



# Advances and Open Problems in Web Tracking

Yohan Beugin - University of Wisconsin-Madison  
Advisor: Prof. Patrick McDaniel

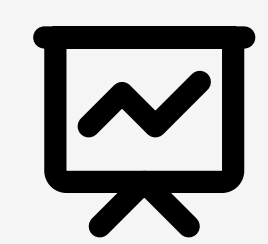


# MADS&P

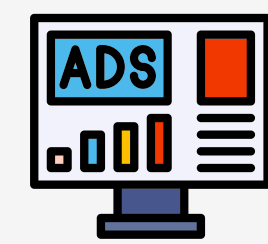
## MOTIVATION

### Web tracking is a pervasive and opaque practice

- Process of collecting, storing, and sharing information about a user or device online.
- Decades of research and regulations, but literature remains fragmented.
- Currently undergoing transformative changes with new protections and regulations.



Analytics



Ad (re)targeting



Conversion measurement



Fraud and bot detection

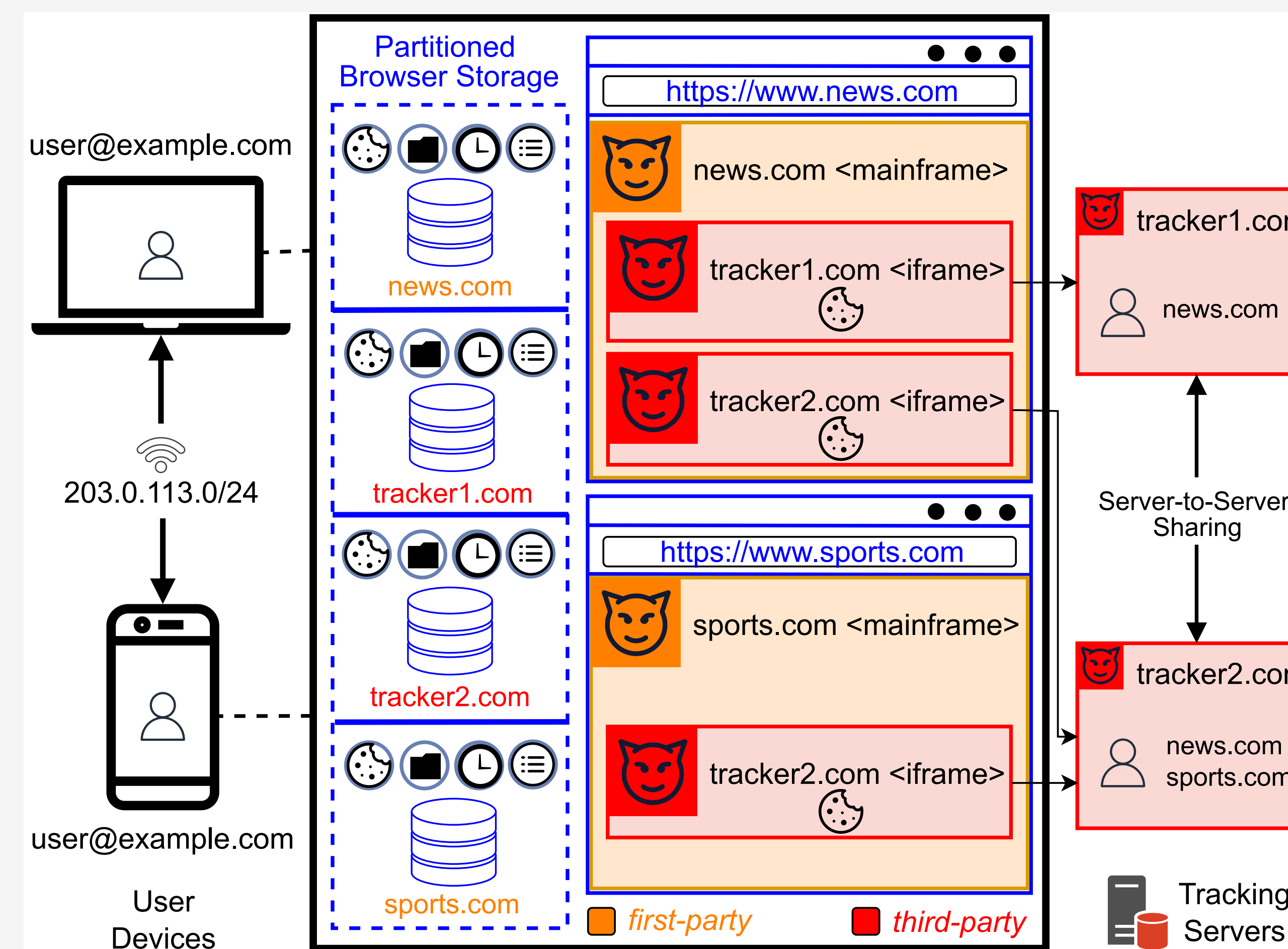
**Need for a comprehensive and systematic study of emerging trends in the evolving tracking landscape to identify crucial research and regulatory gaps**

### Methodology

- Collection of 200+ papers from 2005 and onward on web tracking.
- Top 7 venues: IEEE S&P, USENIX Security, ACM CCS, NDSS, ACM IMC, PETS, & WWW.
- Categorization, synthesis of main findings, and generalization of results.



## THREAT MODEL



### Actors

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

### Adversary goals

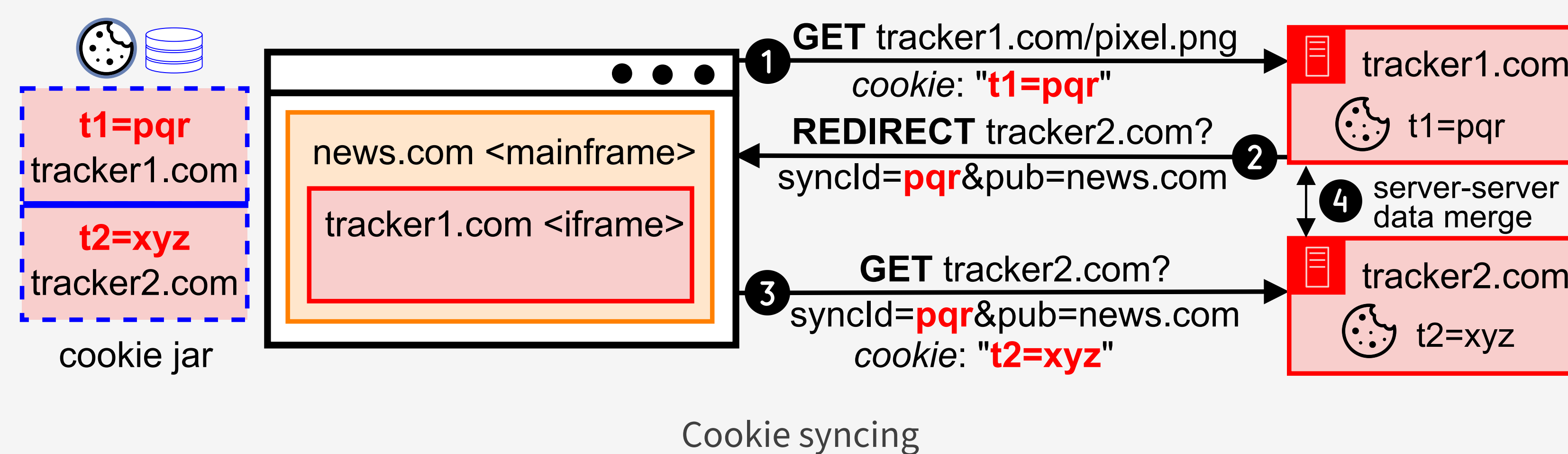
- Same-site
- Cross-site
- Cross-device

### Adversary capabilities

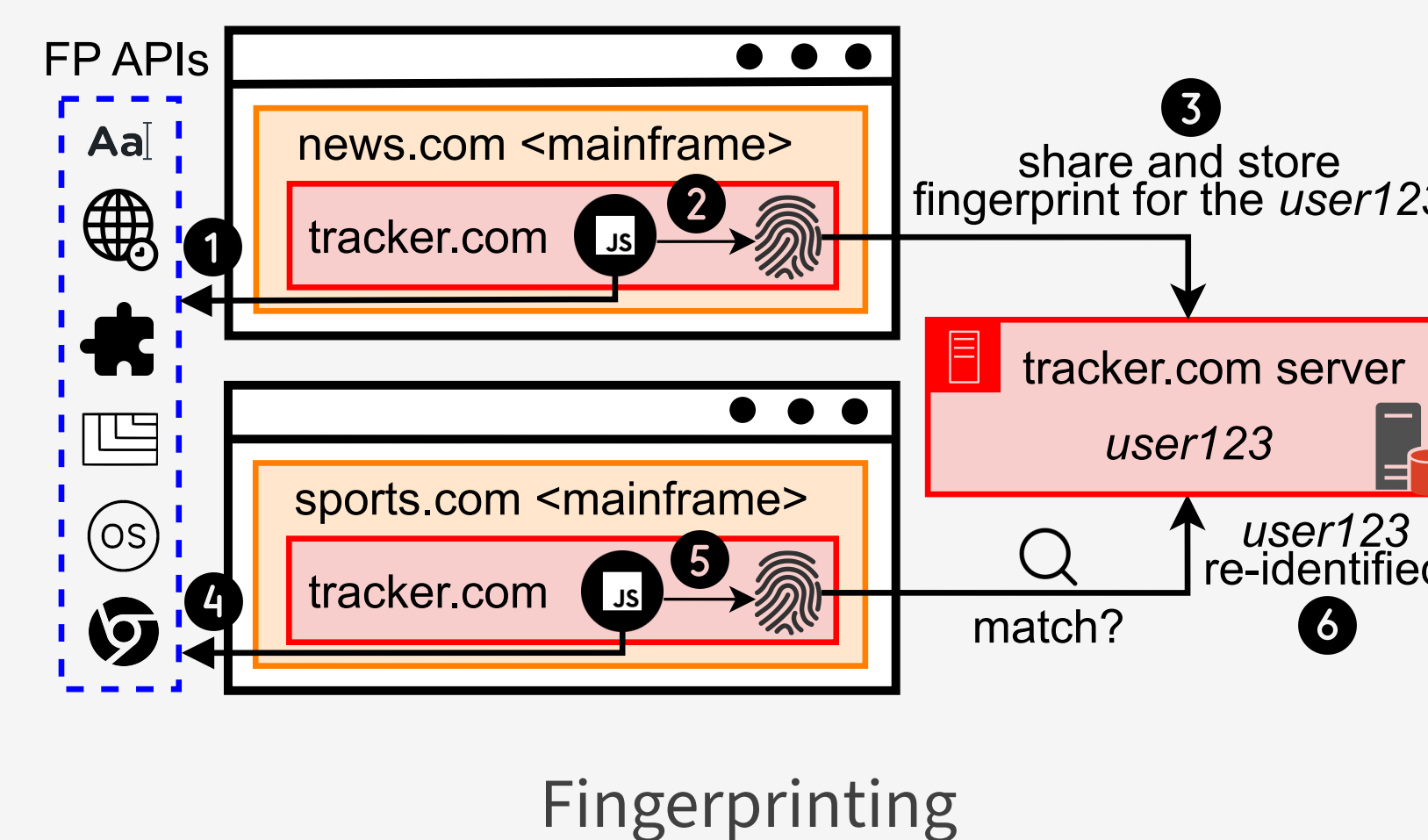
- Inclusion (1st or 3rd-party)
- Collection
- Storage
- Sharing

## WEB TRACKING EVOLUTION

### Shift to first-party tracking & fingerprinting

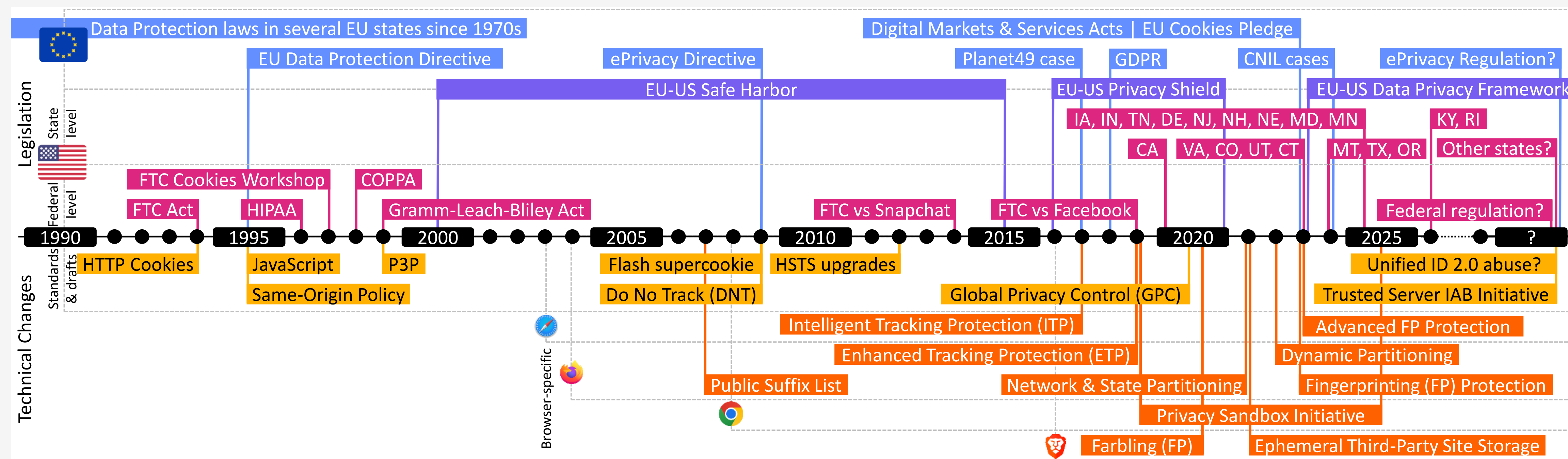


Cookie syncing



Fingerprinting

### Overview of regulatory and technical changes



### Open problems

- User tracking changes:** what alternative forms of user tracking are emerging and how might privacy risk manifest?
- Fingerprinting:** what is its real impact at scale, on vulnerable populations, and with other techniques?
- 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 current?

## 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.

## TYPES OF TRACKING



Stateful tracking



Stateless tracking



Cross-device tracking

## OUR RESEARCH IMPACT

### Reproducible privacy evaluation of web APIs

- Fingerprinting risk demonstrated in Google's Topics API.
- Results quoted by Apple and Firefox, shared with UK CMA.

### Monitoring changes in the advertising ecosystem

- Large scale adoption measurements of new proposals.
- Actionable insights for improving real users' privacy.

### Systematizing web tracking findings

- Collaboration to identify advances and open problems in field.
- Inventorying all results on Google's Privacy Sandbox initiative.

### Privacy Sandstorm research portal

- Call for reproducible analyses and release of artifacts.
- More APIs evaluations to understand possible impacts.
- New research hub: <https://privacysandstorm.github.io>



### Annual state of the web reports

- Cookies chapter for the 2024 and 2025 Web Almanac editions.
- Analysis of measurement data on millions of webpages.

### Largest public dataset of website categories

- Featuring 147 million hosts and 31 million domains.

### Education & service

- Reproducibility efforts in systems security and privacy.
- New web tracking lecture at UW-Madison.

## CONTACT



<https://yohan.beugin.org>



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