

CSE 484 / CSE M 584: Computer Security and Privacy

Web Privacy: Third-Party Tracking

Fall 2017

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

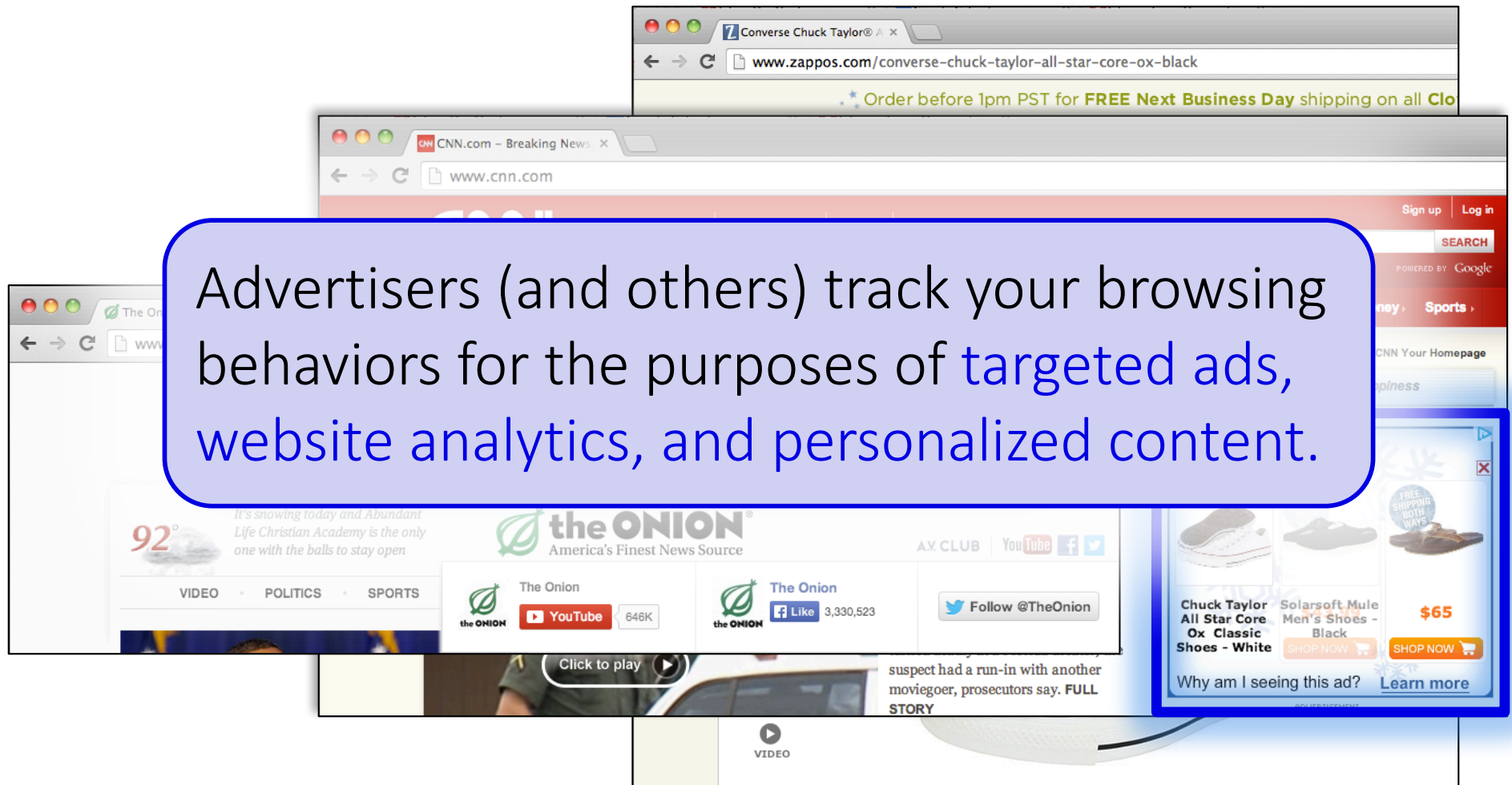
Thanks to Dan Boneh, Dieter Gollmann, Dan Halperin, Yoshi Kohno, Ada Lerner, John Manferdelli, John Mitchell, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials ...

Admin

- Guest lecture on Friday
 - David Molnar from Microsoft Research
 - Will include some background for Homework #3
- Homework #3
 - Out just before or just after Thanksgiving
 - Due Dec 8 (last day of quarter)
 - Hands-on tasks on several topics, but not a full lab
(No lab #3)

Ads That Follow You

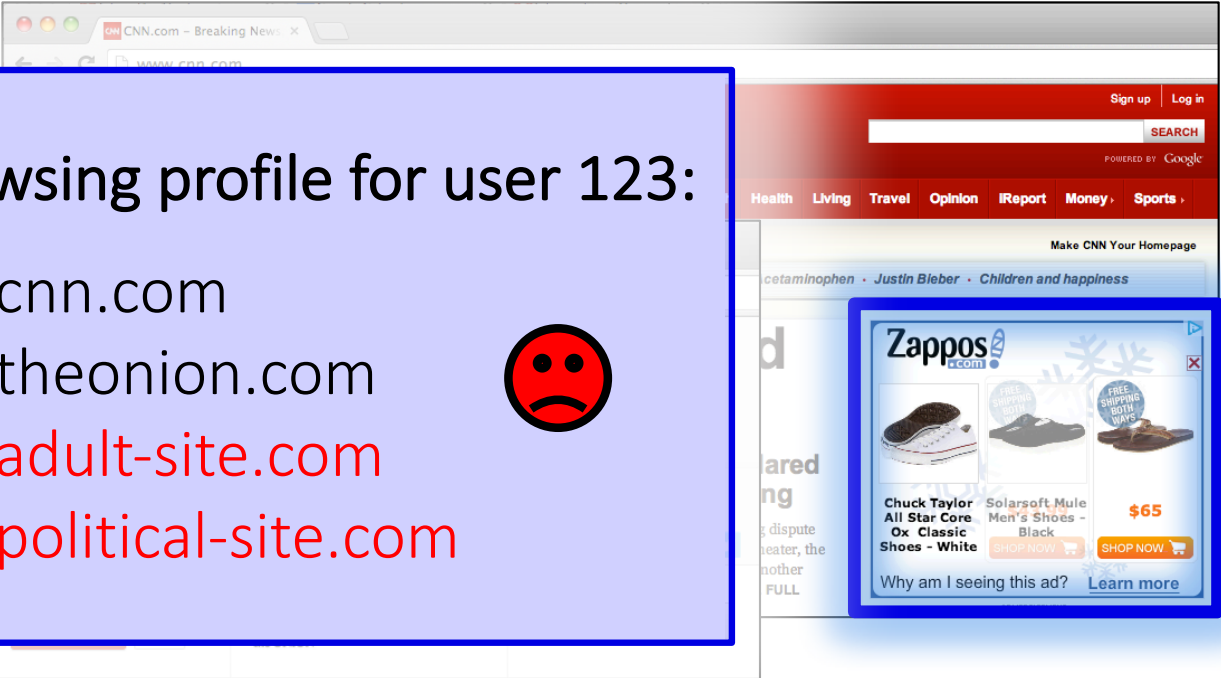
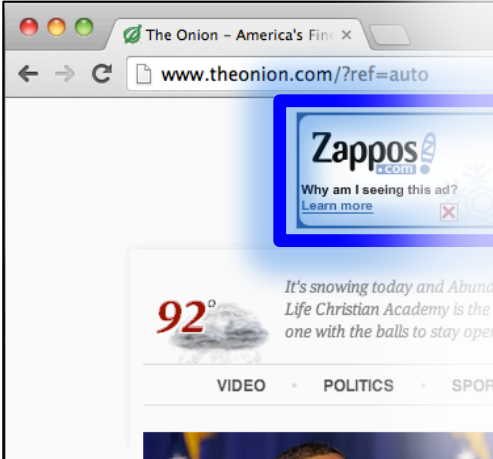

Advertisers (and others) track your browsing behaviors for the purposes of **targeted ads**, website analytics, and personalized content.



Third-Party Web Tracking

Browsing profile for user 123:

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



The image shows a central purple box with a blue border containing a list of websites visited by 'user 123'. To the right of the list is a red sad face icon. Behind the box are two browser window screenshots. The left screenshot shows theonion.com with a Zappos ad highlighted by a blue box. The right screenshot shows cnn.com with a Zappos ad also highlighted by a blue box. Both ads include the text 'Why am I seeing this ad?' and a 'Learn more' link.

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 image is a screenshot of a news article from The New York Times. The article is titled "‘Do Not Track’ Privacy Bill Appears in Congress" and is written by Tanzina Vega. It is dated May 6, 2011, at 5:01 pm and has 3 comments. The article discusses the introduction of two bills in Washington to support a Do Not Track mechanism, which would give users control over how much of their data is collected by advertisers and other online companies. The article is overlaid on a background image of The Wall Street Journal's front page from July 30, 2010, with the headline "The Web's New Gold Mine: Your Secrets".

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

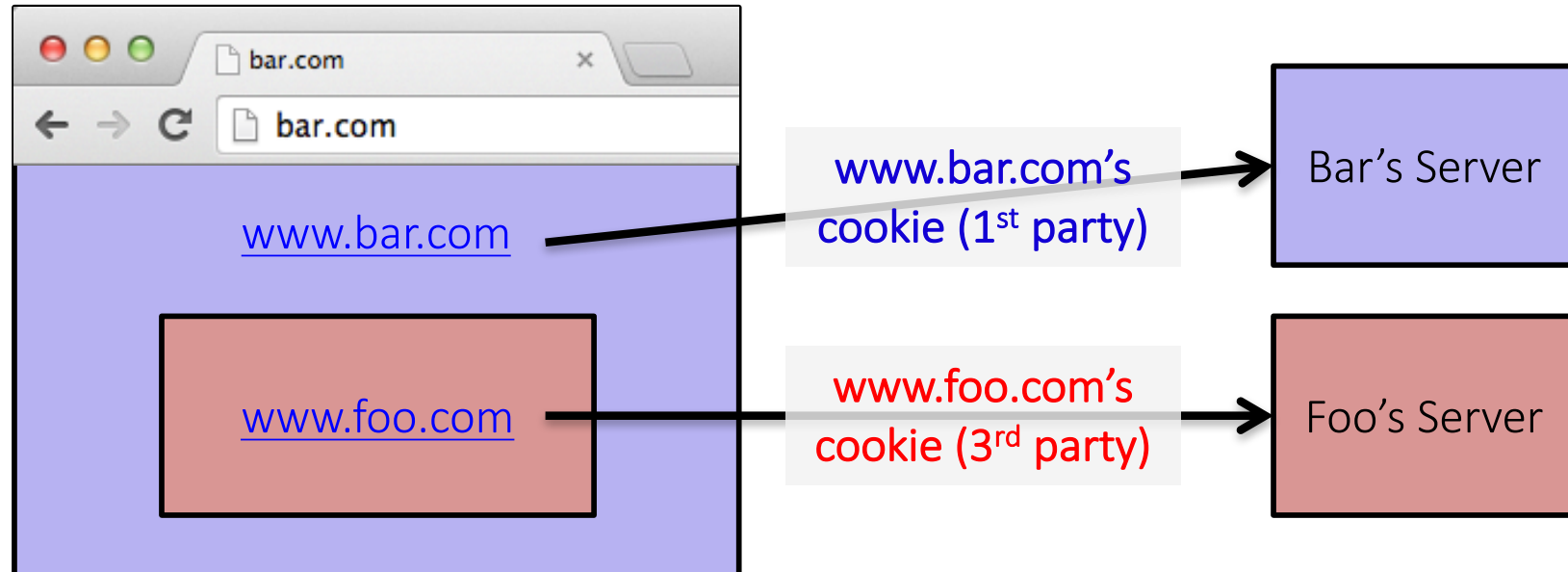
The New York Times
May 6, 2011, 5:01 pm | 3 Comments
‘Do Not Track’ Privacy Bill Appears in Congress
By TANZINA VEGA
And the privacy legislation just keeps on coming.
On Friday, two bills were introduced in Washington in support of a Do Not Track mechanism that would give users control over how much of their data was collected by advertisers and other online companies.

Outline

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

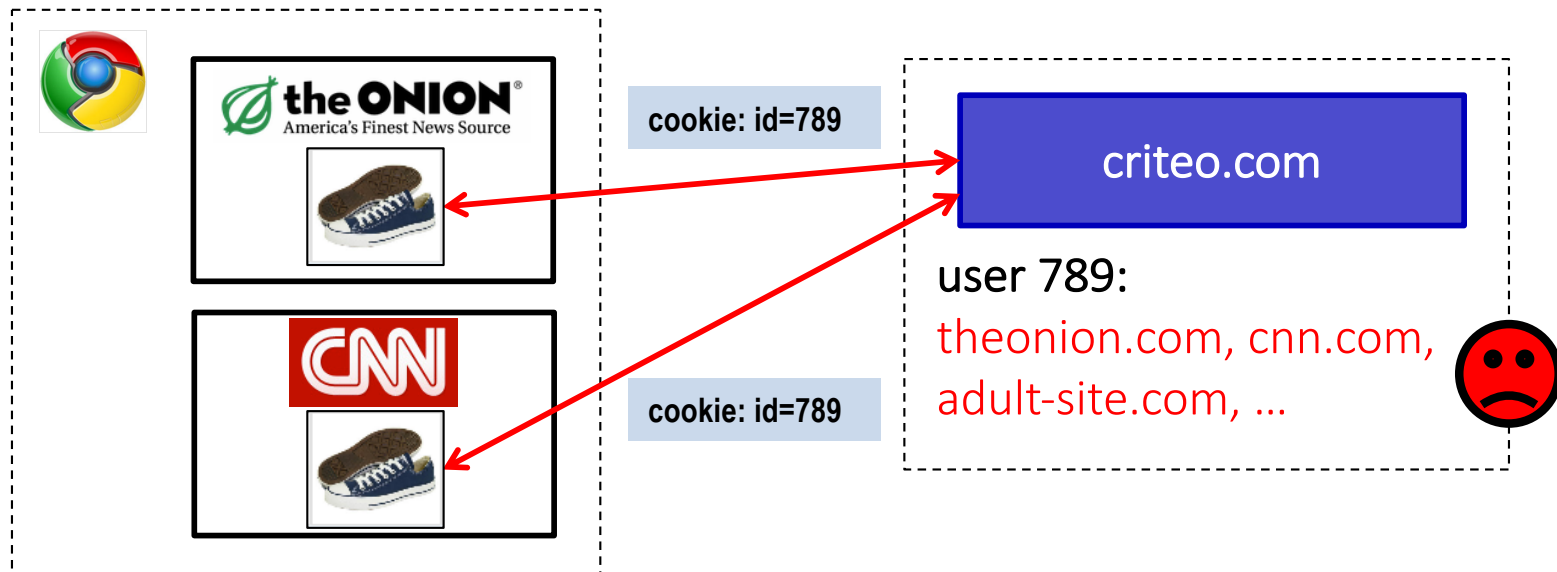
Recall: First and Third Parties

- **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.



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.](#)

Panopticlick Example

Plugin 0: Adobe Acrobat; Adobe Acrobat Plug-In Version 7.00 for Netscape; nppdf32.dll; (Acrobat Portable Document Format; application/pdf; pdf) (Acrobat Forms Data Format; application/vnd.fdf; fdf) (XML Version of Acrobat Forms Data Format; application/vnd.adobe.xfdf; xfdf) (Acrobat XML Data Package; application/vnd.adobe.xdp+xml; xdp) (Adobe FormFlow99 Data File; application/vnd.adobe.xfd+xml; xfd). Plugin 1: Adobe Acrobat; Adobe PDF Plug-In For Firefox and Netscape; nppdf32.dll; (Acrobat Portable Document Format; application/pdf; pdf) (Adobe PDF in XML Format; application/vnd.adobe.pdf+xml; pdfxml). **84% of browsers find some errors**

84% of browser fingerprints are unique

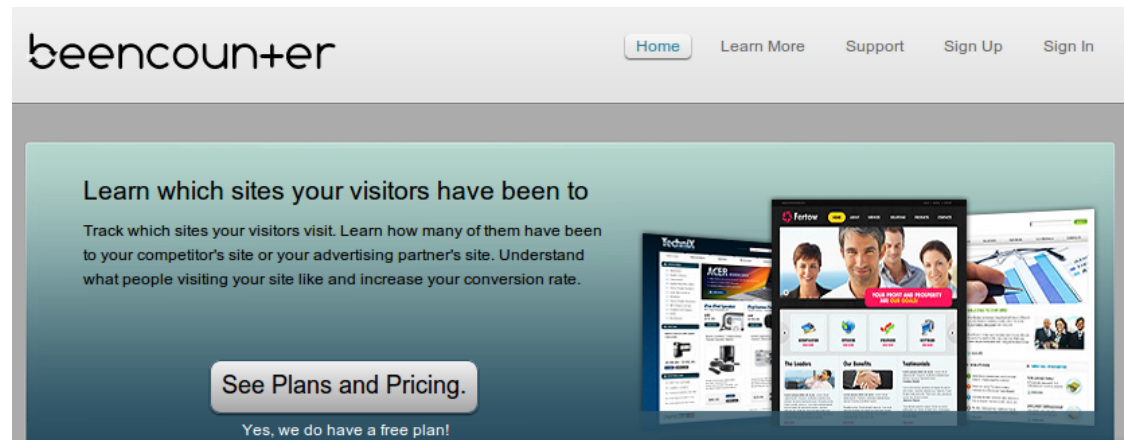
With Flash or Java, 94% are unique

npGoogleOneClick0.dll; (, application/x-vnd.google.oneclicketh.0;). Plugin 3: Microsoft® Windows Media Player Firefox Plugin; np-mswmp; np-mswmp.dll; (np-mswmp; application/x-ms-wmp; *) (; application/asx; *) (; video/x-ms-asf-plugin; *) (; application/x-mplayer2; *) (; video/x-ms-asf; asf,asx,*) (; video/x-ms-wm; wm,*) (; audio/x-ms-wma; wma,*) (; audio/x-ms-wax; wax,*) (; video/x-ms-wmv; wmv,*) (; video/x-ms-wvx; vx,*)). Plugin 4: Move Media Player; npmnqmp 07103010; npmnqmp07103010.dll; (npmnqmp; application/x-vnd.moveplayer.qm; qmx,qpl) (npmnqmp; application/x-vnd.moveplay2.qm;) (npmnqmp; application/x-vnd.movenetworks.qm;). Plugin 5: Mozilla Default Plug-in; Default Plug-in; npnul32.dll; (Mozilla Default Plug-in; *,*). Plugin 6: Shockwave Flash; Shockwave Flash 10.0 r32; NPSWF32.dll; (Adobe Flash movie; application/x-shockwave-flash; swf) (FutureSplash movie; application/futuresplash; spl). Plugin 7: Windows Genuine Advantage; 1.7.0059.0; npLegitCheckPlugin.dll; (npLegitCheckPlugin; application/WGA-plugin; *).

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



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

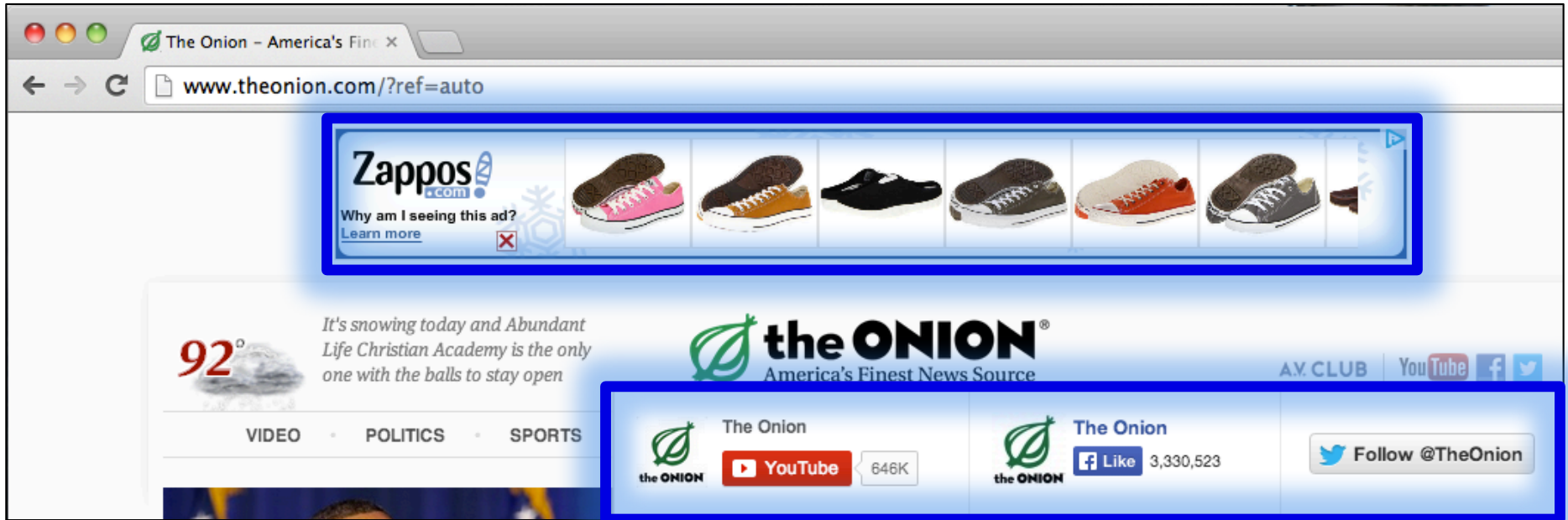
Understanding the Tracking Ecosystem

- 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.

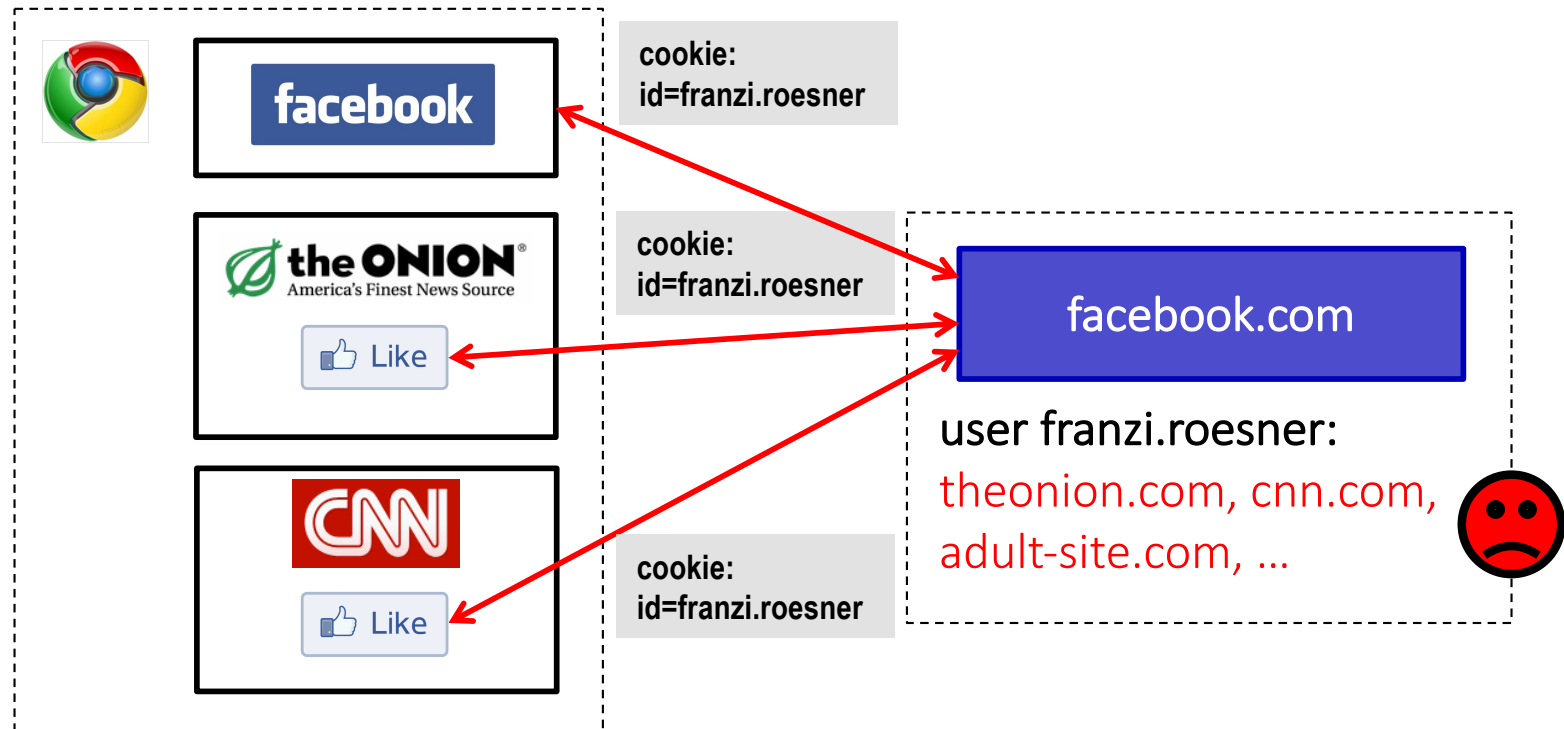
Other Trackers?



“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

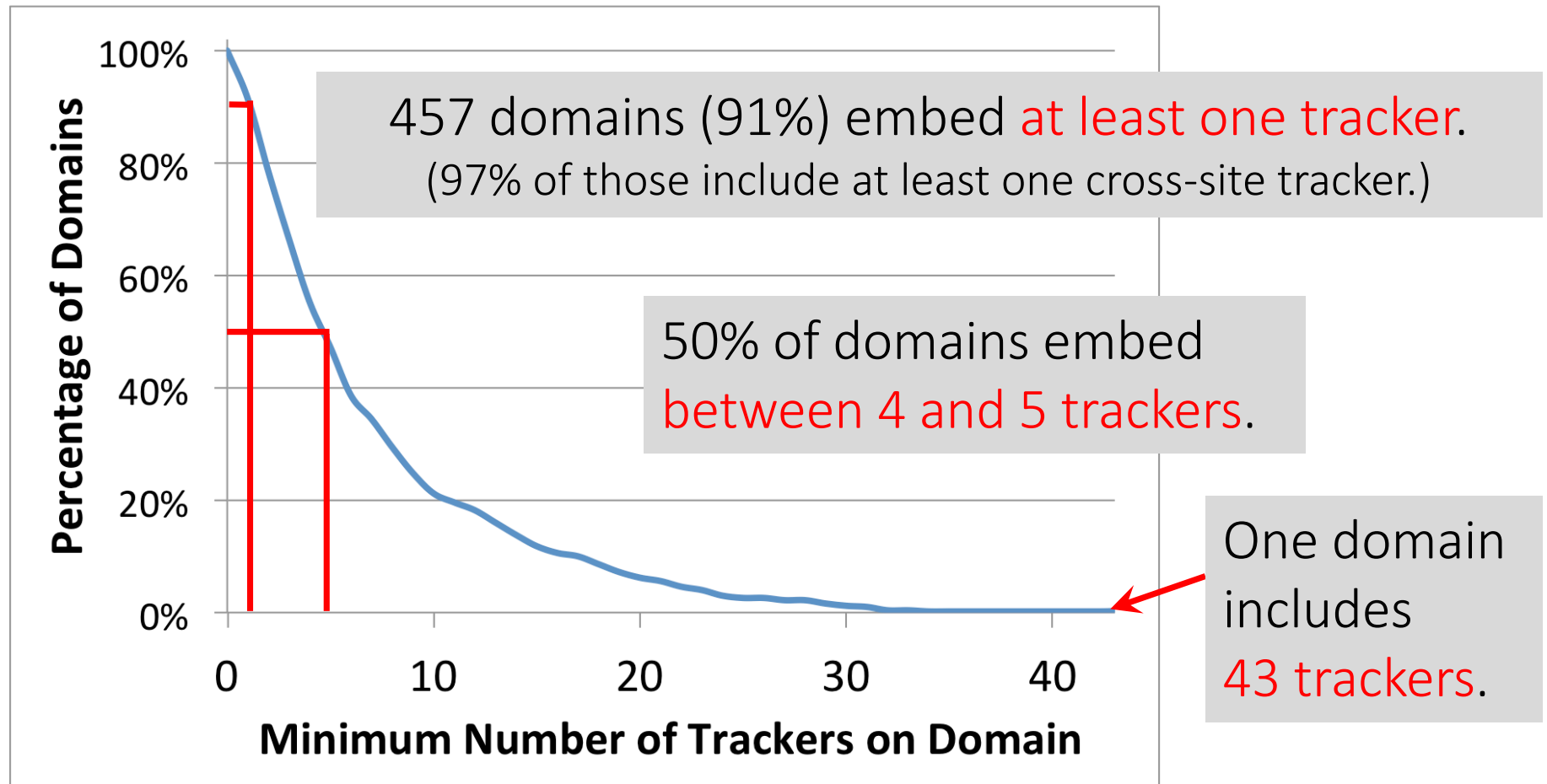
Measurement Study (2011)

- **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.

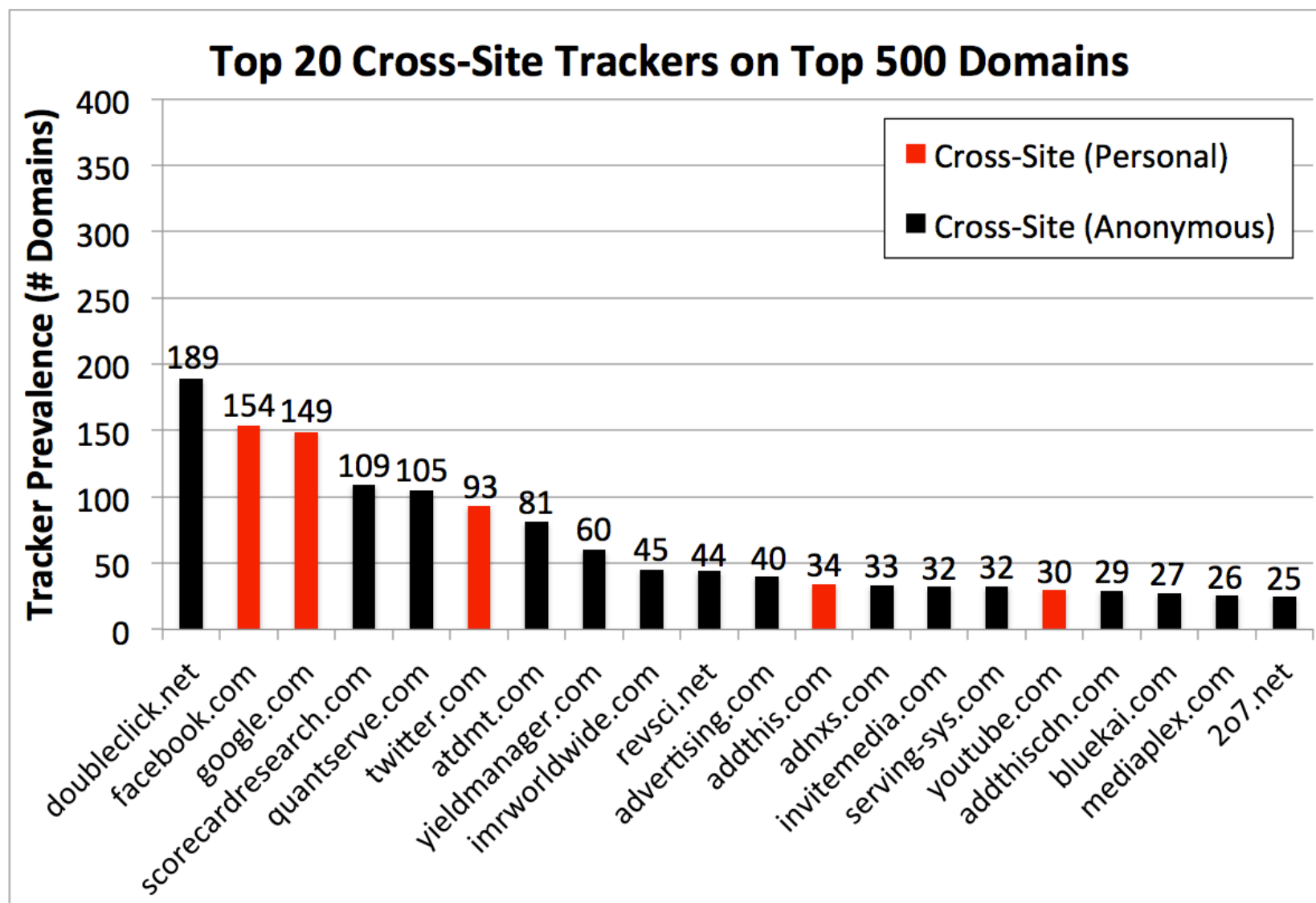
Longitudinal studies since then: **tracking has increased and become more complex.**

How prevalent is tracking?

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

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

How has this changed over time?

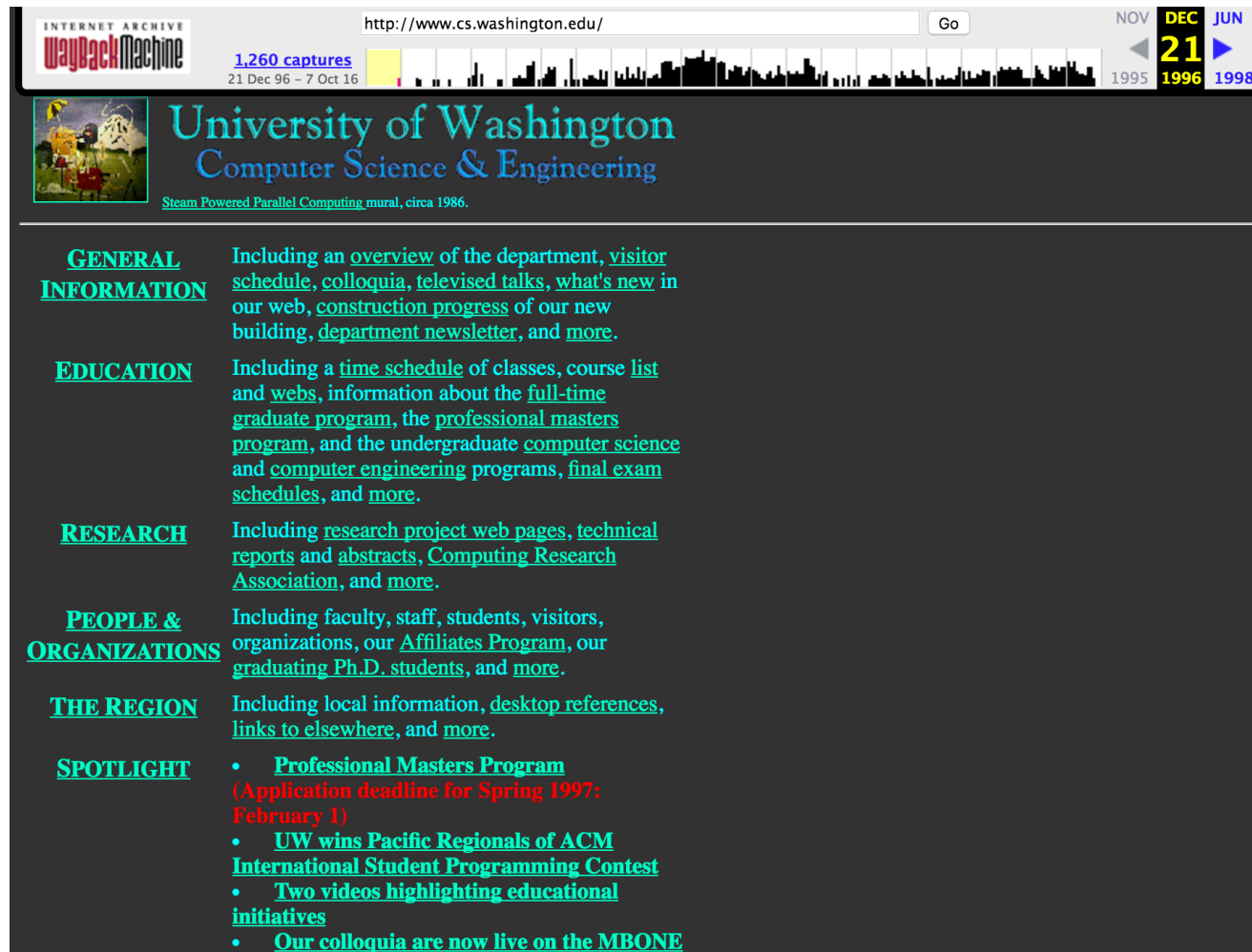
- The web has existed for a while now...
 - What about tracking before 2011? (our first study)
 - What about tracking before 2009? (first academic study)

- Solution: **time travel!**

[USENIX Security '16]



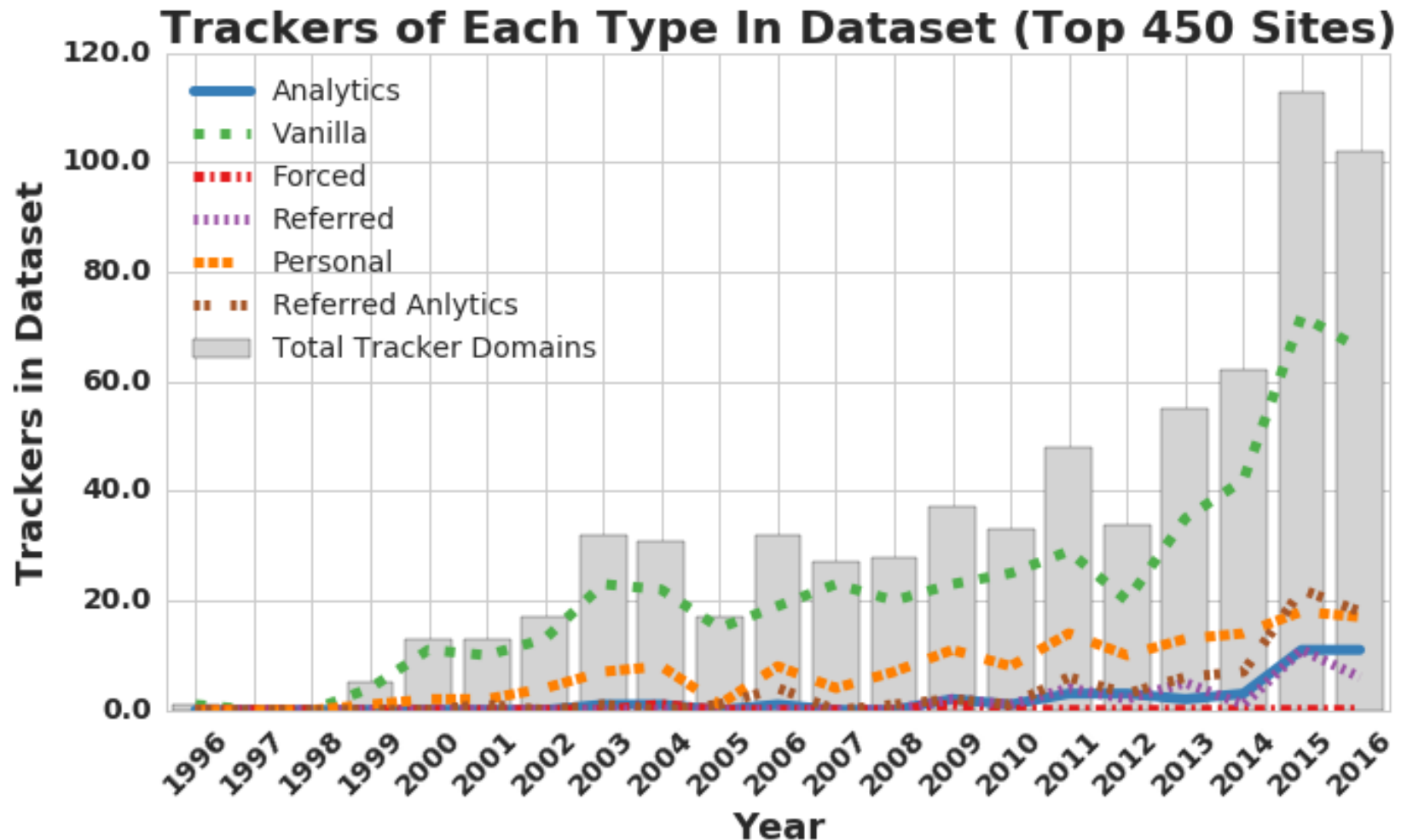
The Wayback Machine to the Rescue



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

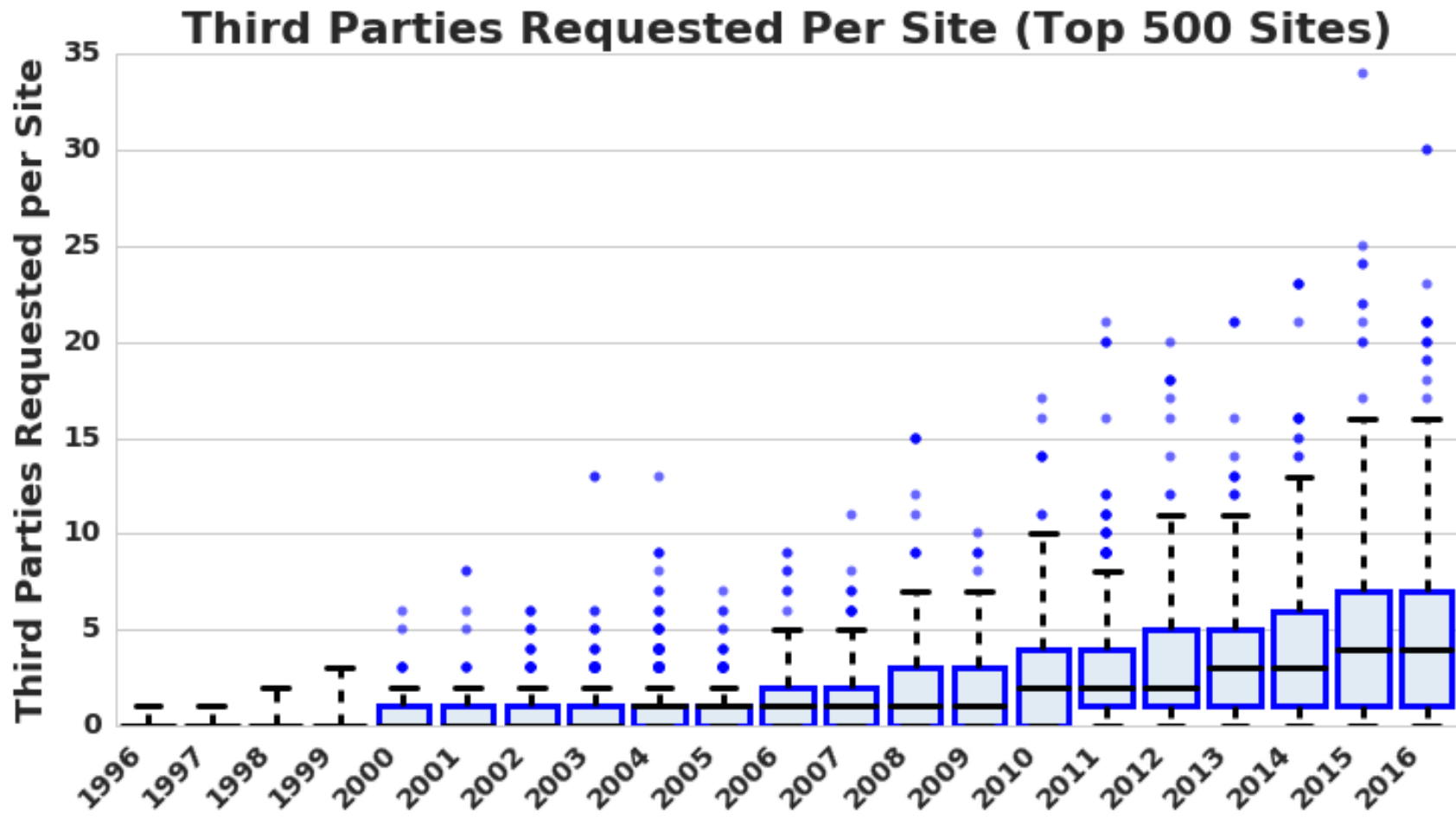
1996-2016: More & More Tracking

- More trackers of more types



1996-2016: More & More Tracking

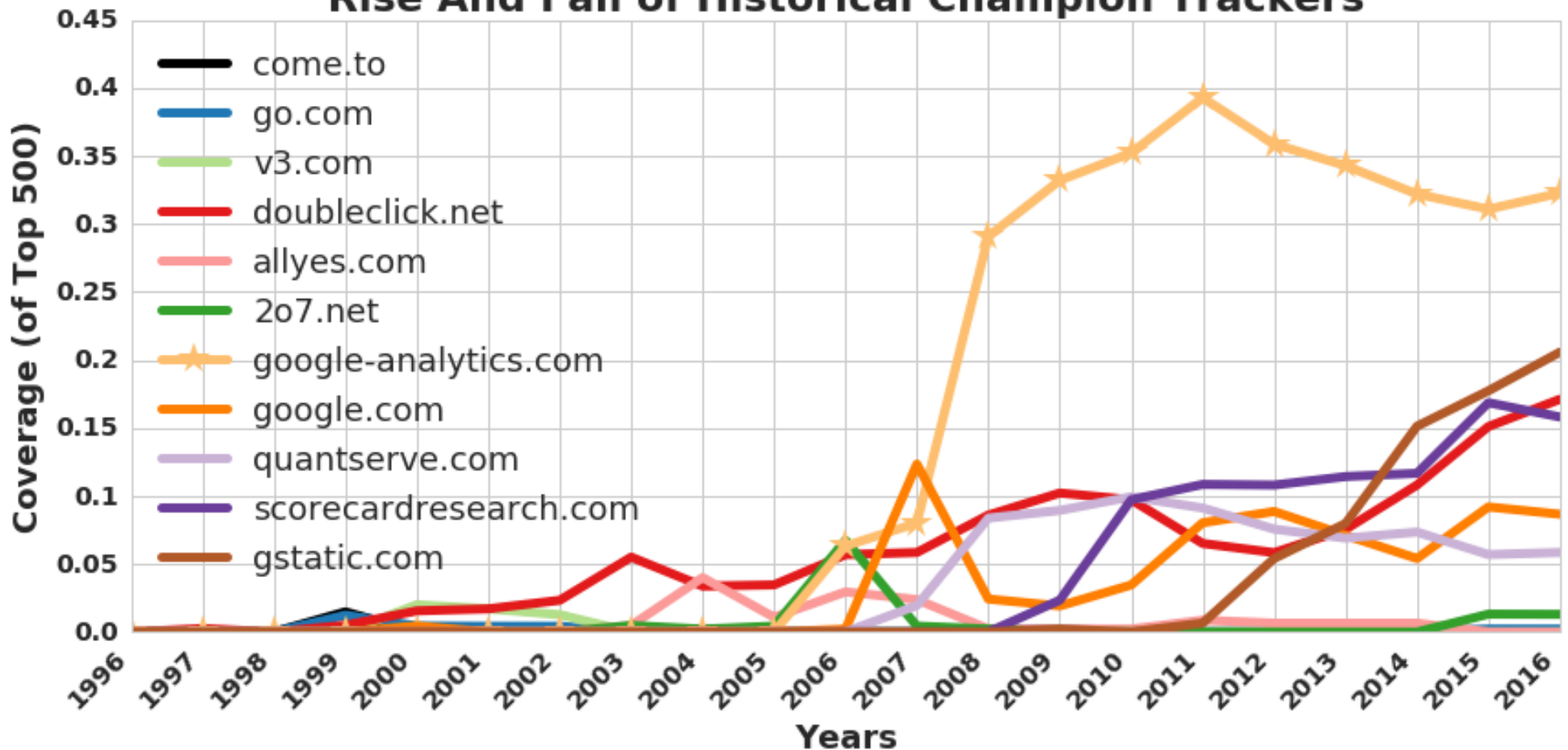
- More trackers of more types, more per site



1996-2016: More & More Tracking

- More trackers of more types, more per site, [more coverage](#)

Rise And Fall of Historical Champion Trackers




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.

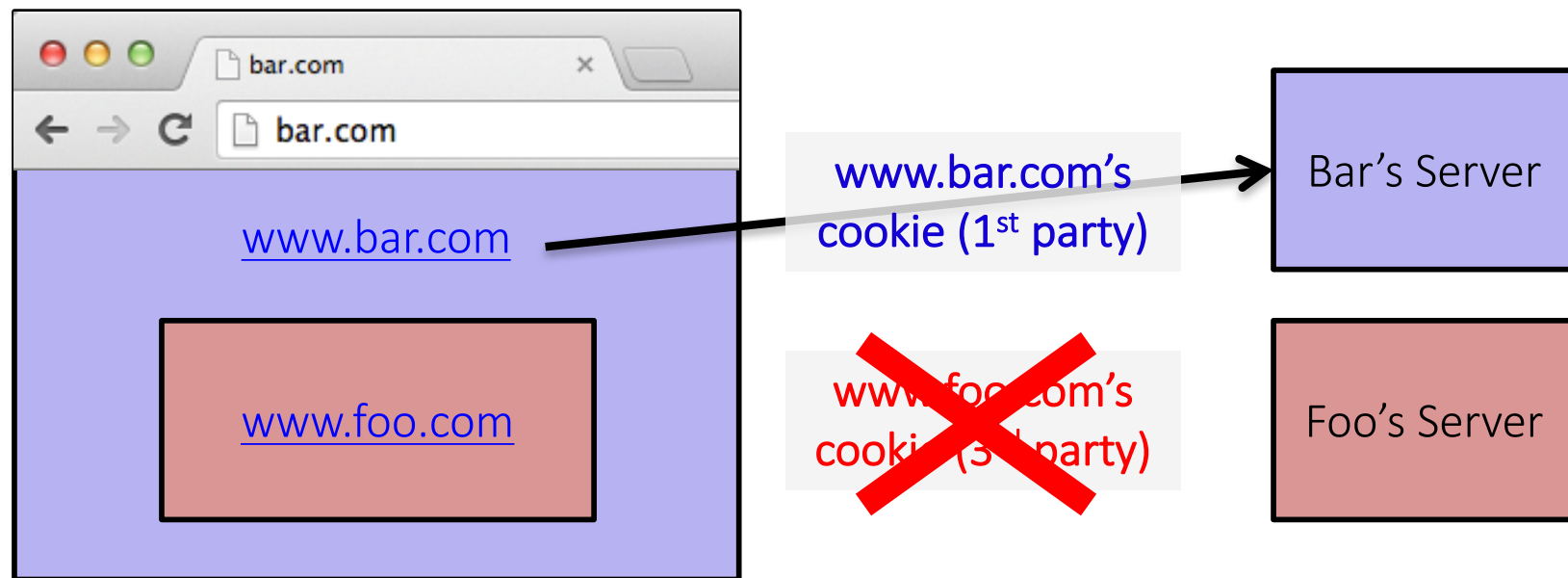
You've gone incognito. Pages you view in incognito tabs won't stick around in your browser's history, cookie store, or search history after you've closed all of your incognito tabs. Any files you download or bookmarks you create will be kept.



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

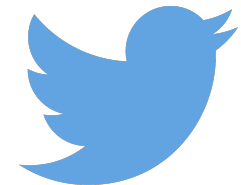
- ☒ Allow local data to be set (recommended)
- ☐ Keep local data only until I quit my browser
- ☐ Block sites from setting any data
- ☒ Block third-party cookies and site data

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

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

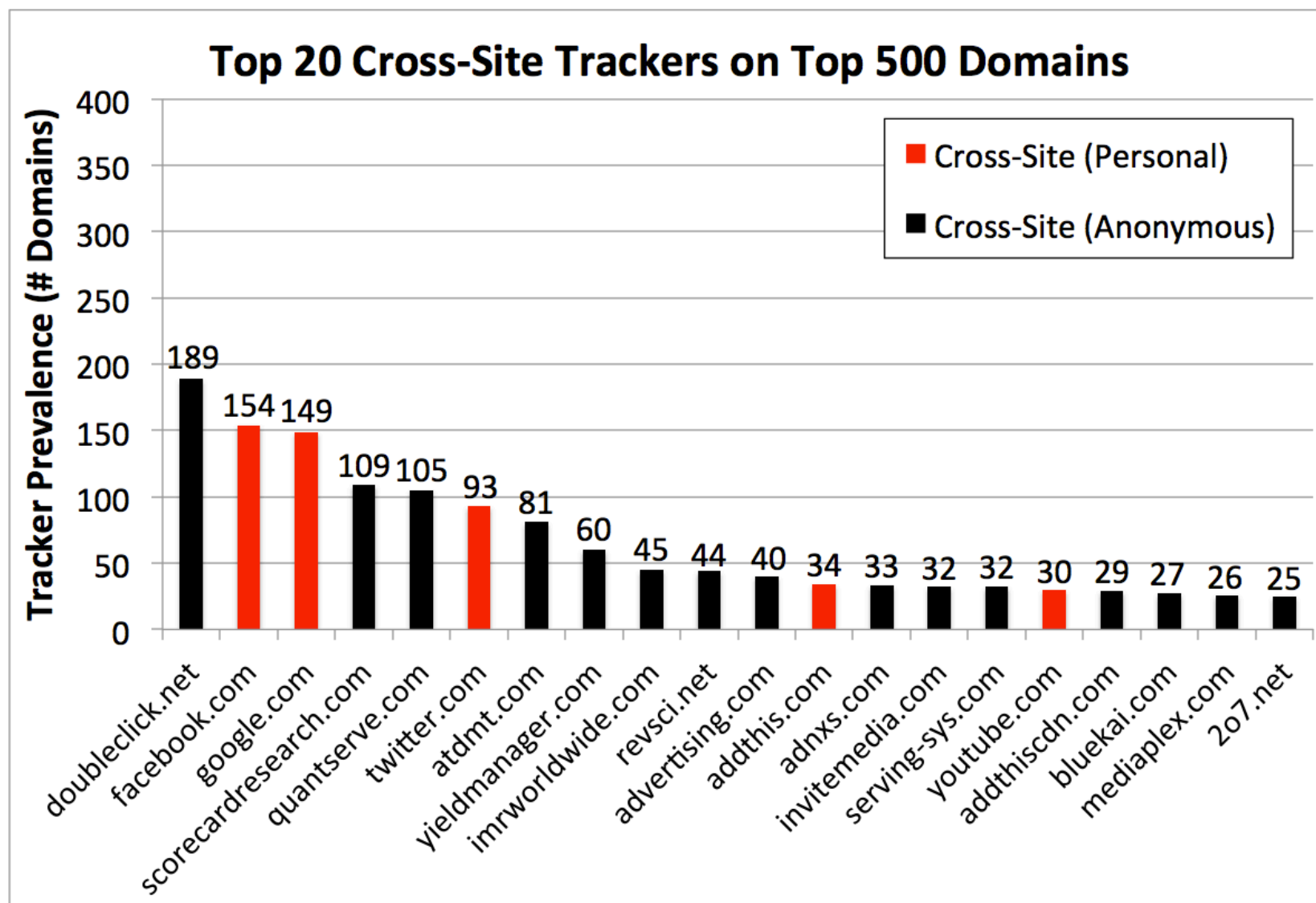
So if a third-party cookie is somehow set, **it can be used.**

How to get a cookie set?
One way: be a first party.

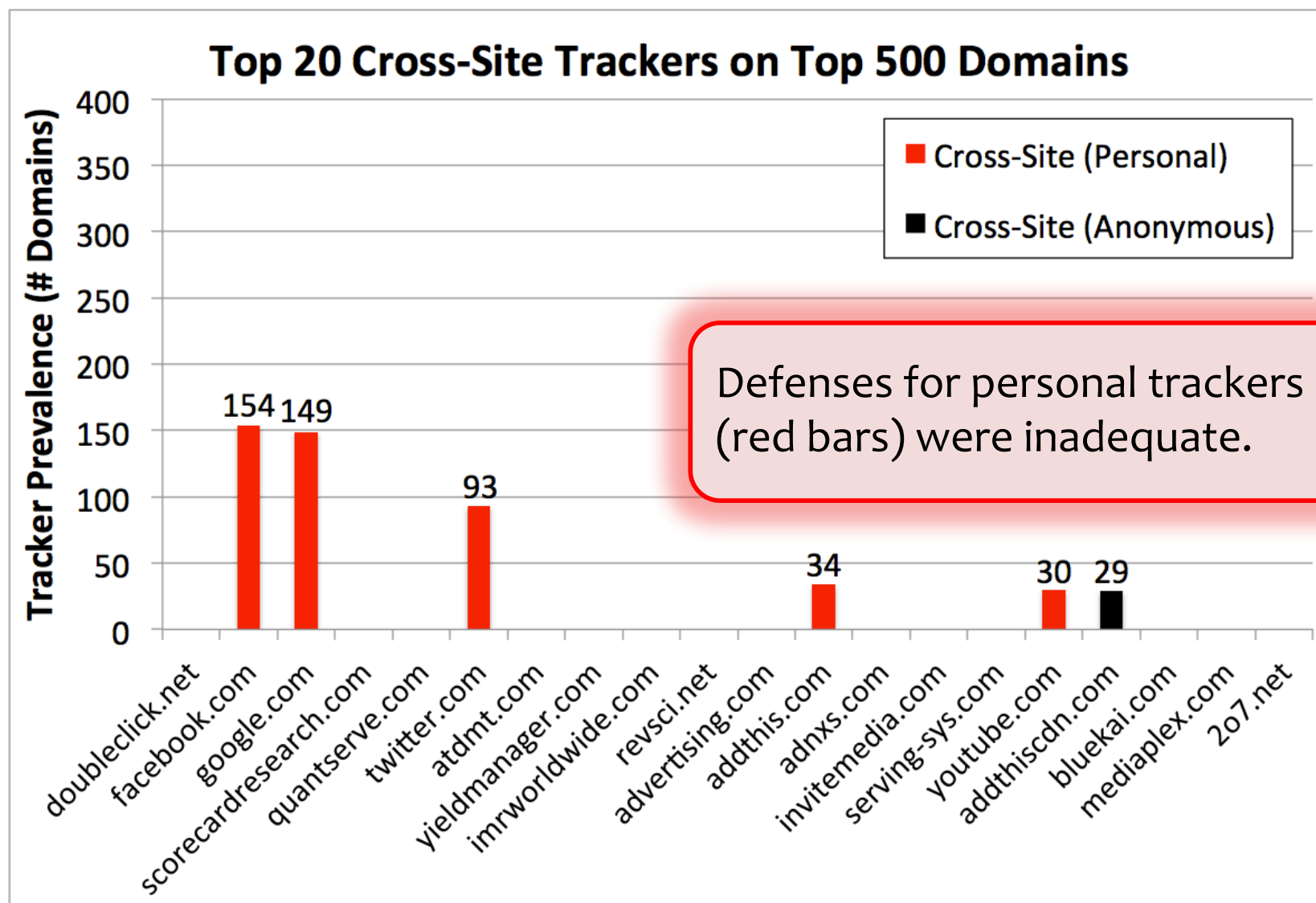


etc.

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.



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

