CS 642: Computer Security and Privacy

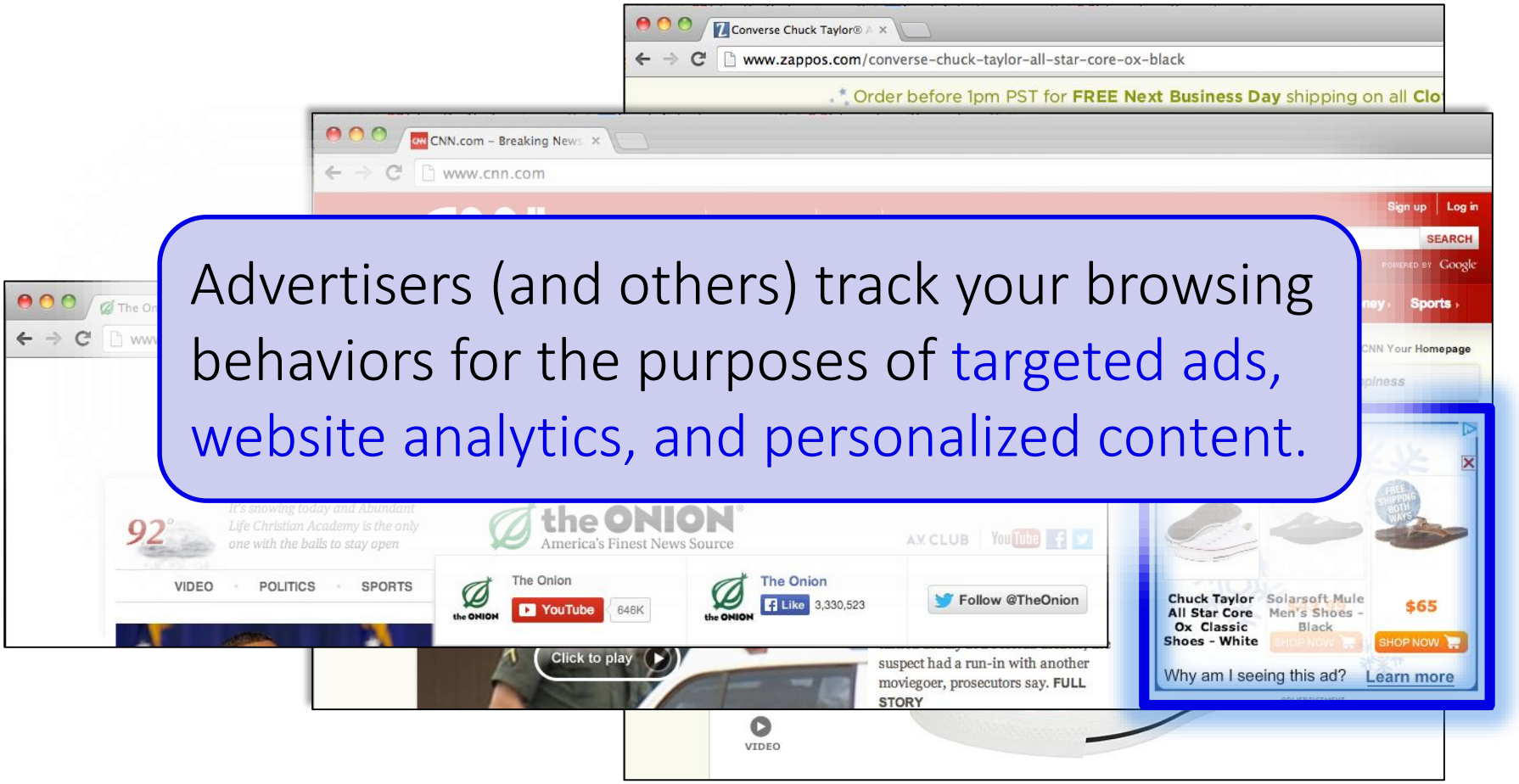
Web Security
[Privacy]

Spring 2020

Earlence Fernandes
earlence@cs.wisc.edu

Thanks to Dan Boneh, Dieter Gollmann, Dan Halperin, Yoshi Kohno, Ada Lerner, John Manferdelli, John Mitchell, Franzi Roesner, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials ...

Ads That Follow You



Third-Party Web Tracking

These ads allow **criteo.com** to link your visits between sites, even if you never click on the ads.

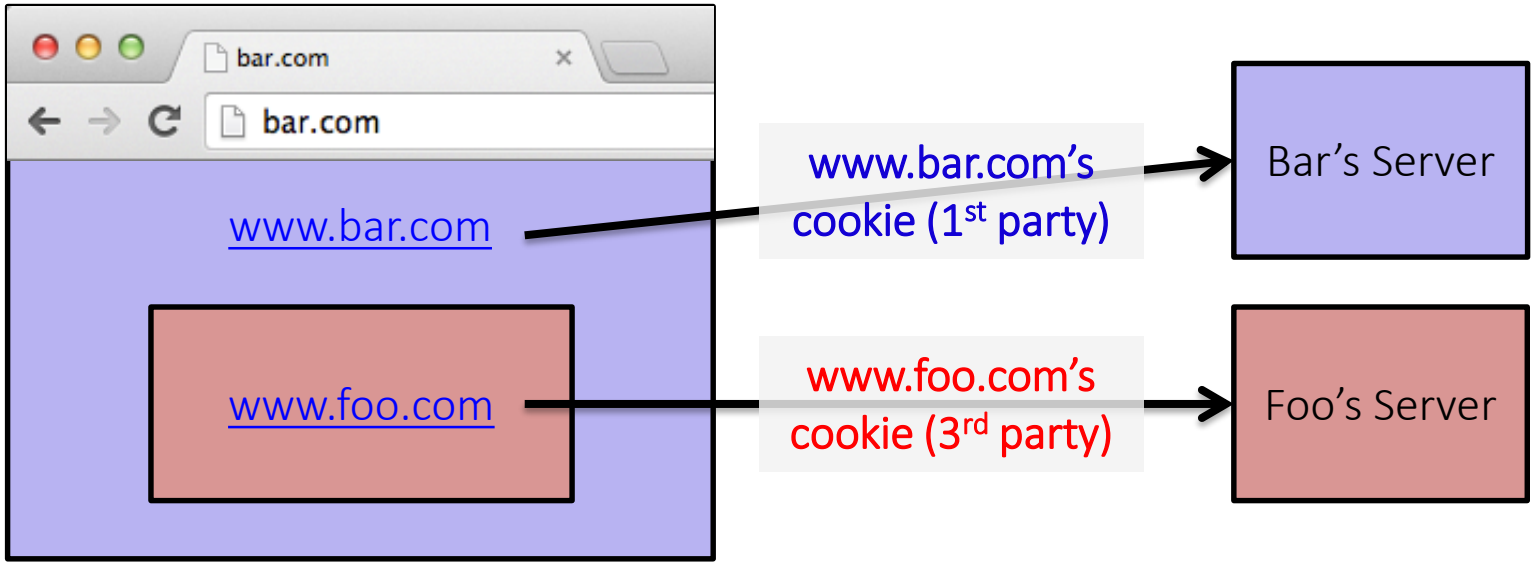
Concerns About Privacy

Outline

1. Understanding web tracking
2. Measuring web tracking
3. Defenses

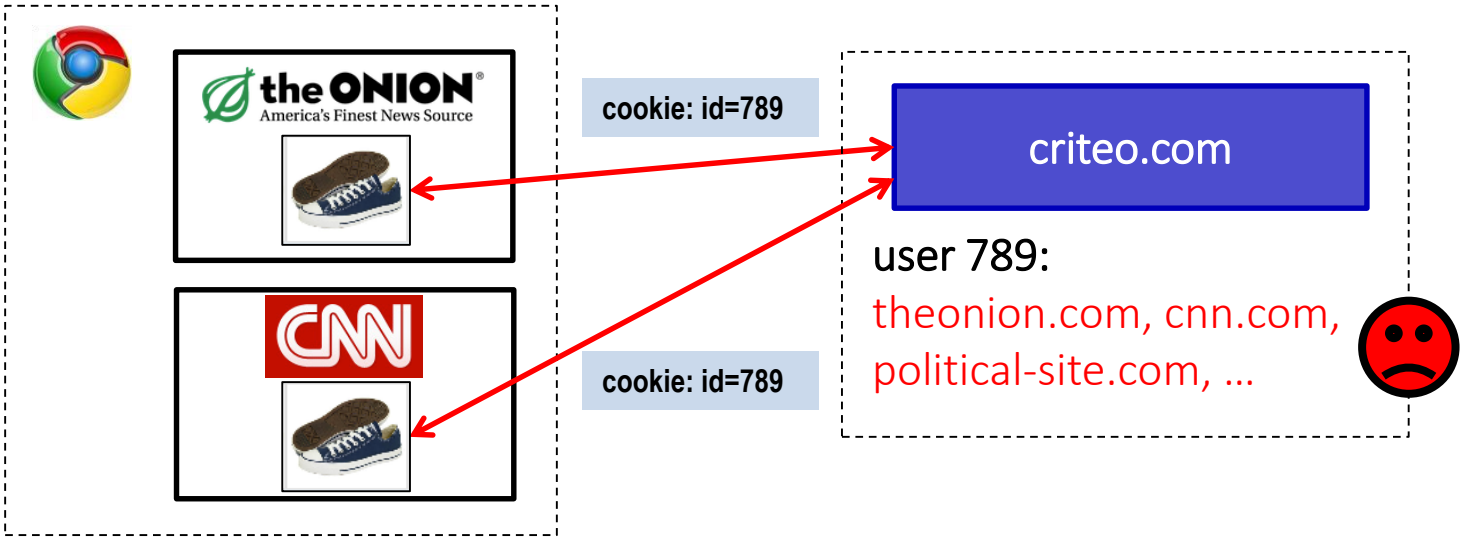
Recall: First and Third Parties

- First-party cookie: belongs to top-level domain.
- Third-party cookie: belongs to domain of embedded content (such as image, iframe).



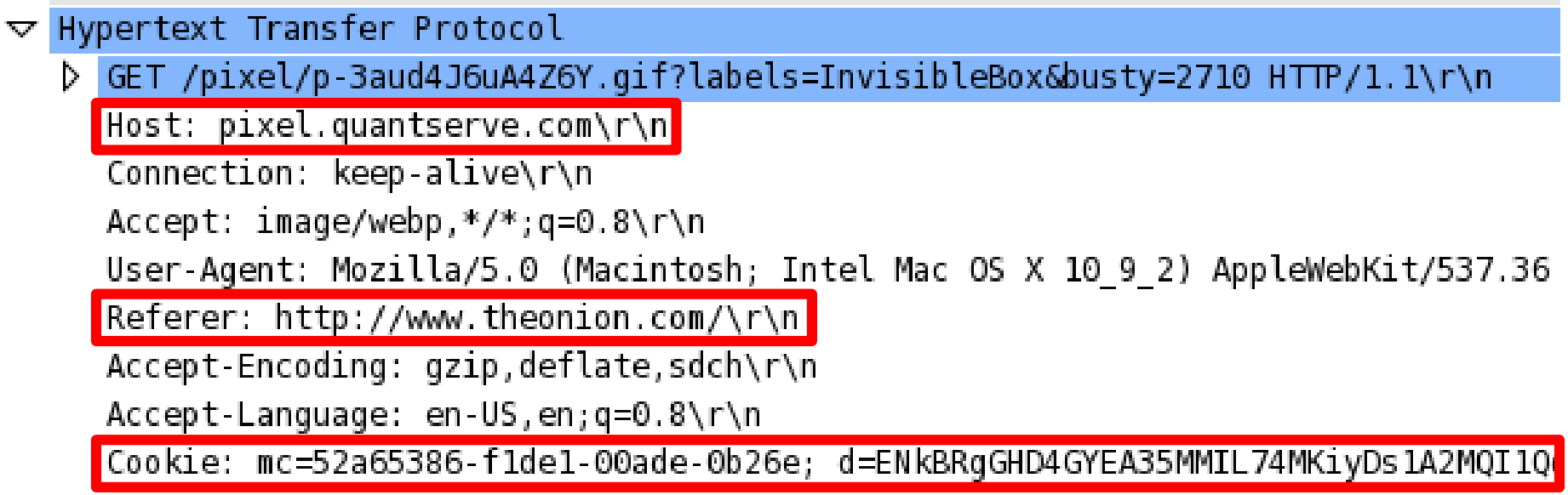
Anonymous Tracking

Trackers included in other sites use third-party cookies containing unique identifiers to create browsing profiles.



Basic Tracking Mechanisms

- Tracking requires:
  (1) re-identifying a user.
  (2) communicating id + visited site back to tracker.



Tracking Technologies

- HTTP Cookies
- HTTP Auth
- HTTP Etags
- Content cache
- IE userData
- HTML5 protocol and content handlers
- HTML5 storage

- Flash cookies
- Silverlight storage
- TLS session ID & resume
- Browsing history
- window.name
- HTTP STS
- DNS cache

- "Zombie" cookies that respawn (http://samy.pl/evercookie)

Fingerprinting Web Browsers

- User agent
- HTTP ACCEPT headers
- Browser plug-ins
- MIME support
- Clock skew
- Installed fonts
- Cookies enabled?
- Browser add-ons
- Screen resolution
- HTML5 canvas (differences in graphics SW/HW!)

A research project of the Electronic Frontier Foundation

Panopticlick

How Unique — and Trackable — Is Your Browser?

Is your browser configuration rare or unique? If so, web sites

Your browser fingerprint **appears to be unique** among the 3,435,834 tested so far

web.

Only anonymous data will be collected by this site.

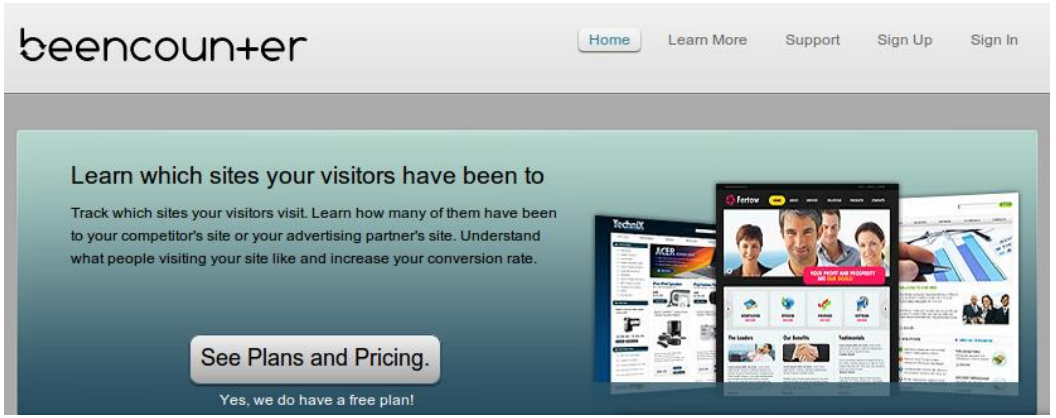
A paper reporting the statistical results of this experiment is now available: How Unique Is Your Browser?, Proceedings of the Privacy Enhancing Technologies Symposium (PETS 2010), Springer Lecture Notes in Computer Science.

Learn about Panopticlick and web tracking. The Panopticlick Privacy Policy. Learn about the Electronic Frontier Foundation.

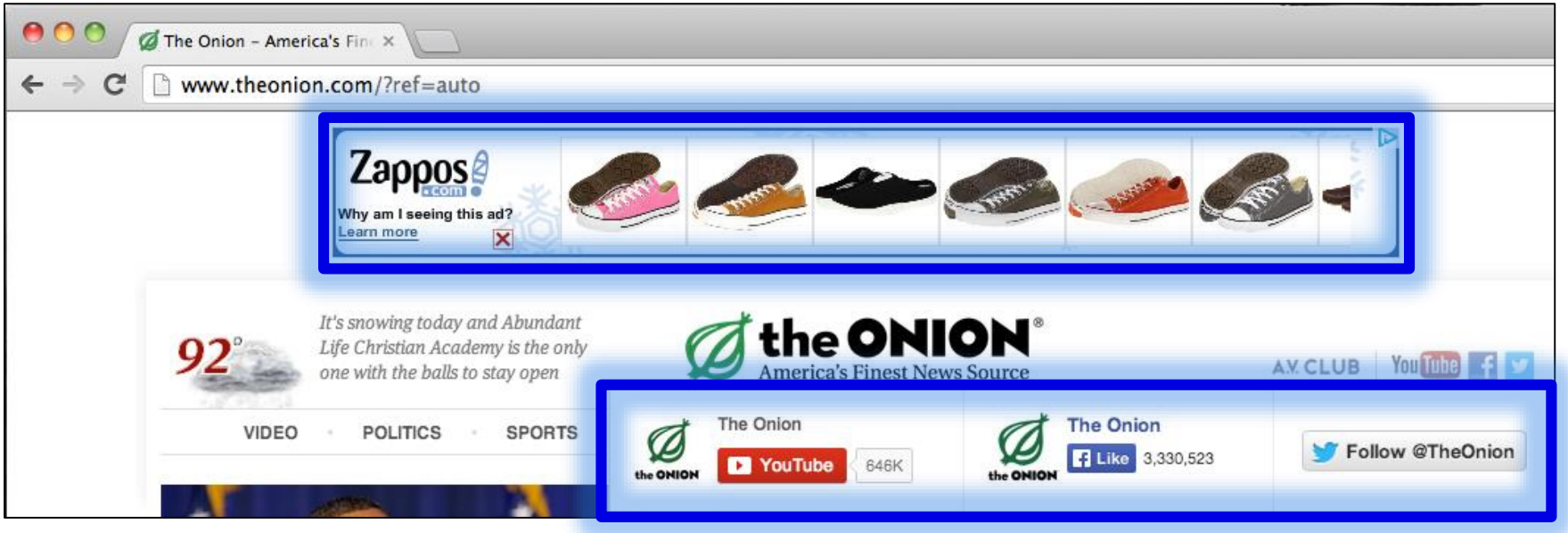
History Sniffing

How can a webpage figure out which sites you visited previously?

- Color of links
  - CSS :visited property
  - getComputedStyle()
- Cached Web content timing
- DNS timing

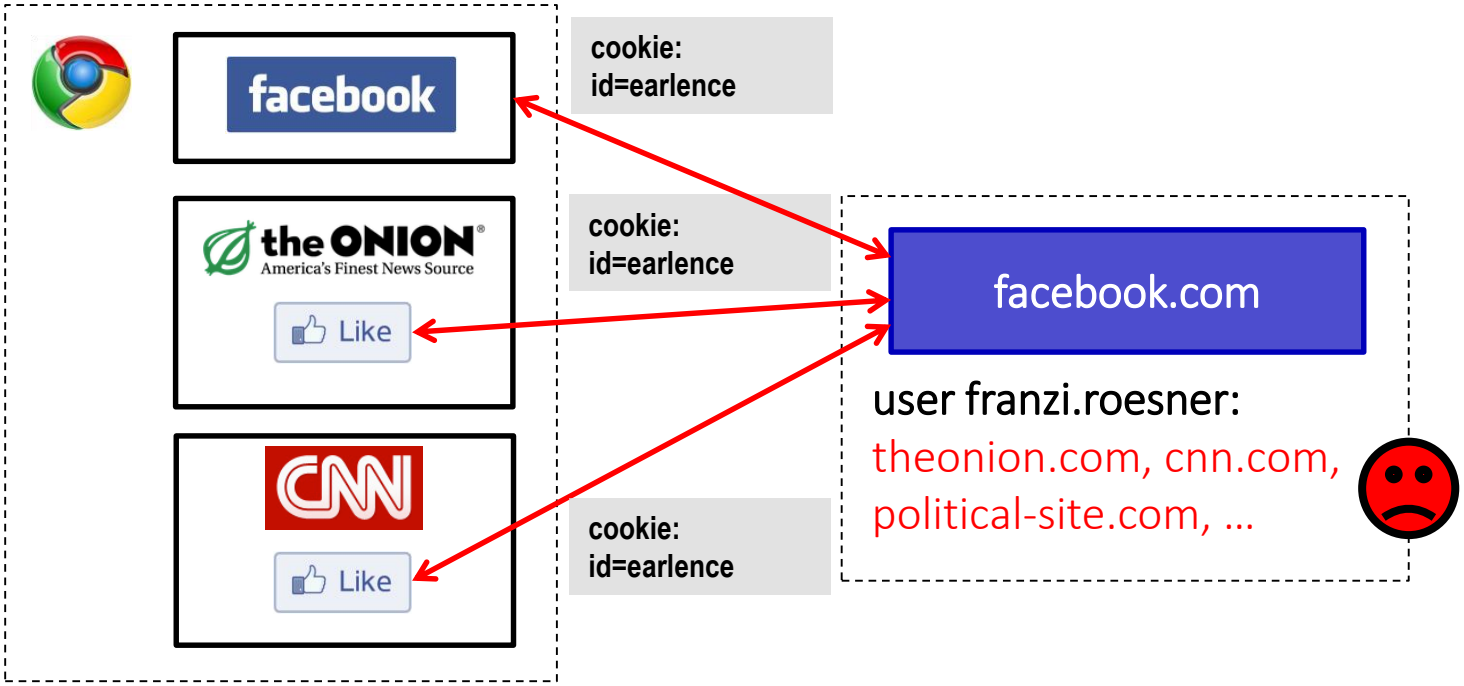


Other Trackers?



"Personal" Trackers

Personal Tracking



- Tracking is not anonymous (linked to accounts).
- Users directly visit tracker's site → evades some defenses.

Outline

1. Understanding web tracking
2. Measuring web tracking
3. Defenses

[NSDI '12]

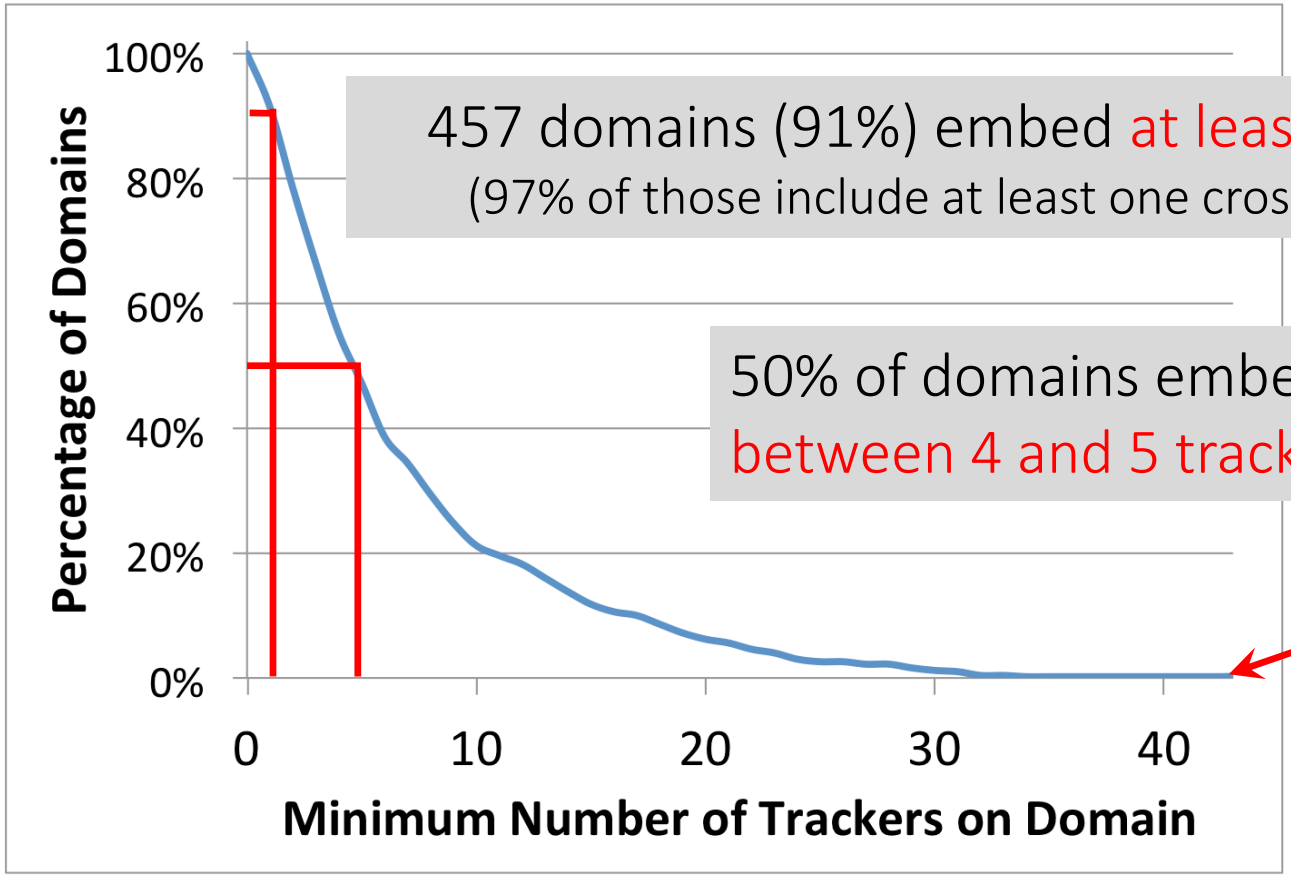
Measurement Study

- **Questions:**
  - How prevalent is tracking (of different types)?
  - How much of a user's browsing history is captured?
  - How effective are defenses?
- **Approach:** Build tool to automatically crawl web, detect and categorize trackers based on our taxonomy.

Longitudinal studies since then: tracking has increased and become more complex.

How prevalent is tracking? (2011)

524 unique trackers on Alexa top 500 websites (homepages + 4 links)

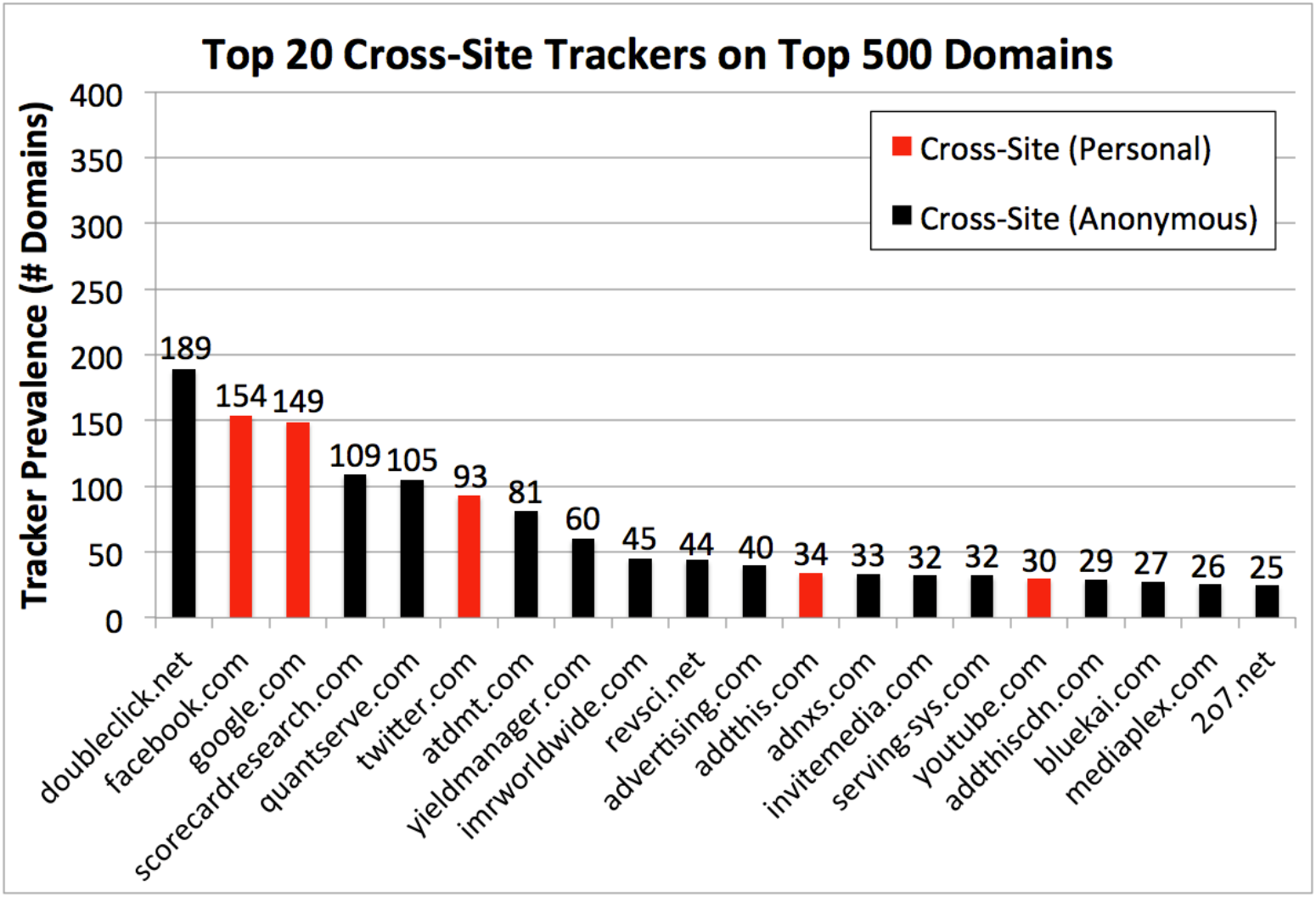


457 domains (91%) embed at least one tracker.
(97% of those include at least one cross-site tracker.)

50% of domains embed between 4 and 5 trackers.

One domain includes 43 trackers.

Who/what are the top trackers? (2011)



How are users affected?

- Question: How much of a real user's browsing history can top trackers capture?
- Measurement challenges:
  - Privacy concerns.
  - Users may not browse realistically while monitored.
- Insight: AOL search logs (released in 2006) represent real user behaviors.

How are users affected?

- Idea: Use AOL search logs to create 30 hypothetical browsing histories.
  - 300 unique queries per user → top search hits.
- Trackers can capture a large fraction:
  - Doubleclick: Avg 39% (Max 66%)
  - Facebook: Avg 23% (Max 45%)
  - Google: Avg 21% (Max 61%)

How are users affected?

POLICY & LAW | US & WORLD | NATIONAL SECURITY

NSA reportedly 'piggybacking' on Google advertising cookies to home in on surveillance targets

By Nathan Ingraham on December 10, 2013 10:41 pm Email @NateIngraham

See also: ADINT (2017)

- Trackers can capture a large fraction:
  - Doubleclick: Avg 39% (Max 66%)
  - Facebook: Avg 23% (Max 45%)
  - Google: Avg 21% (Max 61%)

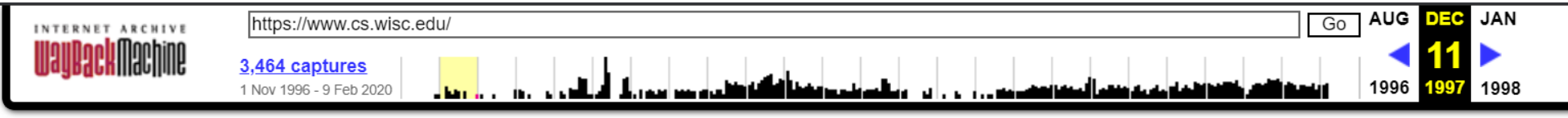
How has this changed over time?

- The web has existed for a while now…
  - What about tracking before 2011?
  - What about tracking before 2009?

- Solution: time travel!

[USENIX Security '16]

The Wayback Machine to the Rescue



UNIVERSITY OF WISCONSIN MADISON
150 YEARS
Your Resource for a Lifetime

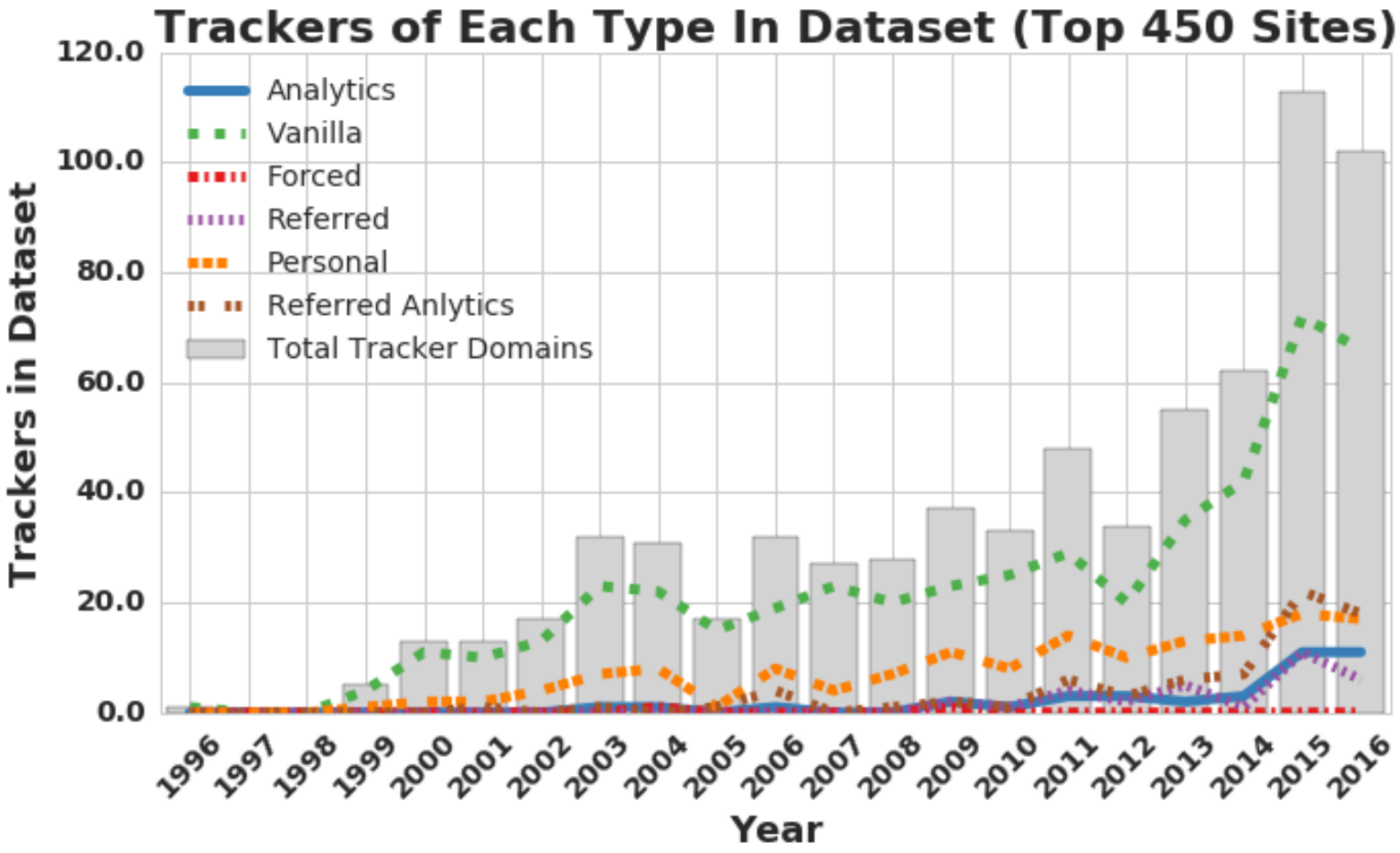
Computer Sciences Department

- About the department
- Research projects and technical reports
- People and organizations
- Academic information
- List of UW CS resources
- The Computer Systems Lab (CSL)
- Online documentation (FAQs)
- Search our website
- Contact Information

Time travel for web tracking: http://trackingexcavator.cs.washington.edu

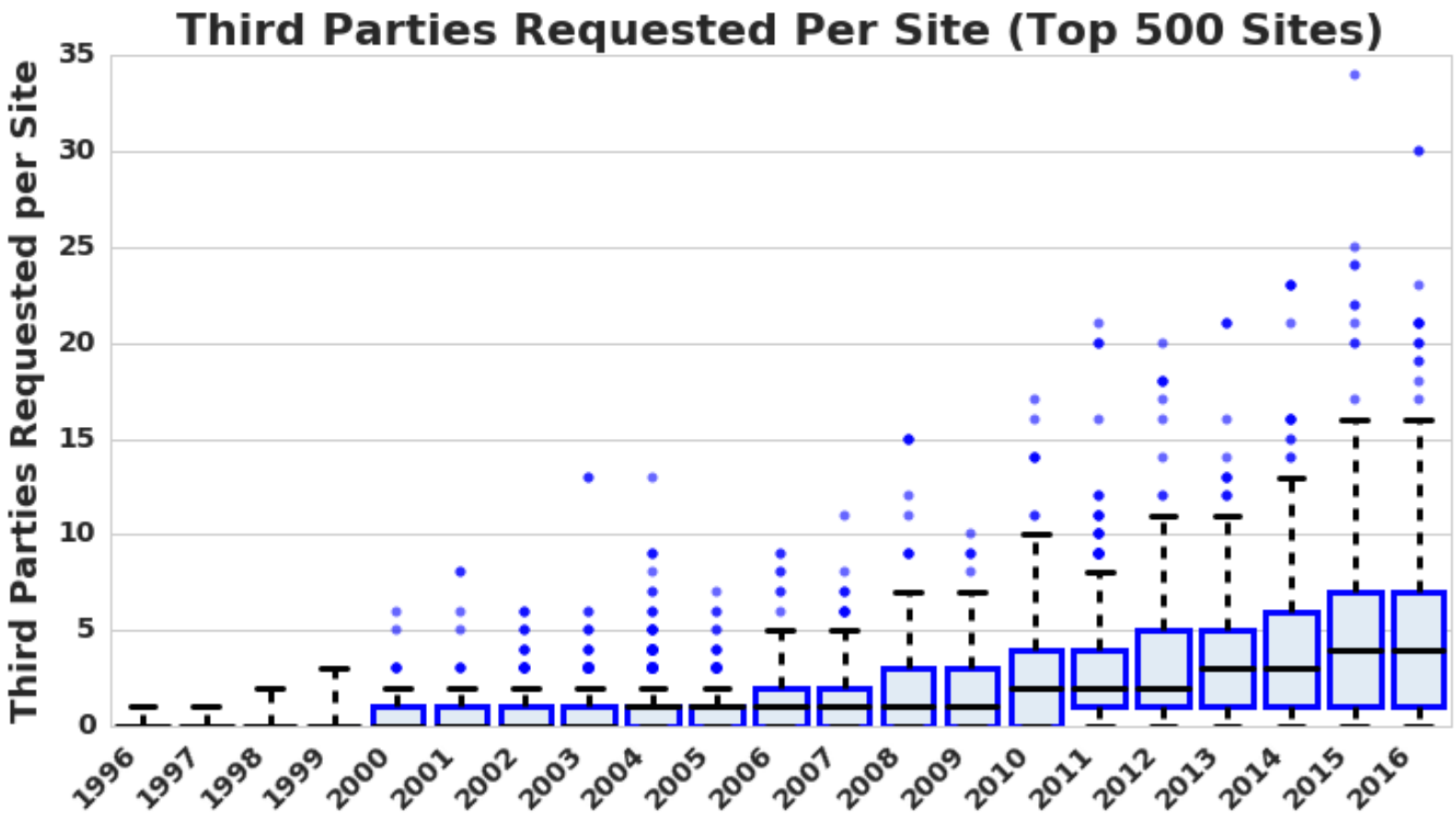
1996-2016: More & More Tracking

- More trackers of more types



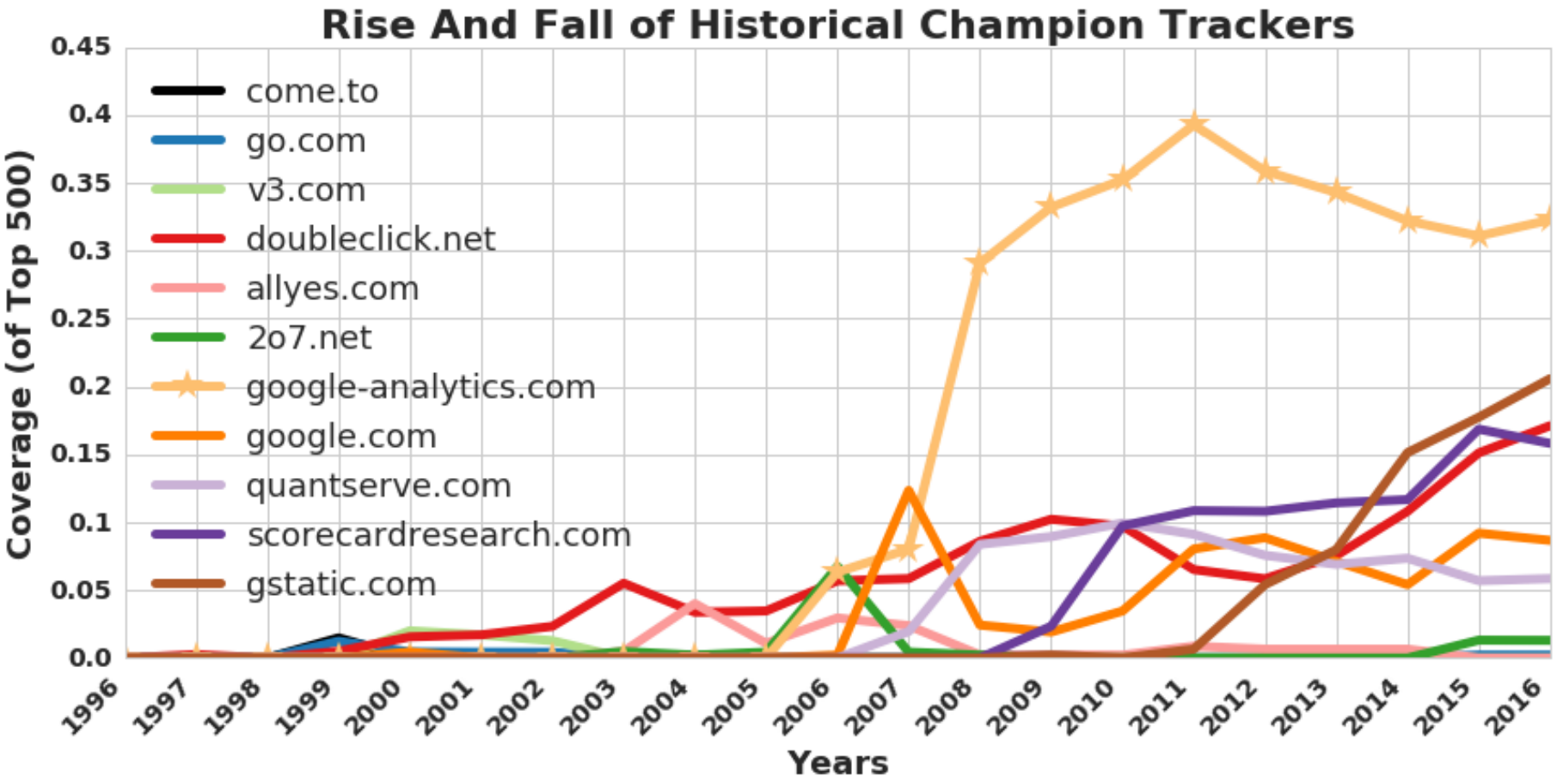
1996-2016: More & More Tracking

- More trackers of more types, more per site



1996-2016: More & More Tracking

- More trackers of more types, more per site, more coverage



Outline

1. Understanding web tracking
2. Measuring web tracking
3. Defenses

Defenses to Reduce Tracking

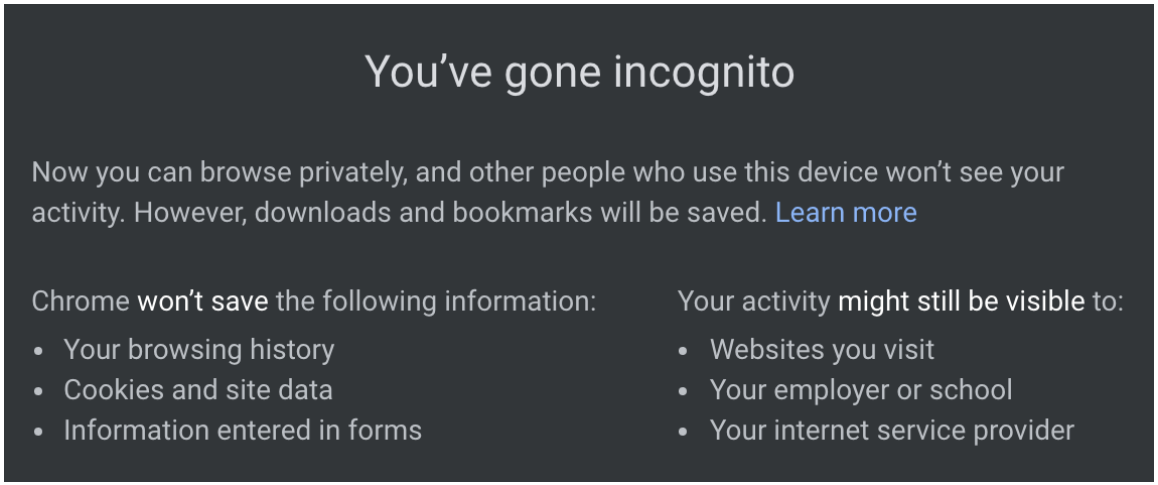
- Do Not Track proposal?

Do Not Track is not a technical defense: trackers must honor the request.

Defenses to Reduce Tracking

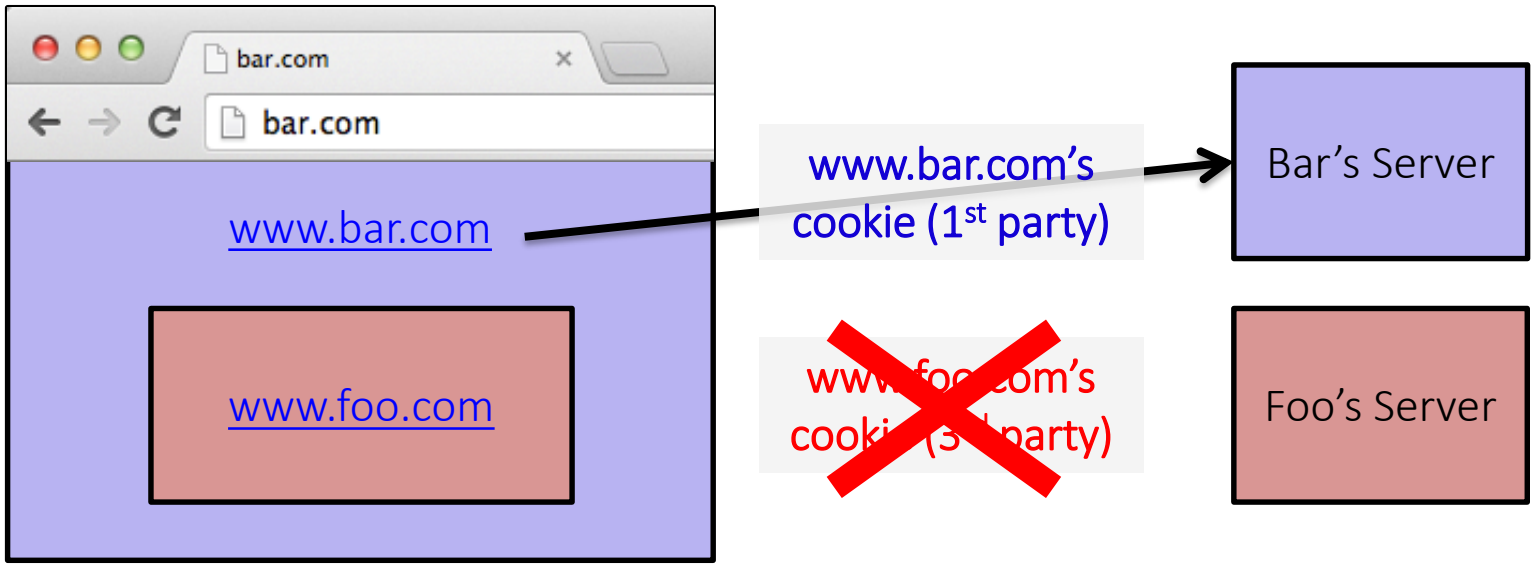
- Do Not Track proposal?
- Private browsing mode?

Private browsing mode protects against local, not network, attackers.



Defenses to Reduce Tracking

- Do Not Track proposal?
- Private browsing mode?
- Third-party cookie blocking?



Quirks of 3rd Party Cookie Blocking

In some browsers, this option means third-party cookies cannot be set, but they CAN be sent.

So if a third-party cookie is somehow set, it can be used.

How to get a cookie set?

One way: be a first party.

Google

etc.

Defenses to Reduce Tracking

- Do Not Track header?
- Private browsing mode?
- Third-party cookie blocking?
- Browser add-ons?

"uses algorithmic methods to decide what is and isn't tracking"

Often rely on blacklists, which may be incomplete.