CSE 484 / CSE M 584: Computer Security and Privacy

# Third-Party Tracking on the Web

Spring 2017

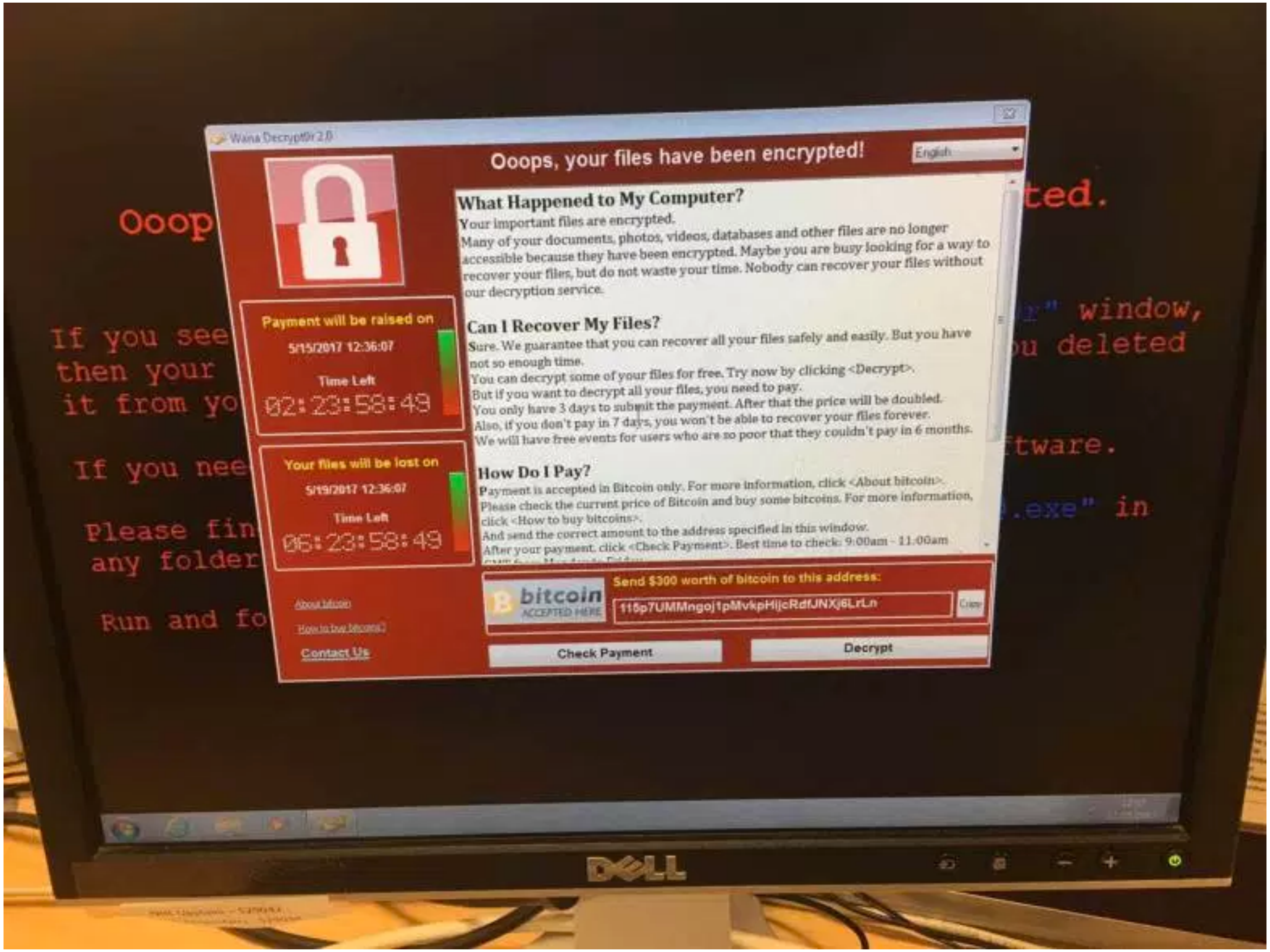
Franziska (Franzi) Roesner
franzi@cs.washington.edu

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

- Reminders
  - My office hours at 9:15am Wednesday
  - Lab #2 due Friday

# Tangent: WannaCry Ransomware



Wana Decrypt0r 2.0

Ooops, your files have been encrypted!

**What Happened to My Computer?**

Your important files are encrypted.

Many of your documents, photos, videos, databases and other files are no longer accessible because they have been encrypted. Maybe you are busy looking for a way to recover your files, but do not waste your time. Nobody can recover your files without our decryption service.

**Can I Recover My Files?**

Sure. We guarantee that you can recover all your files safely and easily. But you have not so enough time.

You can decrypt some of your files for free. Try now by clicking <Decrypt>.

But if you want to decrypt all your files, you need to pay.

You only have 3 days to submit the payment. After that the price will be doubled.

Also, if you don't pay in 7 days, you won't be able to recover your files forever.

We will have free events for users who are so poor that they couldn't pay in 6 months.

**How Do I Pay?**

Payment is accepted in Bitcoin only. For more information, click <About bitcoin>.

Please check the current price of Bitcoin and buy some bitcoins. For more information, click <How to buy bitcoins>.

And send the correct amount to the address specified in this window.

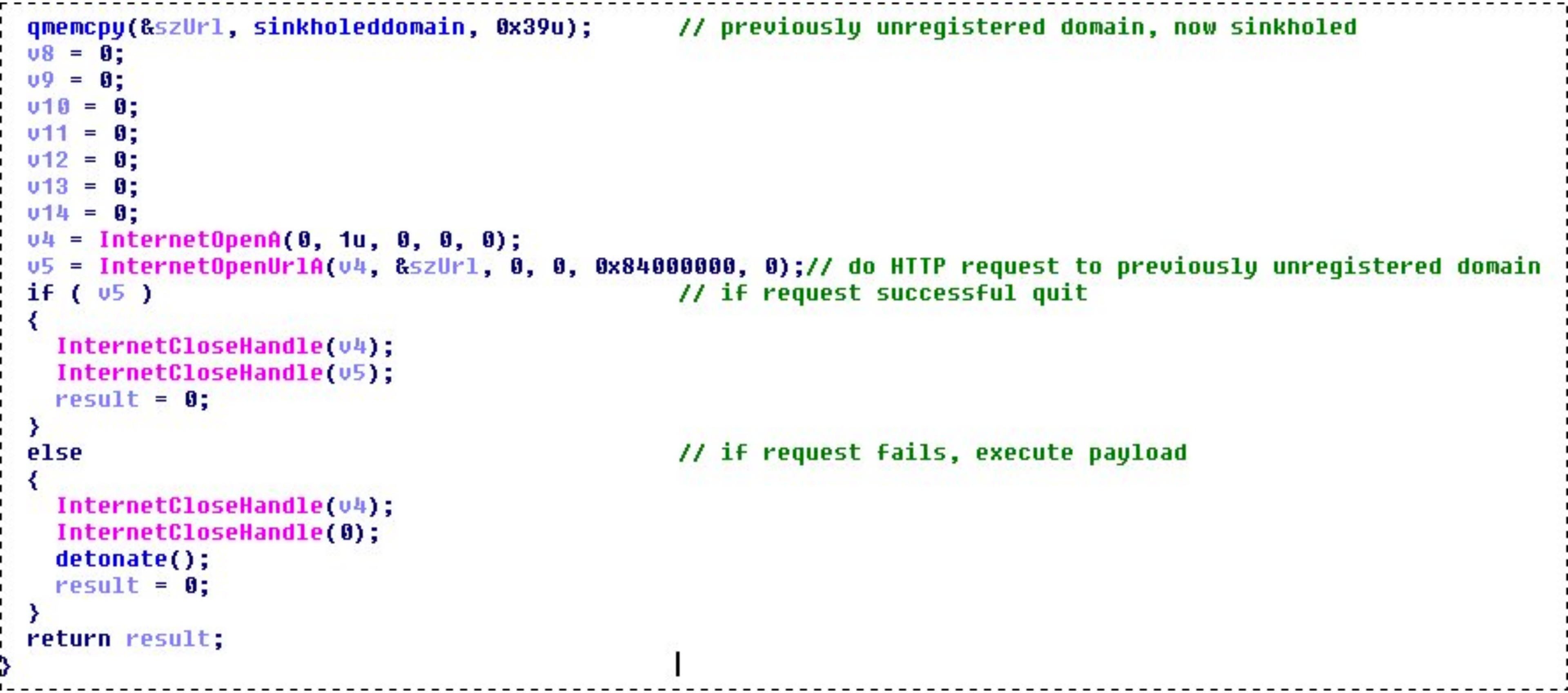
After your payment, click <Check Payment>. Best time to check: 9:00am - 11:00am

Payment will be raised on
5/15/2017 12:36:07
Time Left
02:23:58:49

Your files will be lost on
5/19/2017 12:36:07
Time Left
06:23:58:49

Send $300 worth of bitcoin to this address:
115p7UMMngoj1pMvkpHijcRdfJNXj6LrLn

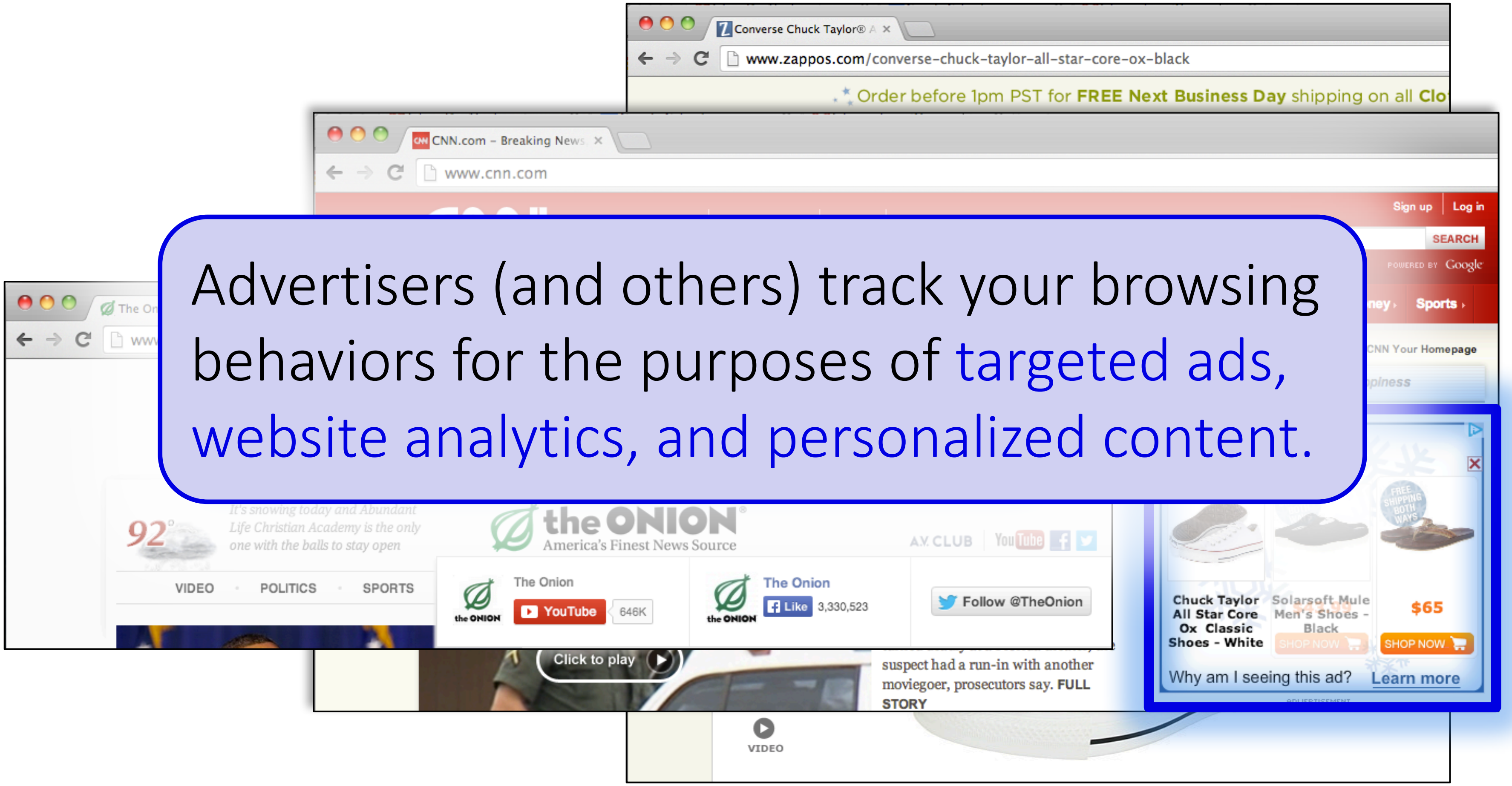
# Tangent: WannaCry Ransomware



```
qmemcpy(&szUrl, sinkholeddomain, 0x39u);      // previously unregistered domain, now sinkholed
v8 = 0;
v9 = 0;
v10 = 0;
v11 = 0;
v12 = 0;
v13 = 0;
v14 = 0;
v4 = InternetOpenA(0, 1u, 0, 0, 0);
v5 = InternetOpenUrlA(v4, &szUrl, 0, 0, 0x84000000, 0);// do HTTP request to previously unregistered domain
if ( v5 )                                     // if request successful quit
{
  InternetCloseHandle(v4);
  InternetCloseHandle(v5);
  result = 0;
}
else                                          // if request fails, execute payload
{
  InternetCloseHandle(v4);
  InternetCloseHandle(0);
  detonate();
  result = 0;
}
return result;
}
```

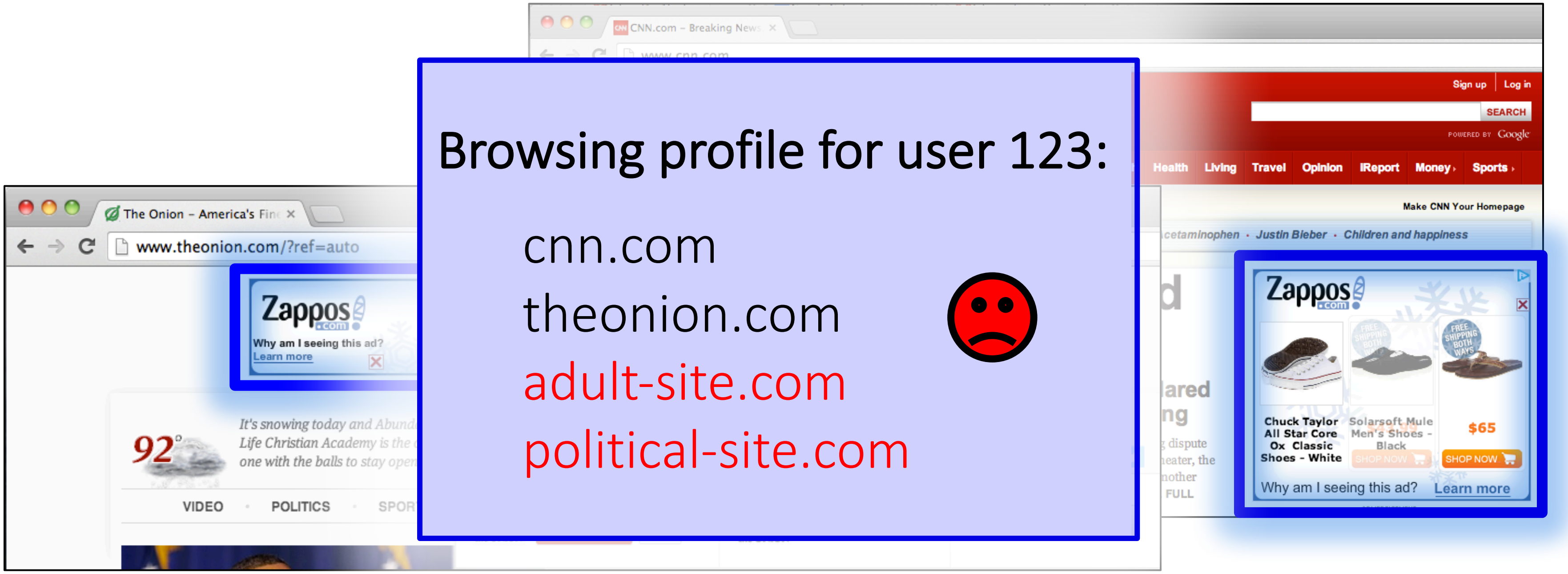
https://www.malwaretech.com/2017/05/how-to-accidentally-stop-a-global-cyber-attacks.html

# Ads That Follow You


Converse Chuck Taylor®
www.zappos.com/converse-chuck-taylor-all-star-core-ox-black
Order before 1pm PST for FREE Next Business Day shipping on all Clo
CNN.com – Breaking News
www.cnn.com
Sign up
Log in
SEARCH
POWERED BY Google
Sports
CNN Your Homepage
The Onion
92°
It's snowing today and Abundant Life Christian Academy is the only one with the balls to stay open
the ONION
America's Finest News Source
A.V. CLUB
VIDEO
POLITICS
SPORTS
YouTube
646K
Like
3,330,523
Follow @TheOnion
Click to play
suspect had a run-in with another moviegoer, prosecutors say. FULL STORY
VIDEO
FREE SHIPPING BOTH WAYS
Chuck Taylor All Star Core Ox Classic Shoes - White
Solarsoft Mule Men's Shoes - Black
$65
SHOP NOW
Why am I seeing this ad?
Learn more

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

# Third-Party Web Tracking

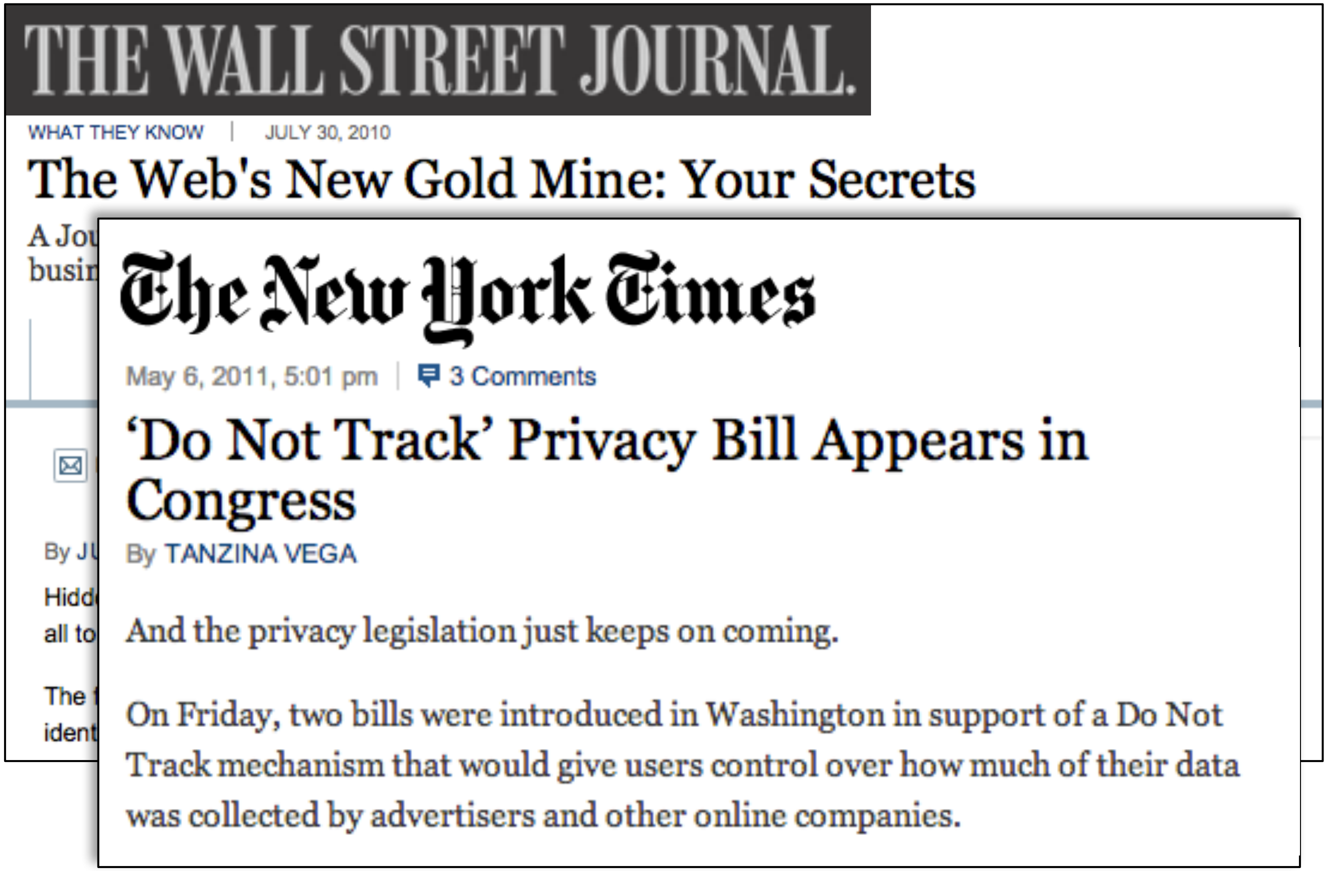


Browsing profile for user 123:

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

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 Jou
busi

By J
Hidd
all to

The
ident

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.

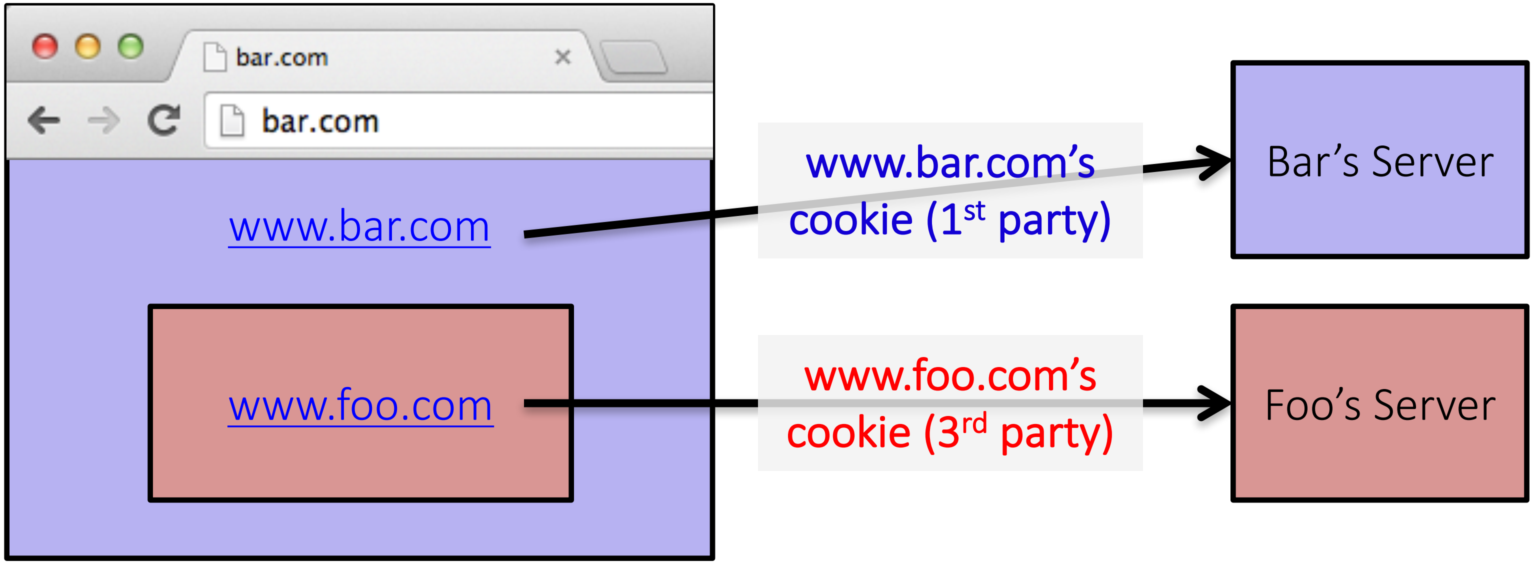
On Friday, two bills were introduced in Washington in support of a Do Not Track mechanism that would give users control over how much of their data was collected by advertisers and other online companies.

# Outline

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

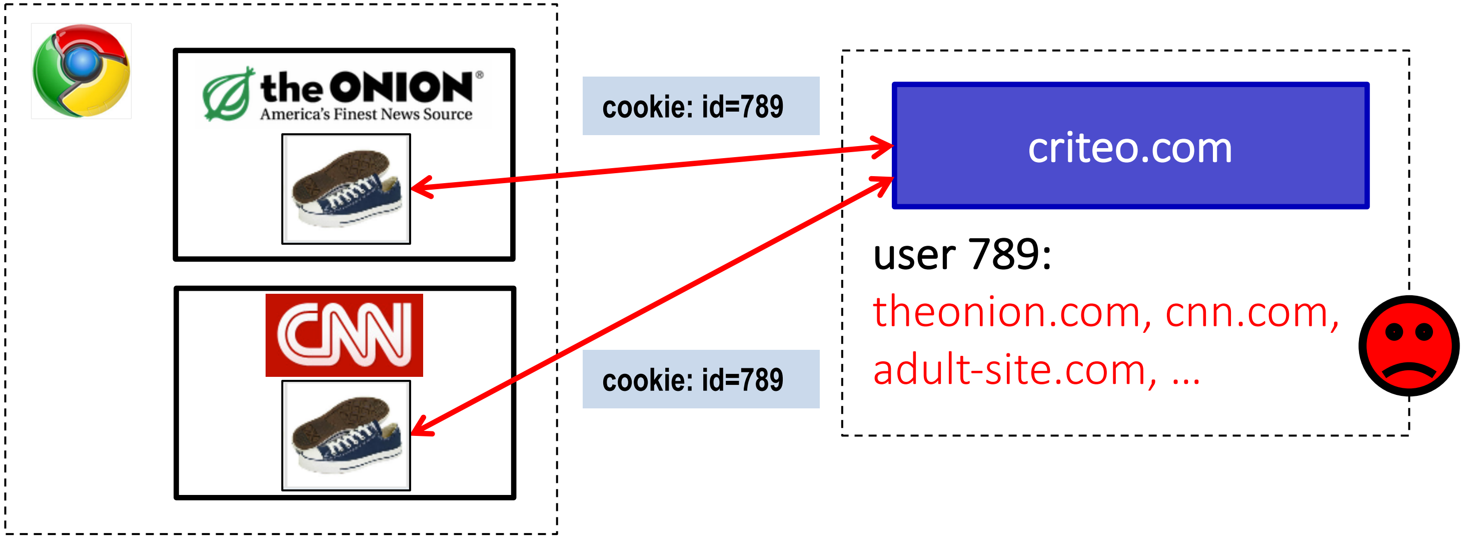
# Recall: First and Third Parties

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


bar.com
bar.com
www.bar.com
www.foo.com
www.bar.com's cookie (1st party)
www.foo.com's cookie (3rd party)
Bar's Server
Foo's Server

# Anonymous Tracking

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

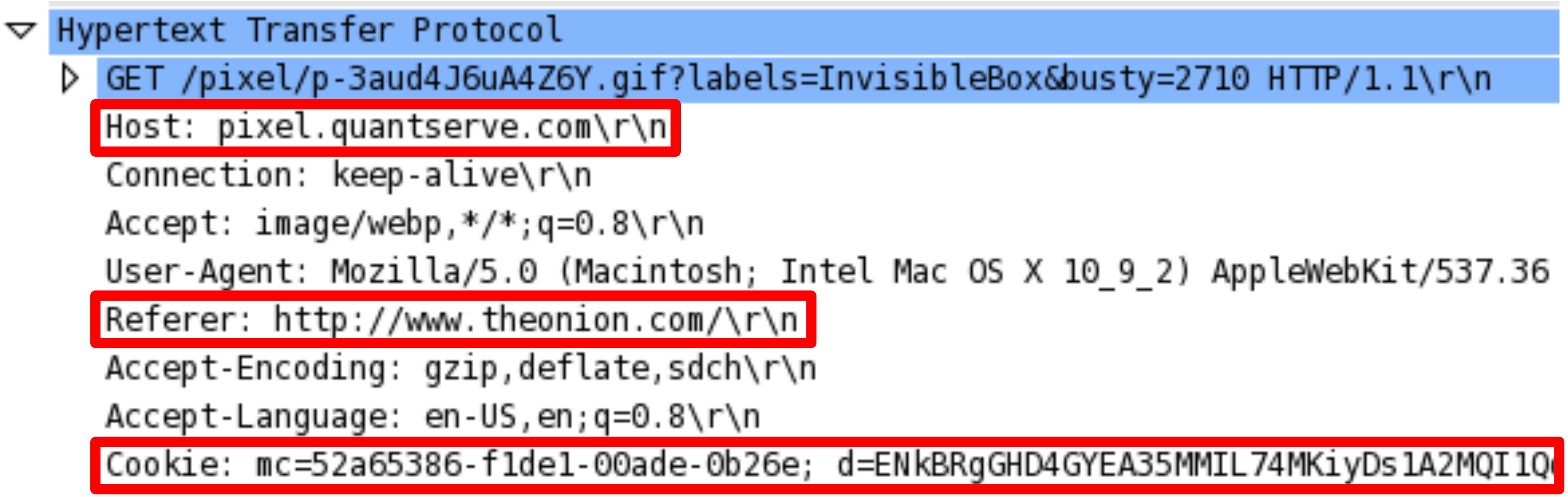


user 789:

theonion.com, cnn.com, adult-site.com, …

# 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

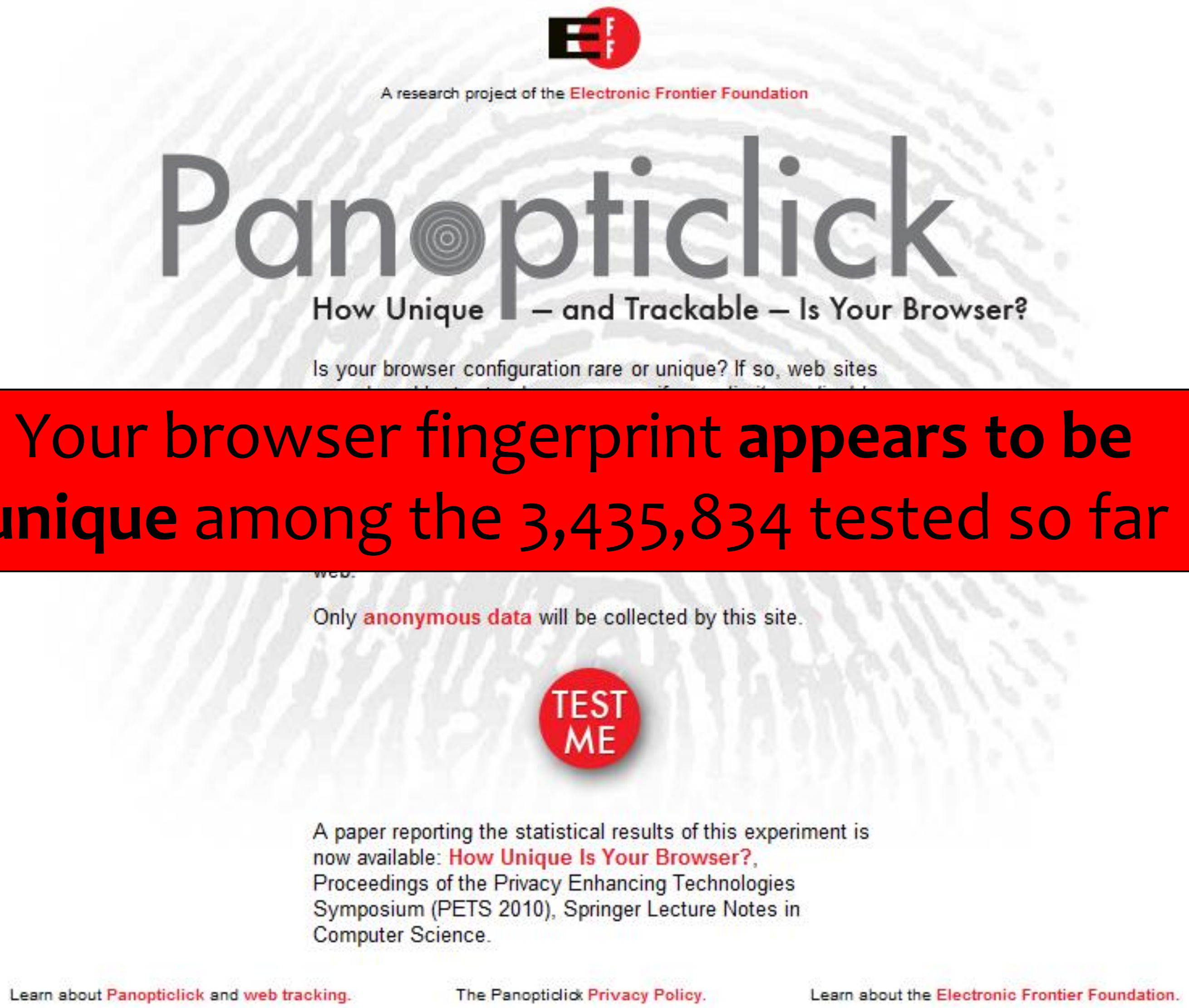
- Flash cookies
- Silverlight storage
- TLS session ID & resume
- Browsing history
- window.name
- HTTP STS
- DNS cache

- "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

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.

# 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; hars)
(Acroba
applicat
FormFlc
npGoogleOneClick8.dll; (; application/x-vnd.google.oneclickctrl.8; ). Plugin 3: MicrosoftÂ® 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; wvx,*). 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; *).

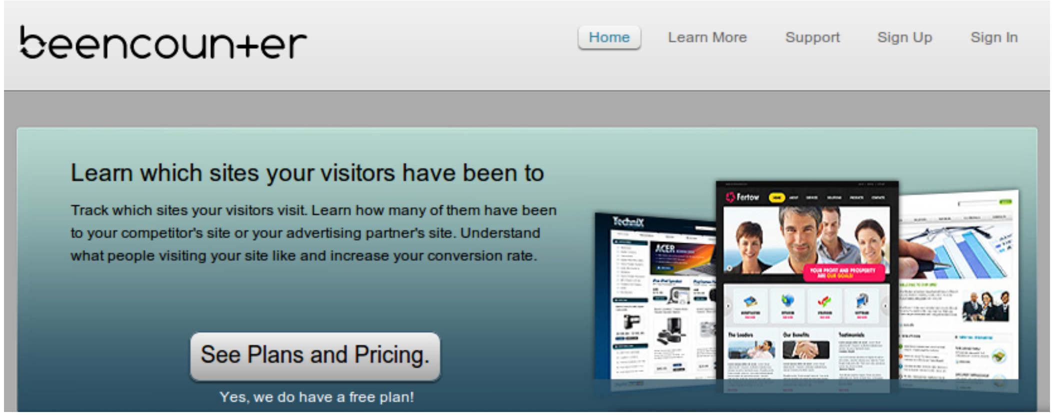
**84% of browser fingerprints are unique**

**With Flash or Java, 94% are unique**

# 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



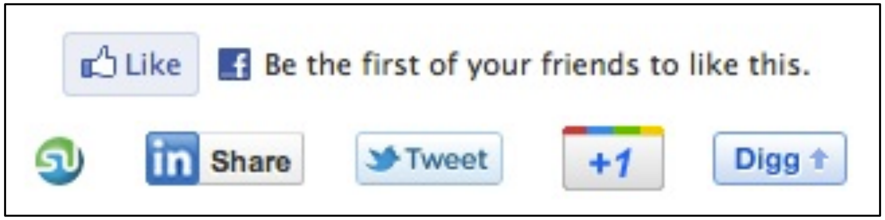
Learn which sites your visitors have been to

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.

Yes, we do have a free plan!

# How Websites Get Your Identity

Personal trackers



Be the first of your friends to like this.

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

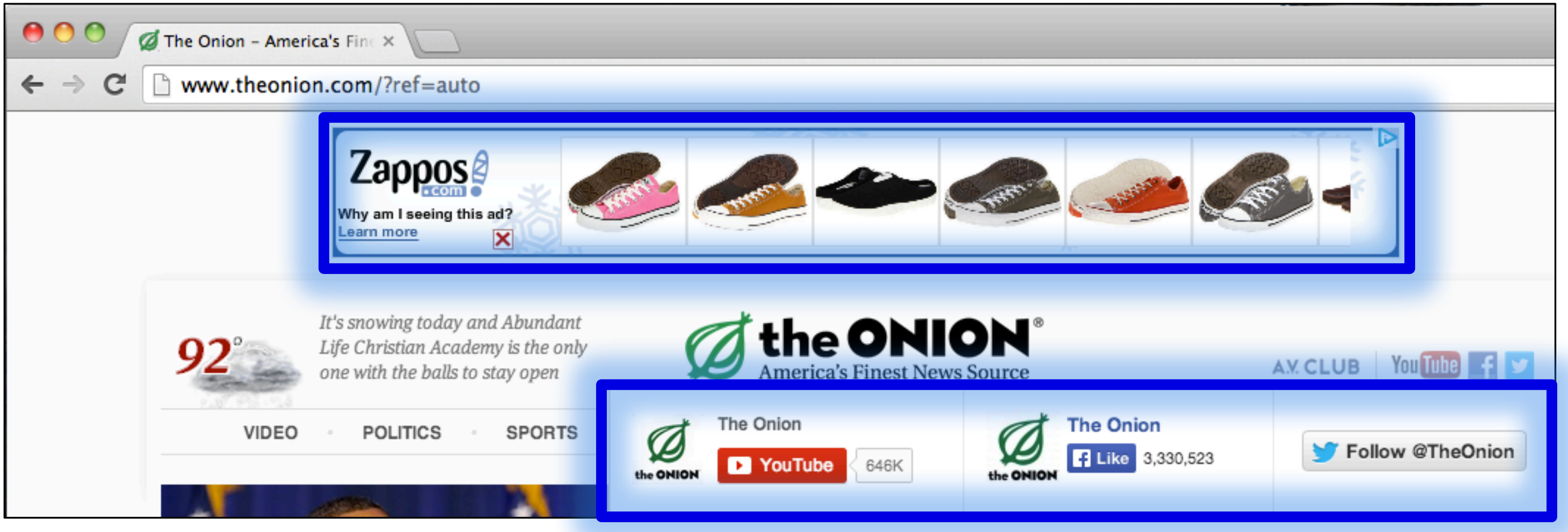
# Understanding the Tracking Ecosystem

- 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]*

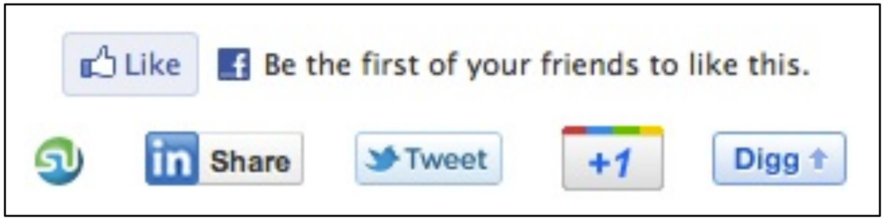
- In the wild, tracking is much more complicated.
- (1) Trackers don't just use cookies.
  - Flash cookies, HTML5 LocalStorage, etc.
- (2) Trackers exhibit different behaviors.
  - Within-site vs. cross-site.
  - Anonymous vs. non-anonymous.
  - Specific behavior types:
    **analytics, vanilla, forced,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

# Other Trackers?



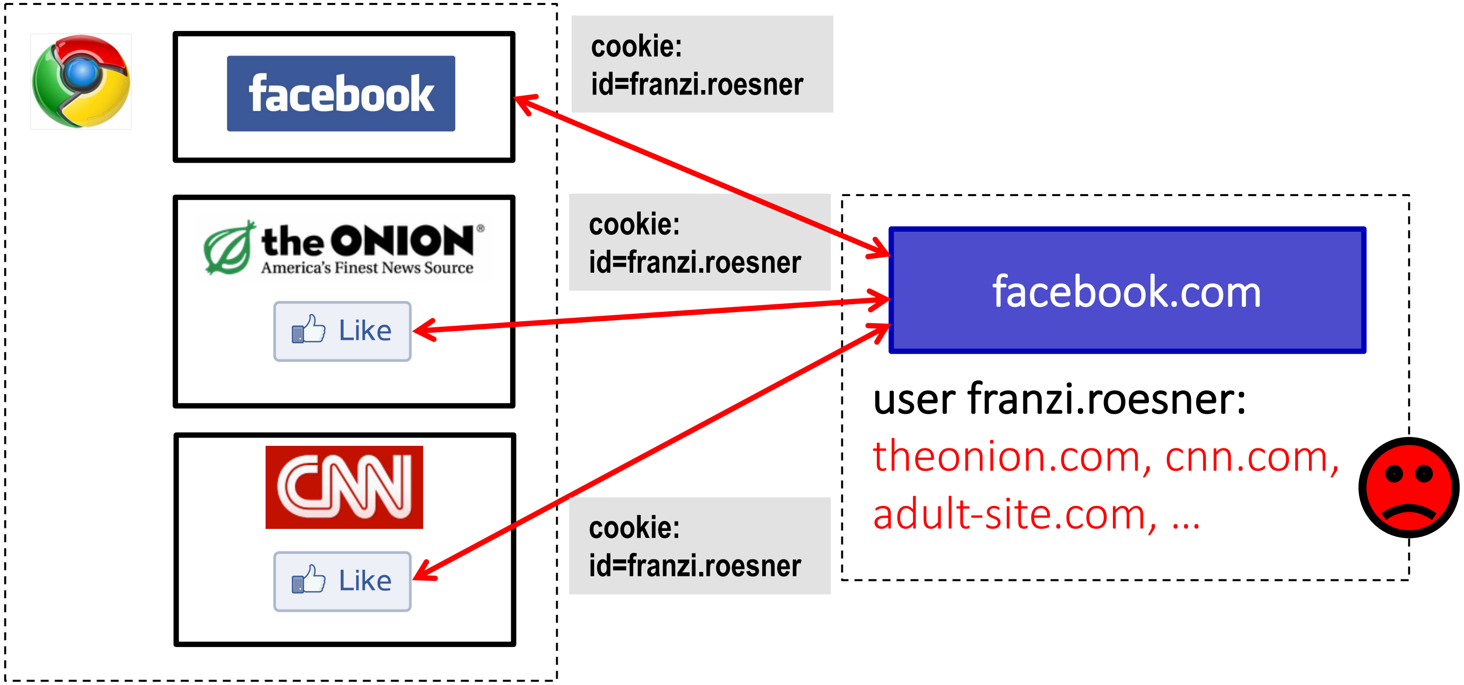
It's snowing today and Abundant Life Christian Academy is the only one with the balls to stay open

“Personal” Trackers



Be the first of your friends to like this.

# 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

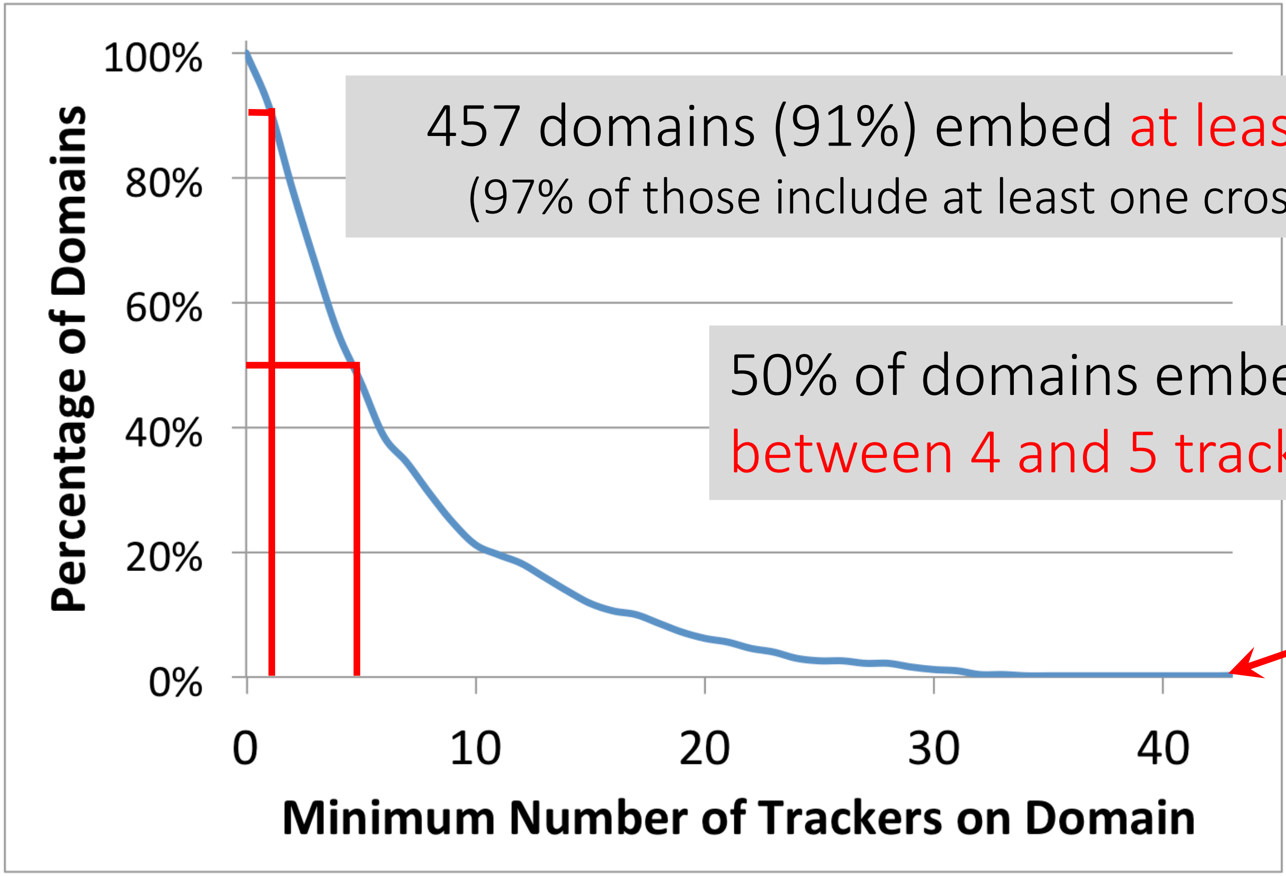
# Measurement Study (2011)

- **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)



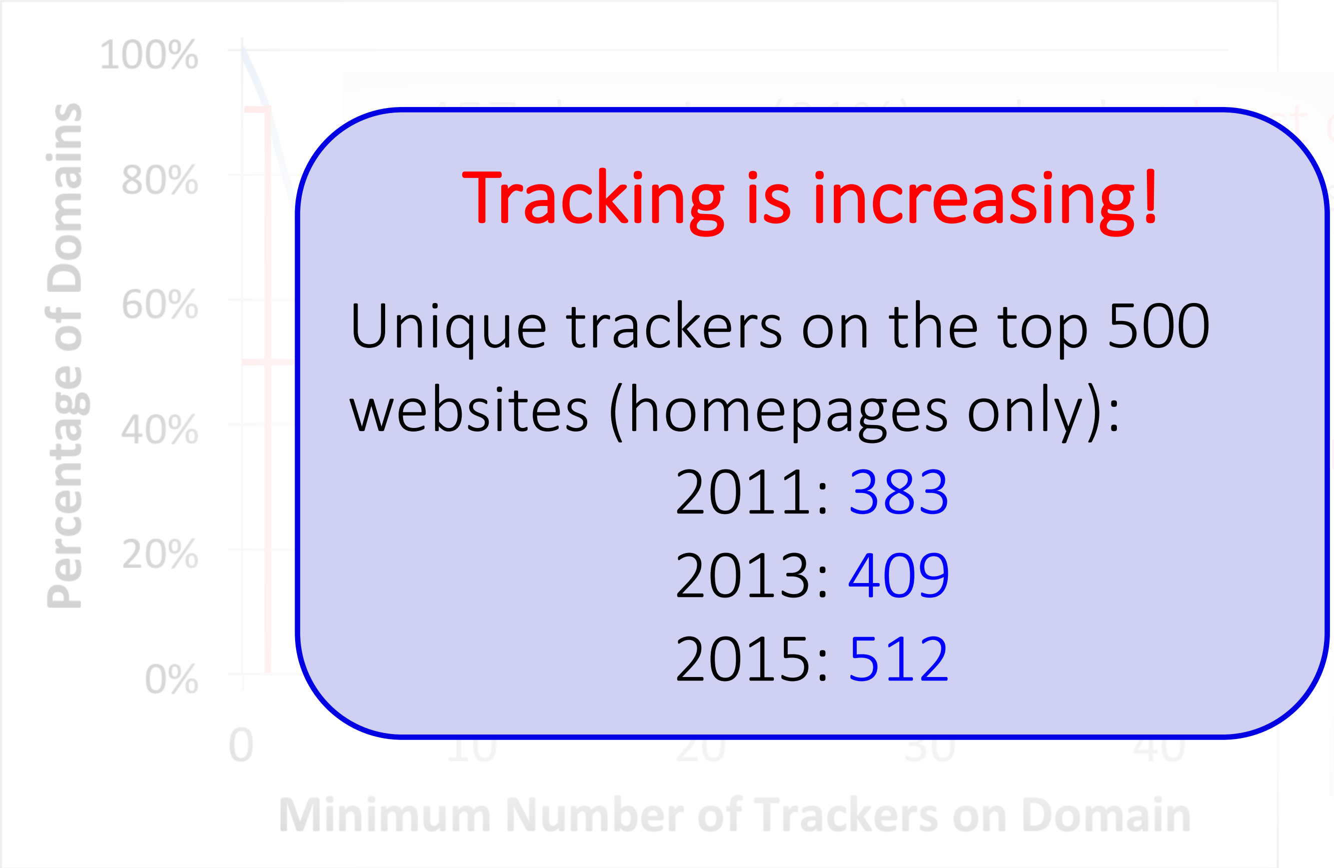
457 domains (91%) embed at least one tracker.
(97% of those include at least one cross-site tracker.)

50% of domains embed between 4 and 5 trackers.

One domain includes 43 trackers.

# How prevalent is tracking?

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



one tracker.

ite tracker.)

rs.

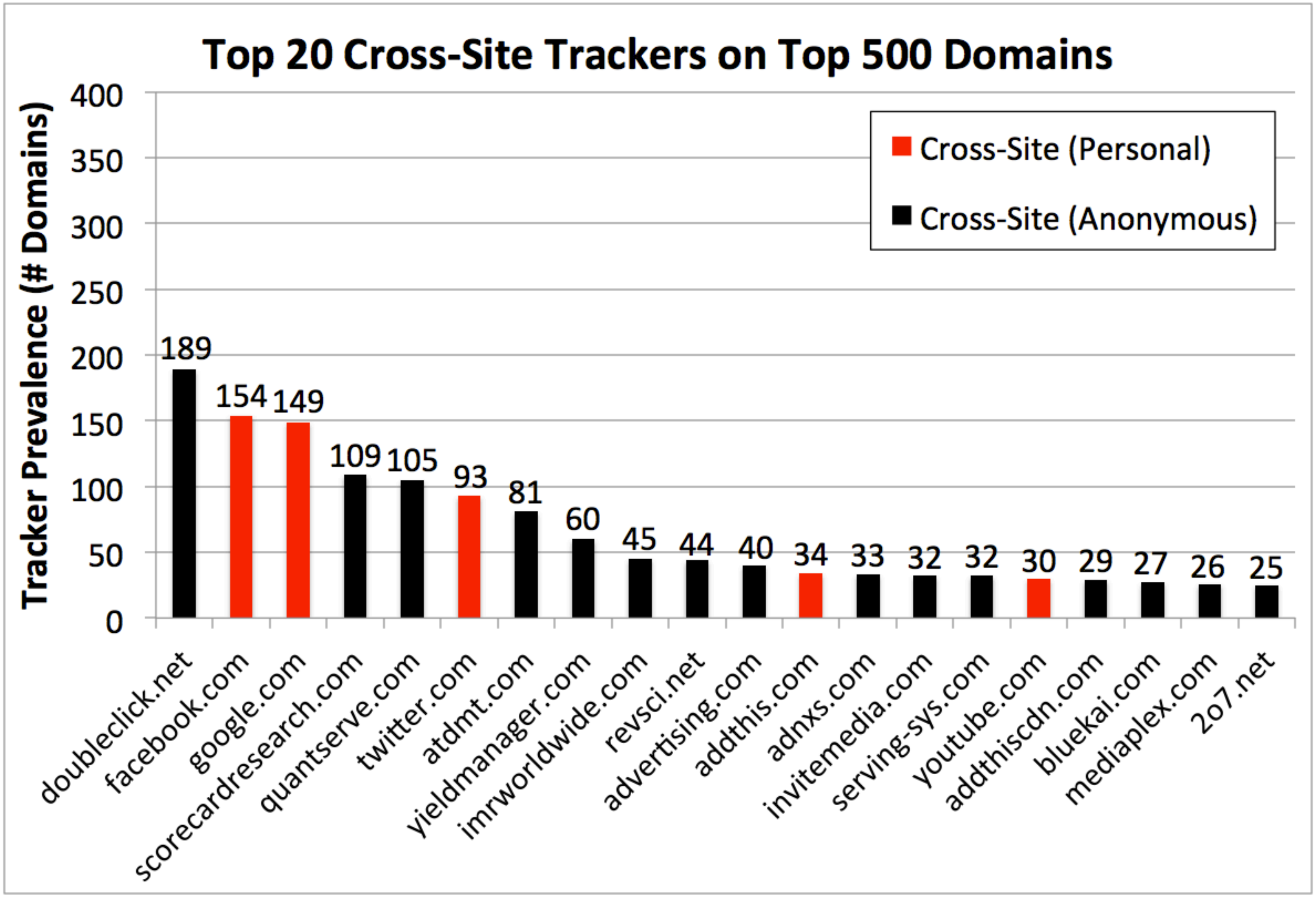
One domain includes 43 trackers.

Tracking is increasing!

Unique trackers on the top 500 websites (homepages only):

2011: 383
2013: 409
2015: 512

# Who/what are the top trackers? (2011)



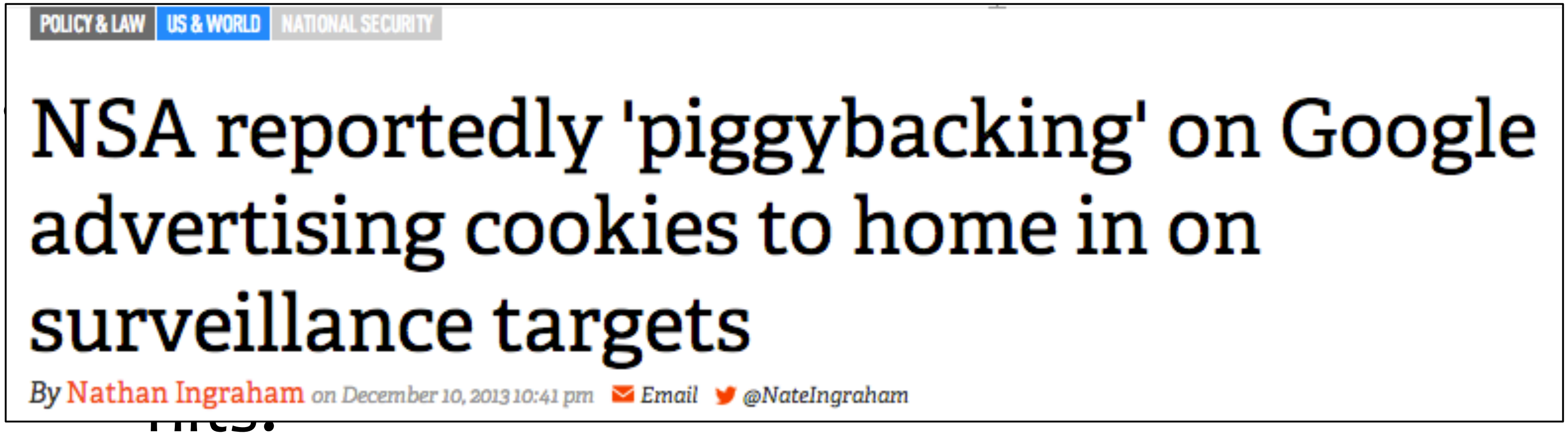
# How are users affected?

- 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

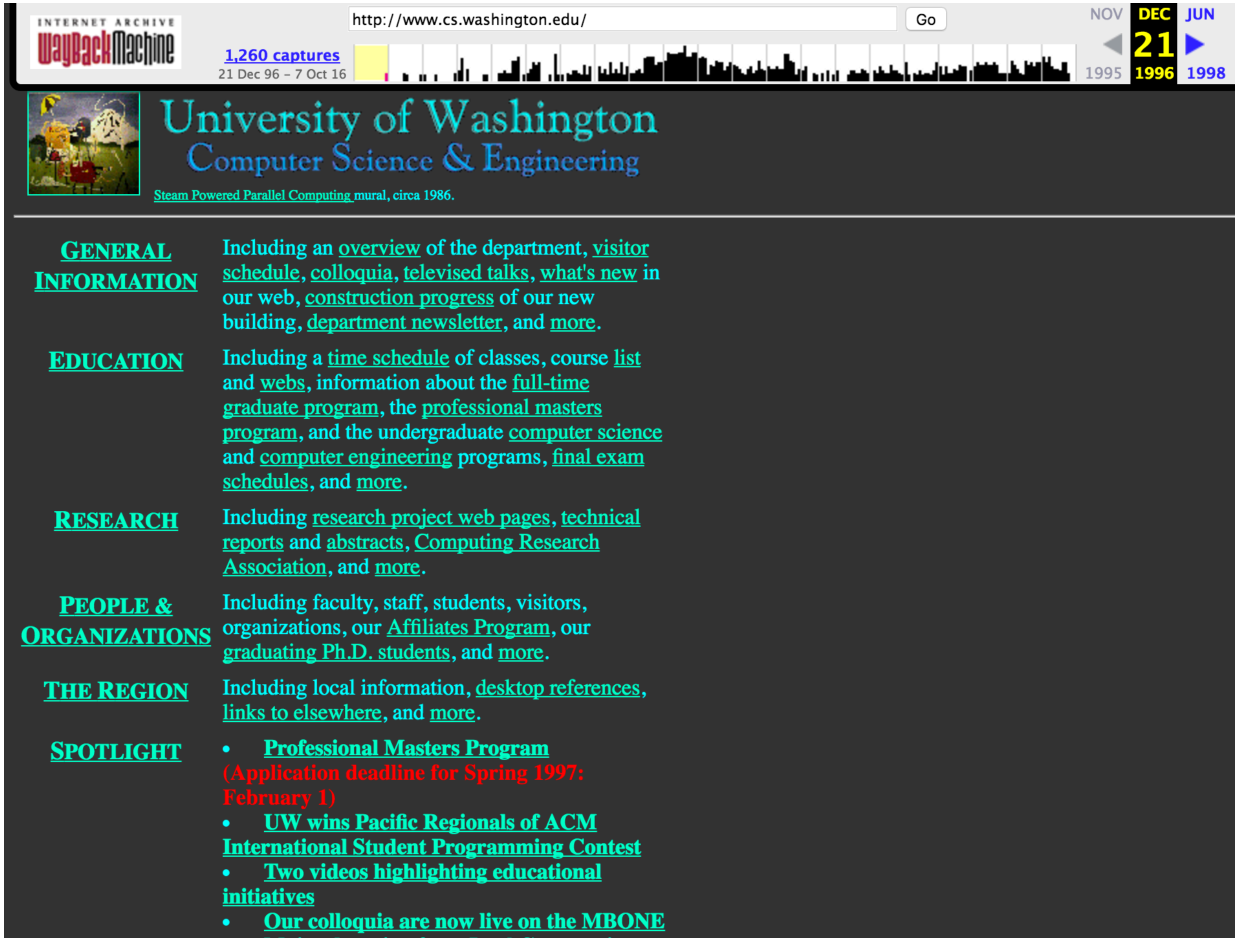
# How has this changed over time?

- 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



University of Washington
Computer Science & Engineering

Steam Powered Parallel Computing mural, circa 1986.

**GENERAL INFORMATION** Including an overview of the department, visitor schedule, colloquia, televised talks, what's new in our web, construction progress of our new building, department newsletter, and more.

**EDUCATION** Including a time schedule of classes, course list and webs, information about the full-time graduate program, the professional masters program, and the undergraduate computer science and computer engineering programs, final exam schedules, and more.

**RESEARCH** Including research project web pages, technical reports and abstracts, Computing Research Association, and more.

**PEOPLE & ORGANIZATIONS** Including faculty, staff, students, visitors, organizations, our Affiliates Program, our graduating Ph.D. students, and more.

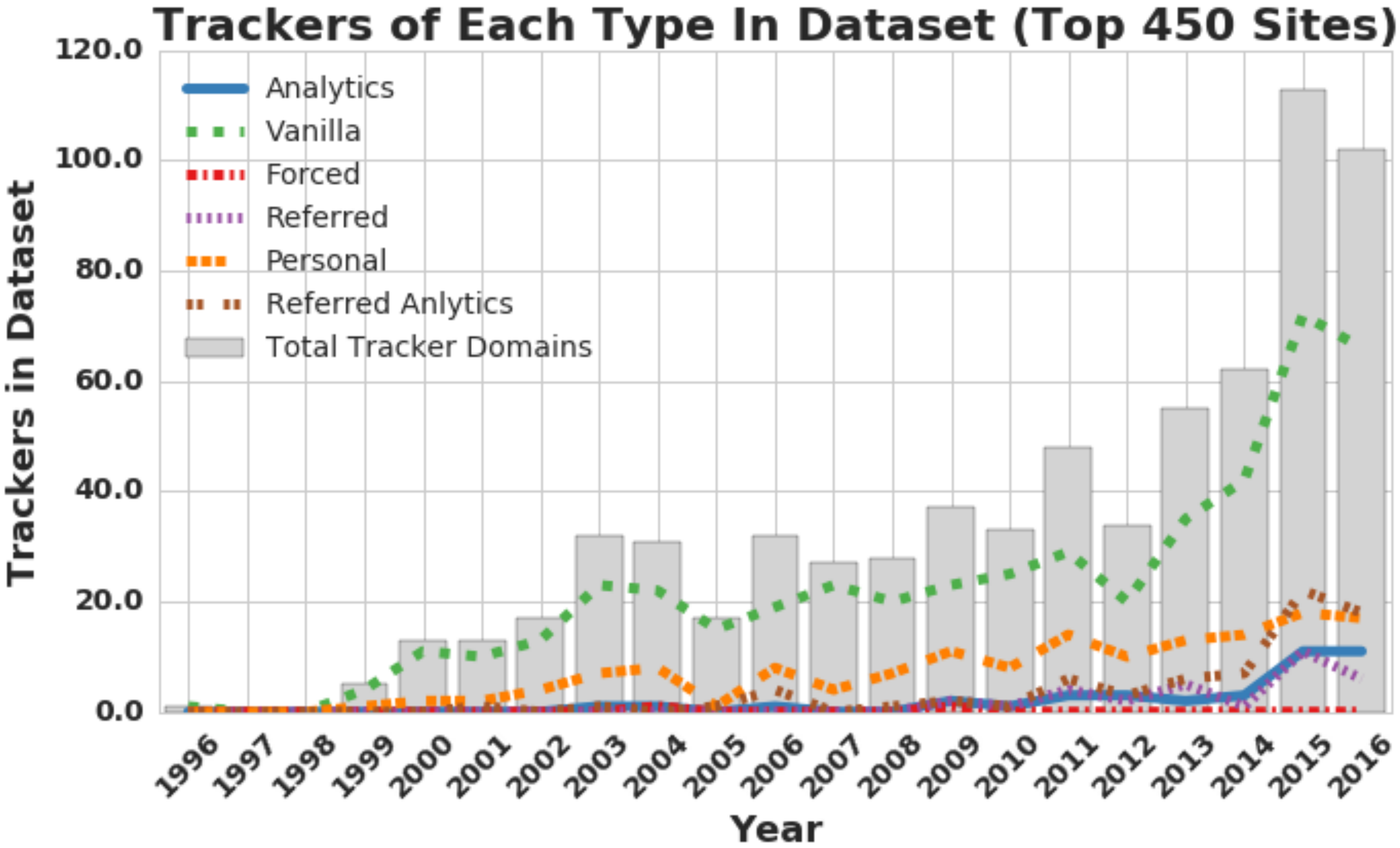
**THE REGION** Including local information, desktop references, links to elsewhere, and more.

**SPOTLIGHT**
- **Professional Masters Program** (Application deadline for Spring 1997: February 1)
- **UW wins Pacific Regionals of ACM International Student Programming Contest**
- **Two videos highlighting educational initiatives**
- **Our colloquia are now live on the MBONE**

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

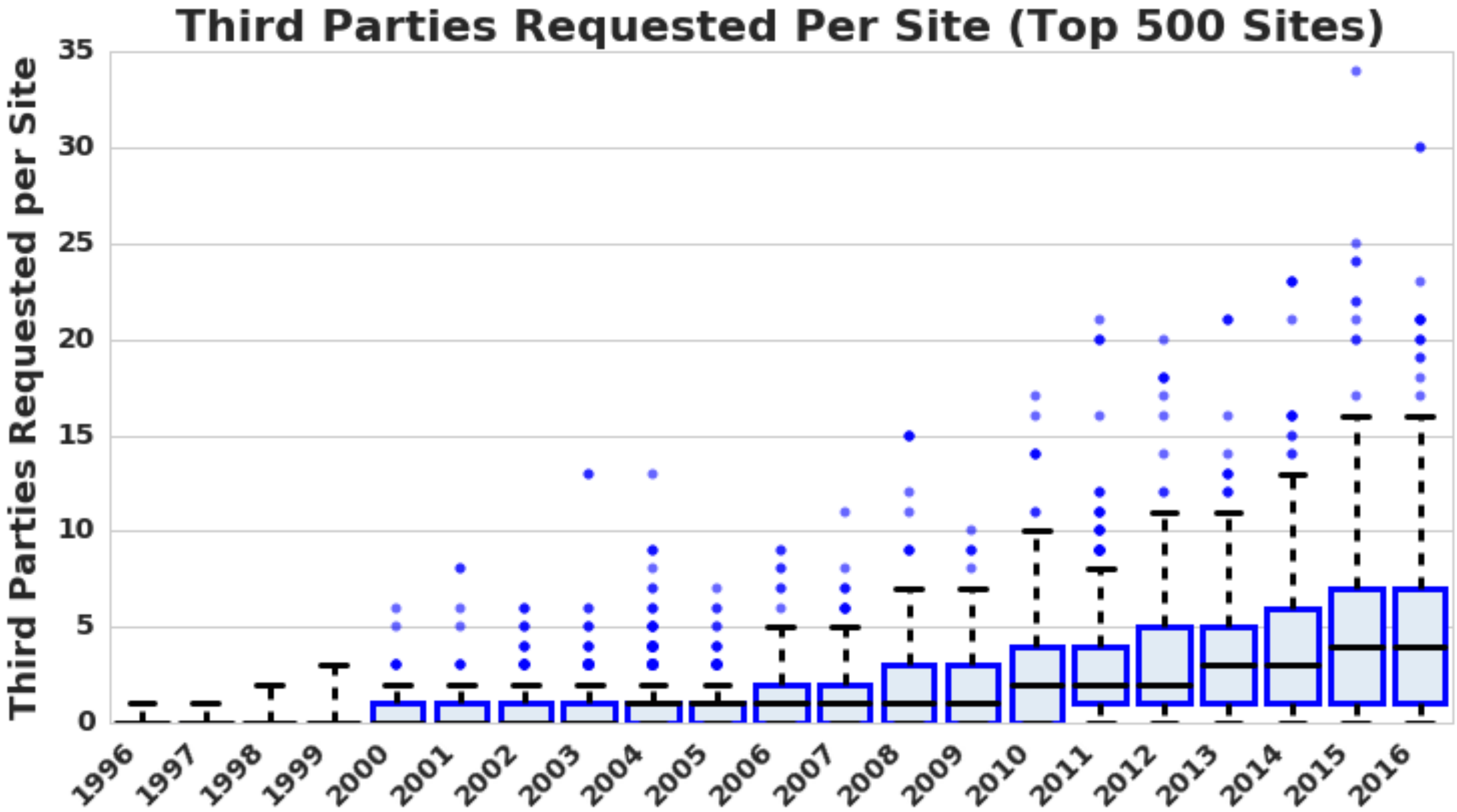
# 1996-2016: More & More Tracking

- More trackers of more types



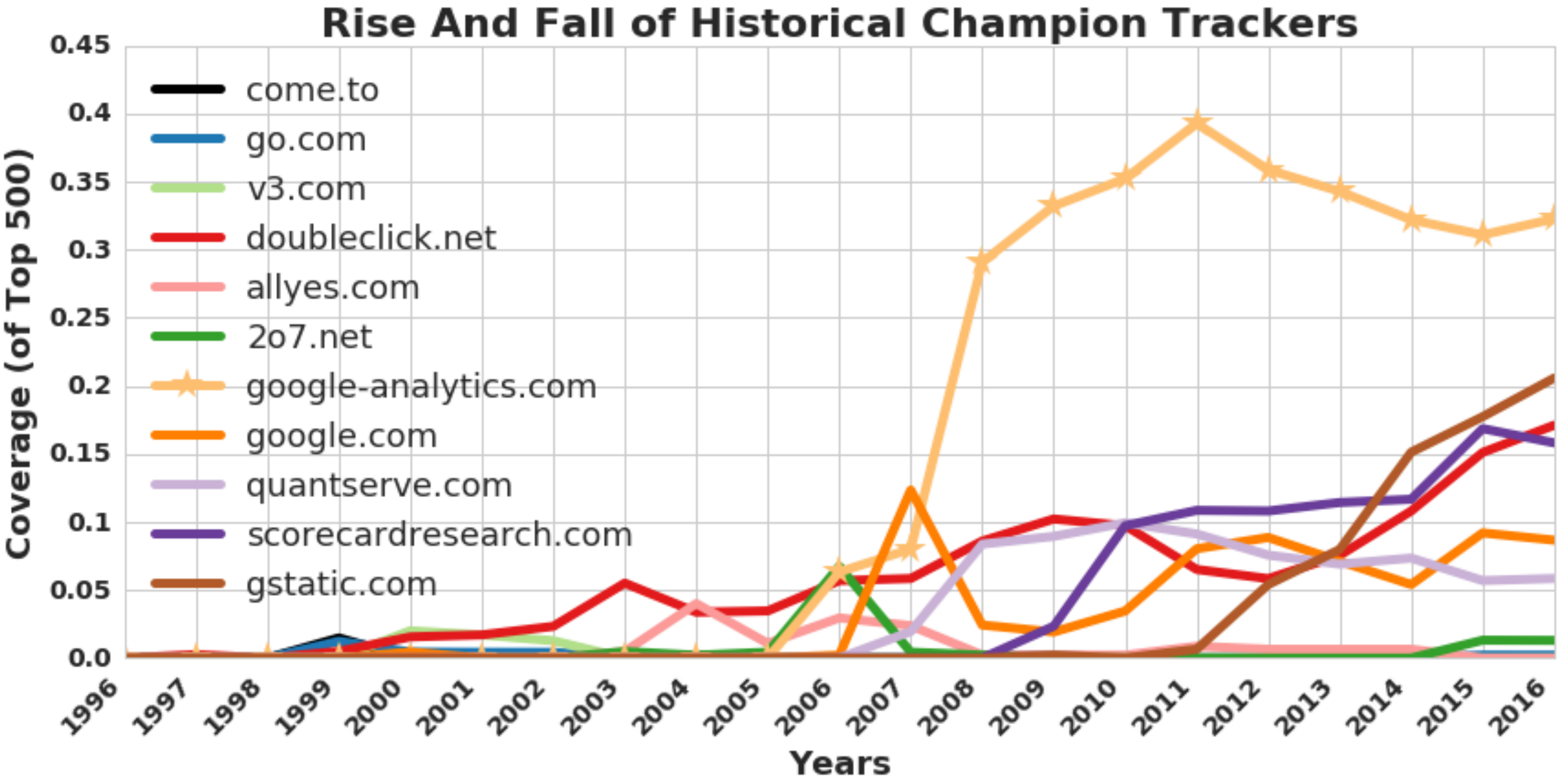
# 1996-2016: More & More Tracking

- More trackers of more types, more per site



# 1996-2016: More & More Tracking

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

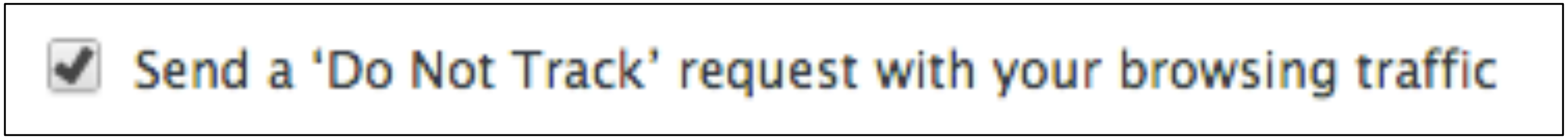


# Outline

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

# Defenses to Reduce Tracking

- Do Not Track proposal?



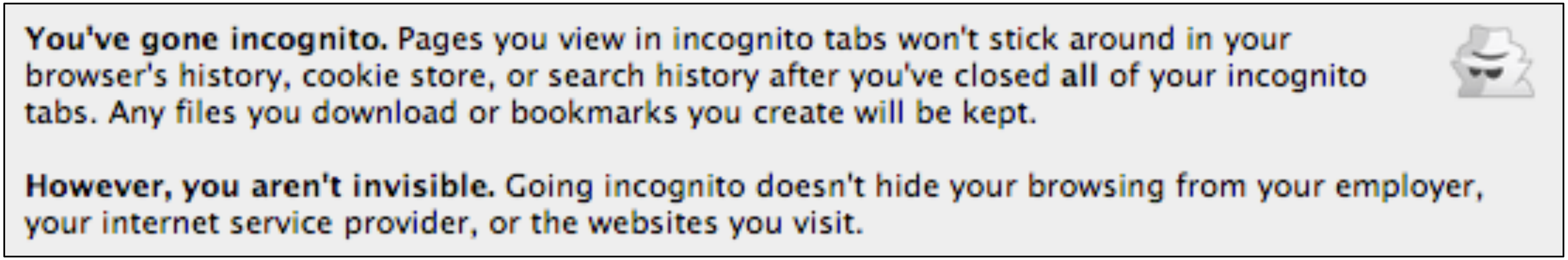
☑ Send a 'Do Not Track' request with your browsing traffic

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.

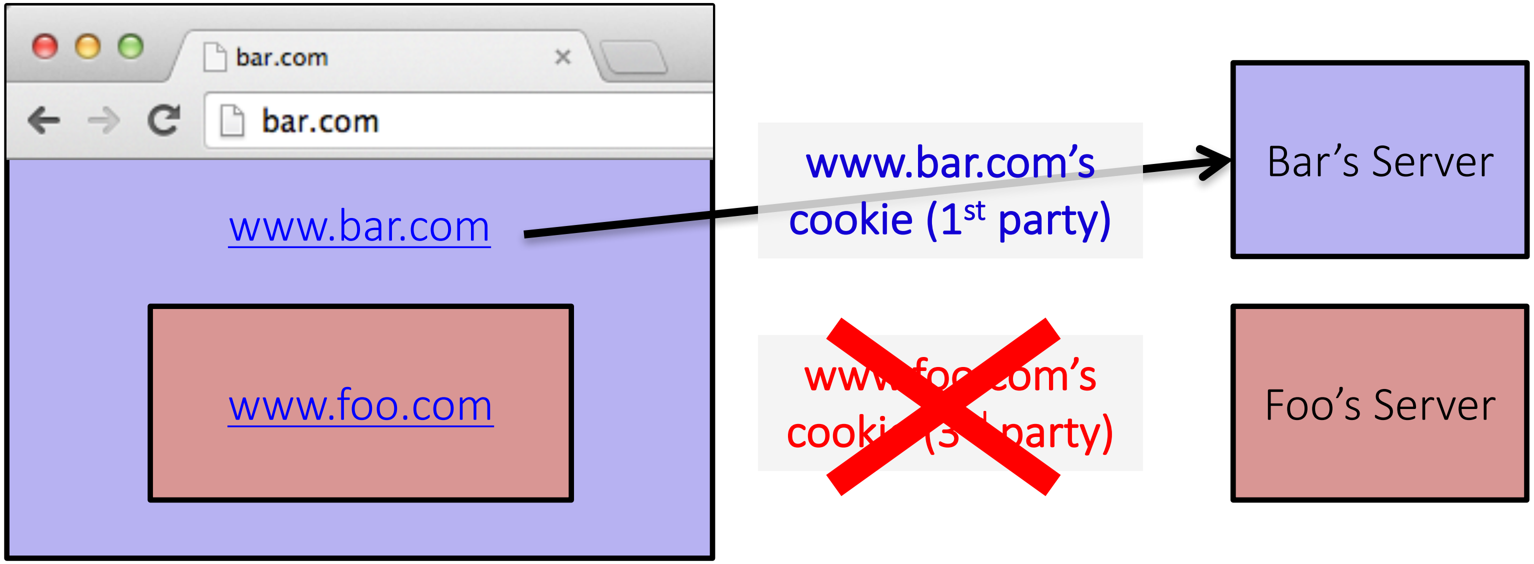


**You've gone incognito.** Pages you view in incognito tabs won't stick around in your browser's history, cookie store, or search history after you've closed all of your incognito tabs. Any files you download or bookmarks you create will be kept.

**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?



www.bar.com

www.foo.com

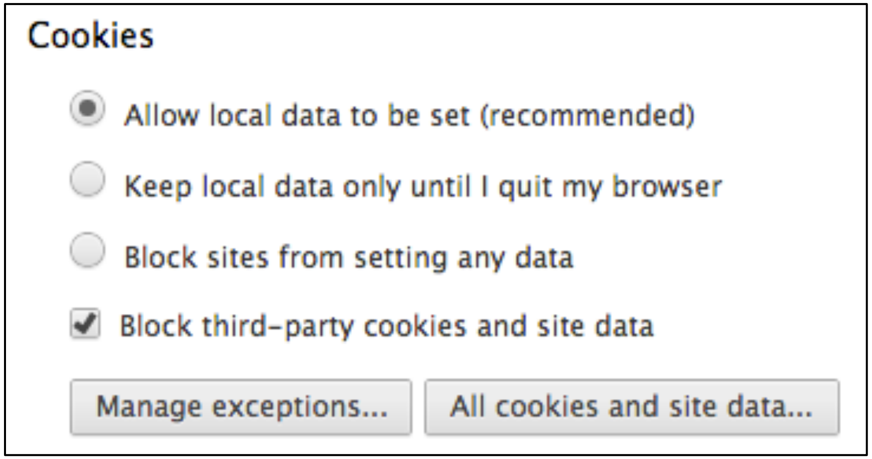
www.bar.com's cookie (1st party)

~~www.foo.com's cookie (3rd party)~~

Bar's Server

Foo's Server

# 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.

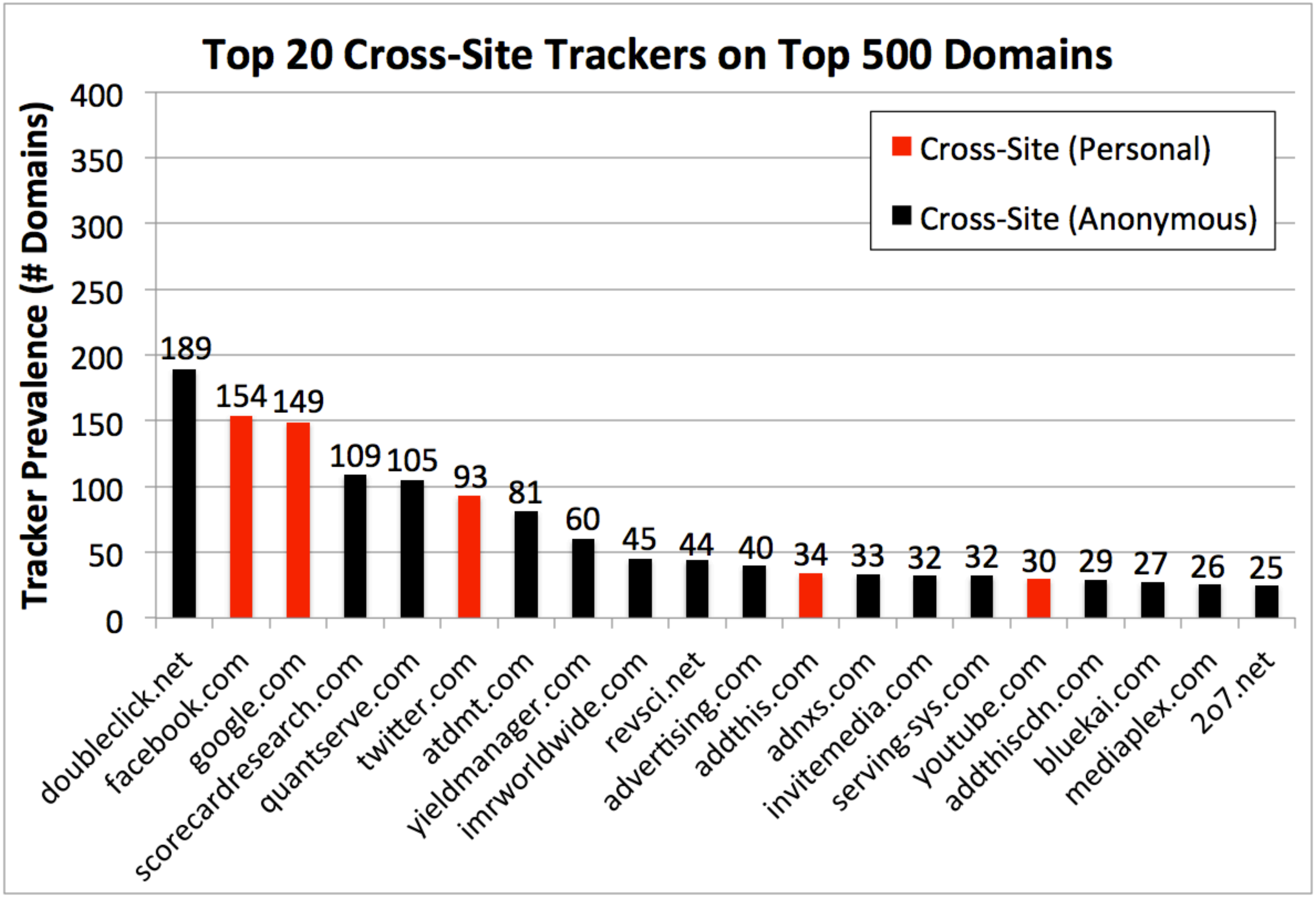
So if a third-party cookie is somehow set, it can be used.

How to get a cookie set?

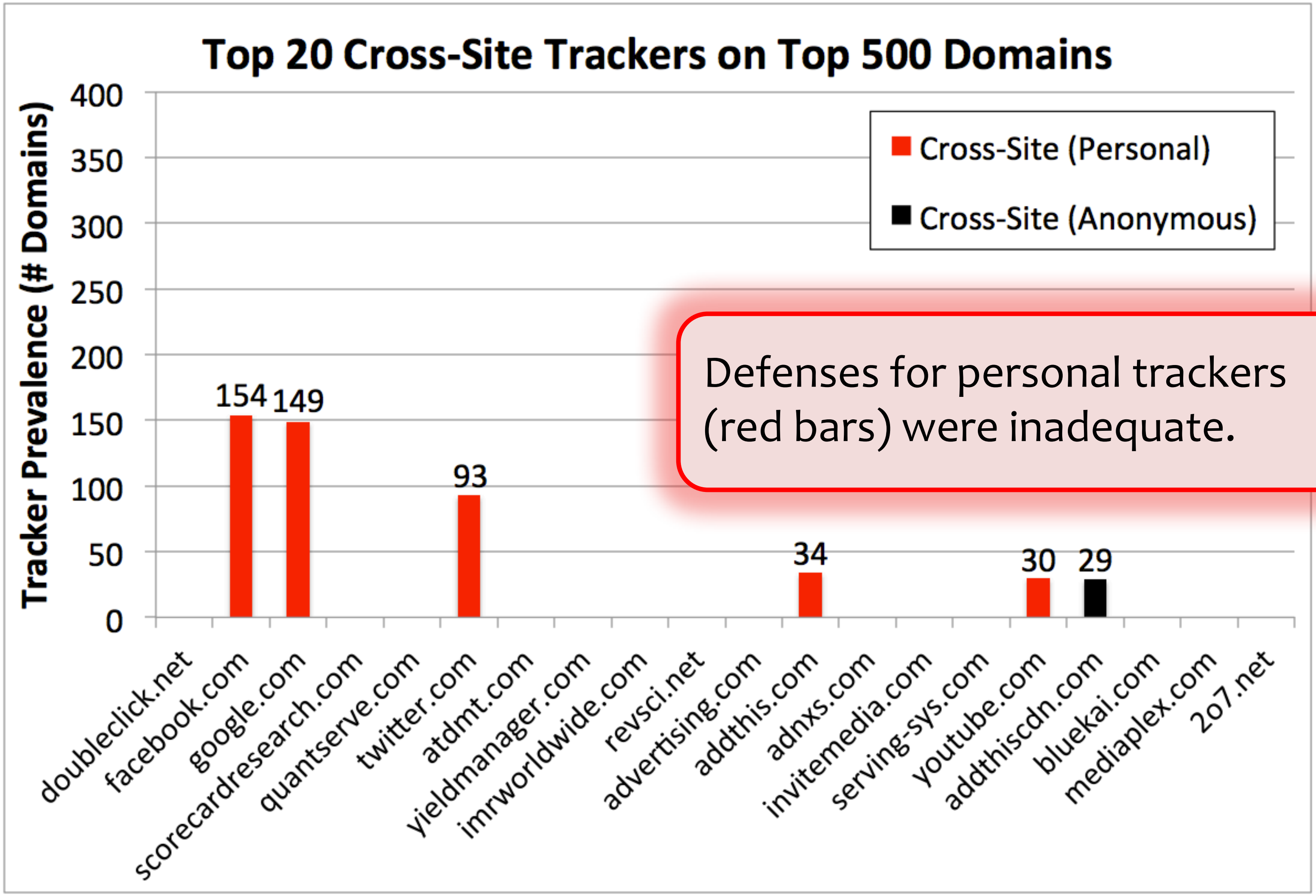
One way: be a first party.

etc.

# What 3rd Party Cookie Blocking Misses



# What 3rd Party Cookie Blocking Misses



Defenses for personal trackers (red bars) were inadequate.

# 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?

Ghostery

Lightbeam for Firefox

Often rely on blacklists, which may be incomplete.

Privacy Badger

BETA

A PROJECT OF THE ELECTRONIC FRONTIER FOUNDATION

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

uBlock Origin