



# Advances and Open Problems in Web Tracking

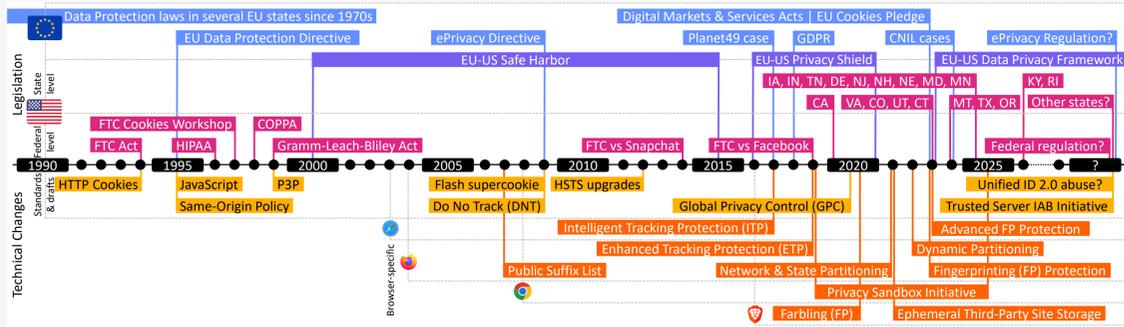
Yohan Beugin - University of Wisconsin-Madison



# MADS&P

## MOTIVATION

### Web tracking is a pervasive and opaque practice

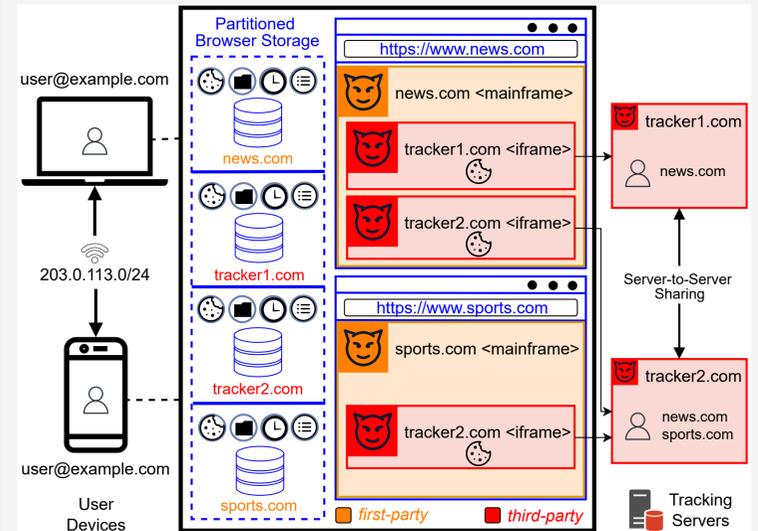


- Enables advertising, retargeting, conversion tracking, etc.
- Decades of research and regulations, but literature remains fragmented.
- Undergoing transformative changes with new protections and regulatory frameworks.

**Need for a comprehensive and systematic study of emerging trends in the evolving tracking landscape to identify crucial research gaps**

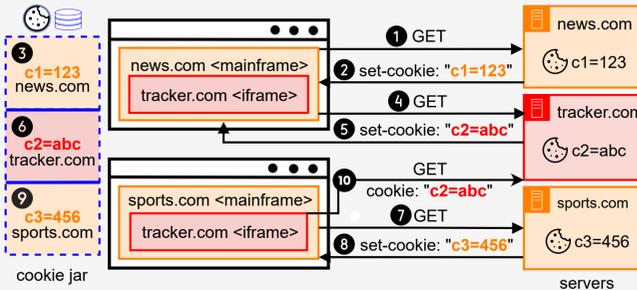
## THREAT MODEL

### Browser's security model: context-origin boundaries

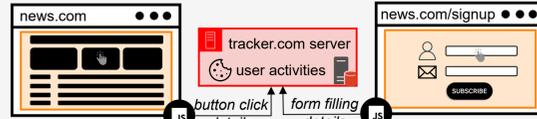


## STATEFUL TRACKING

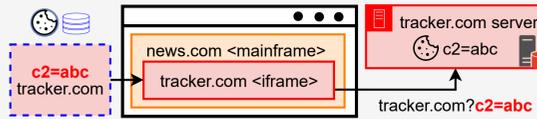
### Cookie-based



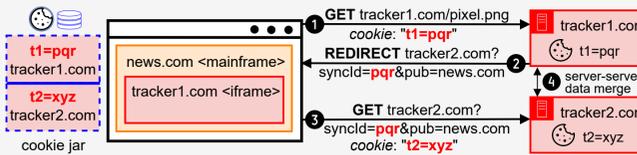
### Tracking scripts



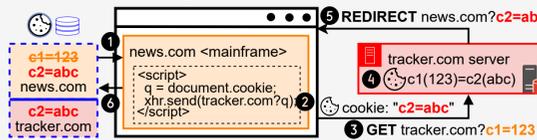
### Navigational tracking



### Cookie syncing



### Bounce tracking



**Defenses:** clearing and restricting cookies, blocking trackers, limiting lifetime of first-party storage, removing identifiers passed in URL parameters.

**Shift to first-party cookies & partitioning:** what alternative forms of user tracking are emerging? that could increase privacy risks, and how might these risk manifest?

**First-party data reliance & identity graphs:** how can we detect or infer opaque first-party data flows within the identity provider ecosystem and quantify its privacy risks?

**Tracking tags:** how are they configured differently across sites, how does it impact tracking?

## CROSS-DEVICE TRACKING

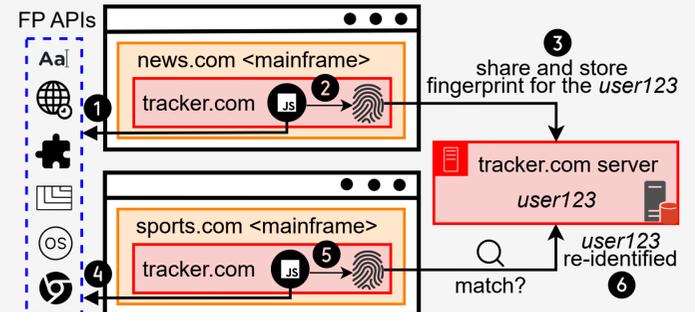
### Deterministic or probabilistic

**Defenses:** account logins inherently limit protections, otherwise traditional defenses apply.

**Theory to practice:** how to characterize cross-device tracking in practice and defend against it?

## STATELESS TRACKING

**Fingerprinting** (browser, side-channel, extension, behavioral, hardware)



**Defenses:** normalization, randomization, and anti-fingerprinting policies.

**Paywalls:** how are paywalls leveraged by publishers for building first-party profiles and perform targeted advertising within their own network?

**Server to server:** how does server-side tracking work, and how to detect and mitigate it effectively?

**Fingerprinting:** what is its real impact at scale, on vulnerable populations, and with other techniques?

**Intent:** how to distinguish between benign and tracking use cases of fingerprinting?

**Hardware:** How to prevent and detect low-level hardware-based fingerprinting?

**Browsers:** How can browsers defend against emerging side-channels?

## OTHER OPEN PROBLEMS

**Current focus on privacy policy and consent:** what about other types of compliance (e.g., EU-US data transfer law)?

**Transdisciplinary studies:** how to reconcile web actors' incentives, responsibilities, and users' expectations?

**Prevention:** can adaptive measurement, monitoring, and disclosure methods be developed to stay ahead of tracking tactics?

**Browsers:** what new privacy-preserving advertising technologies can be built by learning from issues in prior proposals? How can we reliably and at scale automate the evaluation of their potential risks?

**Other ecosystems:** how do tracking mechanisms and protections diverge across web and app ecosystems?

**Generative AI:** how can we counter the privacy, security, and safety risks amplified from the use of generative AI by browsers, websites, and embedded third-parties for tracking, profiling, and personalization?

## TAKEAWAYS

### Archetypal cat-and-mouse game

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

### Purely reactive approaches are insufficient

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

**Paper:** SoK: Advances and Open Problems in Web Tracking



**Authors:** 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)



<https://yohan.beugin.org>

[yohaahaan](https://github.com/yohaahaan)

[yohan@beugin.org](mailto:yohan@beugin.org)