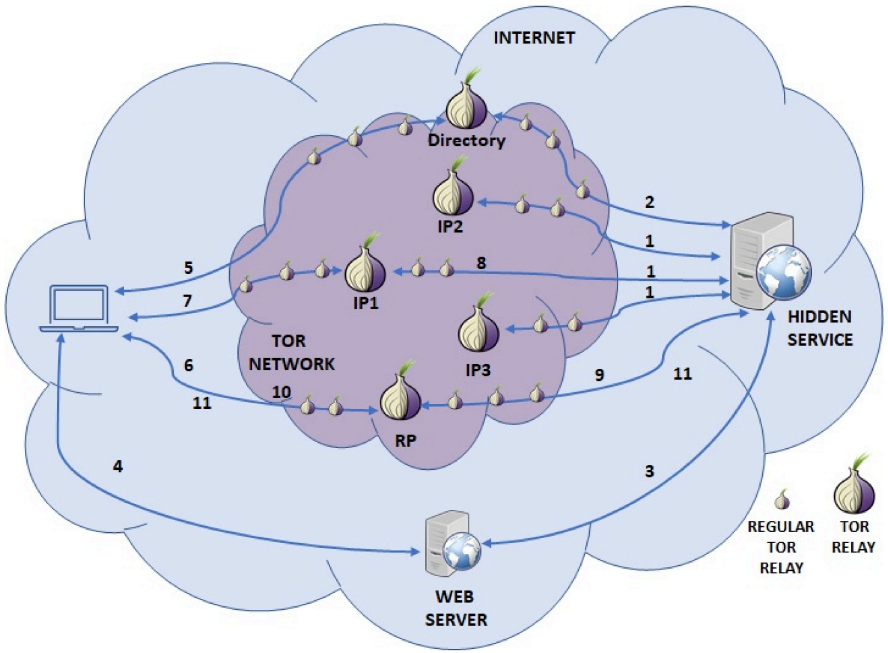
- Hide destination from network attacker watching client
- Destination does not know IP address of client
- Only node before and after is known, not full circuit



Onion Metaphor: Layered Encryption Unpeeled Through the Circuit



Circuit Creation



Hidden Services

You can try it out!

- Browse the web with the Tor browser
- Deploy a node (note: legal implications to be an exit relay!)

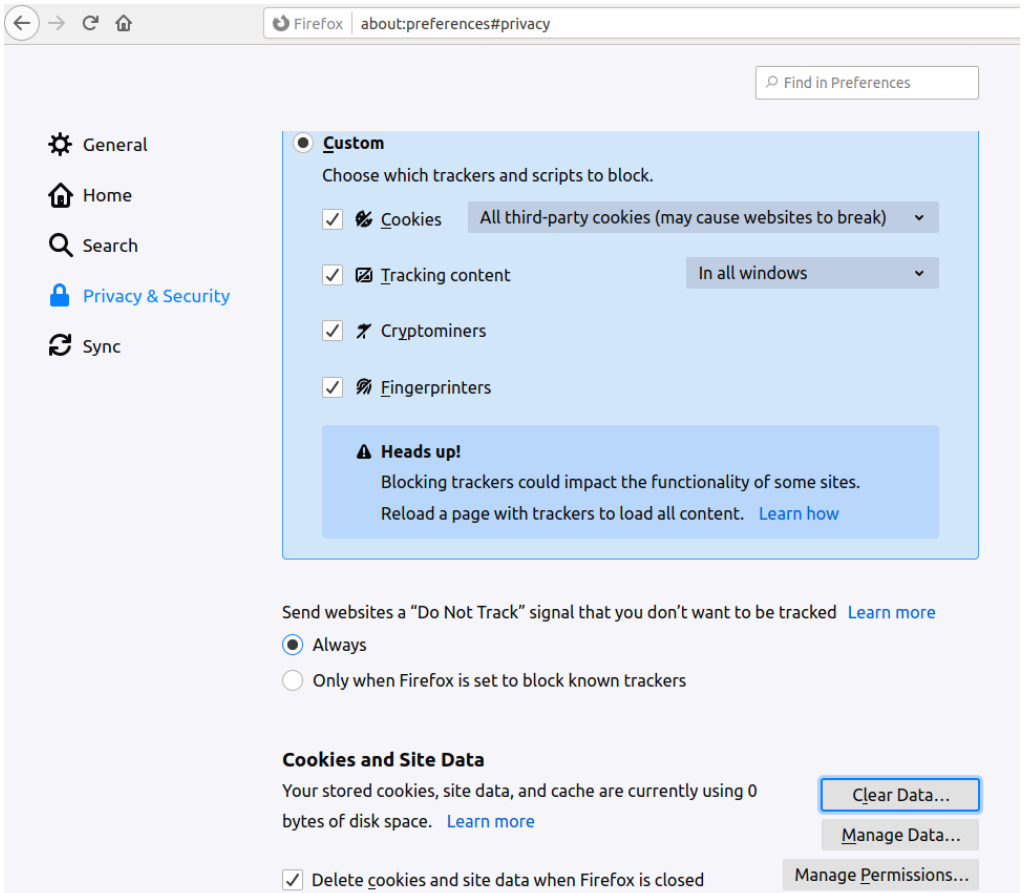
What Can Users Do?

Hard to avoid some platforms on the web

- Behavorial targeted advertising underpins the business model of many web actors
- Facebook profiles users even if they don't have an account

Review Privacy (and Security) Settings

- Services, accounts (Google, Facebook, etc.)
- No to third-party cookies, fingerprinting
- Do not track, GPC signals



See blog posts at https://yohan.beugin.org/posts/index.html

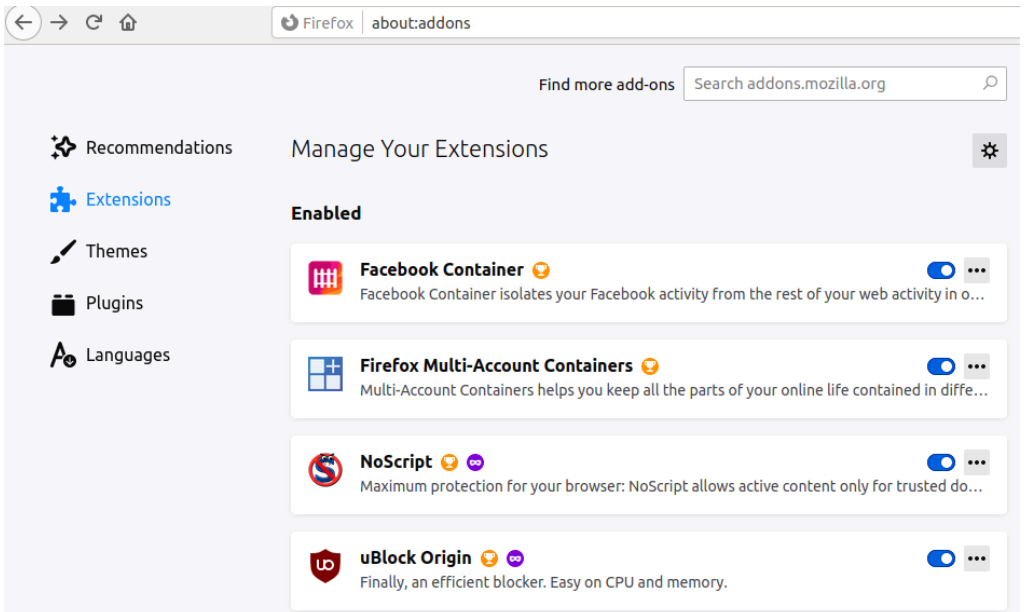
What Can Users Do?

Filter - Block - Isolate

- Ad/script blockers: uBlock Origin, NoScript, EFF Privacy Badger, etc.
- Firefox Multi Containers & Facebook Container

COVER YOUR TRACKS

AM I UNIQUE ?



Search for Alternatives

- Firefox, Tor Browser, Brave, Safari, etc.
- https://alternativeto.net/
- https://www.opensourcealternative.to/
- LineageOS

See blog posts at https://yohan.beugin.org/posts/index.html