



Web Security [Privacy]

Spring 2020

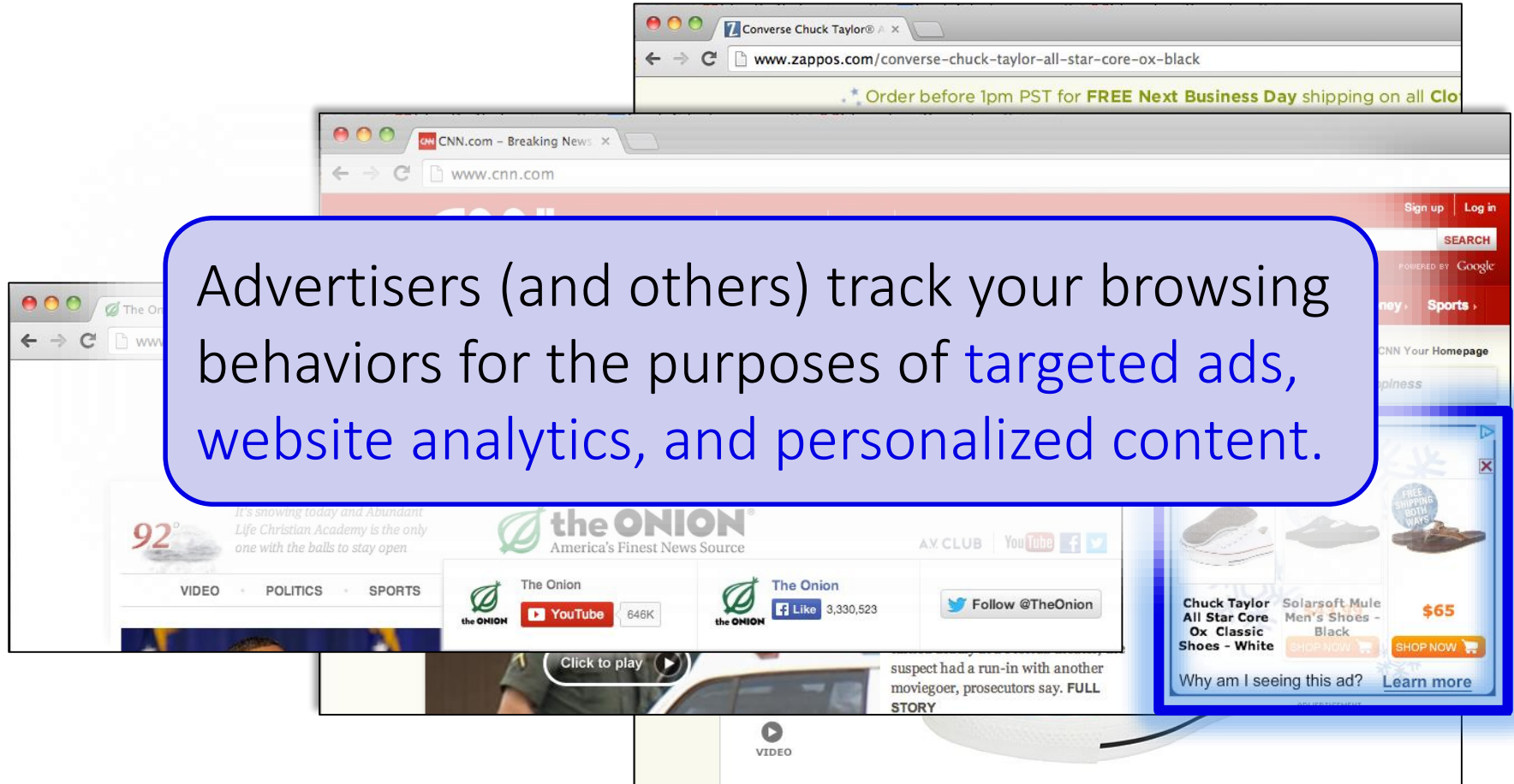
Earlence Fernandes

earlence@cs.wisc.edu

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

Ads That Follow You

Advertisers (and others) track your browsing behaviors for the purposes of **targeted ads**, website analytics, and personalized content.



Third-Party Web Tracking

Browsing profile for user 123:

- cnn.com
- theonion.com
- political-site.com



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

Concerns About Privacy

THE WALL STREET JOURNAL.

WHAT THEY KNOW | JULY 30, 2010

The Wall Street Journal

A Journal investigating business

CNN

Your Privacy

Big dep

By

His

all to be put up

The file consists

identifies her as

The New York Times

May 6, 2011, 5:01 pm | 3 Comments

'Do Not Track' Privacy Bill Appears in Congress

By TANZINA VEGA

And the privacy legislation just keeps on coming.

On Friday, two bills were introduced in Washington in support of a Do Not Track mechanism that would give users control over how much of their data was collected by advertisers and other online companies.

als
ion

By JENNIFER VALENTINO-DEVRIES,
JEREMY SINGER-VINE and ASHKAN SOLTANI

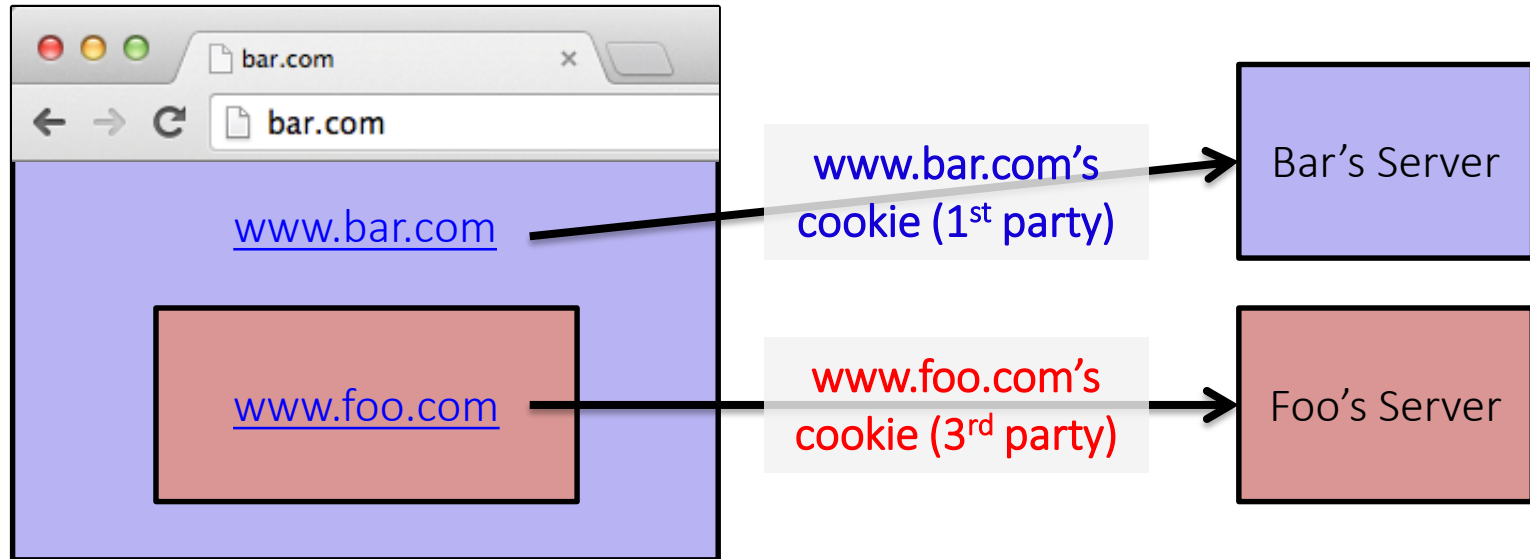
December 24, 2012

Outline

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

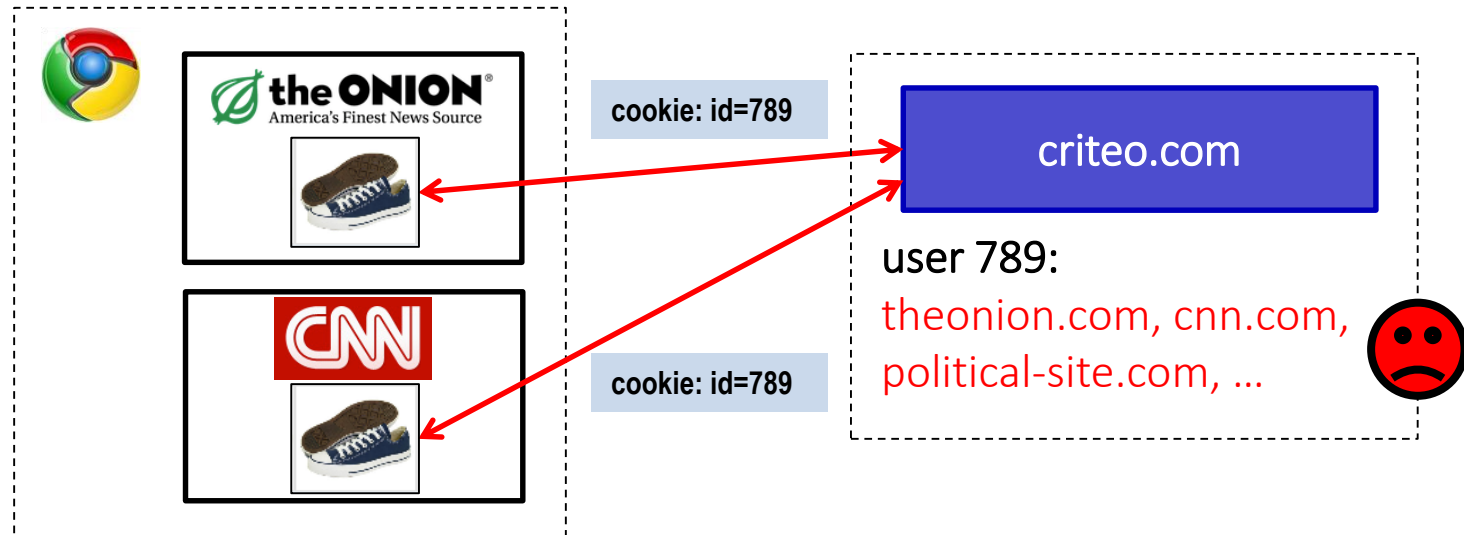
Recall: First and Third Parties

- **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.

▼ Hypertext Transfer Protocol

▷ GET /pixel/p-3aud4J6uA4Z6Y.gif?labels=InvisibleBox&busty=2710 HTTP/1.1\r\n

Host: pixel.quantserve.com\r\n

Connection: keep-alive\r\n

Accept: image/webp,*/*;q=0.8\r\n

User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36

Referer: http://www.theonion.com/\r\n

Accept-Encoding: gzip,deflate,sdch\r\n

Accept-Language: en-US,en;q=0.8\r\n

Cookie: mc=52a65386-f1de1-00ade-0b26e; d=ENkBRgGHD4GYEA35MMIL74MKiyDs1A2MQI1Q

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

Only **anonymous data** will be collected by this site.

TEST
ME

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.](#)

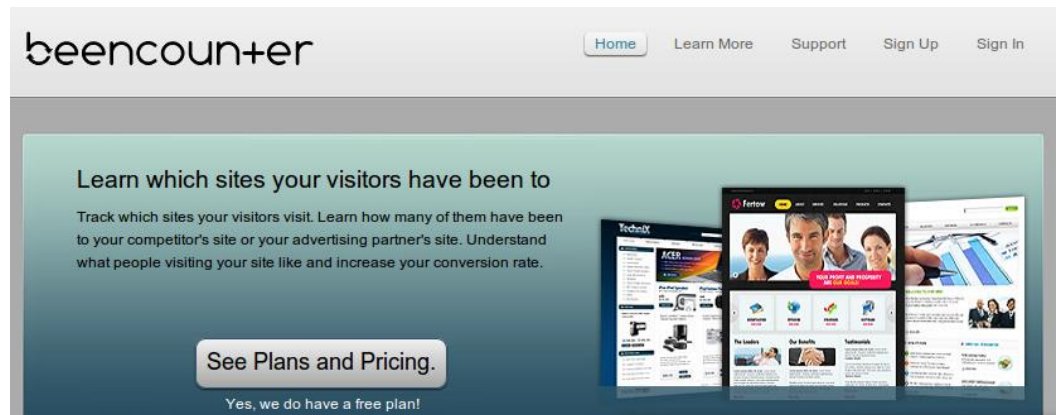
[The Panopticlick Privacy Policy.](#)

[Learn about the Electronic Frontier Foundation.](#)

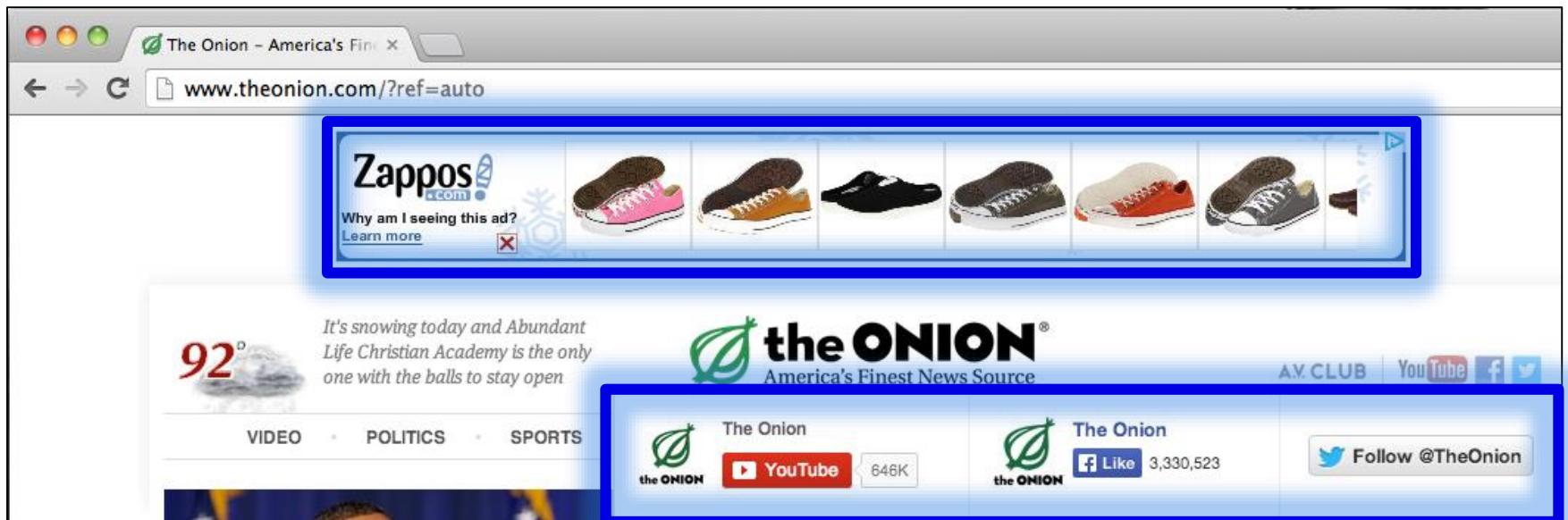
History Sniffing

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

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



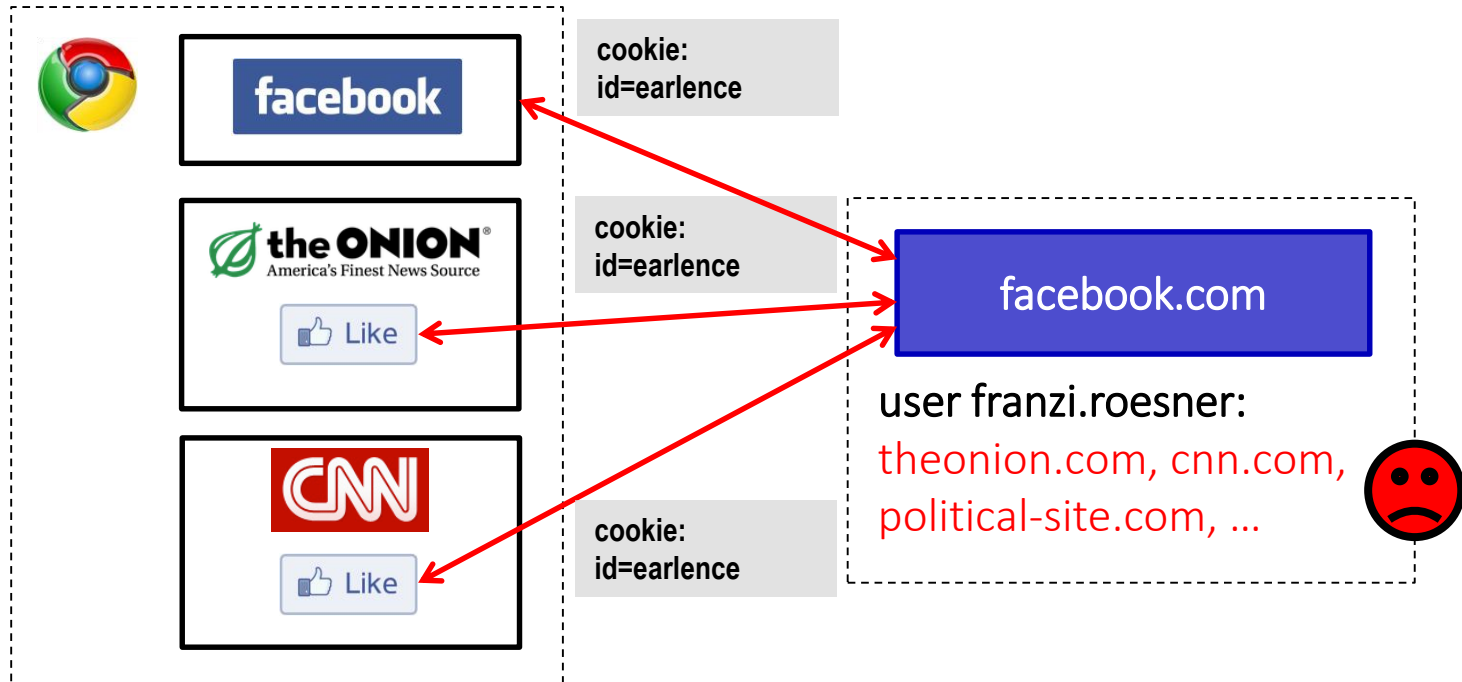
Other Trackers?



“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

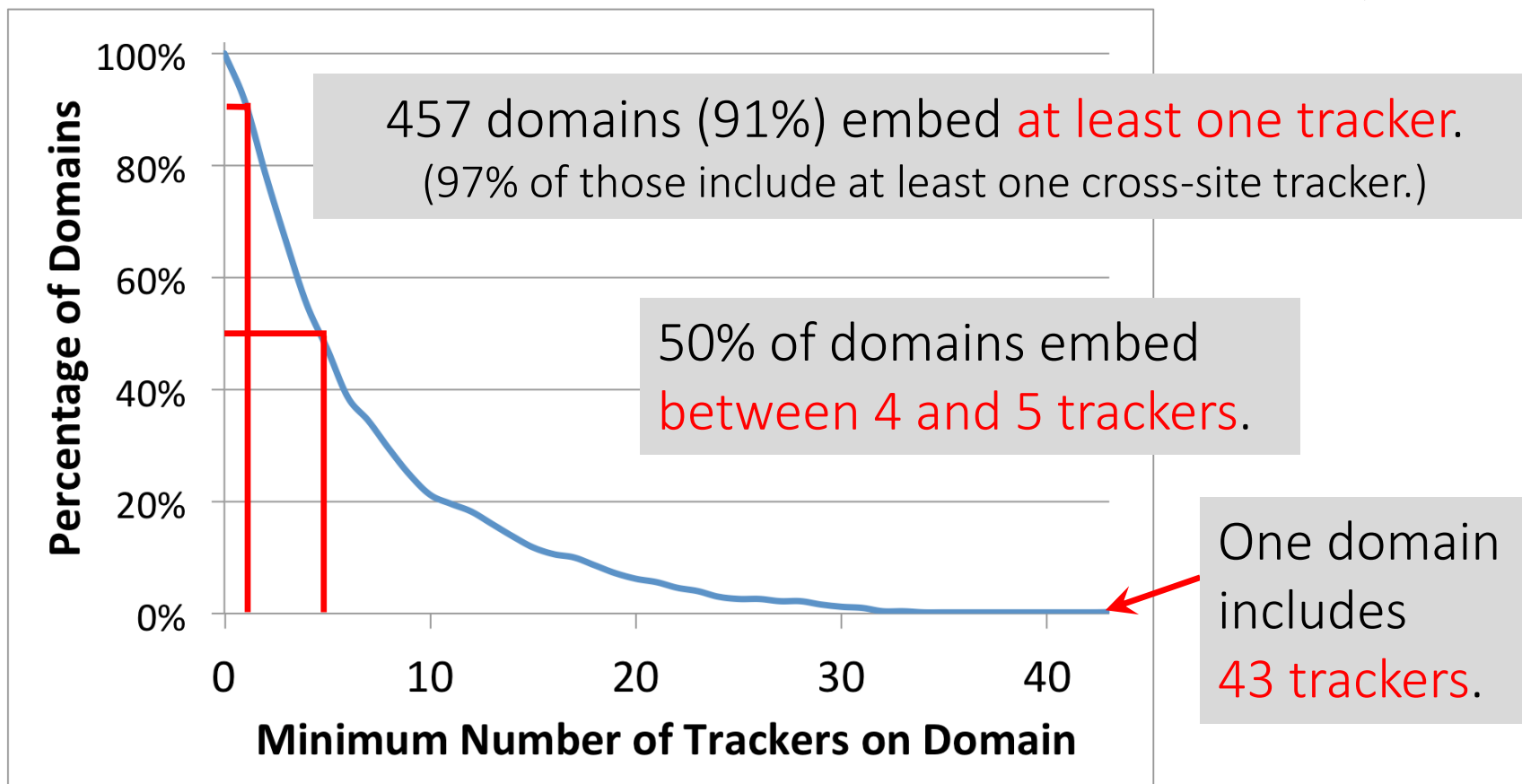
Measurement Study

- **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.

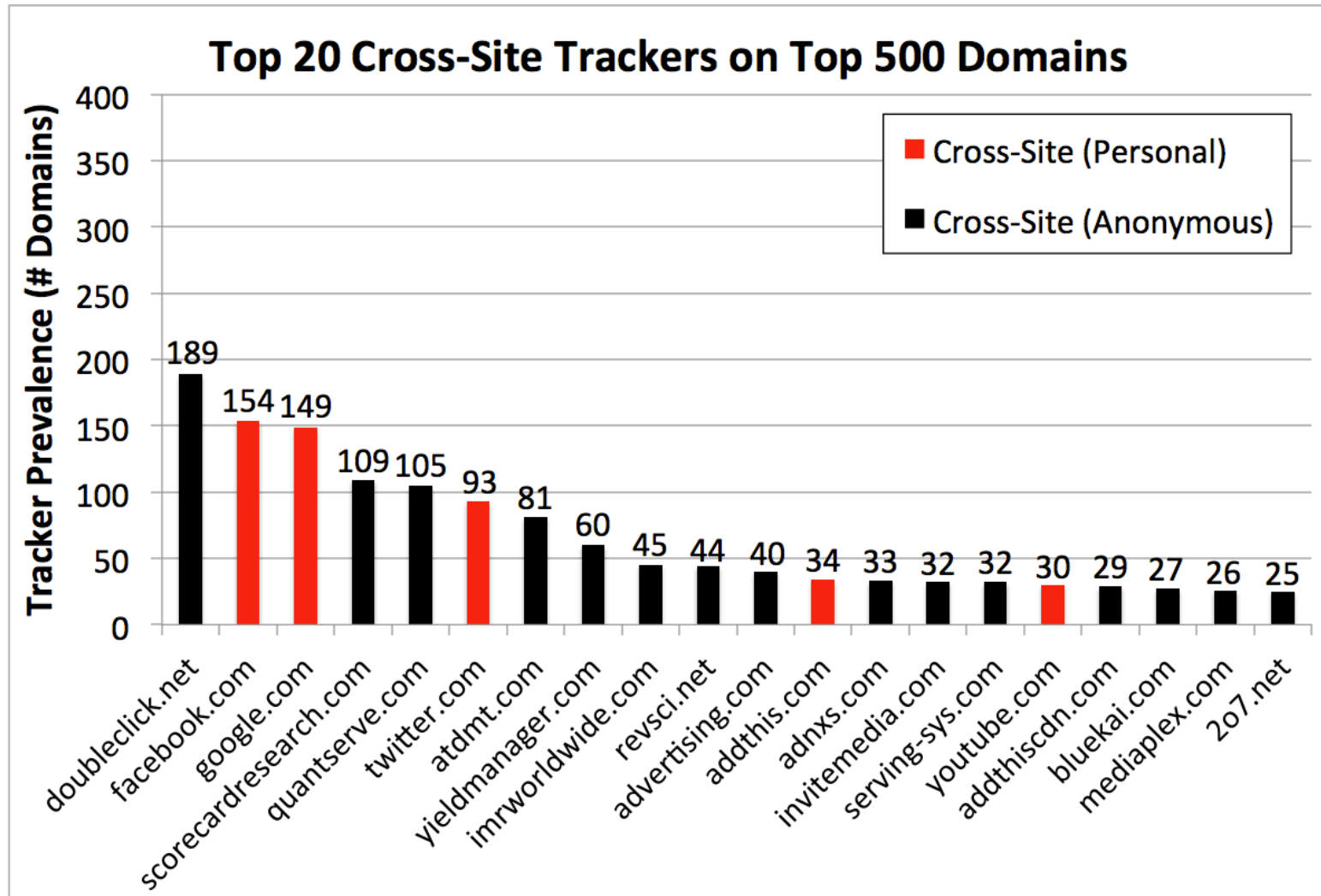
Longitudinal studies since then: **tracking has increased and become more complex.**

How prevalent is tracking? (2011)

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



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

See also: ADINT (2017)

By **Nathan Ingraham** on December 10, 2013 10:41 pm [Email](#) [@NateIngraham](#)

- 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

INTERNET ARCHIVE
Wayback Machine

<https://www.cs.wisc.edu/> AUG **DEC** JAN

3,464 captures
1 Nov 1996 - 9 Feb 2020

1996 **1997** 1998

11



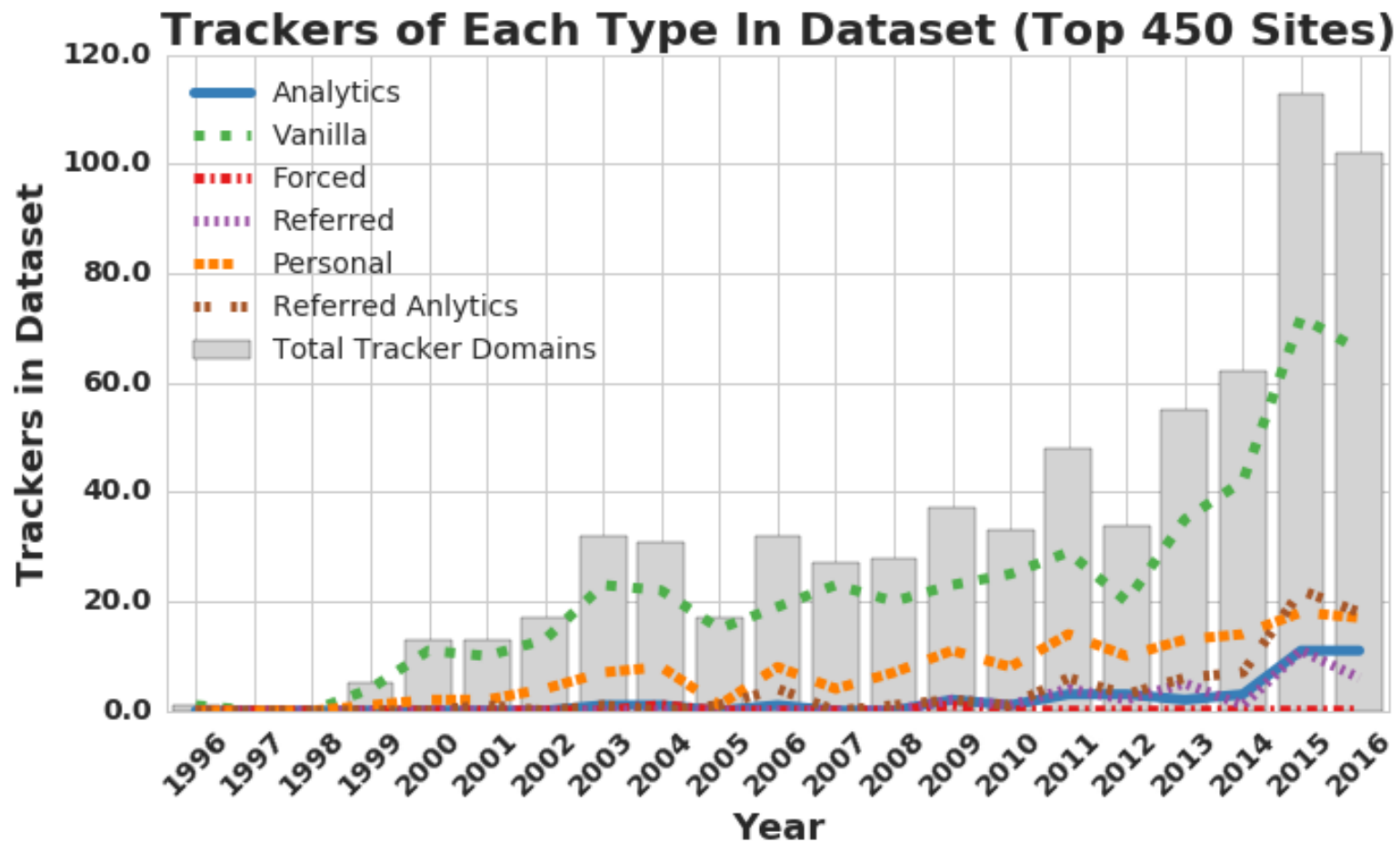
Computer Sciences Department

- [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>

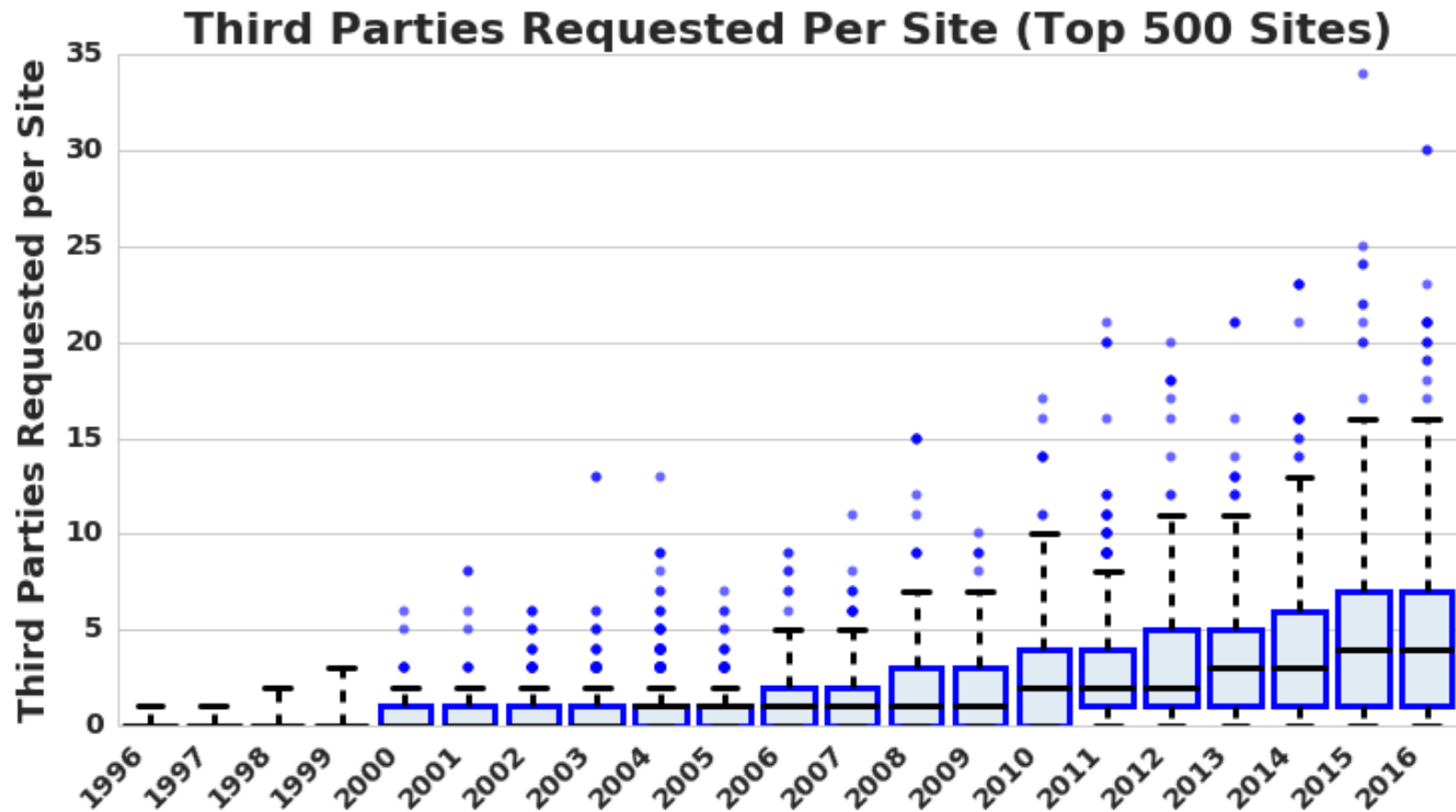
1996-2016: More & More Tracking

- More trackers of more types



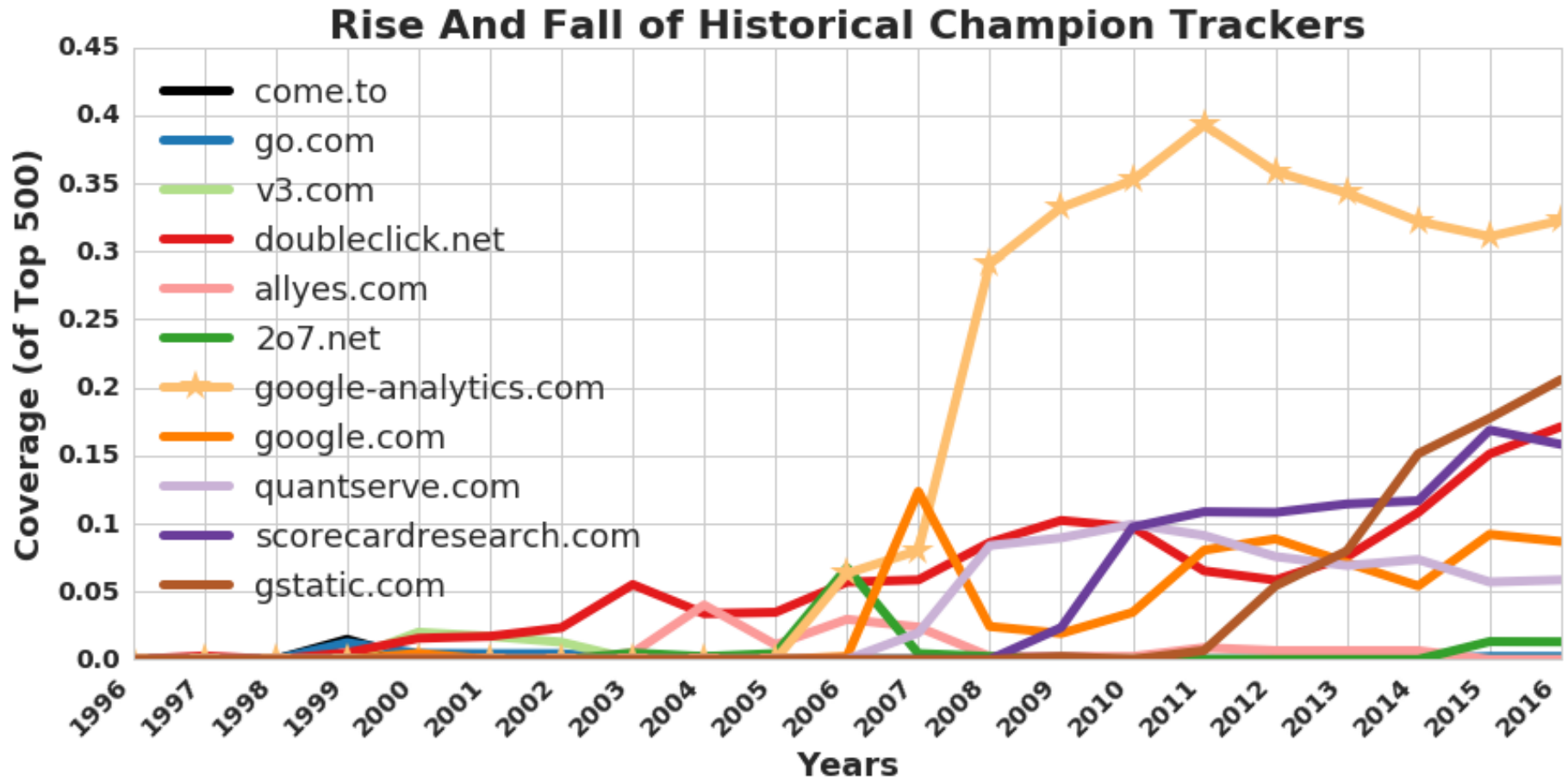
1996-2016: More & More Tracking

- 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?

☒ Send a 'Do Not Track' request with your browsing traffic

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.

You've gone incognito

Now you can browse privately, and other people who use this device won't see your activity. However, downloads and bookmarks will be saved. [Learn more](#)

Chrome won't save the following information:

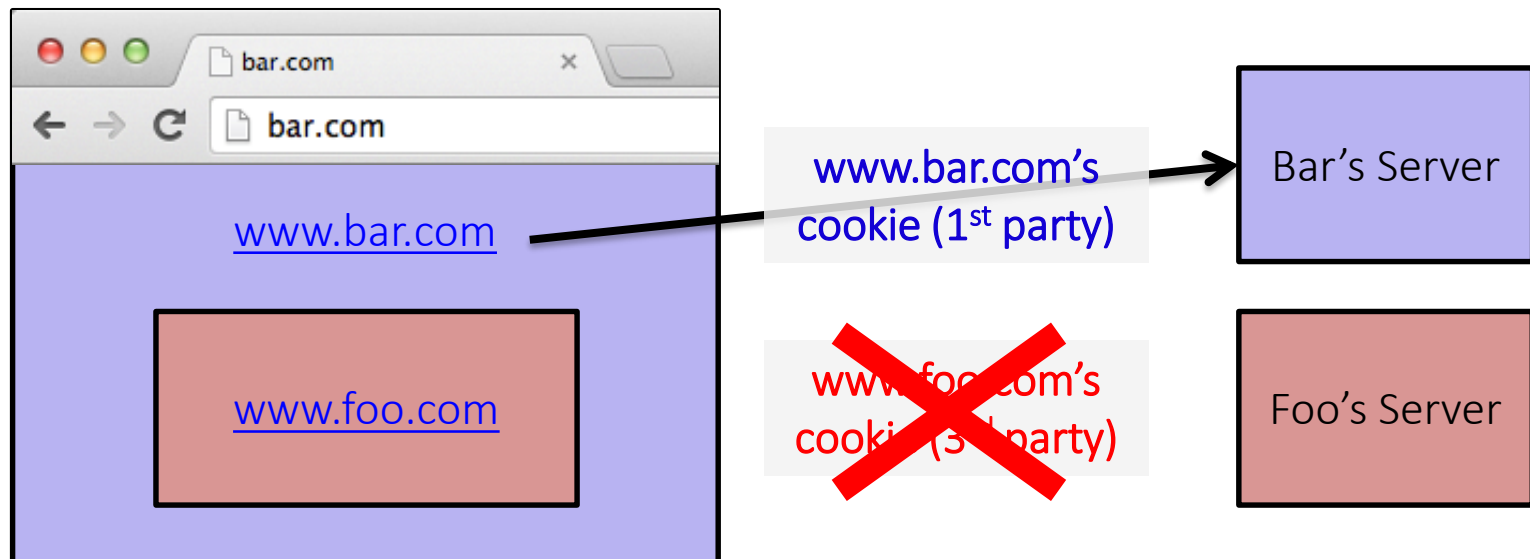
- Your browsing history
- Cookies and site data
- Information entered in forms

Your activity might still be visible to:

- Websites you visit
- Your employer or school
- Your internet service provider

Defenses to Reduce Tracking

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



Quirks of 3rd Party Cookie Blocking

Cookies

- ☒ Allow local data to be set (recommended)
- ☐ Keep local data only until I quit my browser
- ☐ Block sites from setting any data
- ☒ Block third-party cookies and site data

[Manage exceptions...](#) [All cookies and site data...](#)

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.



etc.

Defenses to Reduce Tracking

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



Often rely on blacklists,
which may be incomplete.



*“uses algorithmic
methods to decide what
is and isn't tracking”*

