CSE 484 / CSE M 584: Computer Security and Privacy

Third-Party Tracking on the Web

Fall 2016

Ada (Adam) Lerner
lerner@cs.washington.edu

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

Announcements

- Please form groups for the final project. For the project, you will choose a computer security/privacy topic and explain it twice:
  - Once for a relevant "lay" audience, and
  - Once for a technical audience.

- Submit **who** will be in your group and **what topic you will discuss** by next Monday.

Announcements

- More details on the final project will be landing in the next couple days.

Security Mindset Anecdote

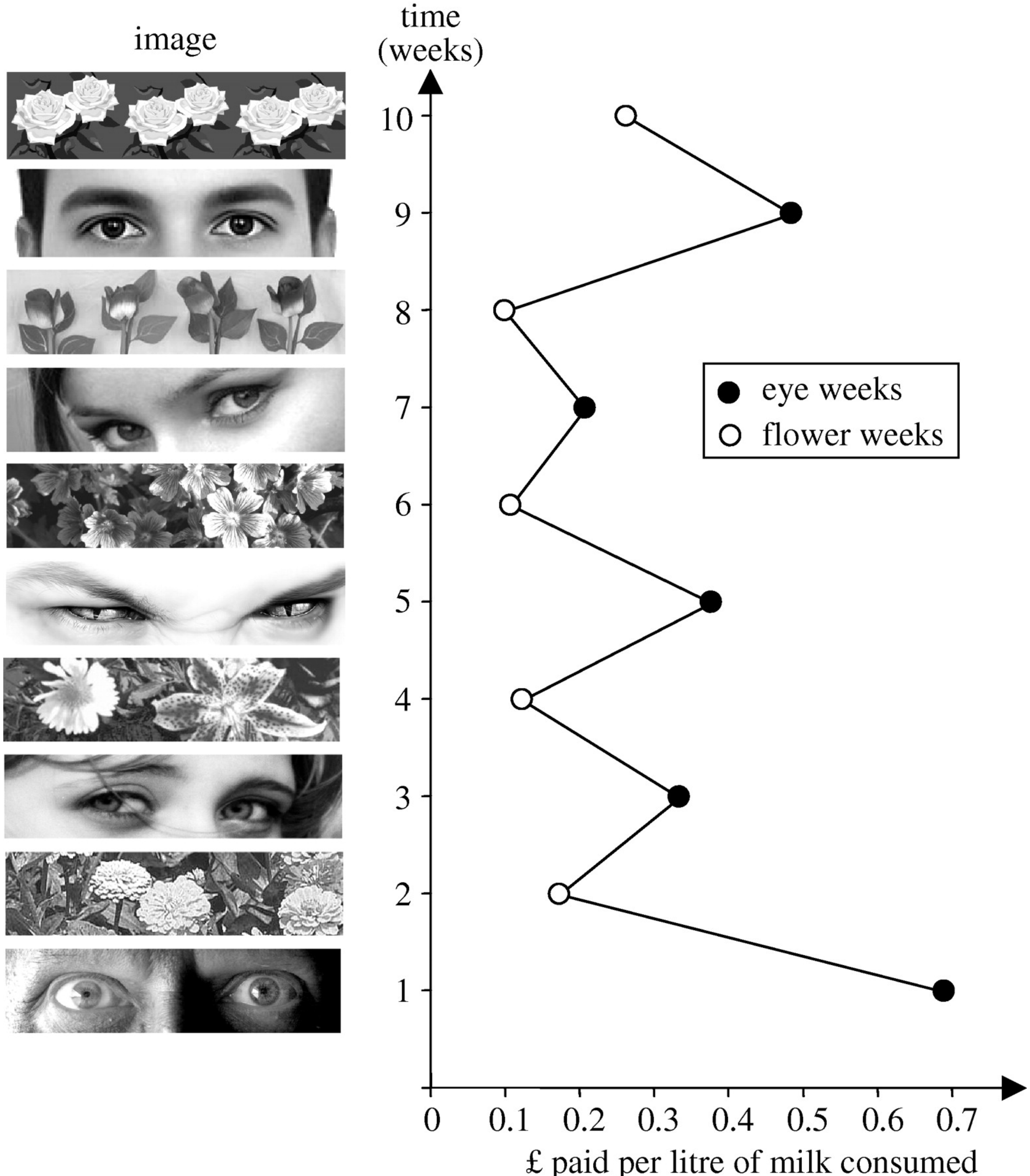


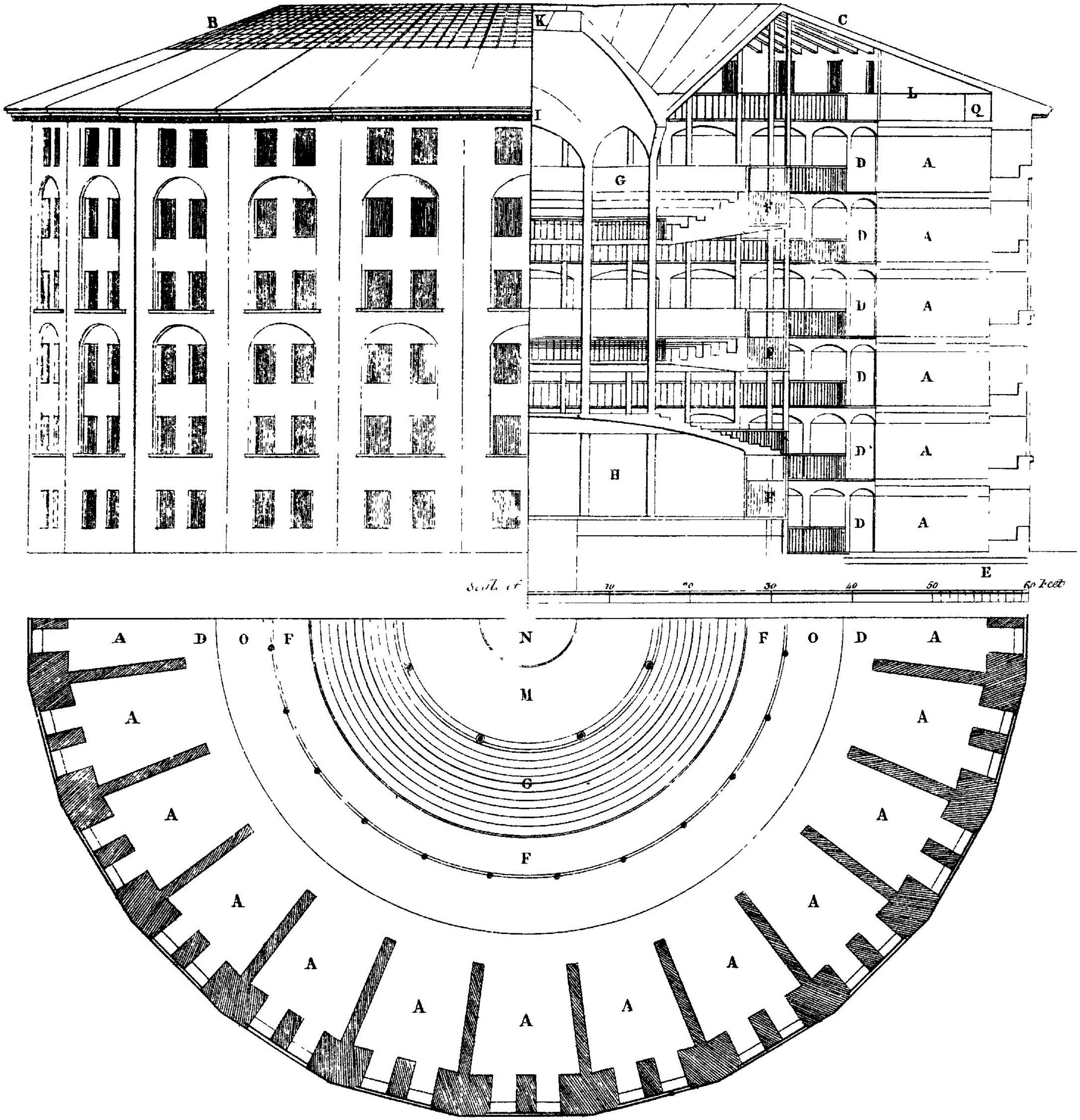
image
time (weeks)
10
9
8
7
6
5
4
3
2
1
● eye weeks
○ flower weeks
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
£ paid per litre of milk consumed

"Panopticon"

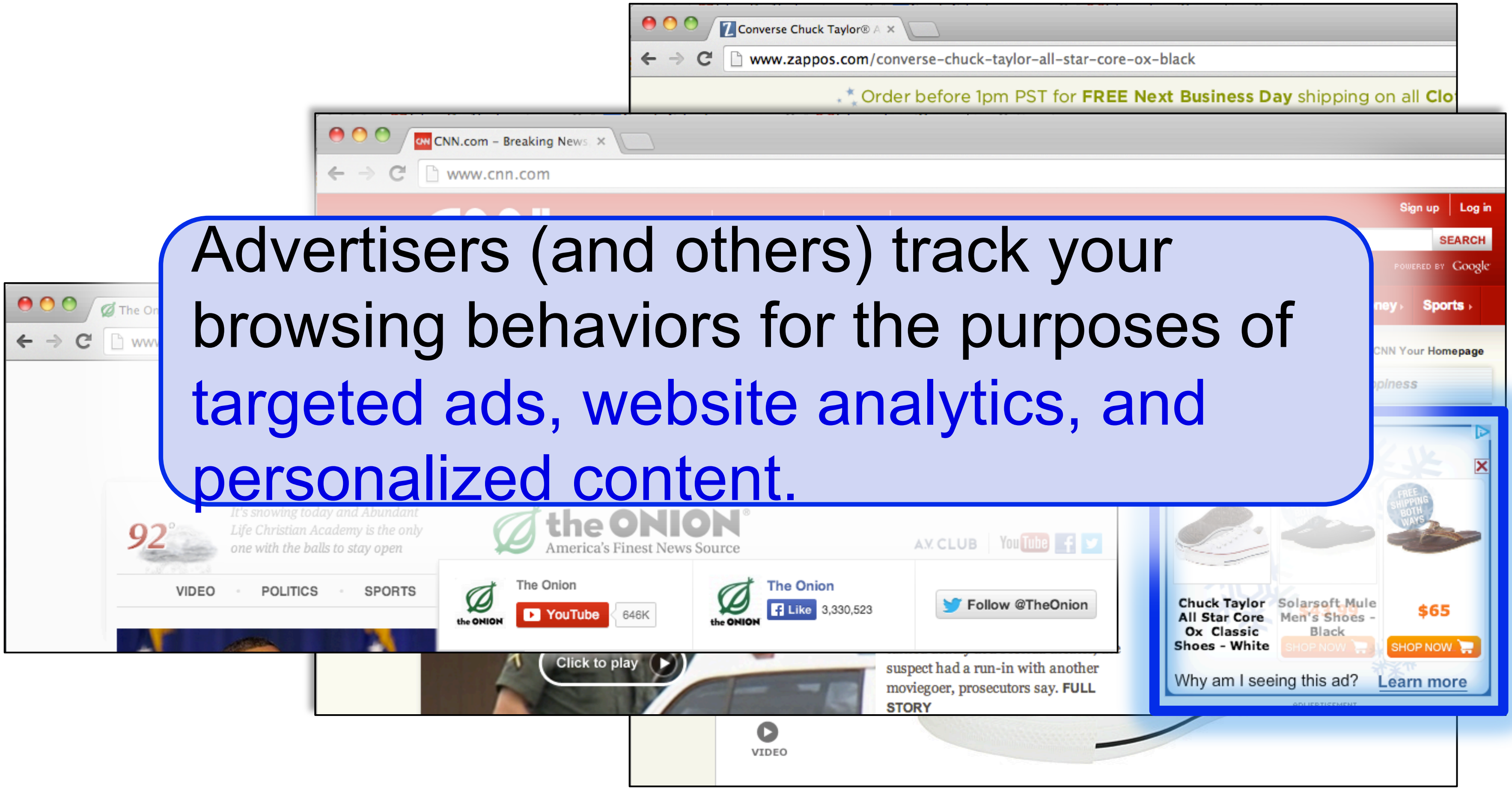
- Prison design by Jeremy Bentham in England in the late 18th century.

- Idea: design a prison so that guard **might be watching** all the prisoners at all times. Prisoners will have to behave as though they are watched all the time, even if they are not.

- Bentham described the Panopticon as
  - "a new mode of obtaining power of mind over mind, in a quantity hitherto without example.”
  - "a mill for grinding rogues honest”
- [see Wikipedia article for citations]

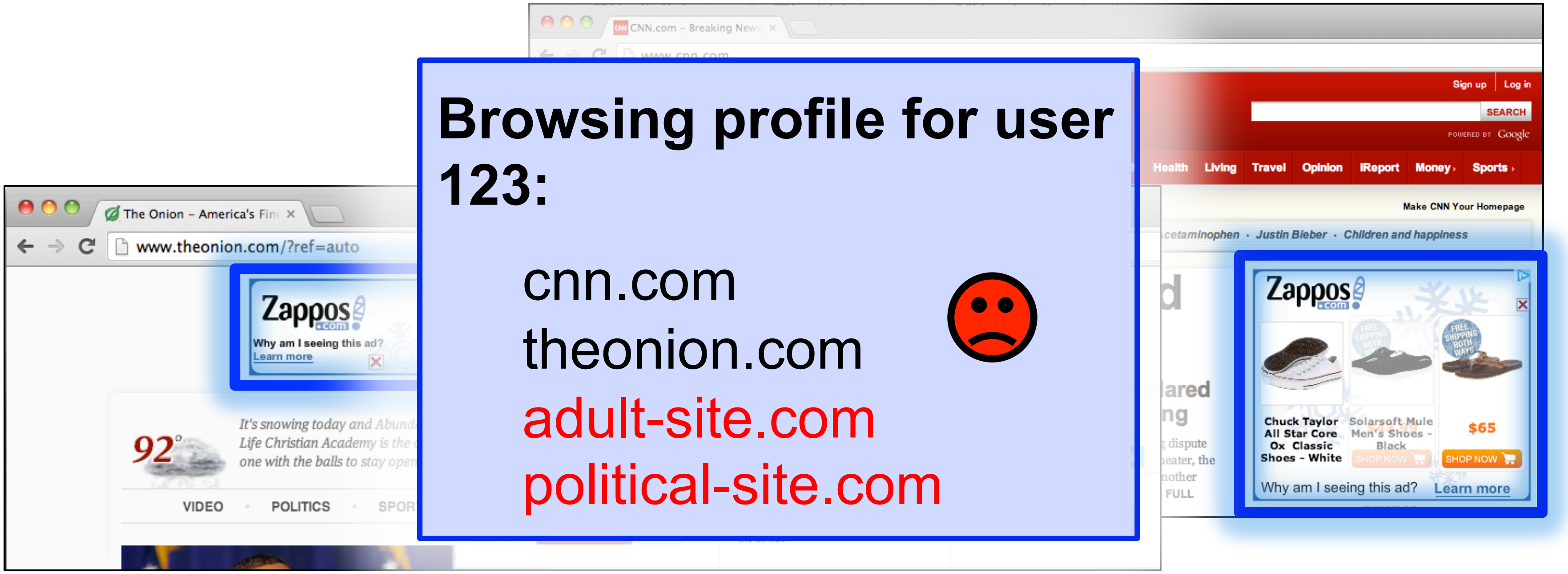

B
K
C
L
Q
I
G
D
A
H
F
E
Scale of
10
20
30
40
50
60 Feet
A D O F
N
F O D A
M
G
F

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
The Onion
YouTube
646K
The Onion
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
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

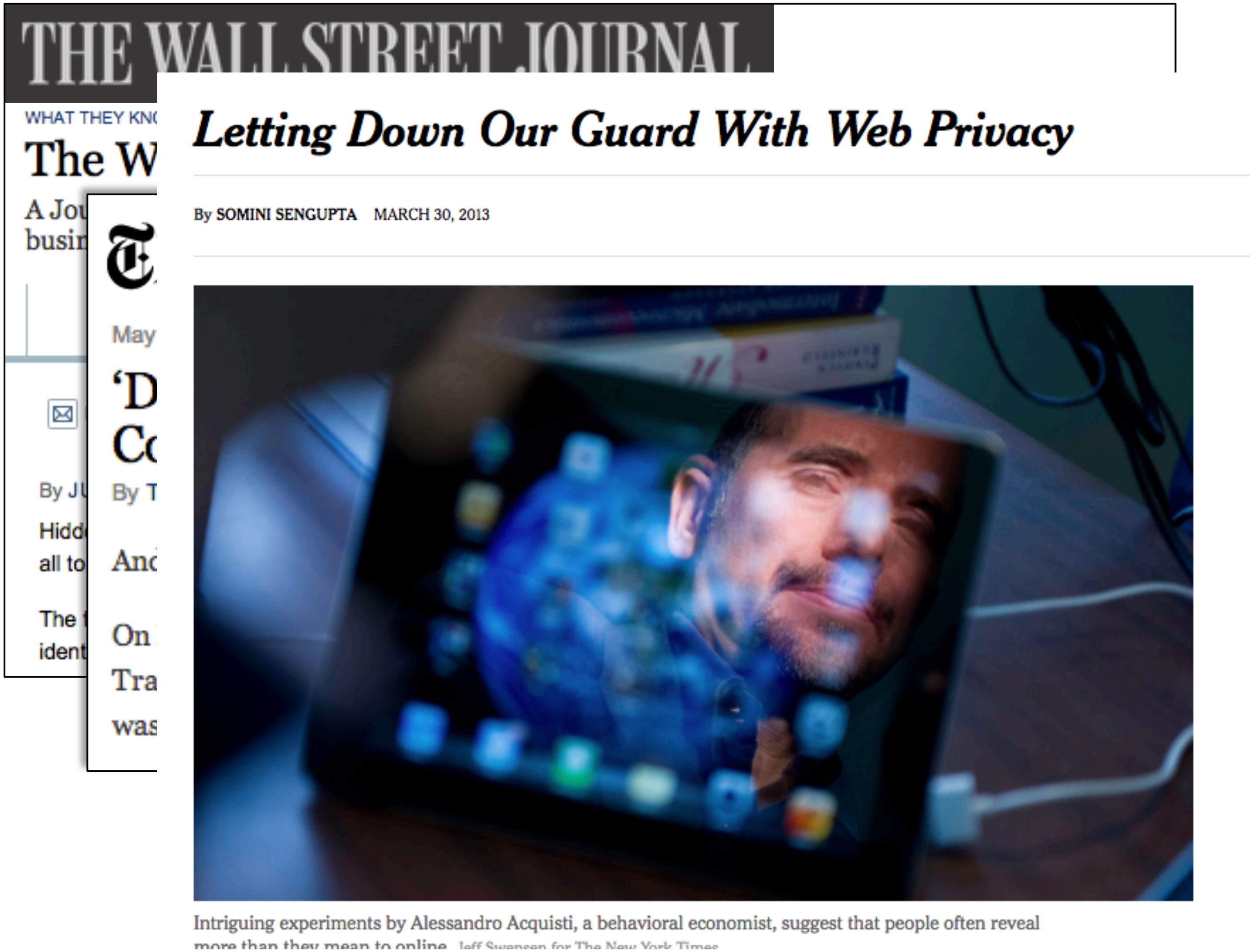


Browsing profile for user 123:

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 KN

The W

A Jou
busi

By J

Hidd
all to

The
ident

May

'D
Co

By T

And

On
Tra
was

Letting Down Our Guard With Web Privacy

By SOMINI SENGUPTA MARCH 30, 2013

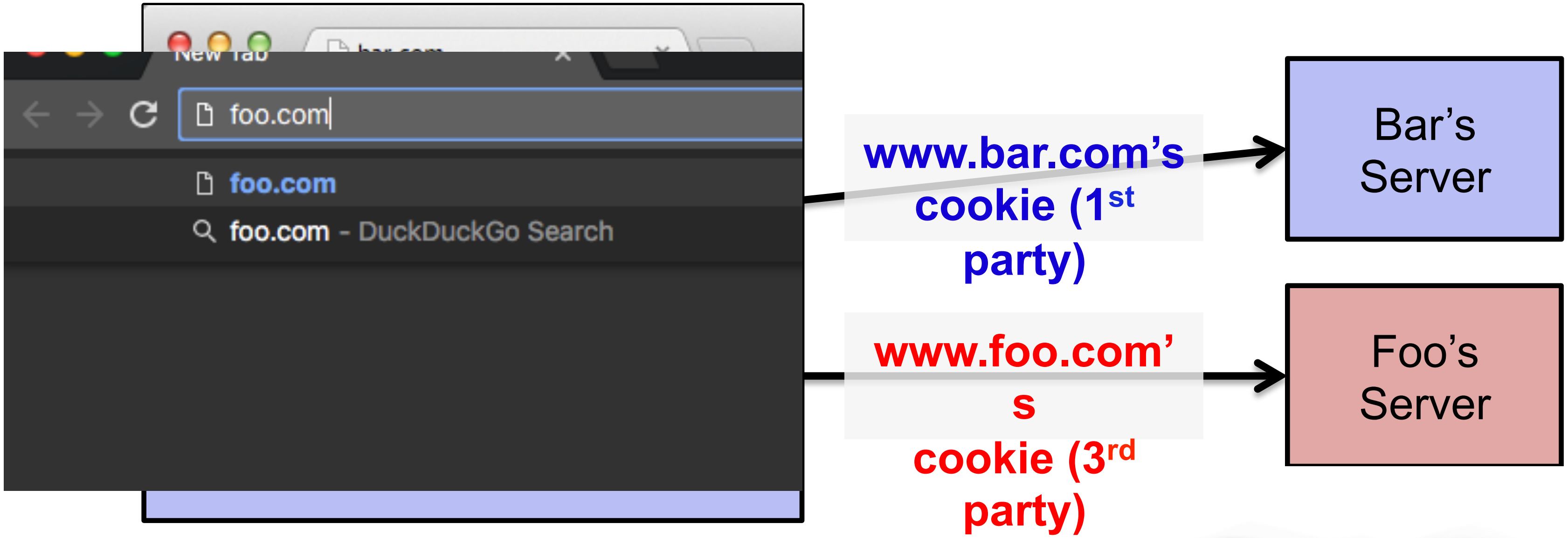
Intriguing experiments by Alessandro Acquisti, a behavioral economist, suggest that people often reveal more than they mean to online. Jeff Swensen for The New York Times

Outline

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

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).



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

Bar's Server

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

Foo's Server

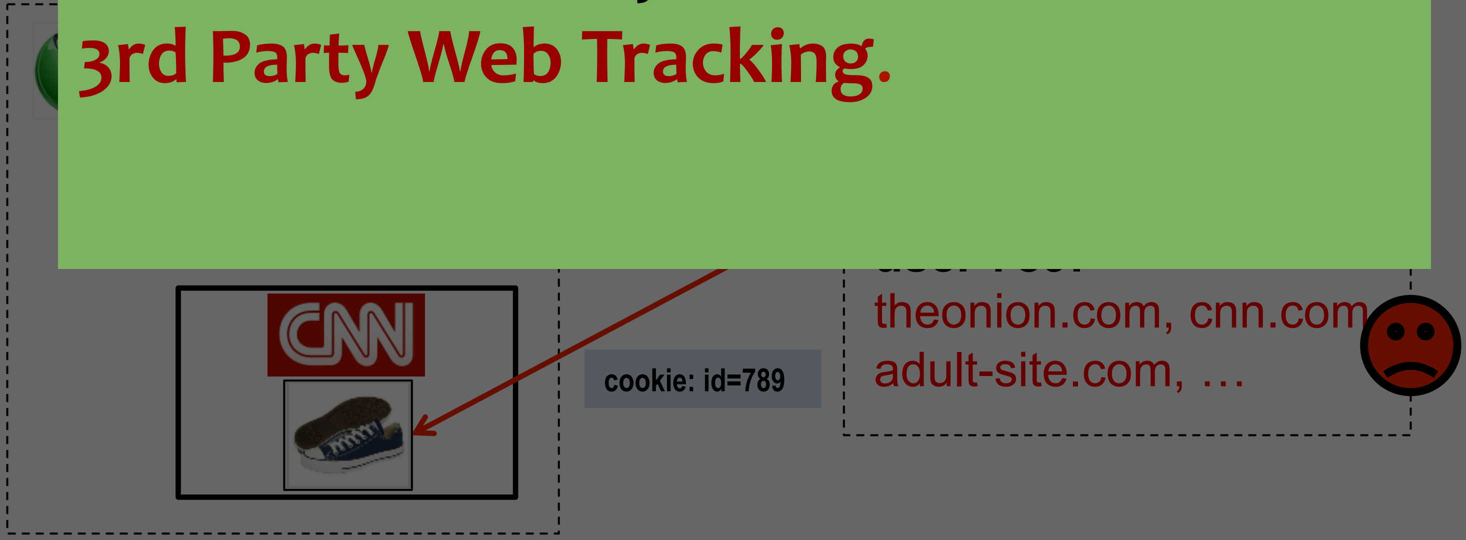
Anonymous Tracking

Track es

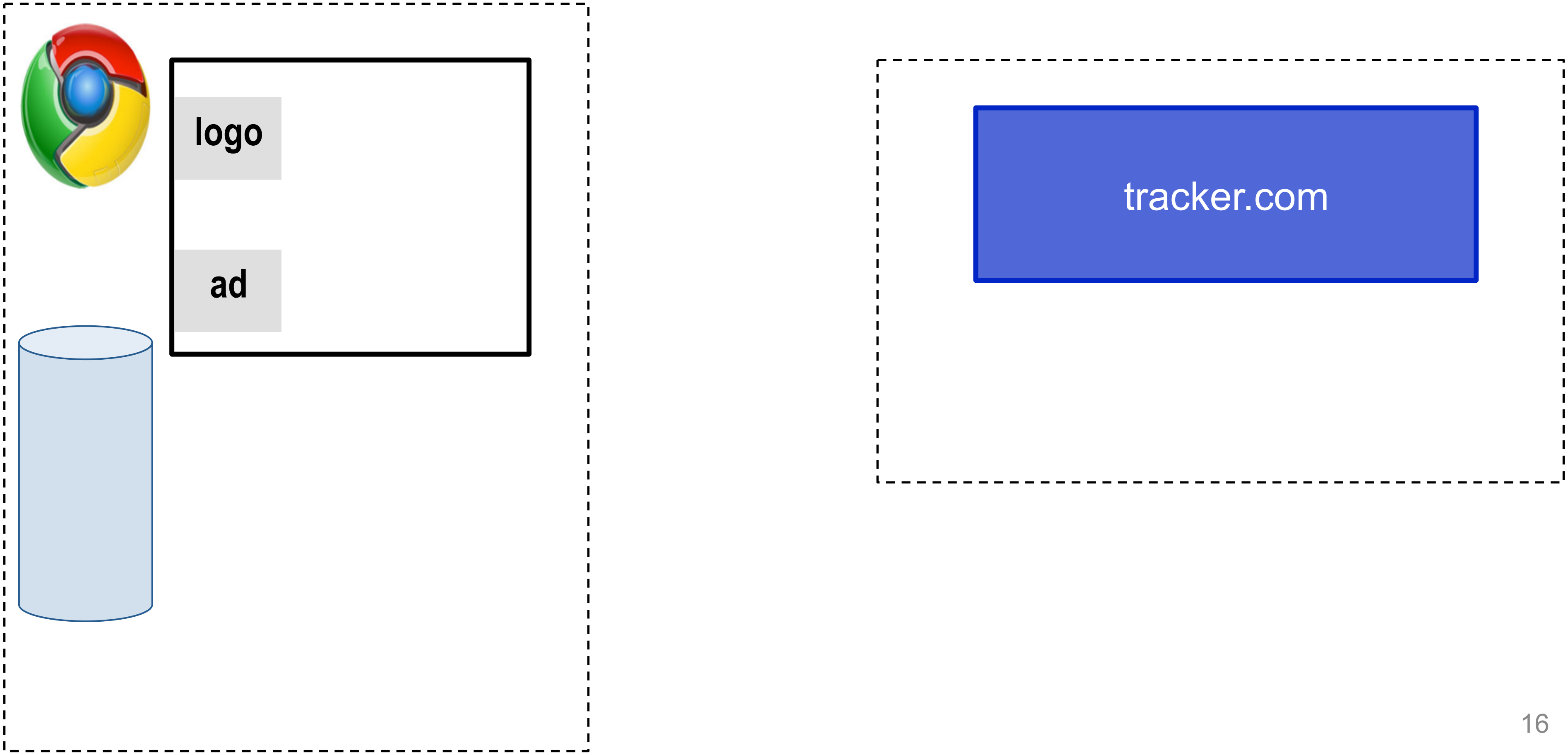
conta files.

If a third party is able to **link together** a subset of a person's **browsing history,** we call this ability

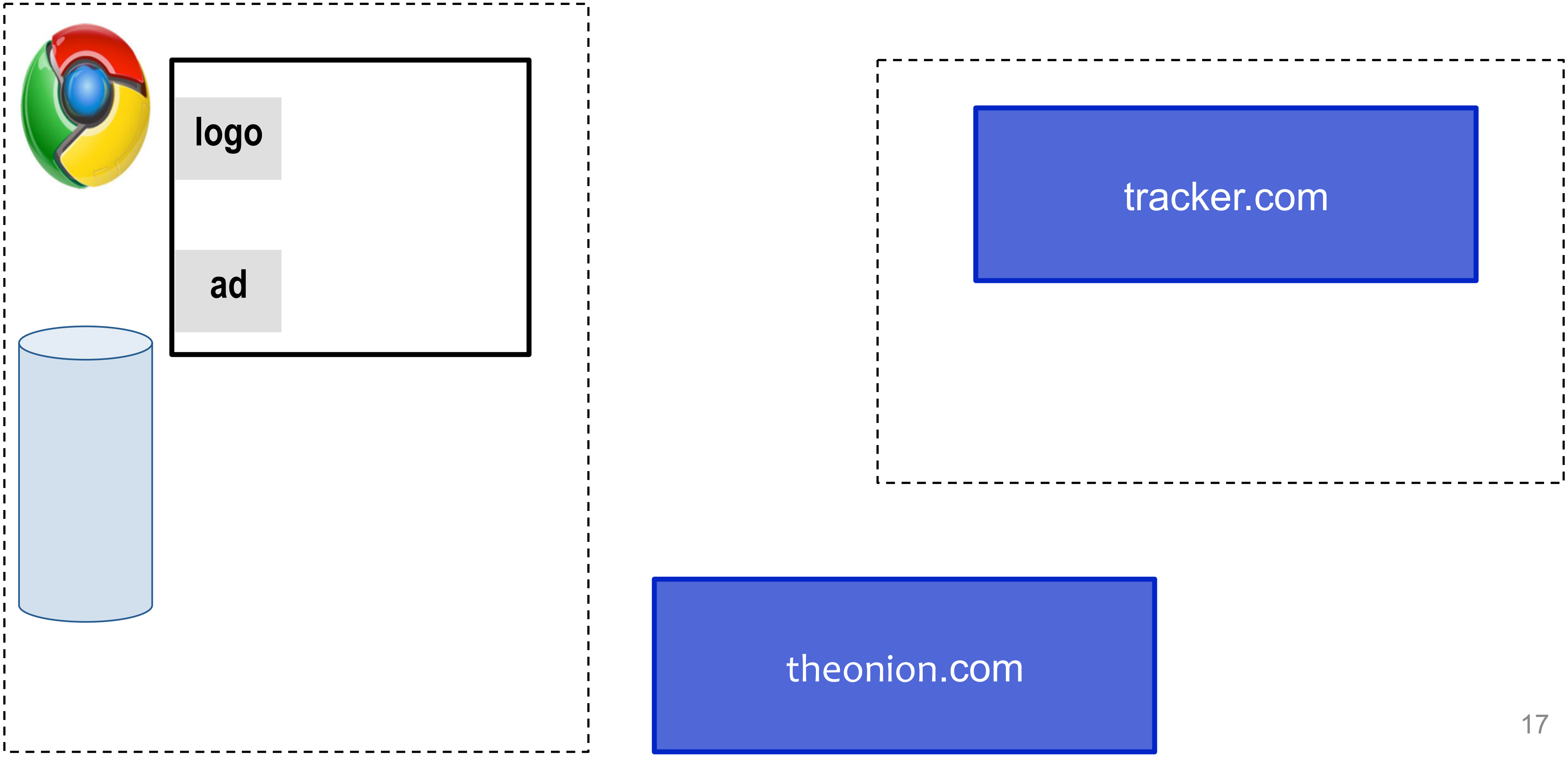
3rd Party Web Tracking.



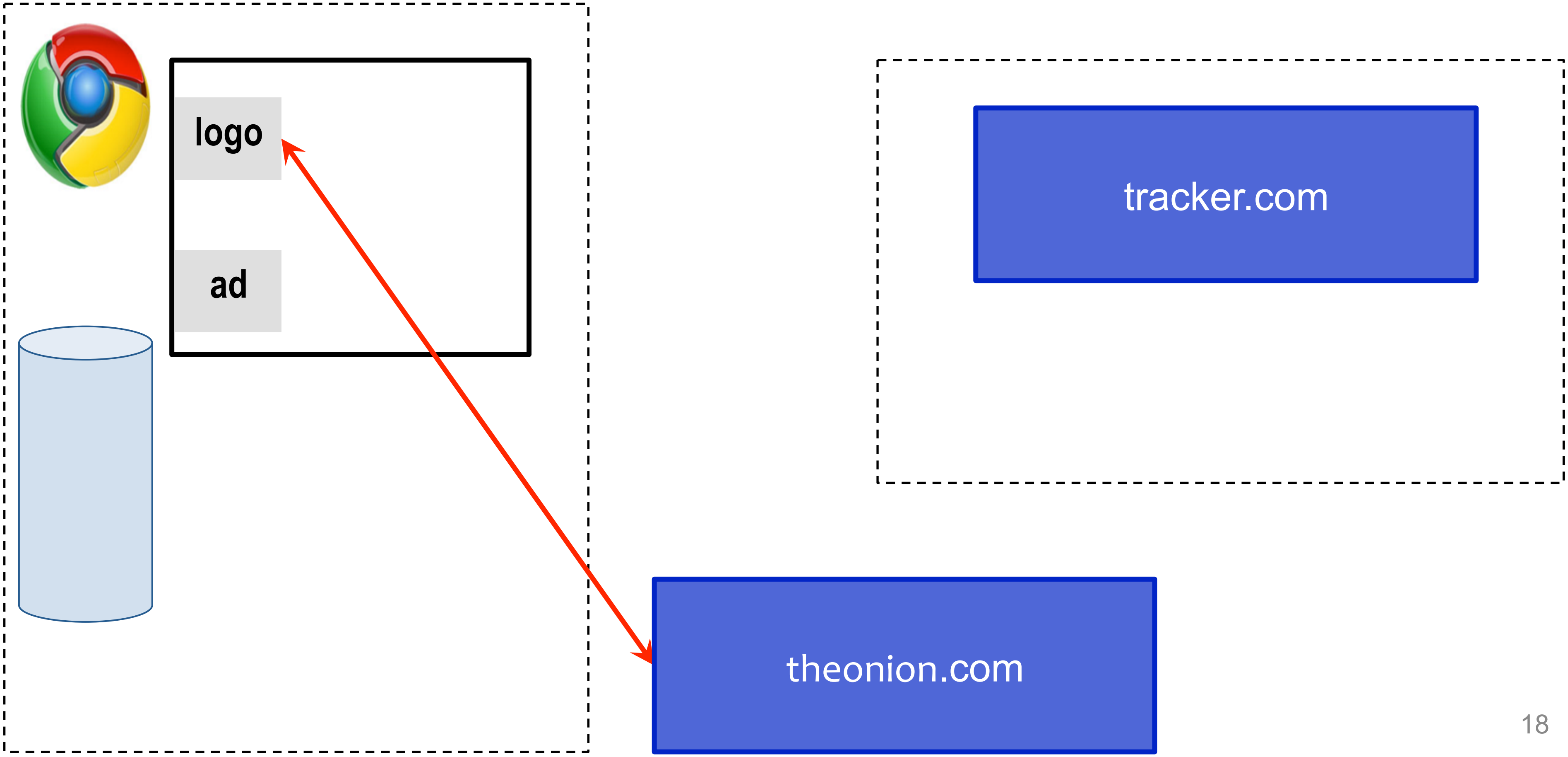
How does Third Party Web Tracking Work?



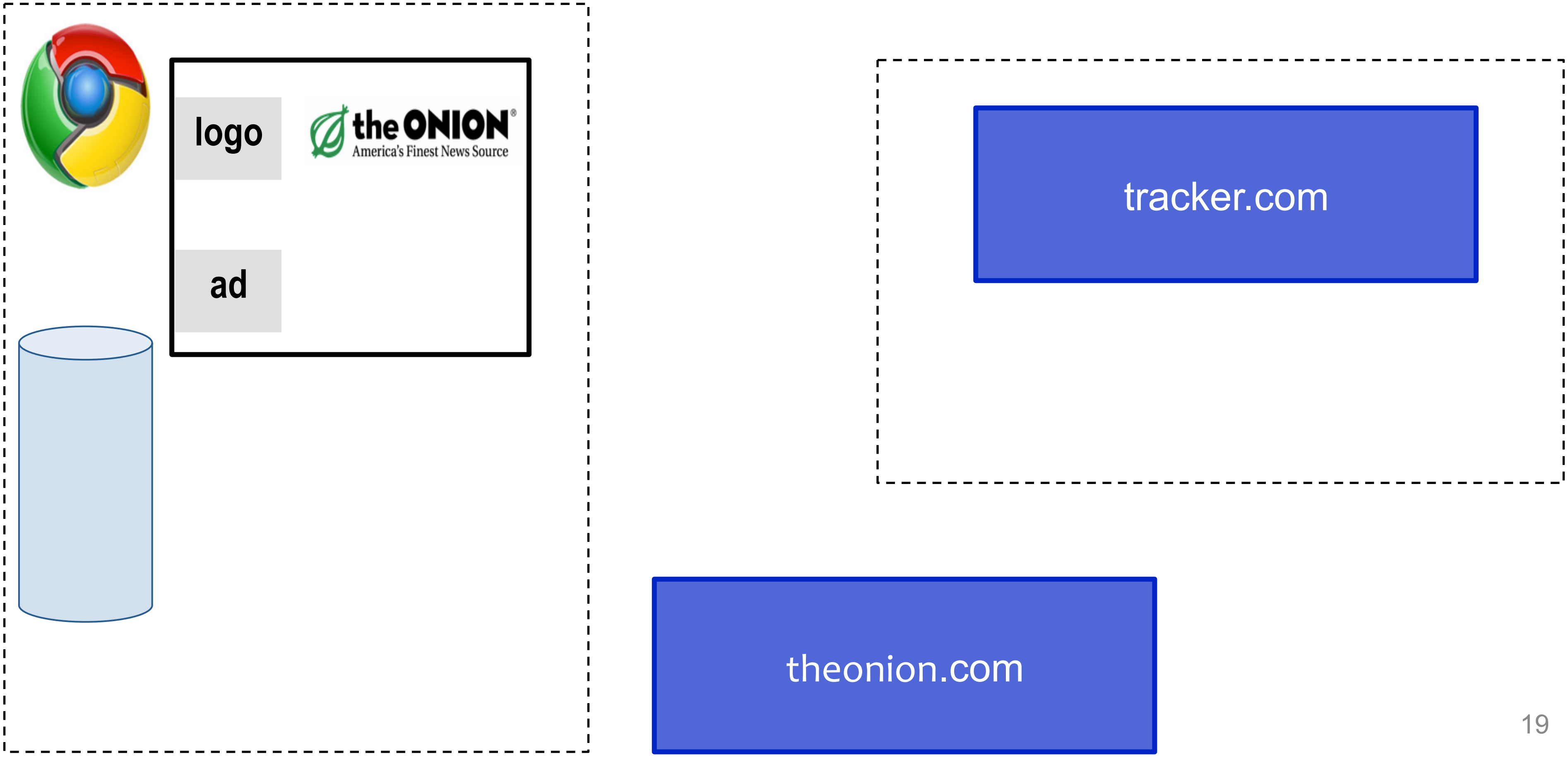
How does Third Party Web Tracking Work?



How does Third Party Web Tracking Work?



How does Third Party Web Tracking Work?



logo

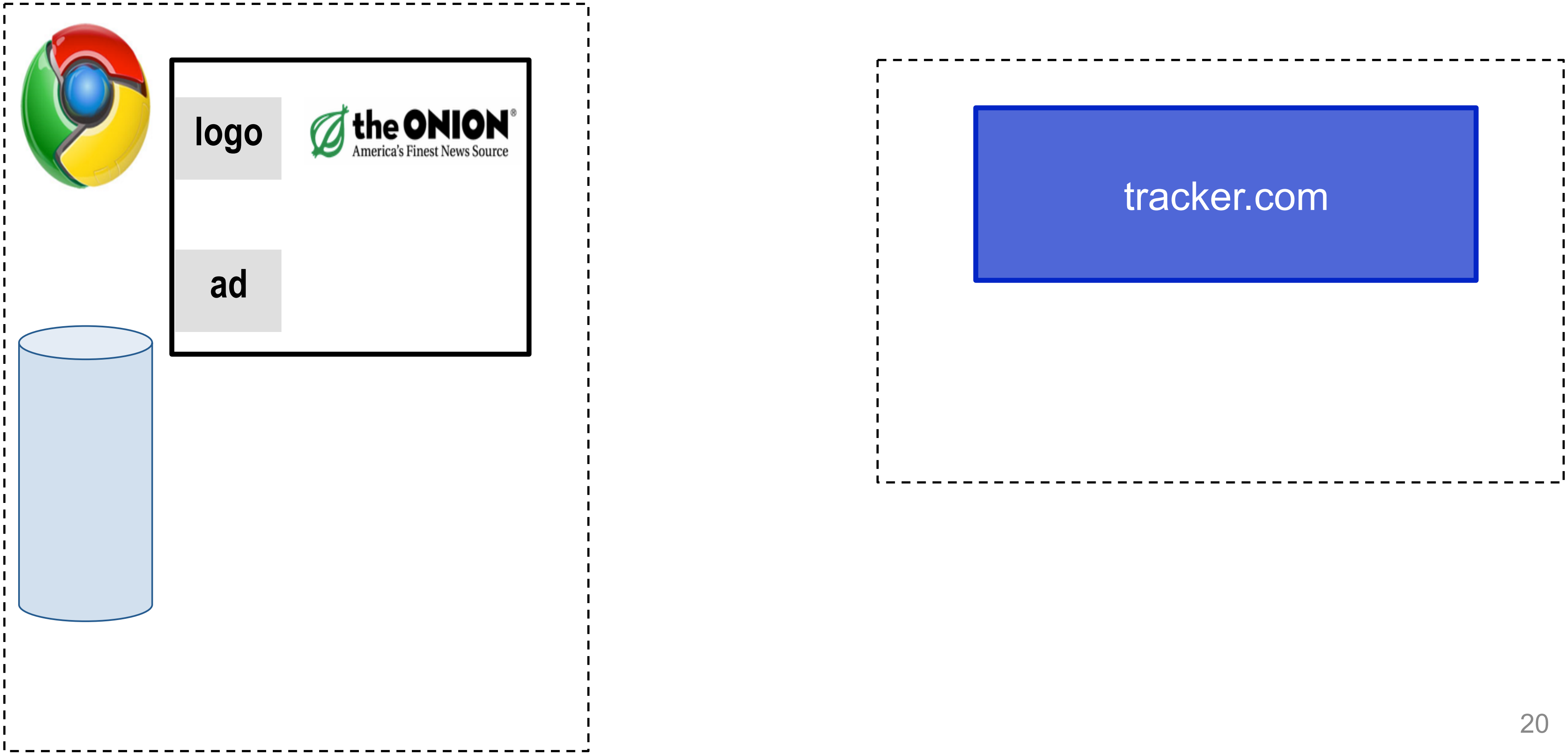
the ONION America's Finest News Source

ad

tracker.com

theonion.com

How does Third Party Web Tracking Work?

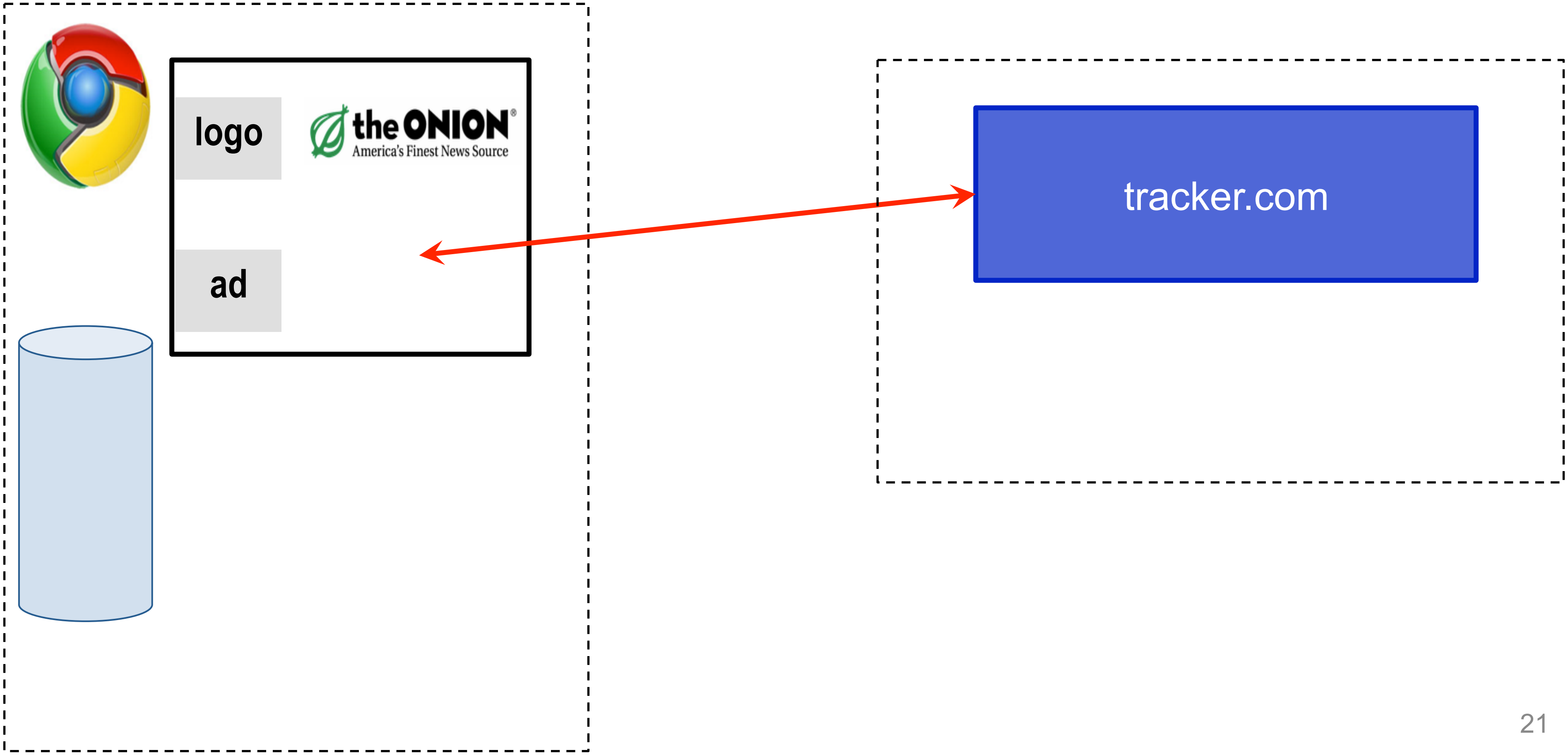


logo

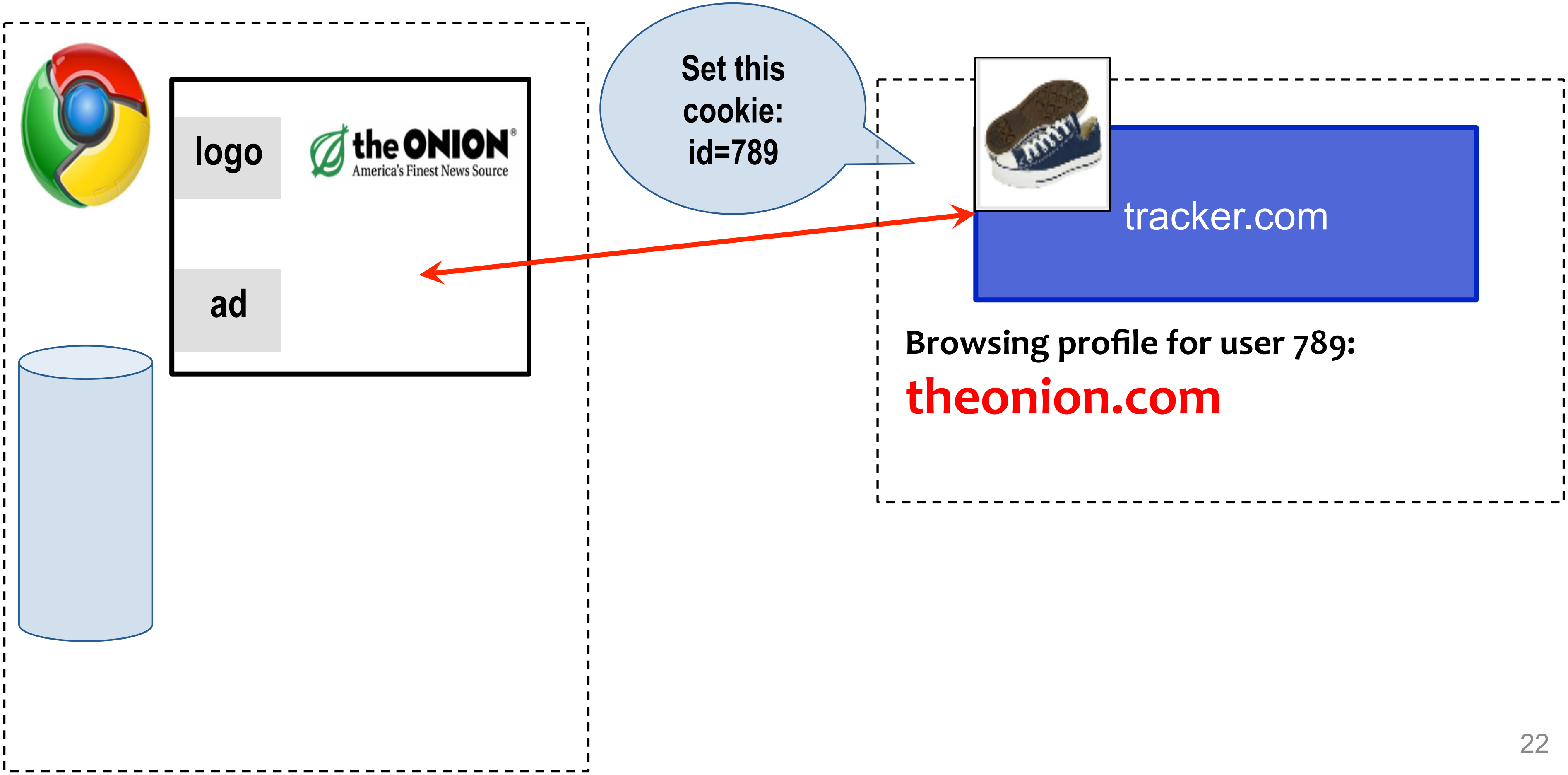
ad

tracker.com

How does Third Party Web Tracking Work?



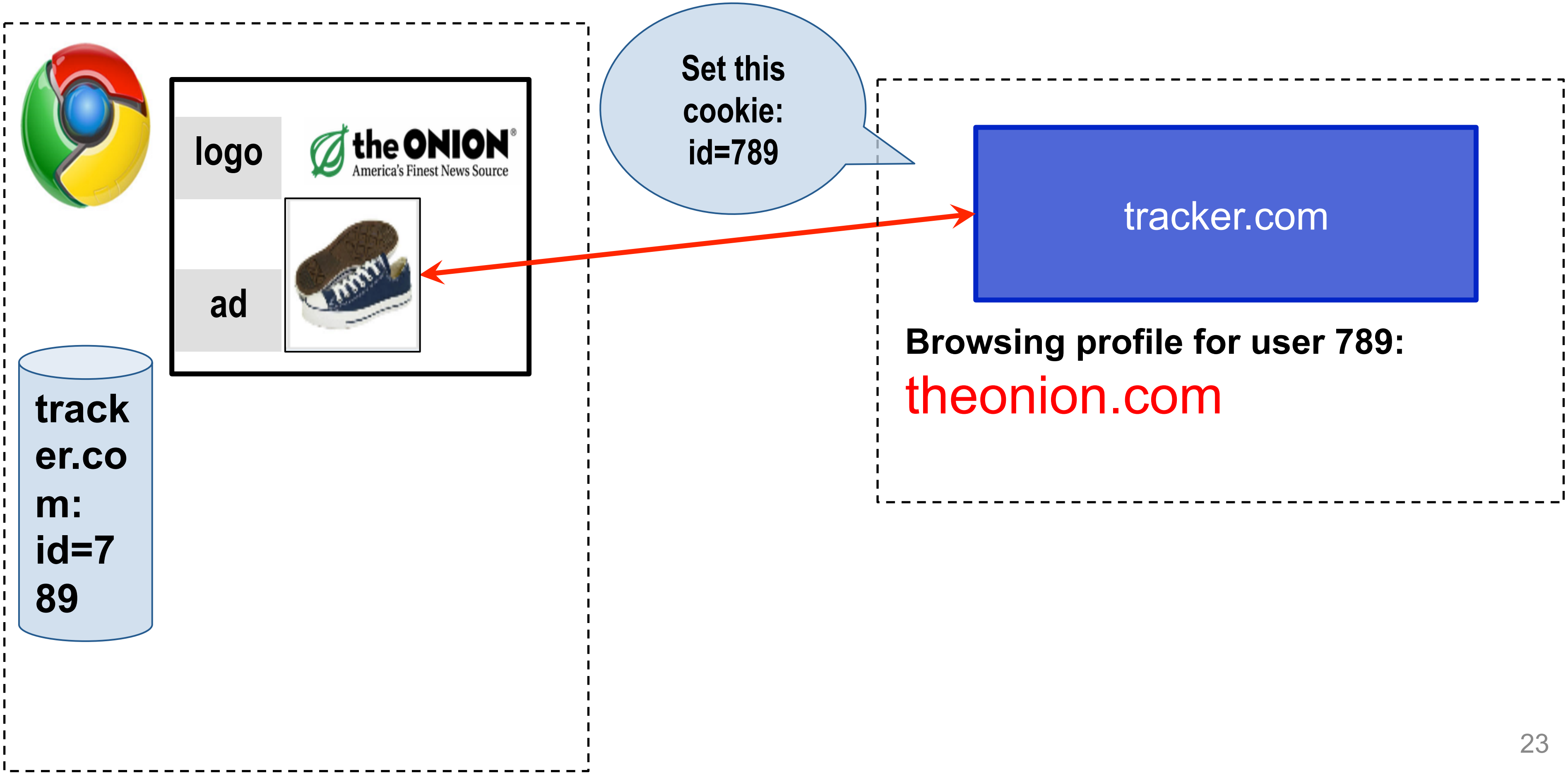
How does Third Party Web Tracking Work?



Browsing profile for user 789:

theonion.com

How does Third Party Web Tracking Work?



logo

the ONION America's Finest News Source

ad

Set this cookie: id=789

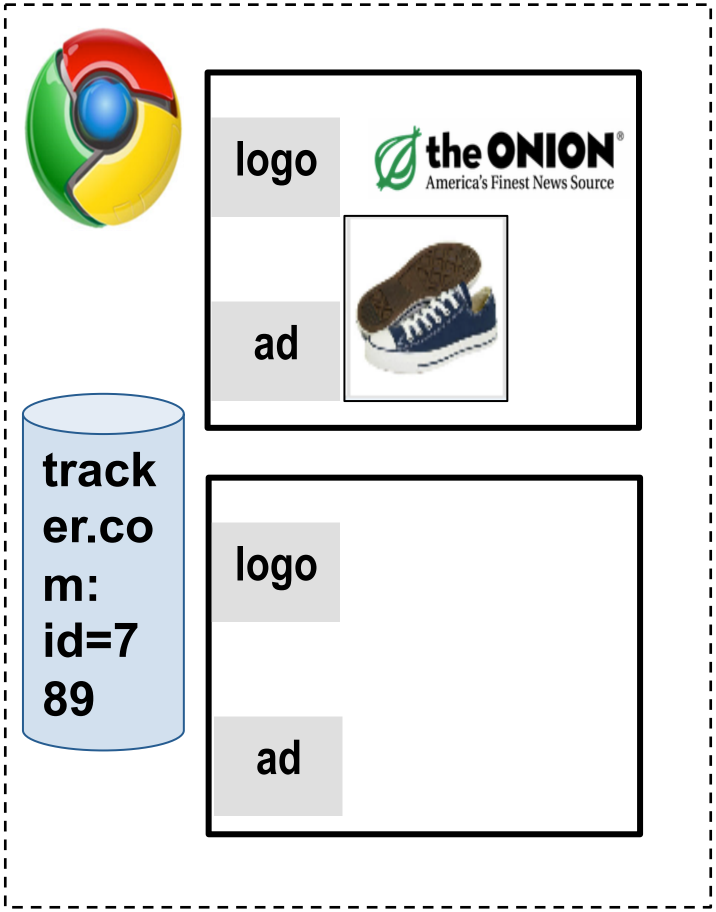
tracker.com

tracker.com: id=789

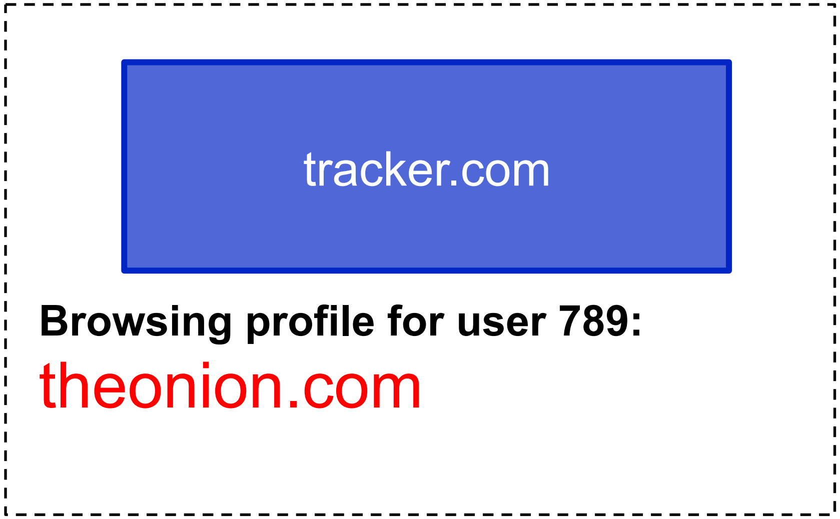
Browsing profile for user 789:

theonion.com

How does Third Party Web Tracking Work?



tracker.com: id=789

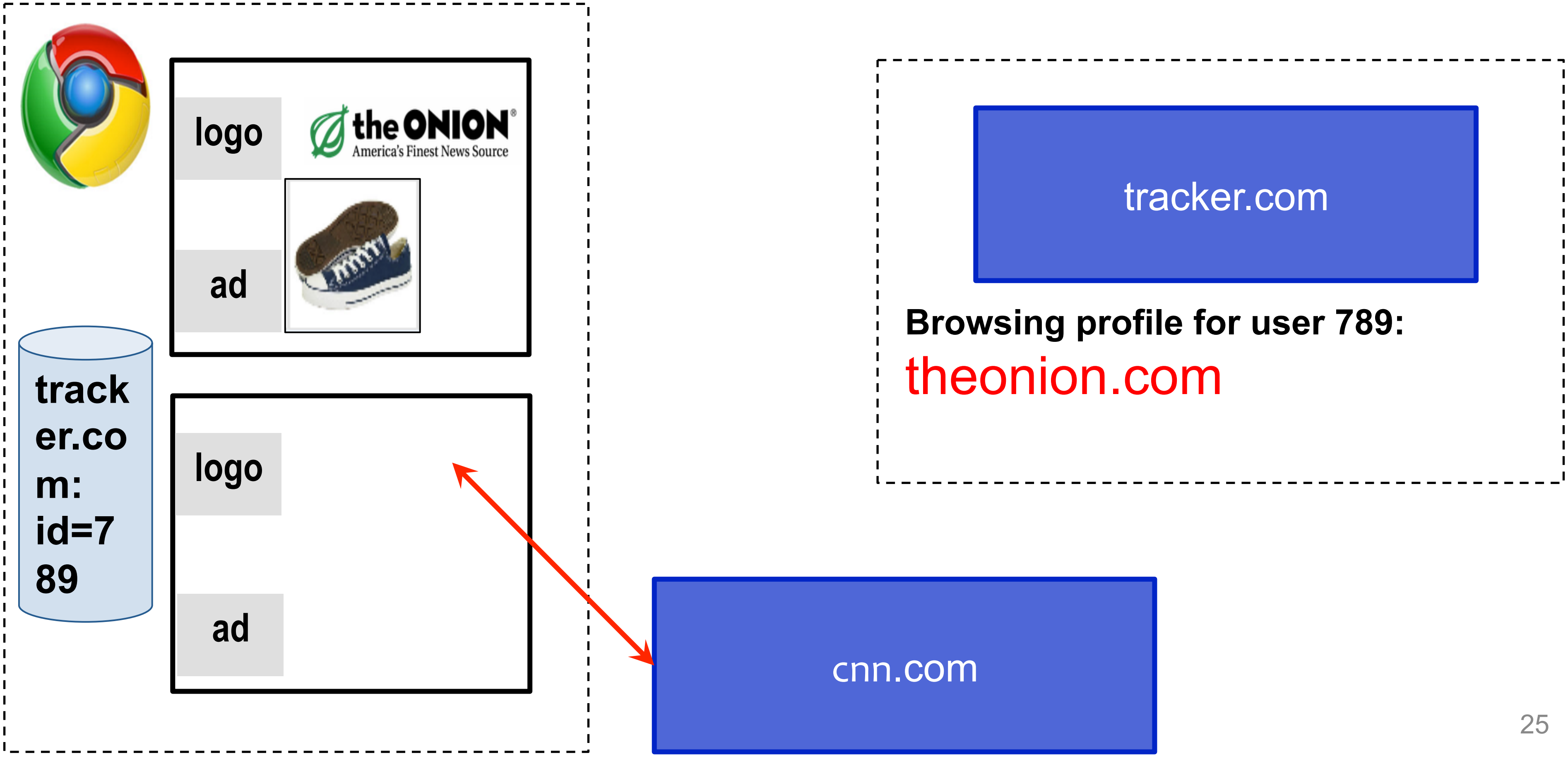


tracker.com

Browsing profile for user 789:

theonion.com

How does Third Party Web Tracking Work?



logo

the ONION America's Finest News Source

ad

tracker.com: id=789

logo

ad

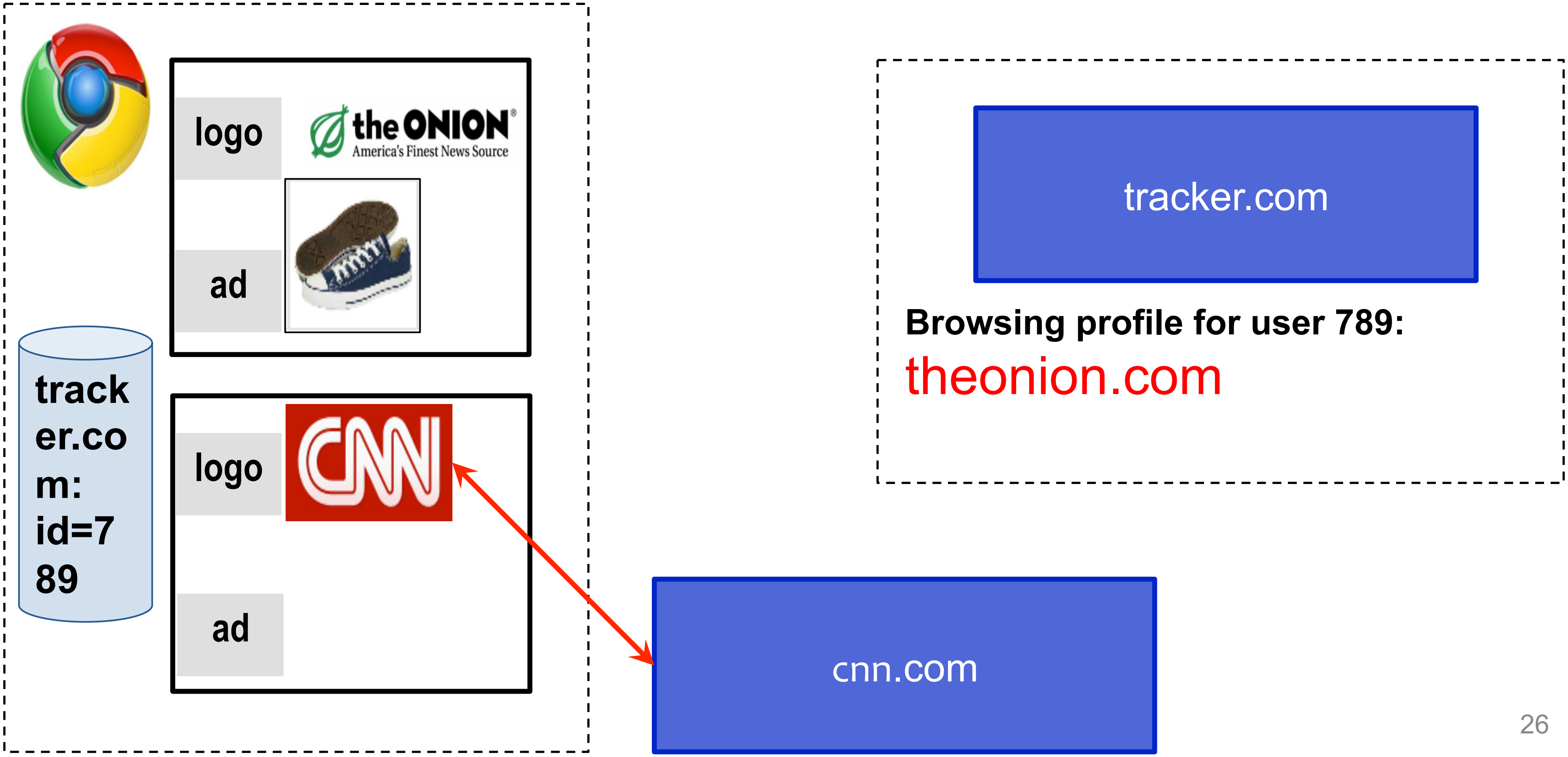
tracker.com

Browsing profile for user 789:

theonion.com

cnn.com

How does Third Party Web Tracking Work?



logo

ad

tracker.com: id=789

logo

ad

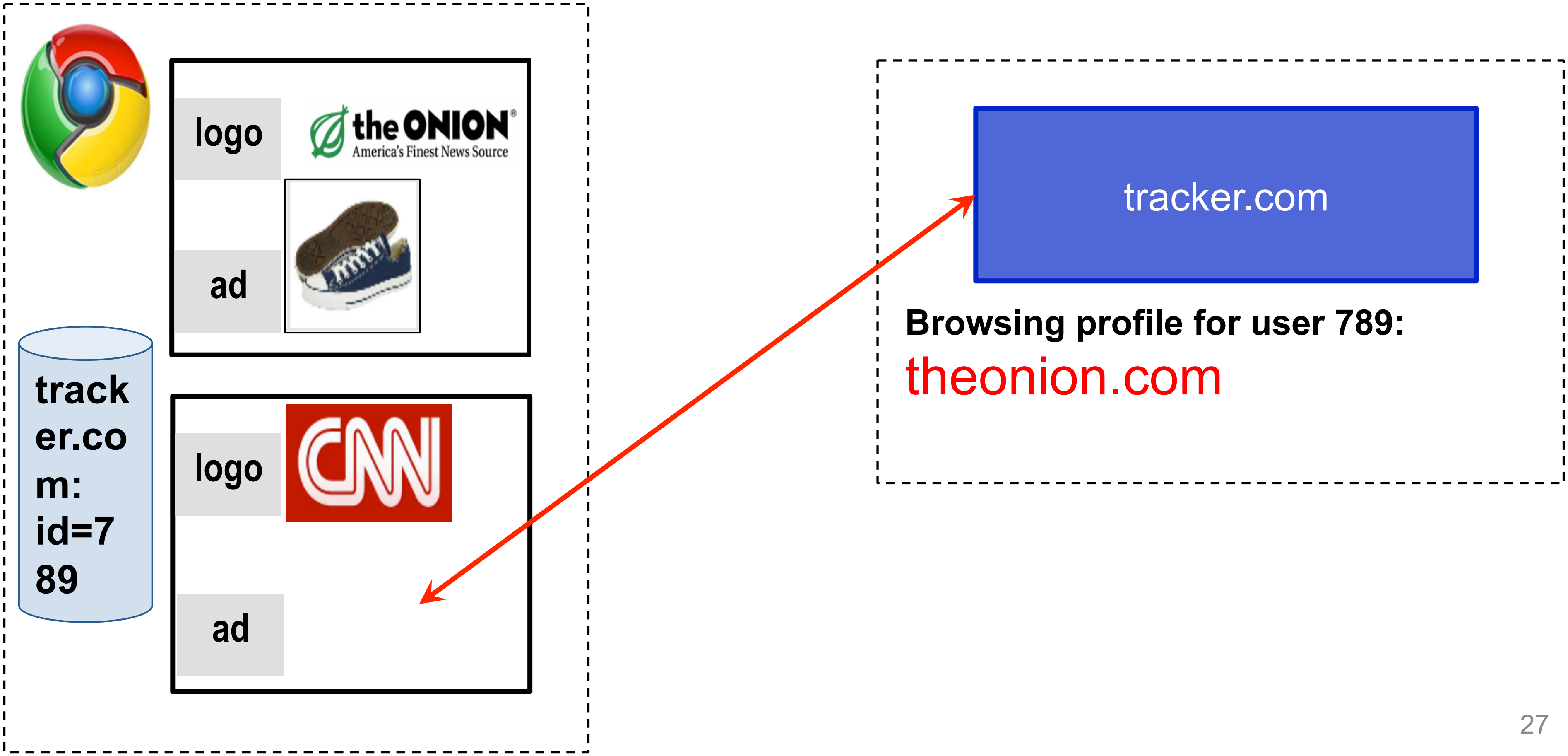
tracker.com

Browsing profile for user 789:

theonion.com

cnn.com

How does Third Party Web Tracking Work?



logo

the ONION America's Finest News Source

ad

tracker.com: id=789

logo

CNN

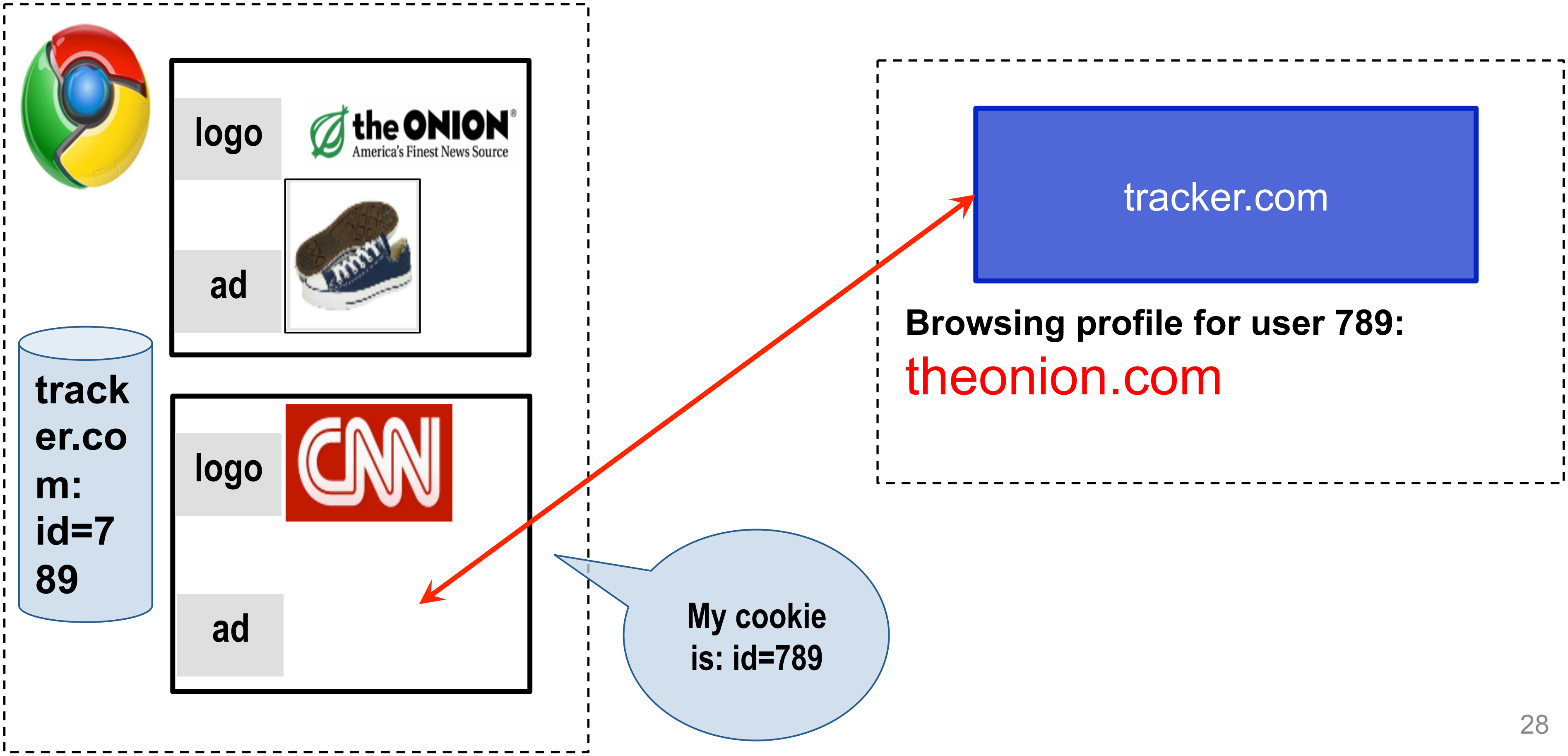
ad

tracker.com

Browsing profile for user 789:

theonion.com

How does Third Party Web Tracking Work?



logo

ad

logo

ad

tracker.com: id=789

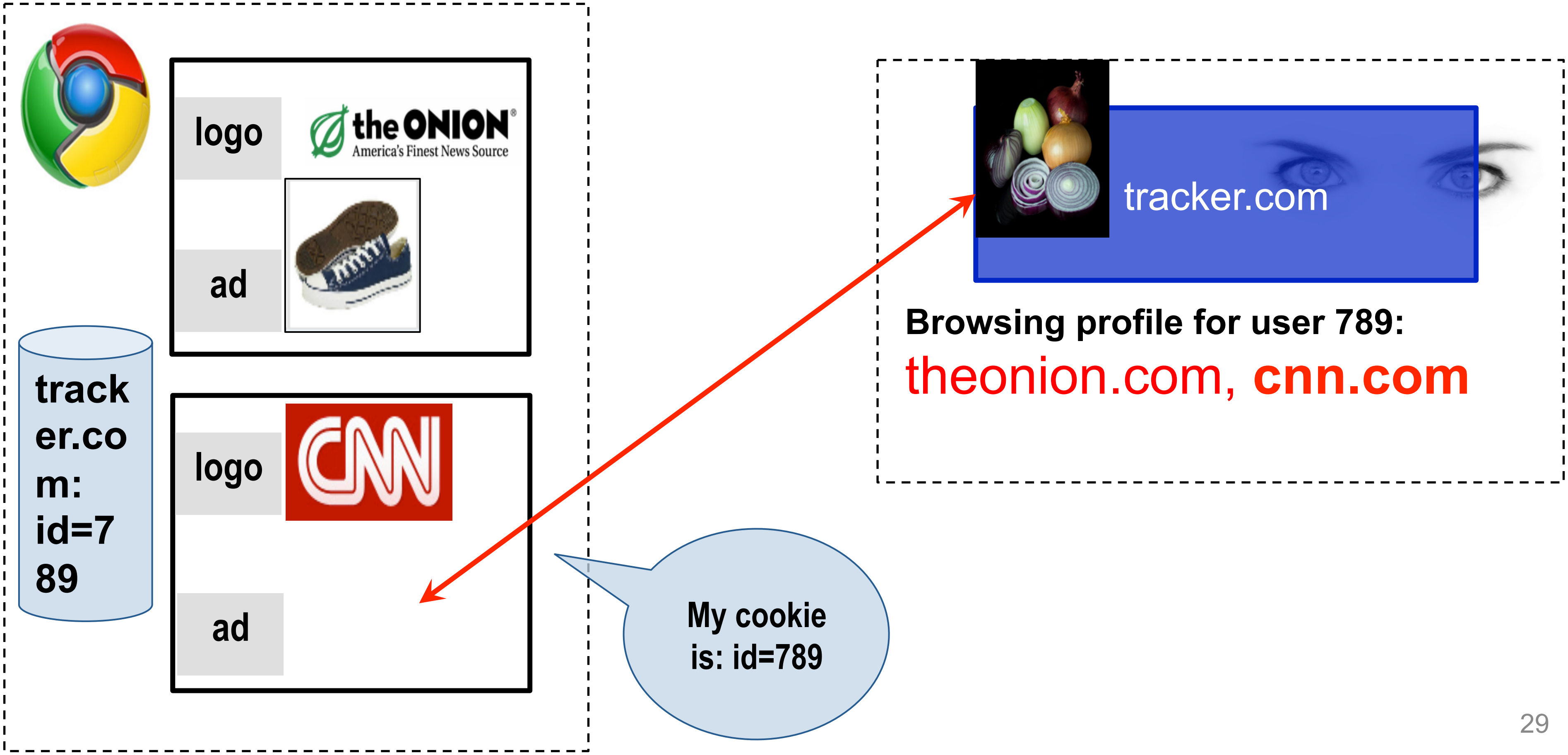
My cookie is: id=789

tracker.com

Browsing profile for user 789:

theonion.com

How does Third Party Web Tracking Work?

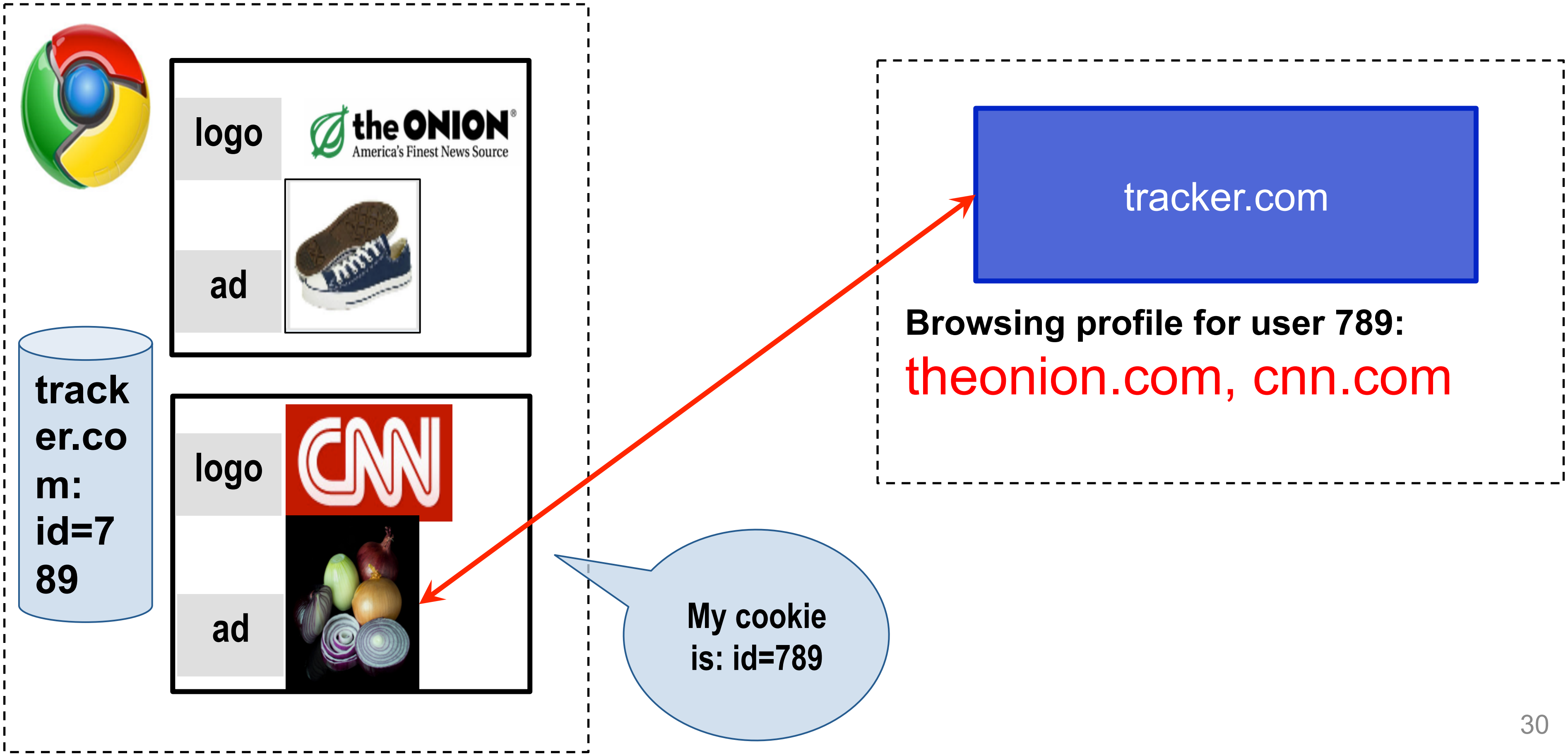


My cookie is: id=789

Browsing profile for user 789:

theonion.com, **cnn.com**

How does Third Party Web Tracking Work?



logo

ad

logo

ad

tracker.com: id=789

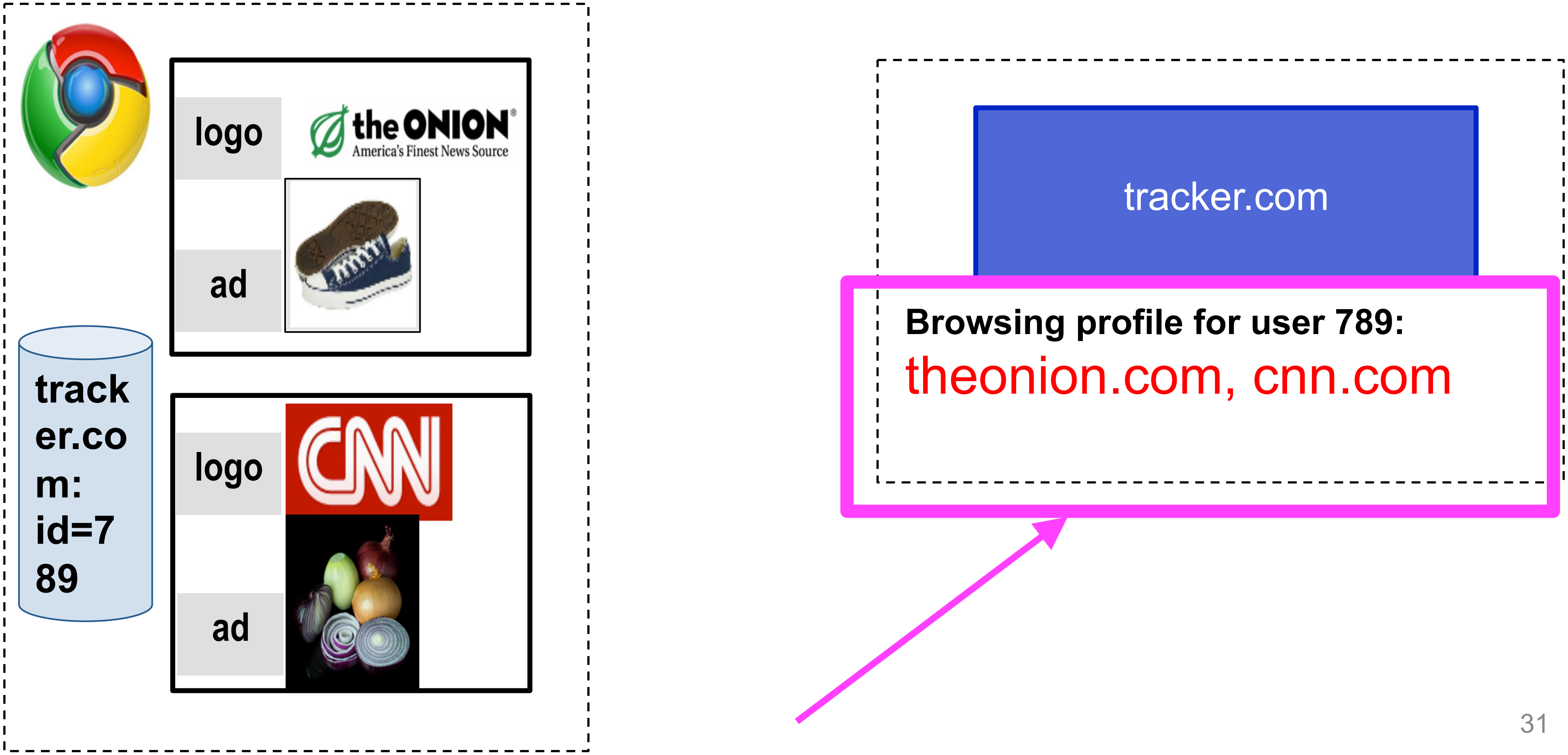
My cookie is: id=789

tracker.com

Browsing profile for user 789:

theonion.com, cnn.com

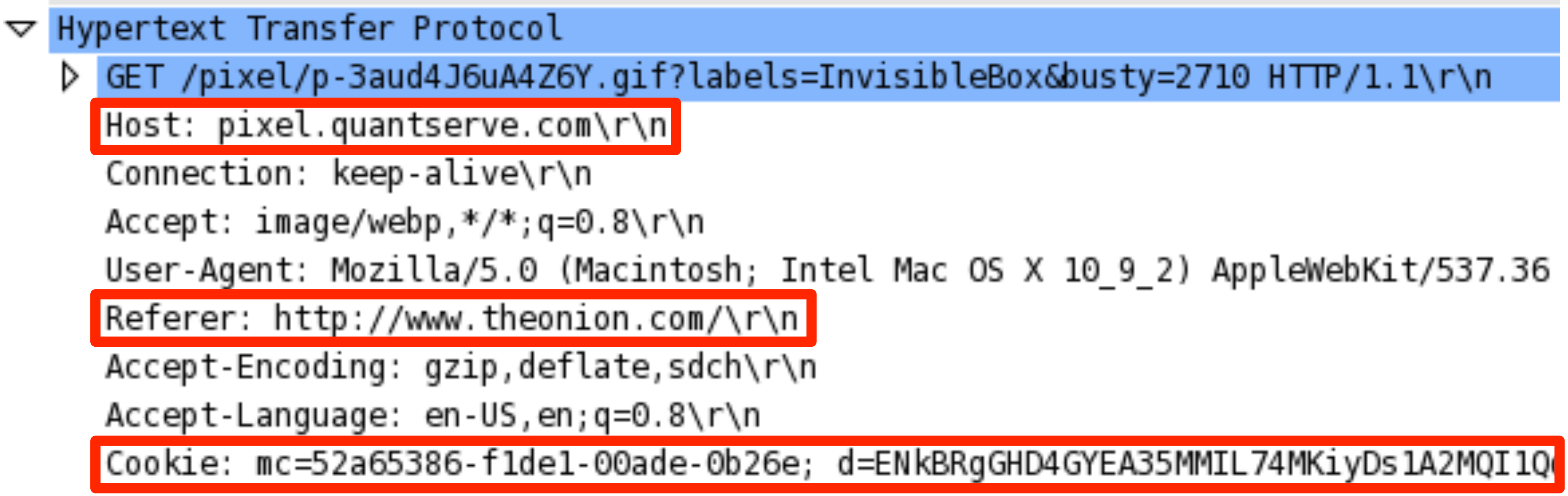
How does Third Party Web Tracking Work?



Browsing profile for user 789:
theonion.com, cnn.com

Basic Tracking Mechanisms

- Tracking requires:
  (1) re-identifying a user.
  (2) telling the tracker which first party site we're on



```
▽ 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
```

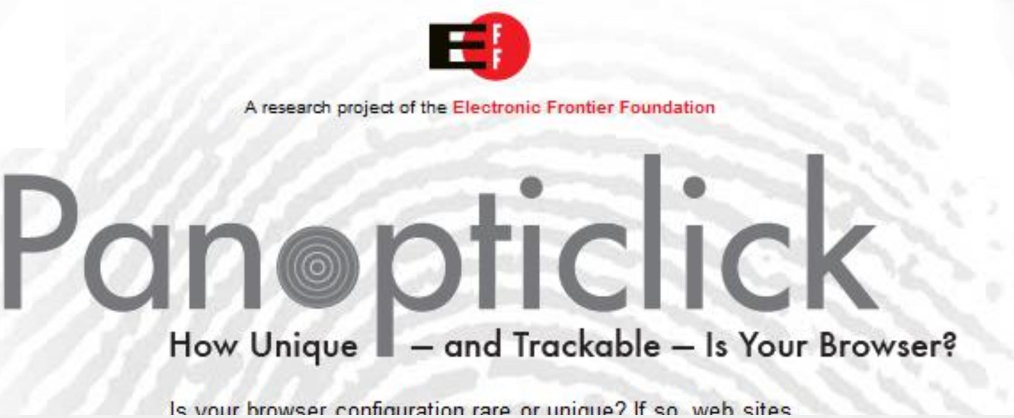
Tracking Technologies

- 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!)

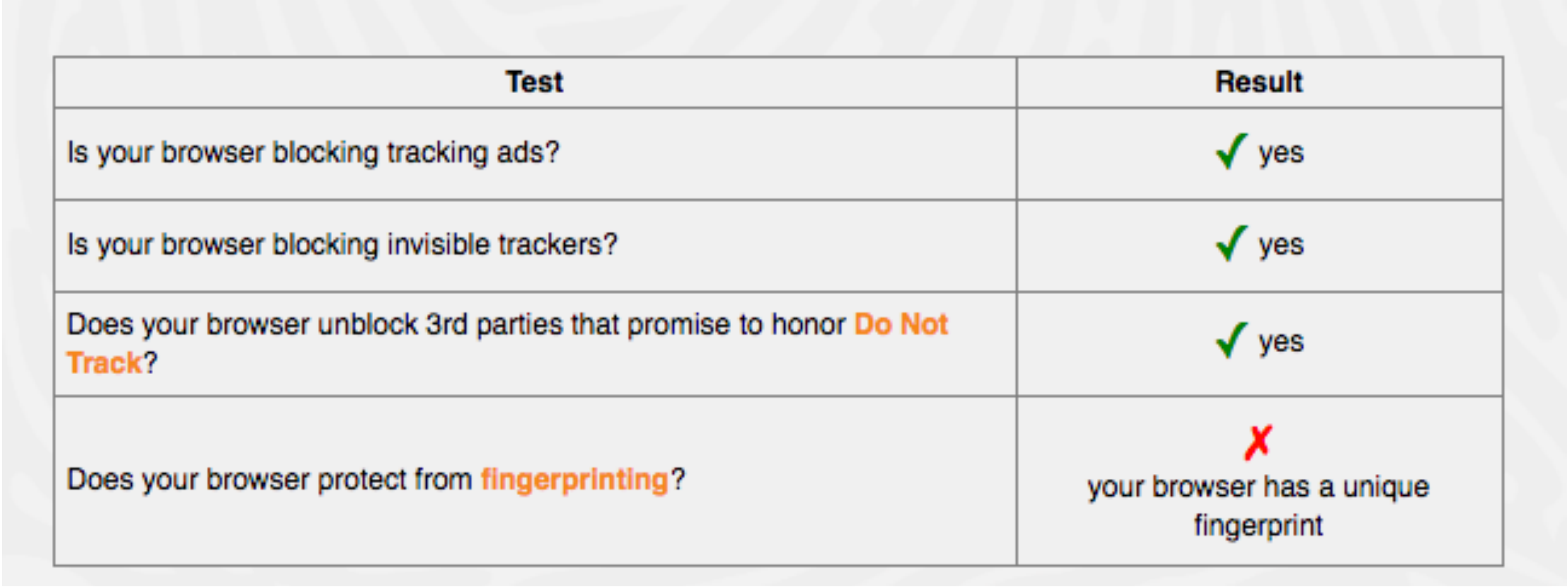


A research project of the Electronic Frontier Foundation

Panopticlick

How Unique — and Trackable — Is Your Browser?

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



Test	Result
Is your browser blocking tracking ads?	✓ yes
Is your browser blocking invisible trackers?	✓ yes
Does your browser unblock 3rd parties that promise to honor Do Not Track?	✓ yes
Does your browser protect from fingerprinting?	✗ your browser has a unique fingerprint

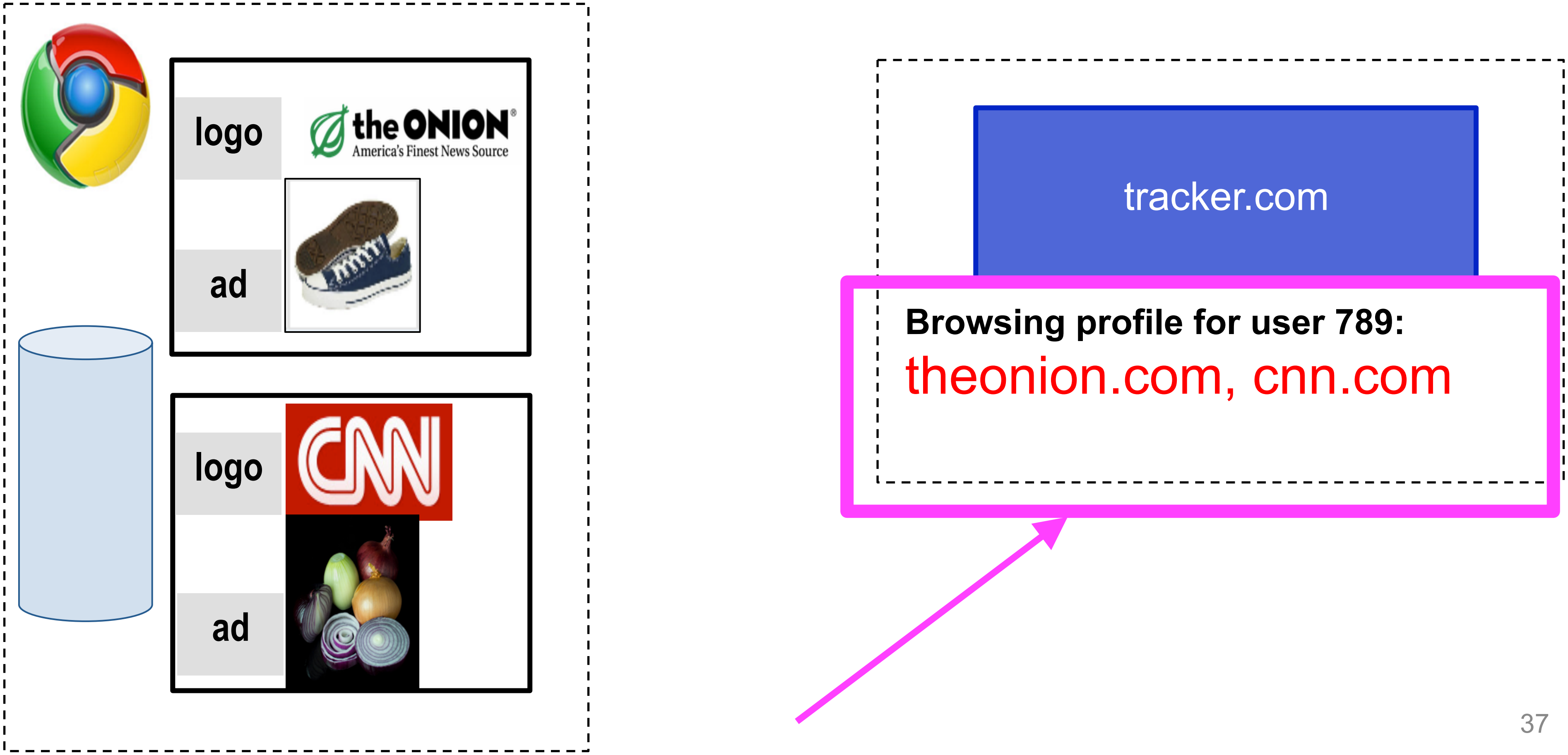
Panopticlick Example

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; ars)
(Acroba on/
vnd.ado 9 Data
File; app ;
application/x-vnd.google.oneclickctrl.8;). Plugin 3: MicrosoftA® 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

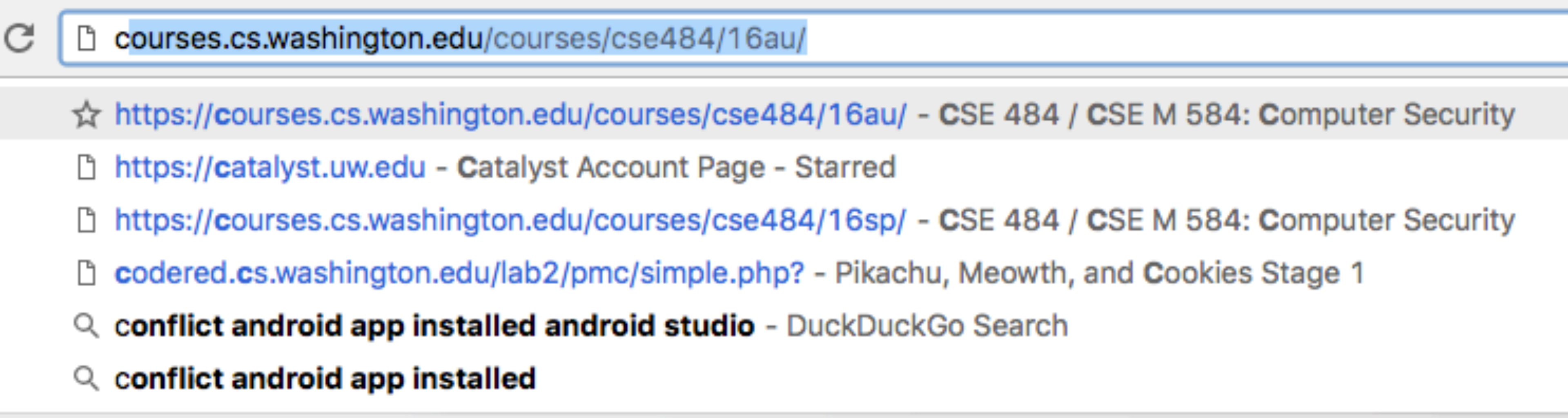
With Flash or Java, 94% are unique

How does Third Party Web Tracking Work?



Browsing profile for user 789:
theonion.com, cnn.com

History Sniffing



courses.cs.washington.edu/courses/cse484/16au/

☆ https://courses.cs.washington.edu/courses/cse484/16au/ - CSE 484 / CSE M 584: Computer Security

https://catalyst.uw.edu - Catalyst Account Page - Starred

https://courses.cs.washington.edu/courses/cse484/16sp/ - CSE 484 / CSE M 584: Computer Security

codered.cs.washington.edu/lab2/pmc/simple.php? - Pikachu, Meowth, and Cookies Stage 1

conflict android app installed android studio - DuckDuckGo Search

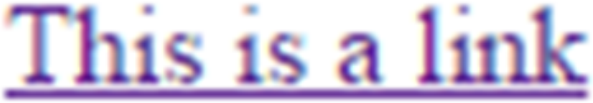
conflict android app installed

History Sniffing

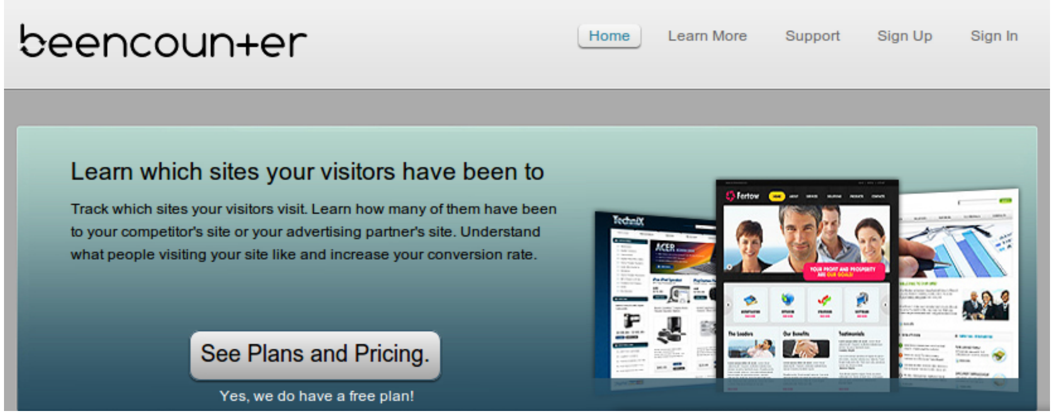
How can a webpage figure out which sites you visited previously?

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

This is a link



This is a link



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!

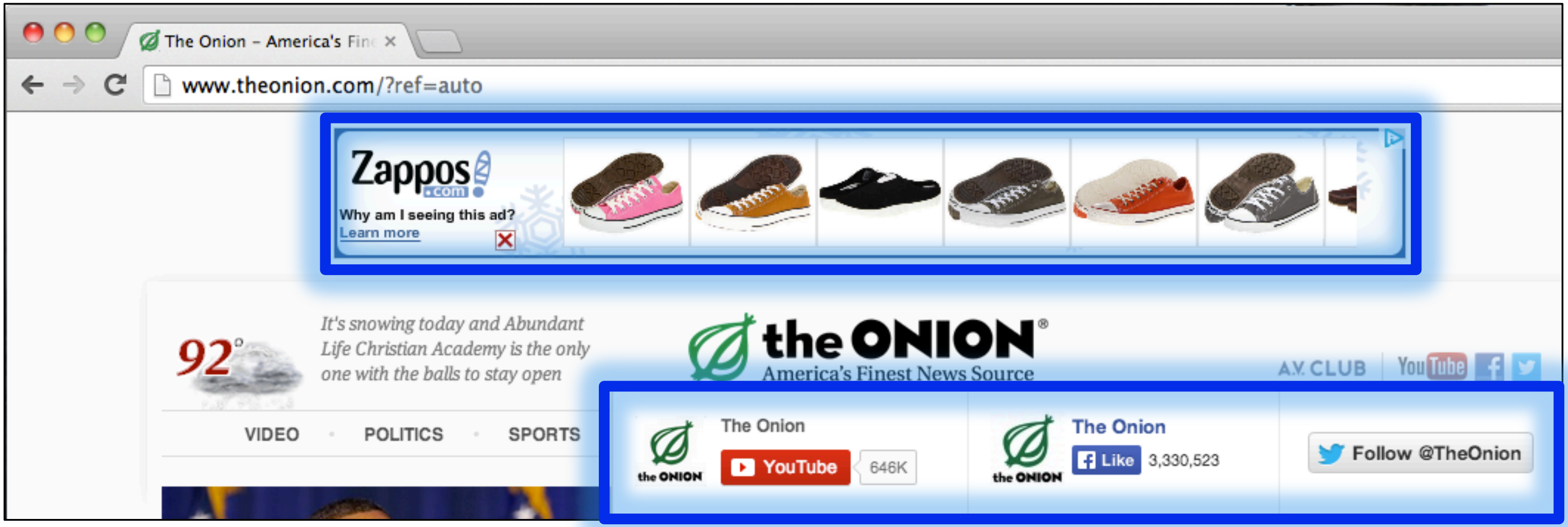
Understanding the Tracking Ecosystem

- 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]*

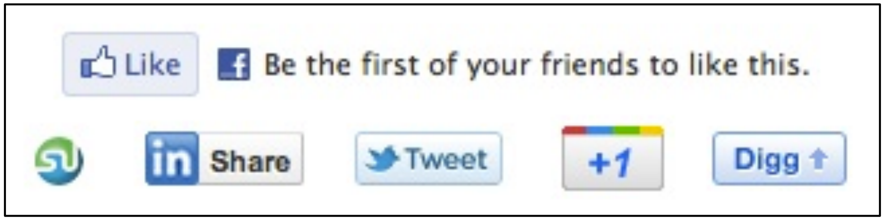
- In the wild, tracking is much more complicated.
- (1) Trackers don't just use cookies.
  - Flash cookies, HTML5 LocalStorage, etc.
- (2) Trackers exhibit different behaviors.
  - Within-site vs. cross-site.
  - Anonymous vs. non-anonymous.
  - Specific behavior types:
    analytics, vanilla, forced,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

Other Trackers?

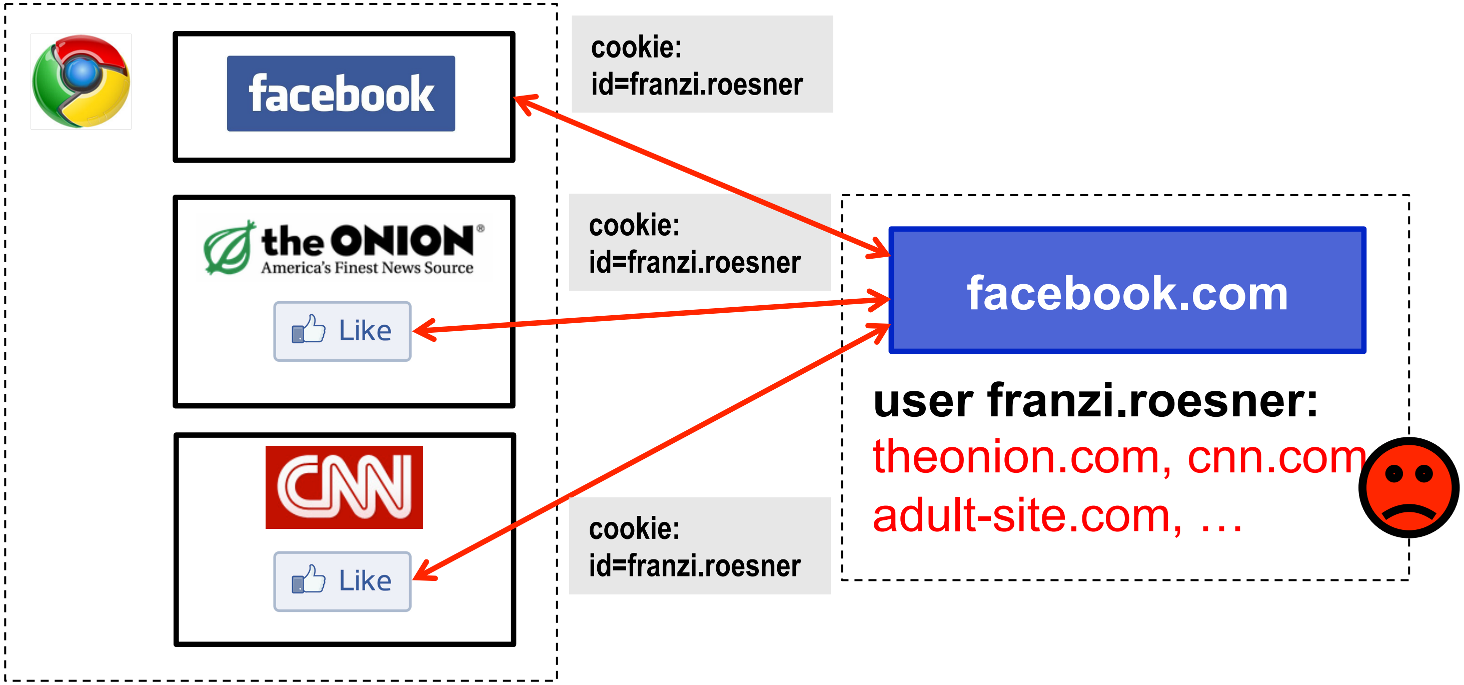


It's snowing today and Abundant Life Christian Academy is the only one with the balls to stay open

“Personal” Trackers



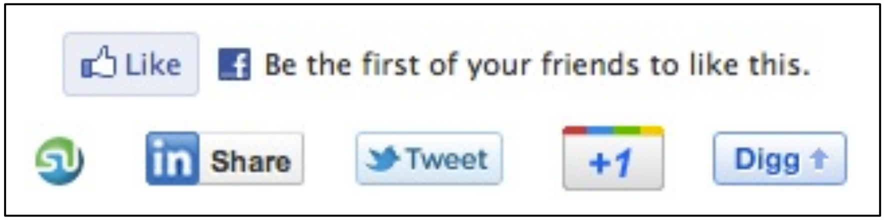
Personal Tracking



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

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

Outline

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

Measurement Study (2011)

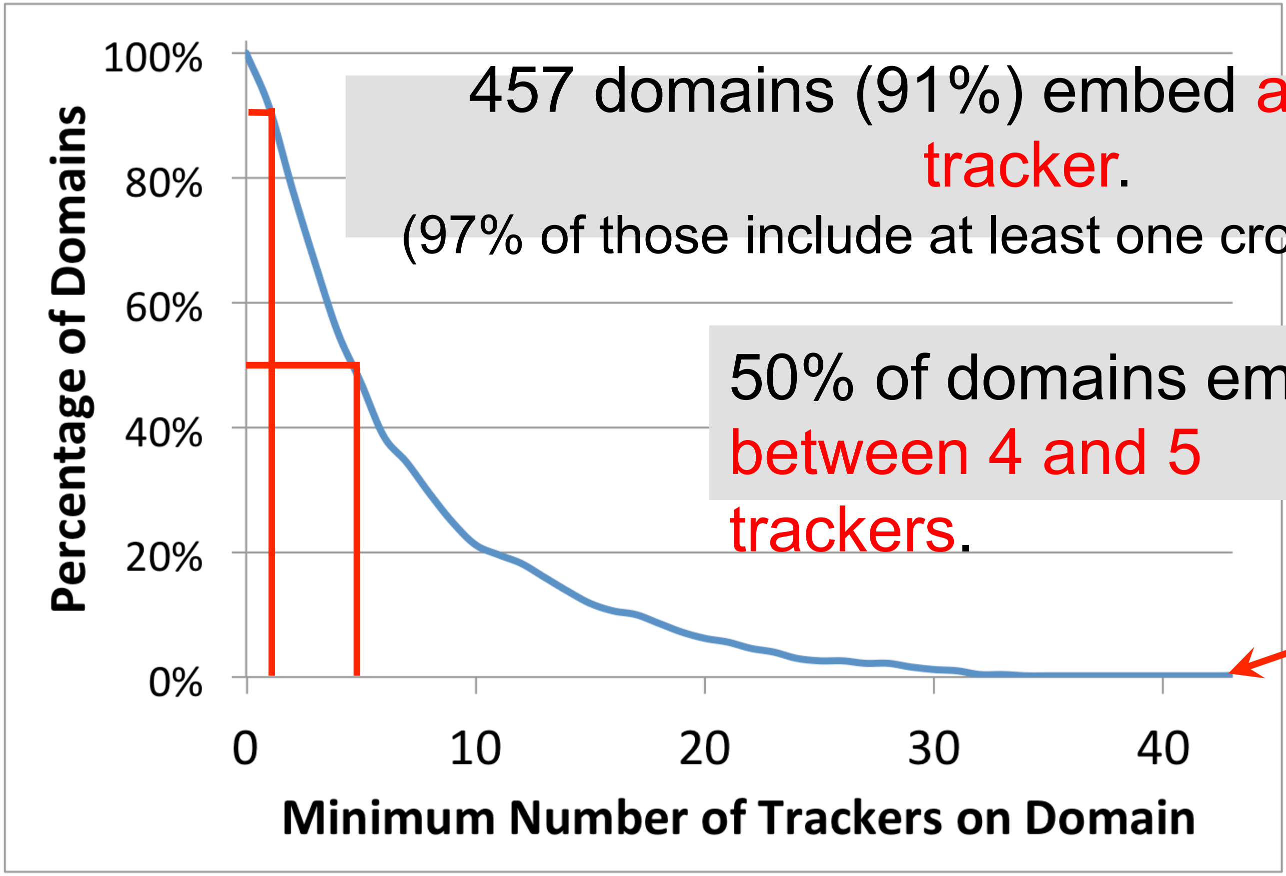
- **Questions:**
  - How prevalent is tracking (of different types)?
  - How much of a user's browsing history is captured?
  - How effective are defenses?

- **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)



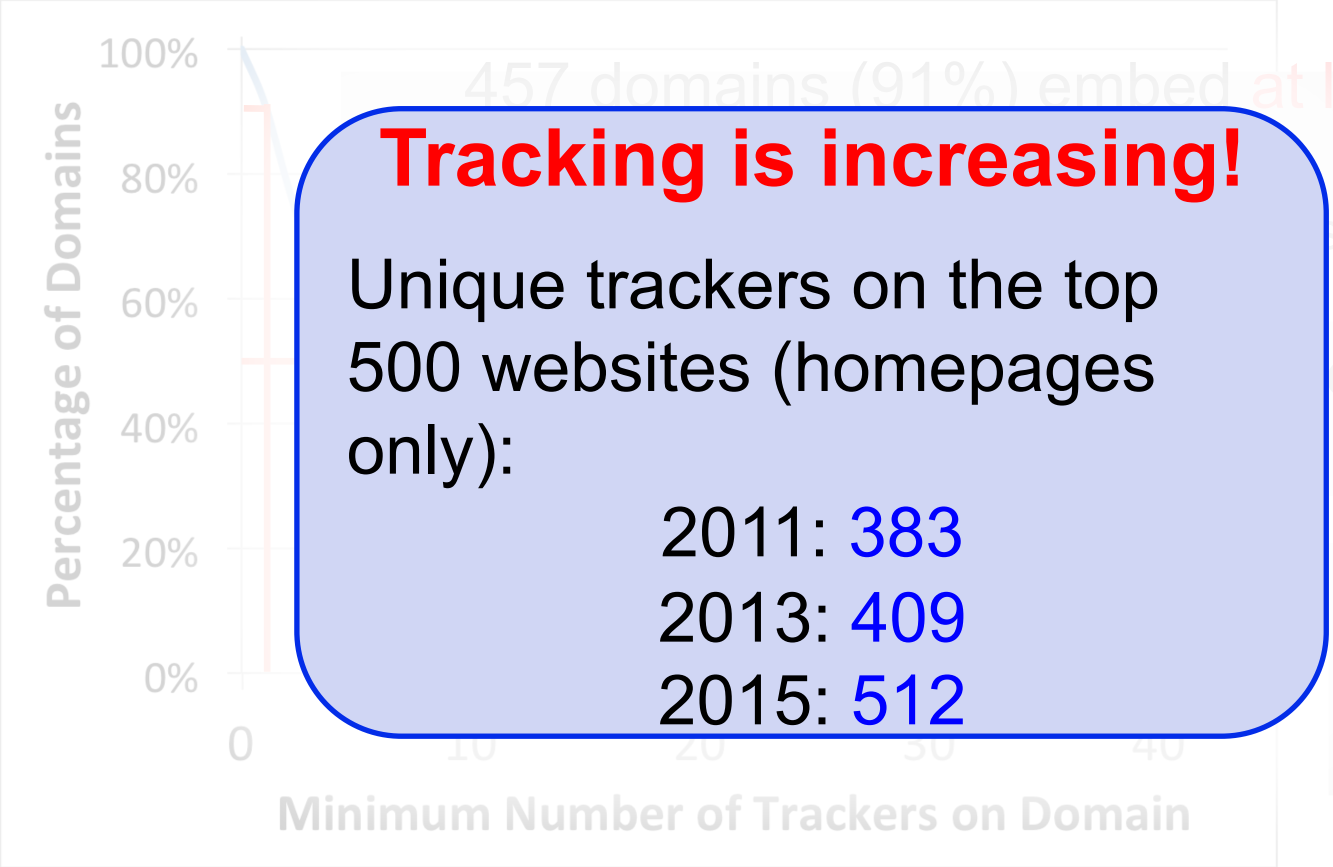
457 domains (91%) embed at least one tracker.
(97% of those include at least one cross-site tracker.)

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)



457 domains (91%) embed at least one

-site tracker.)

ed

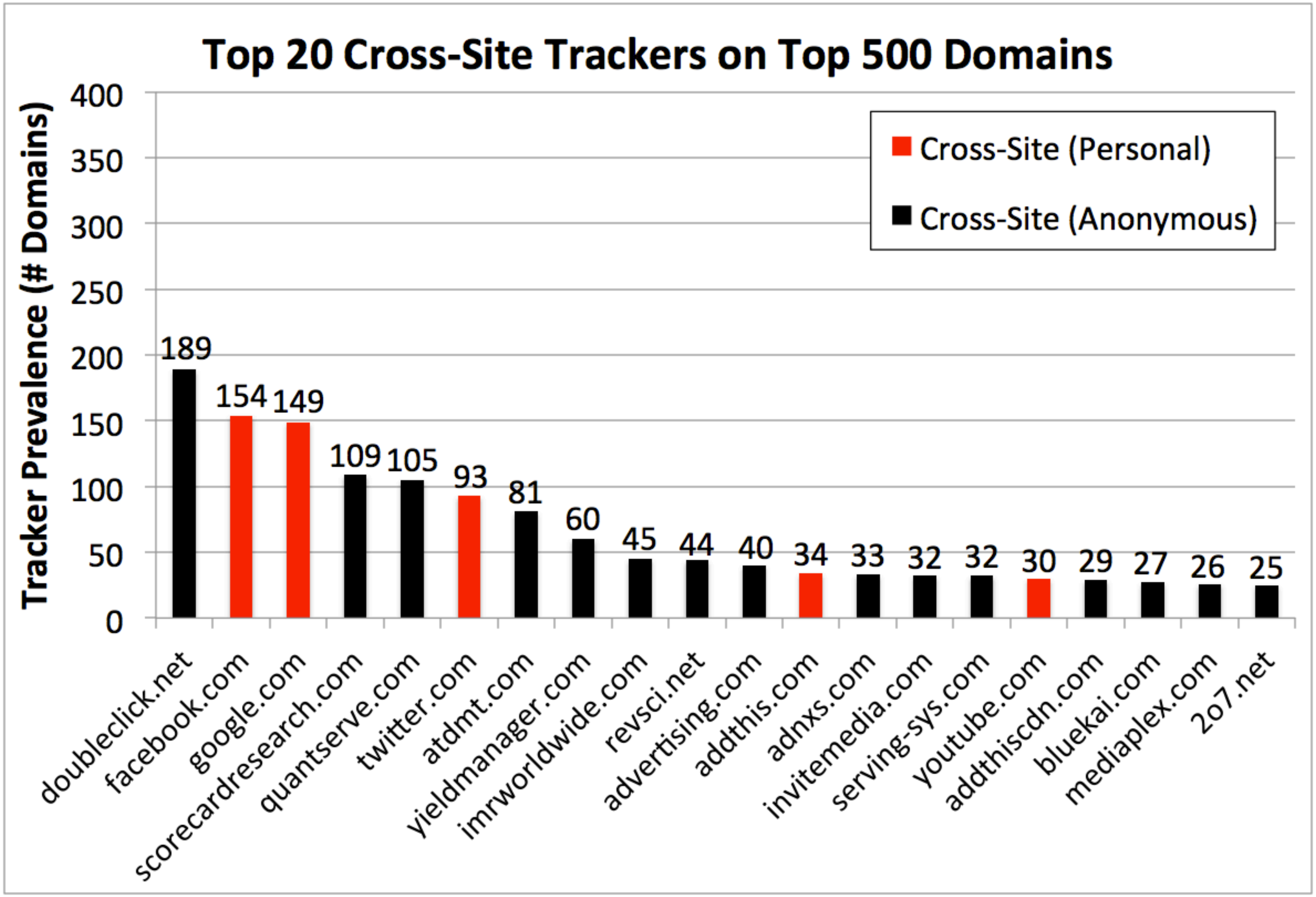
One domain includes 43 trackers.

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)



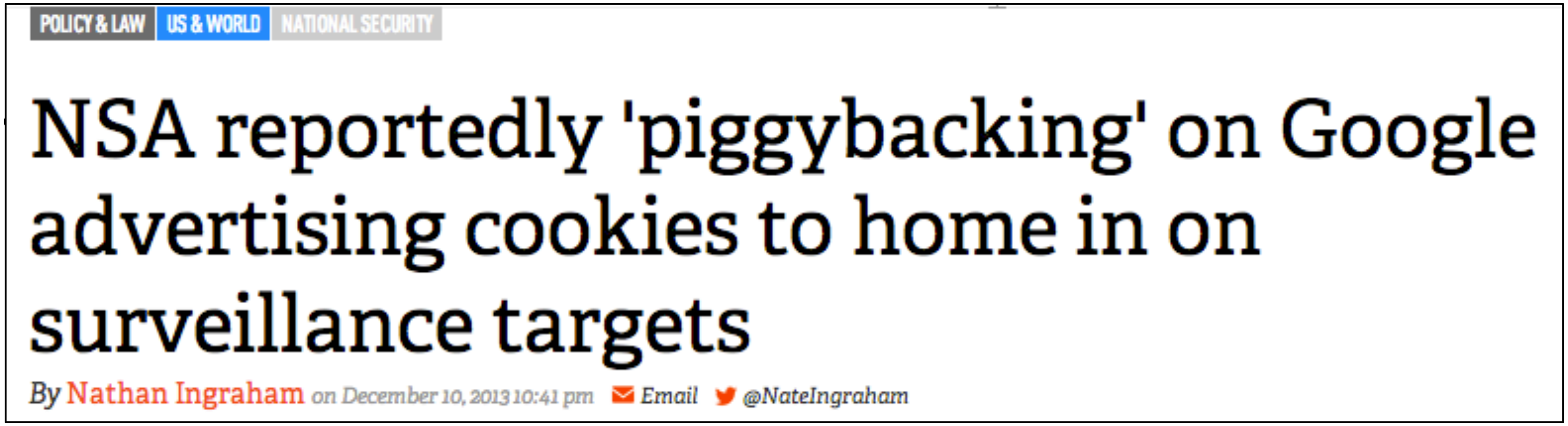
How are users affected?

- 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

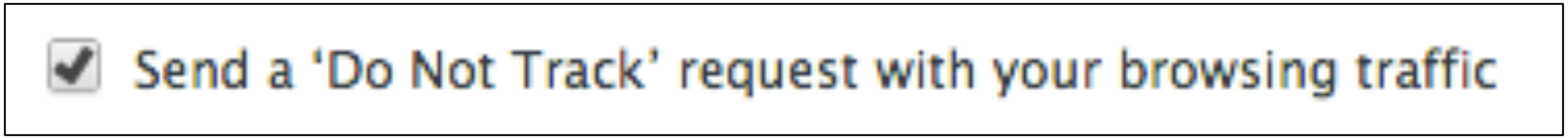
- Surprisingly little use of these mechanisms!
- Of 524 trackers on Alexa Top 500:
  - Only 5 set unique identifiers in LocalStorage
  - 35 set unique identifiers in Flash cookies
- Respawning:
  - LS → Cookie: 1 case; Cookie → LS: 3 cases
  - Flash→ Cookie: 6 cases; Cookie → Flash: 7 cases

Outline

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

Defenses to Reduce Tracking

- Do Not Track proposal?



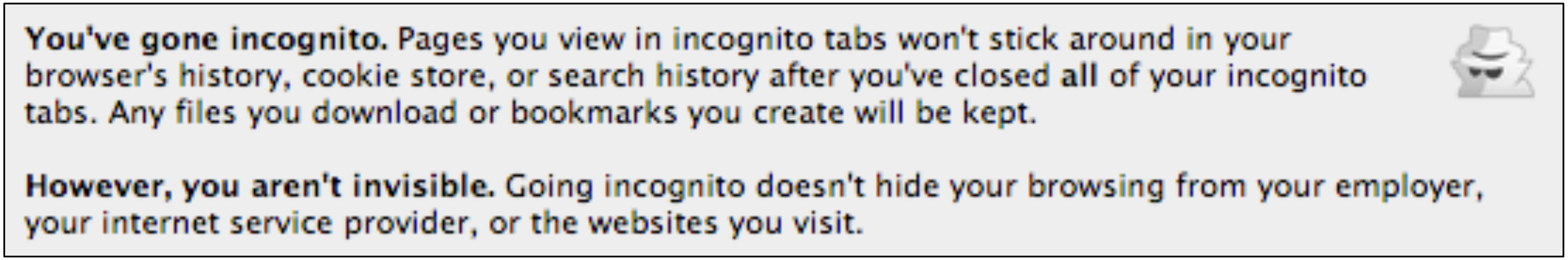
☑ Send a 'Do Not Track' request with your browsing traffic

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.

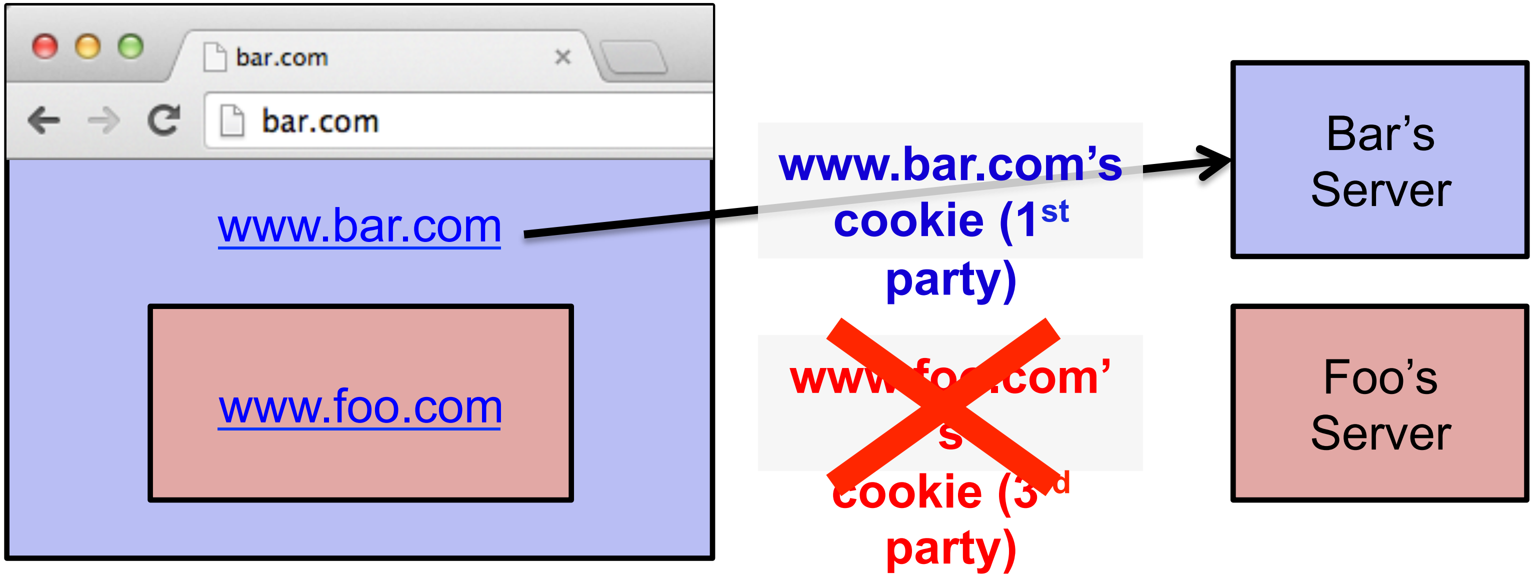


You've gone incognito. Pages you view in incognito tabs won't stick around in your browser's history, cookie store, or search history after you've closed all of your incognito tabs. Any files you download or bookmarks you create will be kept.

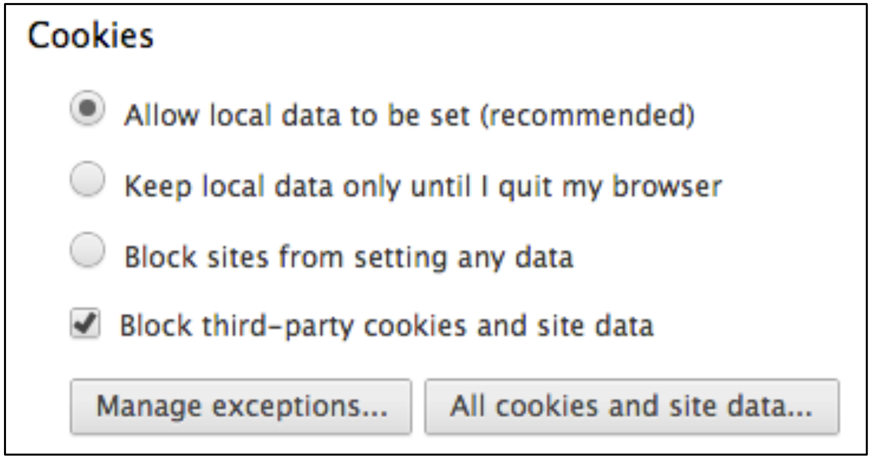
However, you aren't invisible. Going incognito doesn't hide your browsing from your employer, your internet service provider, or the websites you visit.

Defenses to Reduce Tracking

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


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

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

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

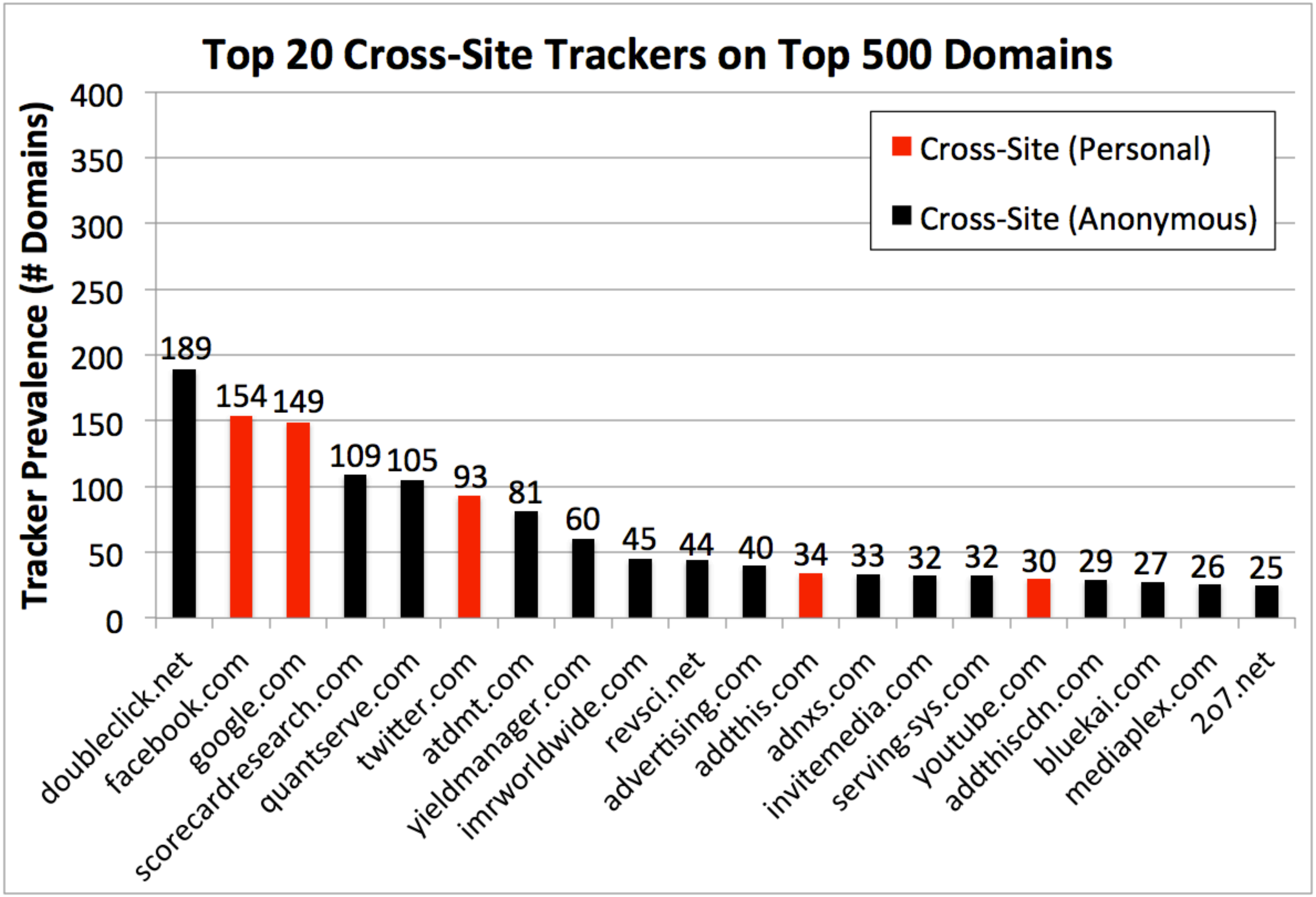
So if a third-party cookie is somehow set, it can be used.

How to get a cookie set?

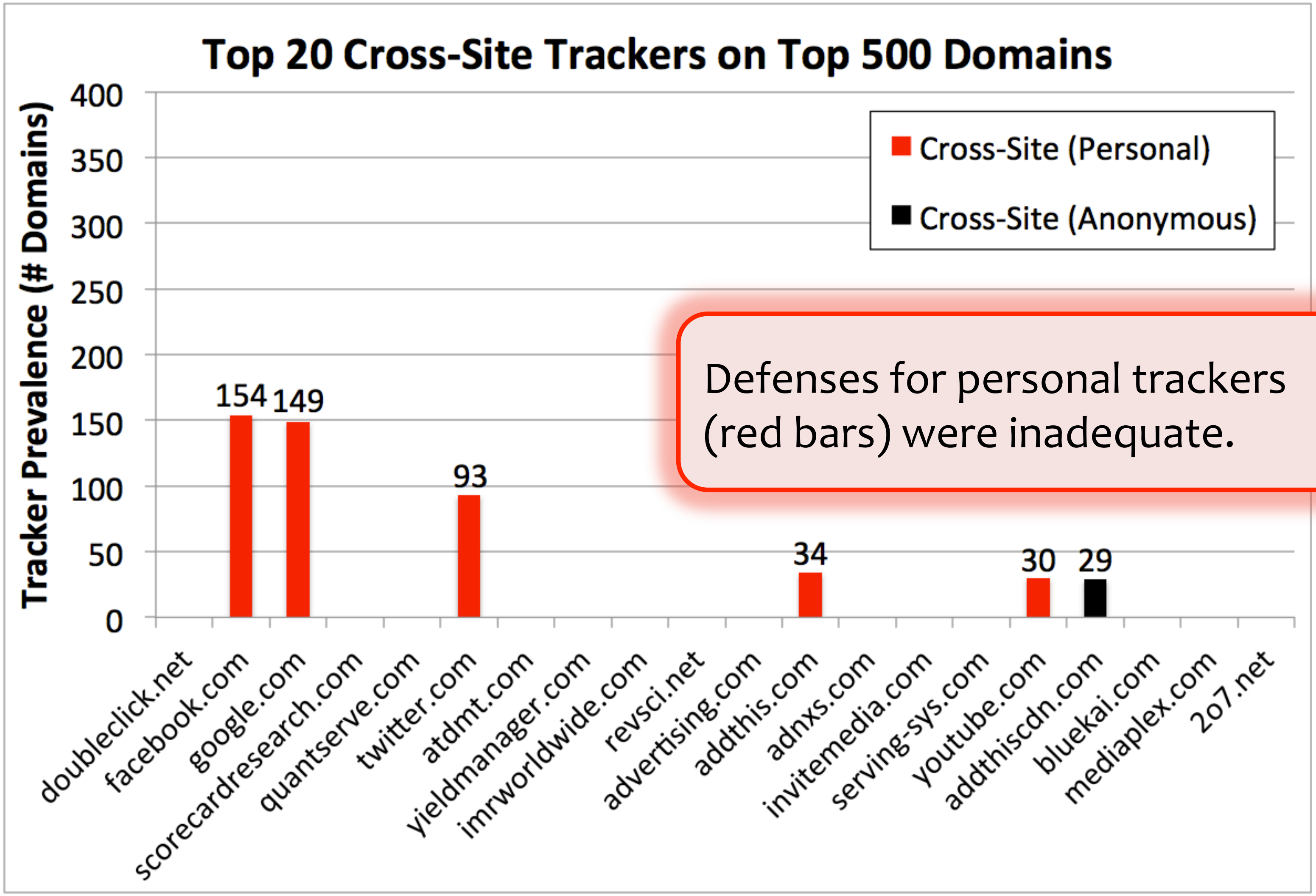
One way: be a first party.

etc.

What 3rd Party Cookie Blocking Misses



What 3rd Party Cookie Blocking Misses



Defenses for personal trackers (red bars) were inadequate.

Defenses to Reduce Tracking

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

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"

https://www.eff.org/privacybadger