

CSE 484 / CSE M 584: Computer Security and Privacy

Third-Party Tracking on the Web

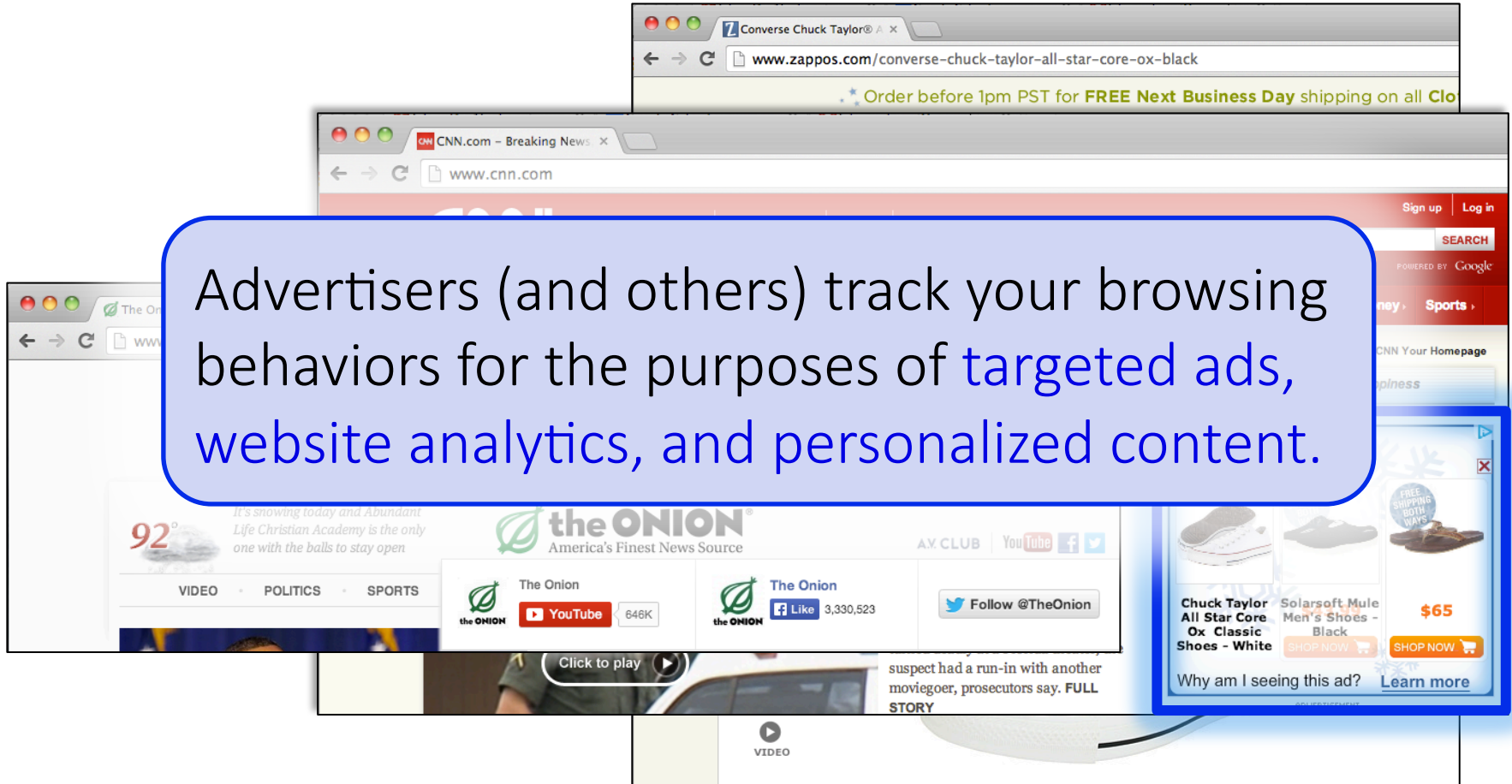
Spring 2015

Franziska (Franzi) Roesner
franzi@cs.washington.edu

Thanks to Dan Boneh, Dieter Gollmann, Dan Halperin, Yoshi Kohno, John Manferdelli, John Mitchell, 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
- adult-site.com
- political-site.com

The image shows a browser window with a Zappos advertisement. The ad is for men's shoes and includes a 'Learn more' link. A blue box highlights the ad and a list of visited sites. A sad face icon is next to the list.

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

Concerns About Privacy (2010 – 2011)

THE WALL STREET JOURNAL.
WHAT THEY KNOW | JULY 30, 2010
The Web's New Gold Mine: Your Secrets

A Job
busin

The New York Times
May 6, 2011, 5:01 pm | 3 Comments

'Do Not Track' Privacy Bill Appears in Congress
By TANZINA VEGA

By J
Hidd
all to
The
ident

And the privacy legislation just keeps on coming.

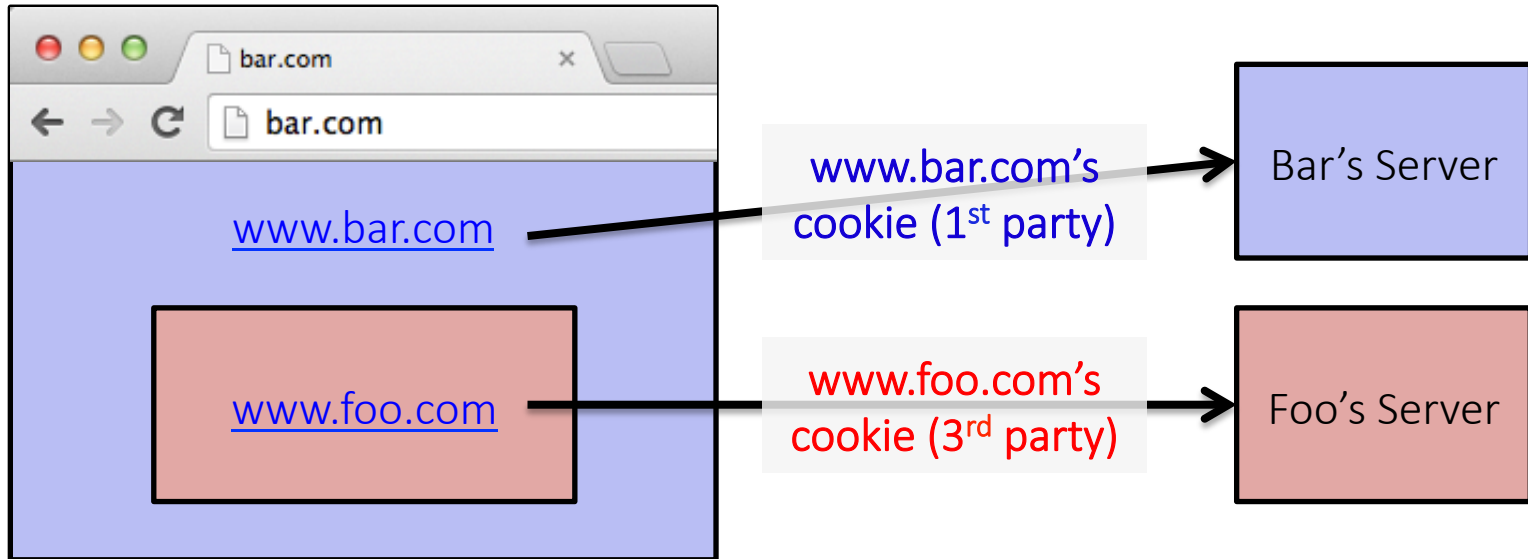
On Friday, two bills were introduced in Washington in support of a Do Not Track mechanism that would give users control over how much of their data was collected by advertisers and other online companies.

Outline

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

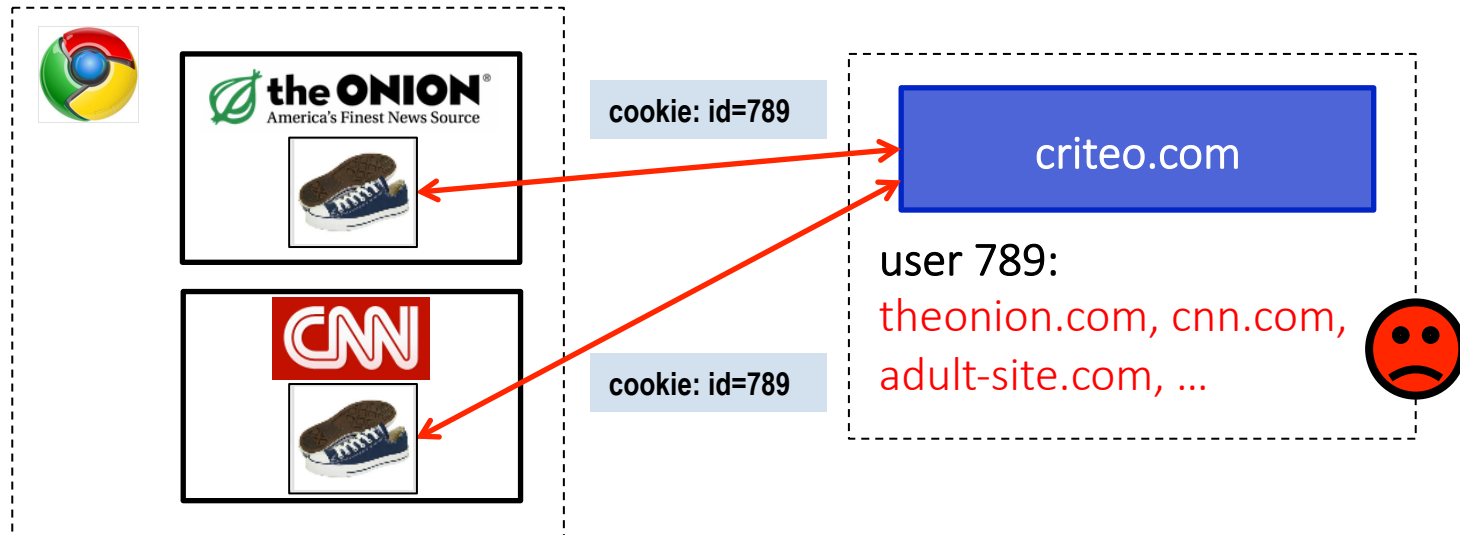
Recall: First and Third Parties

- **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

Fingerprinting Web Browsers

- User agent
- HTTP ACCEPT headers
- Browser plug-ins
- MIME support
- Clock skew
- Installed fonts
- Cookies enabled?
- Browser add-ons
- Screen resolution



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.



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.](#)

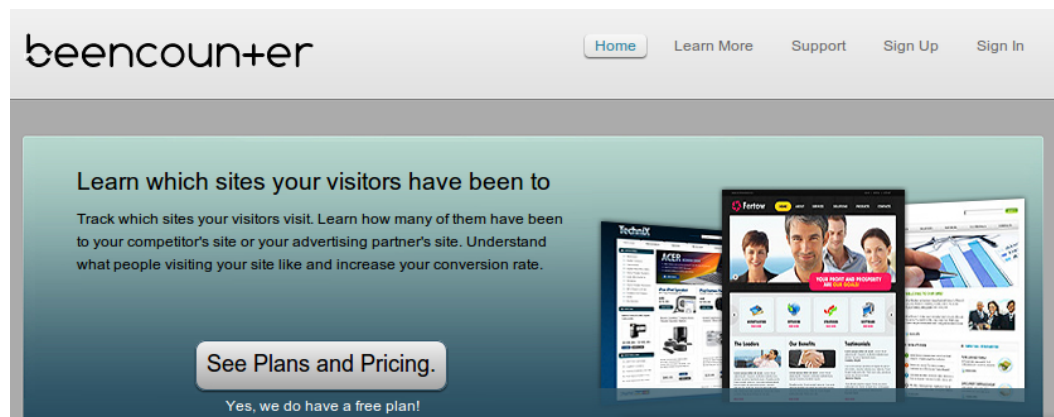
[The Panopticlick Privacy Policy.](#)

[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



The screenshot shows the website for beencounter. The header includes the logo 'beencounter' and navigation links: Home, Learn More, Support, Sign Up, and Sign In. The main content area features a promotional banner with the following text:

Learn which sites your visitors have been to

Track which sites your visitors visit. Learn how many of them have been to your competitor's site or your advertising partner's site. Understand what people visiting your site like and increase your conversion rate.

[See Plans and Pricing.](#)

Yes, we do have a free plan!

The banner also includes several small images of various websites, such as 'Techdix' and 'Fertox', illustrating the types of sites that can be tracked.

How Websites Get Your Identity

Personal trackers



Leakage of identifiers

GET http://ad.doubleclick.net/adj/...

Referer: http://submit.SPORTS.com/...?email=jdoe@email.com

Cookie: id=35c192bcfe0000b1...

Security bugs

Third party buys your identity

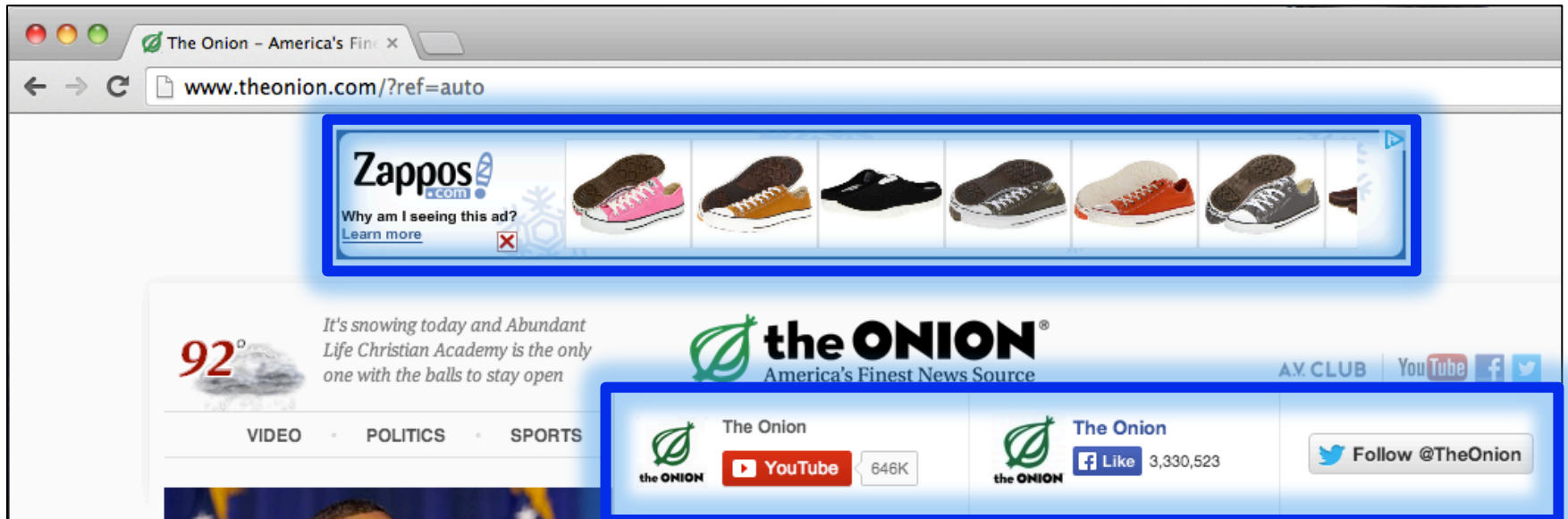
Understanding the Tracking Ecosystem

- In 2011, much discussion about tracking, but limited understanding of how it actually works.
- Our Goal: systematically study web tracking ecosystem to inform policy and defenses.
- Challenges:
 - No agreement on definition of tracking.
 - No automated way to detect trackers.
(State of the art: blacklists)

Our Tracking Taxonomy *[NSDI '12]*

- In the wild, tracking is much more complicated.
- (1) Trackers don't just use cookies.
 - Flash cookies, HTML5 LocalStorage, etc.
- (2) Trackers exhibit different behaviors.
 - Within-site vs. cross-site.
 - Anonymous vs. non-anonymous.
 - Specific behavior types:
analytics, vanilla, forced, referred, personal.

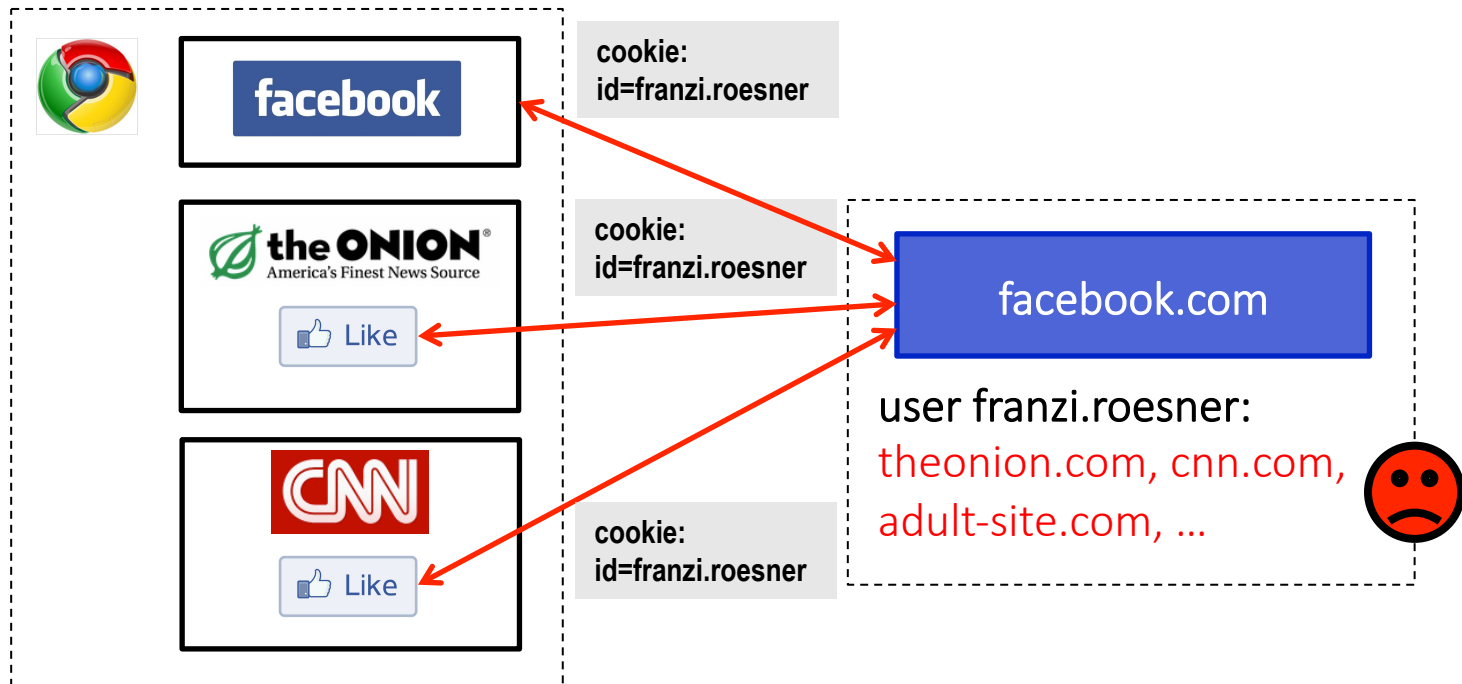
Other Trackers?



“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

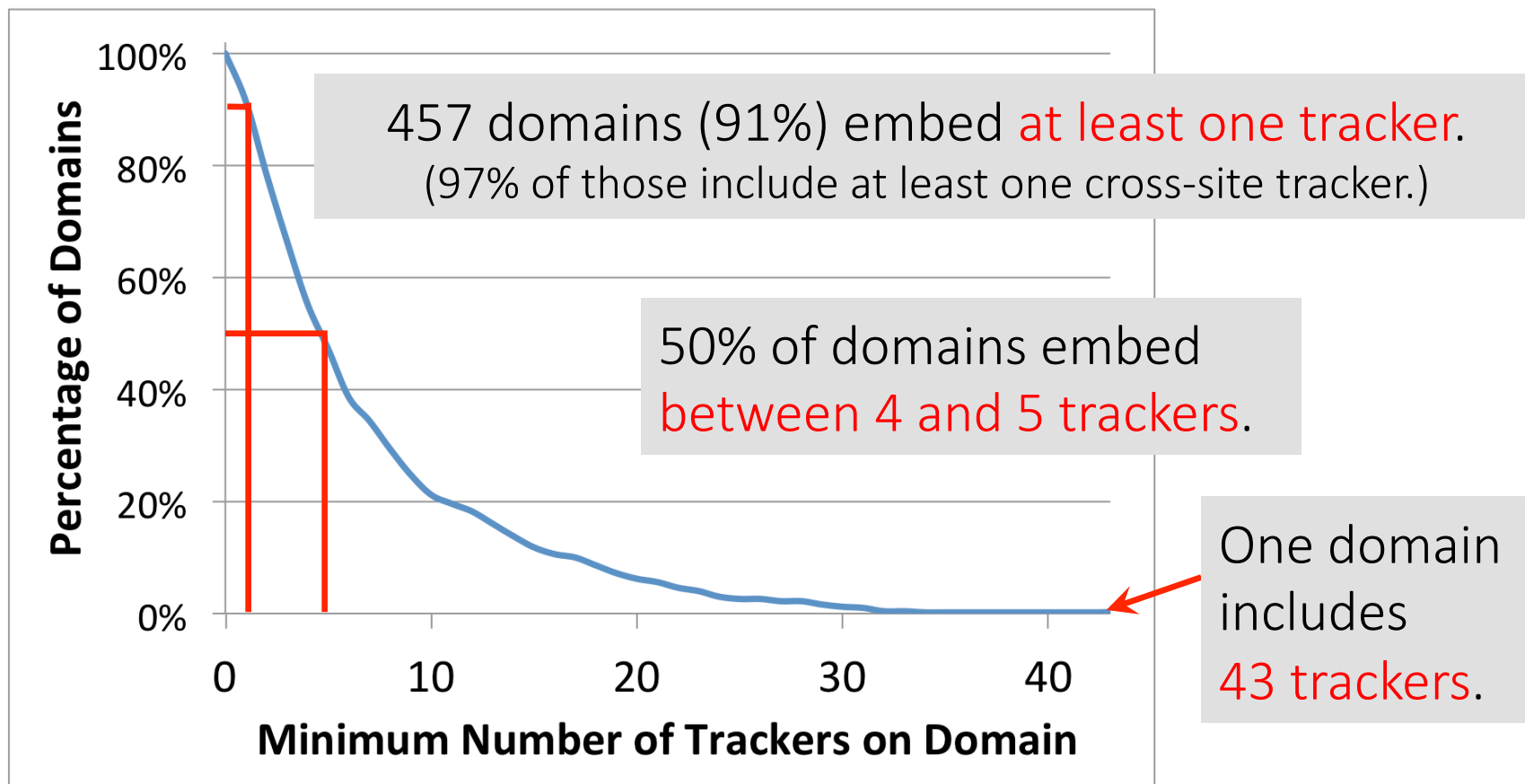
Measurement Study (2011)

- 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.

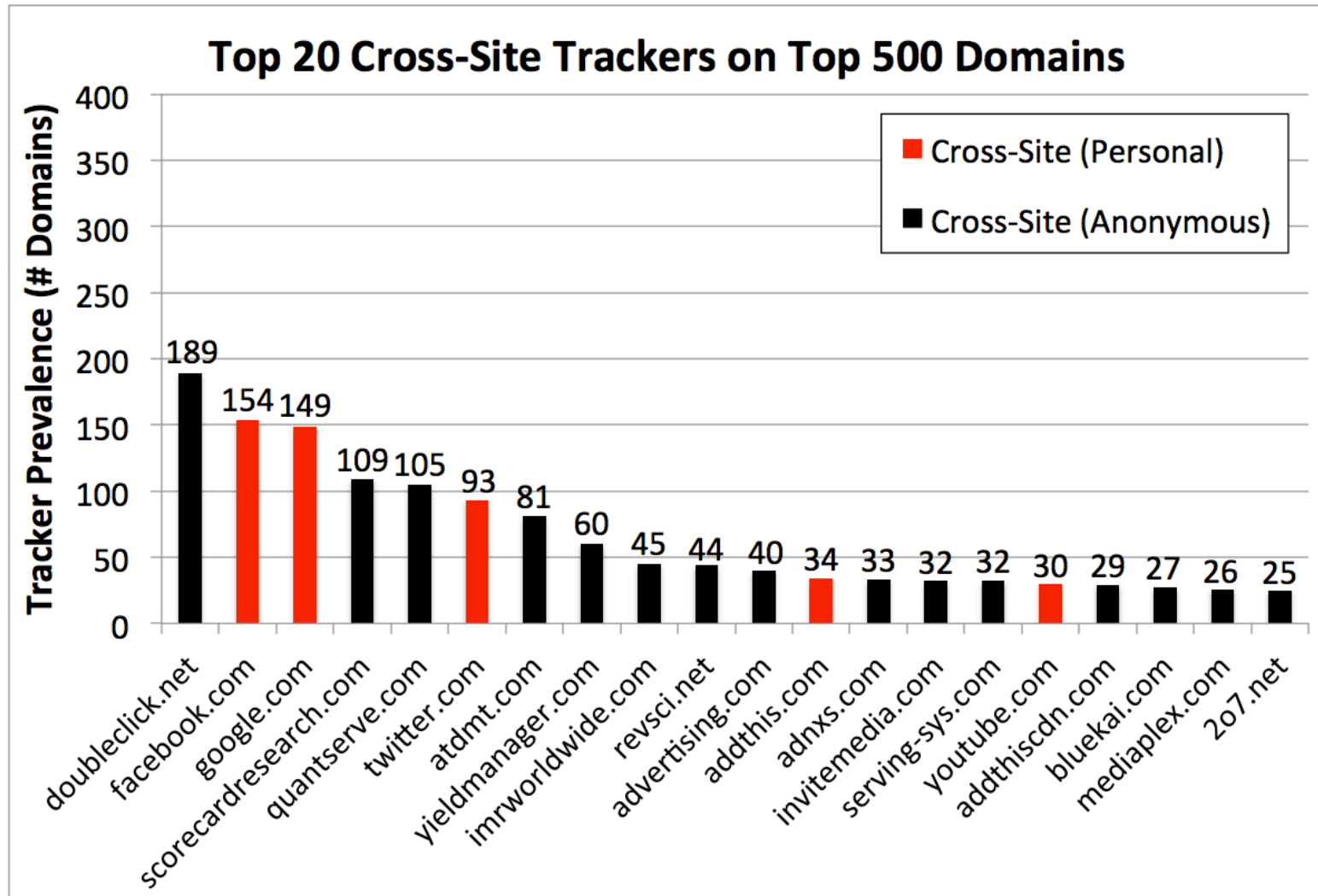
Longitudinal study in 2013 showed that the tracking ecosystem has **not substantially changed since 2011.**

How prevalent is tracking?

524 unique trackers on Alexa top 500 websites (2011).



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](#)

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

LocalStorage and Flash Cookies

- Surprisingly little use of these mechanisms!
- Of 524 trackers on Alexa Top 500:
 - Only 5 set unique identifiers in LocalStorage
 - 35 set unique identifiers in Flash cookies
- Respawning:
 - LS → Cookie: 1 case; Cookie → LS: 3 cases
 - Flash → Cookie: 6 cases; Cookie → Flash: 7 cases

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.

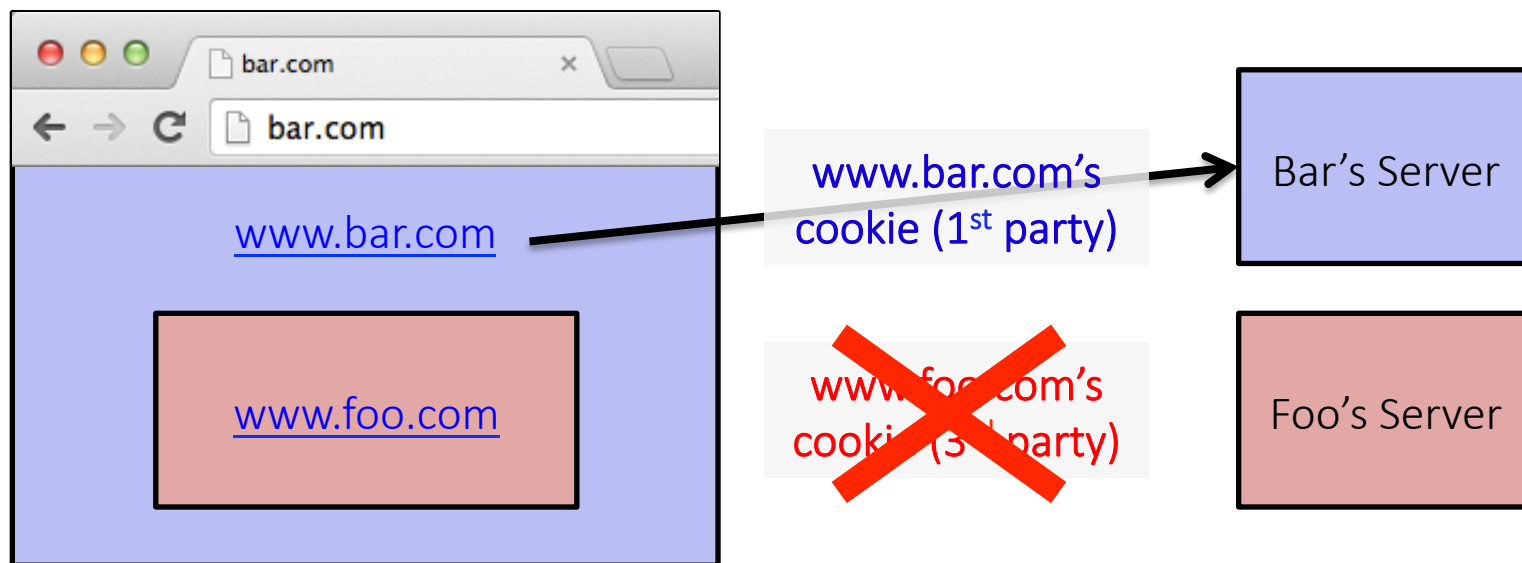
You've gone incognito. Pages you view in incognito tabs won't stick around in your browser's history, cookie store, or search history after you've closed all of your incognito tabs. Any files you download or bookmarks you create will be kept.



However, you aren't invisible. Going incognito doesn't hide your browsing from your employer, your internet service provider, or the websites you visit.

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.**

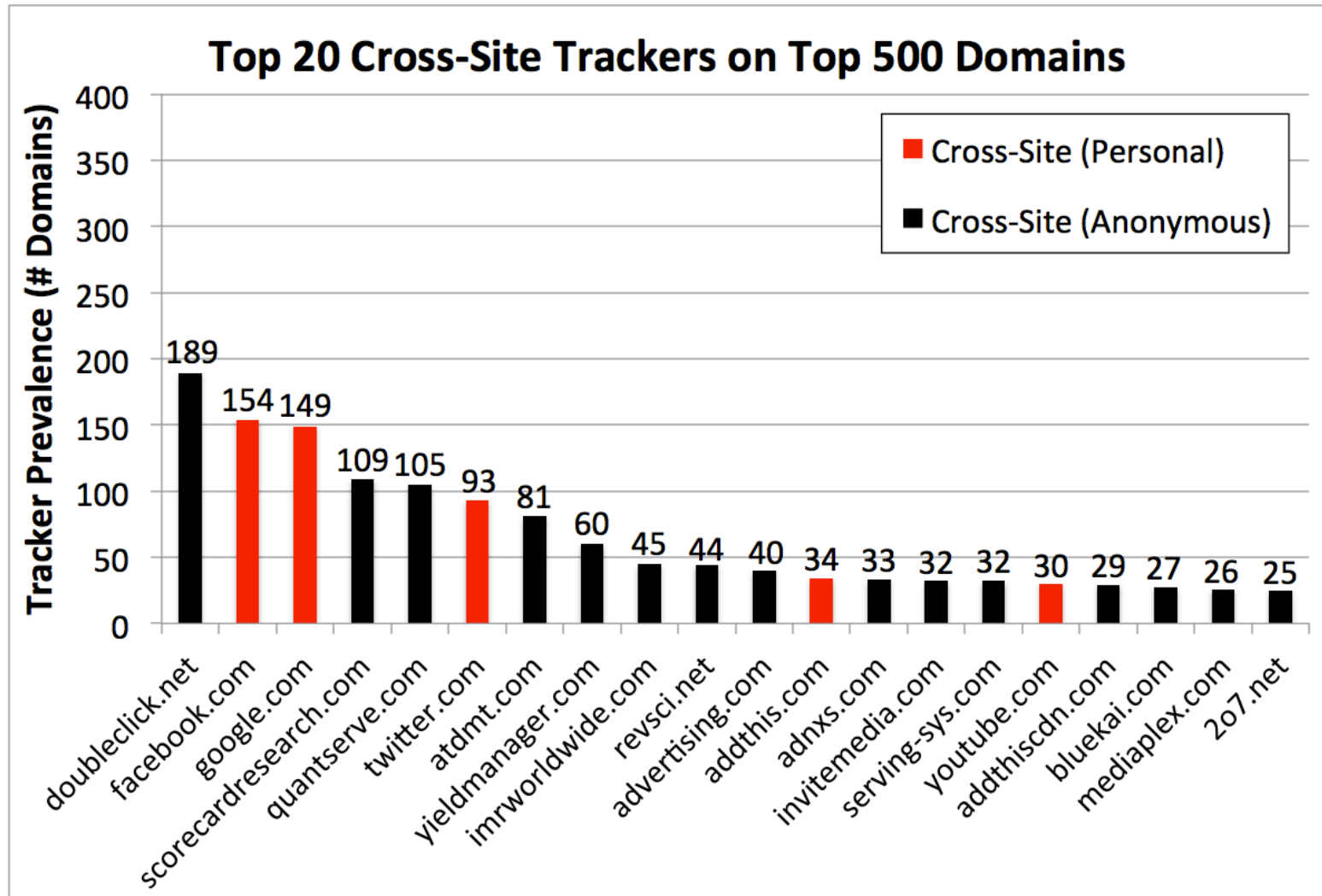
How to get a cookie set?

One way: be a first party.

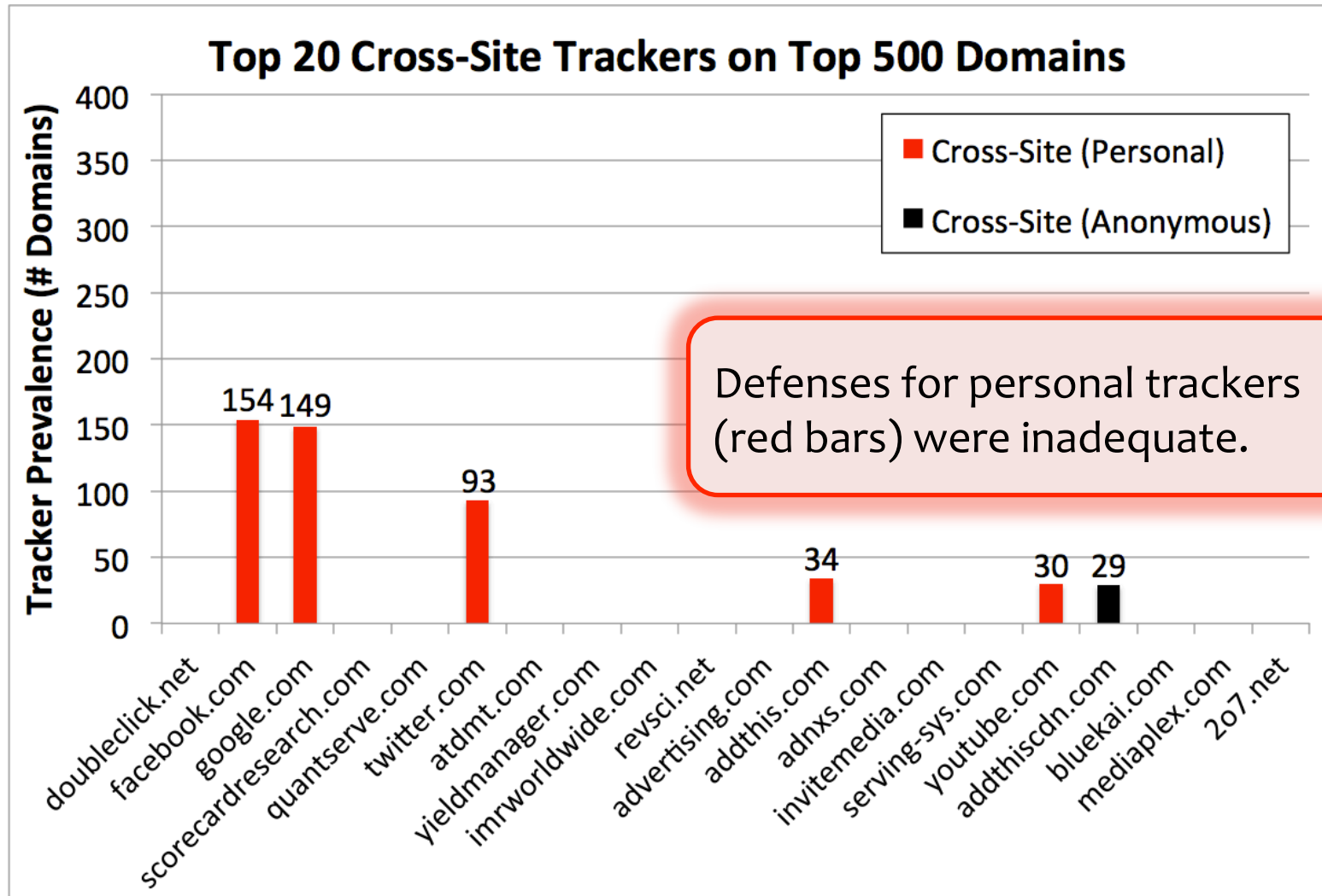


etc.

What 3rd Party Cookie Blocking Misses



What 3rd Party Cookie Blocking Misses



Our Defense: ShareMeNot



- Prior defenses for personal trackers: ineffective or completely removed social media buttons.
- Our defense:
 - ShareMeNot (for Chrome/Firefox) protects against tracking **without compromising button functionality**.
 - Blocks requests to load buttons, **replaces with local versions**. On click, shares to social media as expected.
 - Techniques adopted by **Ghostery and the EFF**.

<http://sharemenot.cs.washington.edu>

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.



Recommended

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

<https://www.eff.org/privacybadger>